

Robust tests for the common principal components model

I. M. Rodrigues¹, G. Boente² and A. M. Pires¹

¹ Departamento de Matemática, Instituto Superior Técnico and CEMAT, Av. Rovisco Pais, 1049-001, Lisboa, Portugal

² Departamento de Matemática and Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1. 1428, Buenos Aires, Argentina.

Keywords: Common principal components model, Proportionality model, Robustness.

1 Introduction

In multivariate analysis we often deal with situations involving several populations, such as discriminant analysis, where the assumption of equality of scatter matrices is usually assumed. Yet sometimes, this assumption is not adequate but problems related to an excessive number of parameters will arise if we estimate the scatter matrices separately for each population. In many practical situations this problem can be avoided if the scatter matrices of the different populations exhibit some common structure. Several authors, as for instance Flury (1988), have studied models for common structure dispersion. One such basic common structure assumes that the k scatter matrices have different eigenvalues but identical eigenvectors, *i.e.*, the matrices are pairwise commutable and may be simultaneously diagonalizable. That is, there is a $(p \times p)$ -orthogonal matrix $\beta : \Sigma_i = \beta \Lambda_i \beta^T$, $i = 1, \dots, k$ where Σ_i is the $(p \times p)$ -scatter matrix of the i -th population and $\Lambda_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ip})$. This model, proposed by Flury (1984), became known as the Common Principal Components (CPC) model. The more restrictive proportionality model assumes that the scatter matrices are equal up to a proportionality constant, *i.e.*, $\Sigma_i = \rho_i \Sigma_1$, for $1 \leq i \leq k$ and $\rho_1 = 1$. In Flury (1988) a unified study of the maximum likelihood estimators under a CPC model and under a proportionality model is given and likelihood ratio tests for a hierarchy of models is studied. However, as it is well known, the likelihood ratio test are in most situations affected by anomalous observations. In this work we propose robust procedures for testing the relationship between scatter matrices. An robust statistic for testing proportionality against a common principal components model is considered. Also, the null hypothesis of a CPC model versus no restrictions on the scatter matrices is studied.

2 The proposal for testing proportionality

Suppose that we wish to test

$$H_0 : \Sigma_i = \rho_i \Sigma_1, \text{ with } \rho_1 = 1, \text{ versus } H_1 : \Sigma_i = \beta \Lambda_i \beta^T, \quad 1 \leq i \leq k.$$

Let $c_{ij} = \frac{\lambda_{ij}}{\lambda_{i1}}$, and $\mathbf{c}_i = (c_{i2}, \dots, c_{ip})$, then under H_0 $c_{ij} = c_{1j}$, for all i and j . On the other hand, under H_1 , $c_{ij} = c_{1j}$ for all i and j entail that H_0 is true. Our test will be based on robust estimators, obtained by plug-in or by projection-pursuit (see Boente and Orellana, 2001 and Boente, Pires and Rodrigues, 2002, 2005), of the eigenvalues of the matrices Σ_i assuming a CPC model.

3 The proposal for testing CPC

The next level in the hierarchy discussed in Flury (1988) corresponds to test

$$H_0 : \Sigma_i = \beta \Lambda_i \beta^T, \text{ versus } H_1 : \Sigma_i \text{ arbitrary positive definite scatter matrices, } \quad 1 \leq i \leq k.$$

The test statistic to test this hypothesis is given by

$$\sum_{i=1}^k n_i \log \left[\frac{\det(\widehat{\mathbf{\Lambda}}_i)}{\det(\mathbf{S}_i)} \right]$$

where $\widehat{\mathbf{\Lambda}}_i$ is the diagonal matrix of the eigenvalues estimated under H_0 , and \mathbf{S}_i is the covariance matrix of the i -th population.

The idea beyond this statistic is that under H_0 it should be expected that $\widehat{\boldsymbol{\beta}}^T \mathbf{S}_i \widehat{\boldsymbol{\beta}}$ will be approximately a diagonal matrix. This idea can be used to robustify the test statistic by considering

$$T_{\text{CPC}} = \sum_{i=1}^k n_i \log \left[\frac{\det \left(\text{diag} \left(\widehat{\boldsymbol{\beta}}^T \mathbf{V}_i \widehat{\boldsymbol{\beta}} \right) \right)}{\det \left(\widehat{\boldsymbol{\beta}}^T \mathbf{V}_i \widehat{\boldsymbol{\beta}} \right)} \right] = \sum_{i=1}^k n_i \log \left[\frac{\det \left(\text{diag} \left(\widehat{\boldsymbol{\beta}}^T \mathbf{V}_i \widehat{\boldsymbol{\beta}} \right) \right)}{\det(\mathbf{V}_i)} \right]$$

where \mathbf{V}_i are independent robust affine equivariant scatter estimates which are asymptotically normally distributed and spherically invariant.

4 Monte Carlo study

Through a Monte Carlo study, we evaluate the finite sample behaviour, in level an power, of both tests statistics for normal and contaminated samples.

References

- G. Boente and L. Orellana (2001). A Robust Approach to Common Principal Components. In *Statistics in Genetics and in the Environmental Sciences*, eds. L. T. Fernholz, S. Morgenthaler, and W. Stahel, pp. 117-147. Basel: Birkhauser.
- G. Boente, A. M. Pires and I. M. Rodrigues (2002). Influence functions and outlier detection under the common principal components model: A robust approach. *Biometrika*, 89, 861–875.
- G. Boente, A. M. Pires and I. M. Rodrigues (2005). General projection–pursuit estimators for the common principal components model: Influence functions and Monte Carlo study. To appear in *Journal of Multivariate Analysis*.
- B. Flury (1984). Common principal components in k groups. *Journal of the American Statistical Association*, 79, 892–898.
- B. Flury (1988). *Common Principal Components and Related Multivariate Models*. New York: John Wiley.