Spatial sign tests for canonical correlations and independence between two vectors

S. Taskinen, A. Kankainen, E. Ollila and H. Oja[1]

[1] Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, FIN-40351 Jyväskylä, Finland

Keywords: Affine equivariant signs, Efficiency, Robustness.

1 Introduction

Let

$$\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^T = \begin{pmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_n \\ \boldsymbol{y}_1 & \ldots & \boldsymbol{y}_n \end{pmatrix}^T = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^T$$

be a k-variate data set of n observations from distribution with cdf F and partitioned covariance matrix

$$\Sigma = \Sigma(F) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

We wish to test the null hypothesis H_0^1: $\Sigma_{12} = 0$ or the null hypothesis H_0^2: $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent. Note that H_0^1 implies that all the canonical correlations are zero, but does not imply the independence between two vectors unless their joint distribution is multinormal.

Several tests for correlations and independence are found in the literature. Wilks (1935) derived the likelihood ratio test for H_0: $\Sigma_{12} = 0$ when the observations are multinormal. An other classical test statistic is introduced for example by Pillai (1955). If $S = S(\boldsymbol{z})$ is a partitioned sample covariance matrix, then the Pillai's test statistic is expressed as $T_0 = \mathrm{Tr}(S_{11}^{-1} S_{12} S_{22}^{-1} S_{21})$. Nonparametric approaches to the problem were given by Puri and Sen (1971) and Gieser and Randles (1997). The latter based their test on interdirections. They found limiting distribution of the test statistic under the null hypothesis of independence and under contiguos alternatives and compared the efficiency and robustness of their test with the classical ones.

2 Tests based on signs

Assume now that F is elliptic. For testing H_0^1, a shape matrix $V = V(\boldsymbol{z})$ based on standardized spatial signs (Randles, 2000, Hettmansperger and Randles, 2001) is first constructed: The spatial sign of $\boldsymbol{z}$ is $\boldsymbol{S}(\boldsymbol{z}) = \|\boldsymbol{z}\|^{-1}\boldsymbol{z}$, that is, the unit vector in the direction of $\boldsymbol{z}$. The location and shape estimates $\hat{\boldsymbol{\mu}}$ and $\widehat{V}$ are chosen so that the spatial signs of the transformed observations $\widehat{V}^{-1/2}(\boldsymbol{z}_i - \hat{\boldsymbol{\mu}})$, $i = 1, \ldots, n$ are standardized, that is,

$$\mathrm{ave}_i\{\widehat{\boldsymbol{S}}_i\} = \boldsymbol{0} \quad \text{and} \quad \mathrm{ave}_i\{\widehat{\boldsymbol{S}}_i \widehat{\boldsymbol{S}}_i^T\} = \frac{1}{k} I_k.$$

Note that $\widehat{V}$ is known as Tyler's M-estimate (1987). The test statistic is then $T_1 = \mathrm{Tr}(V_{11}^{-1} V_{12} V_{22}^{-1} V_{21})$ and the standardized version of the test statistic has a limiting chi-squared distribution.

For testing H_0^2, standardized spatial sign vectors S_i, $i = 1, \ldots, n$ for $\boldsymbol{x}$ and R_i, $i = 1, \ldots, n$ for $\boldsymbol{y}$ are constructed separately. Then $H_{12} = \mathrm{ave}_i\{S_i R_i^T\}$ is the covariance matrix between $\boldsymbol{x}$- and $\boldsymbol{y}$-signs and the test statistic is $T_2 = \mathrm{Tr}(H_{12} H_{12}^T)$. Again (under general assumptions) the limiting distribution of the test statistic is chi-squared distribution.

The tests are compared with the classical ones through limiting Pitman efficiencies in the multivariate normal and t-distribution cases. The theory is illustrated by an example.

References

P.W. Gieser and R.H. Randles (1997). A nonparametric test of independence between two vectors. *Journal of the American Statistical Association*, 92, 561–567.

T.P. Hettmansperger and R.H. Randles (2001). A practical affine equivariant robust estimate of multivariate location. Submitted.

K.C.S. Pillai (1955). Some new test criteria in multivariate analysis. *Annals of Mathematical Statistics*, 26, 117–121.

M.L. Puri and P.K. Sen (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley & Sons, New York.

R.H. Randles (2000). A simpler, affine equivariant, multivariate, distribution-free sign test. *Journal of the American Statistical Association*, 95, 1263–1268.

D.E. Tyler (1987). A distribution free M-estimator of multivariate scatter. *Annals of Statistics*, 15, 234–251.

S.S. Wilks (1935). On the independence of k sets of normally distributed statistical variates. *Econometrica*, 3, 309–326.

Please fill in this form and mail it together with your abstract.

My abstract fits best to topic number 14

List of Topics:

1. Algorithms
2. Applications
3. Biostatistics
4. Computing and graphics
5. Data analysis
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression
21. Time series analysis
22. Wavelets
23. Other (please specify)