ON NONPARAMETRIC TESTS OF INDEPENDENCE AND ROBUST CANONICAL CORRELATION ANALYSIS

Sara Taskinen

Department of Mathematics and Statistics
University of Jyväskylä

November 13, 2003

Abstract

In this thesis, multivariate statistical methods that deal with measuring association between two random vectors are considered. At first, several new test statistics based on the multivariate sign and rank concepts are proposed for testing whether two random vectors are independent. In the second part of the thesis, the use of the so called robust scatter and shape matrices in the canonical correlation analysis is examined. The statistical properties (limiting distributions, efficiencies, robustness) of new methods are studied, and the use of the methods is illustrated by several examples.

Acknowledgements

I wish to express my sincere gratitude to my supervisors Professor Hannu Oja and PhD Annaliisa Kankainen for their expert guidance and constant support during my work. Besides Hannu and Annaliisa, I am deebly indebted to all those who have participated in this project. Especially, I wish to thank my co-authors Professor Ronald H. Randles, Professor Christophe Croux and PhD Esa Ollila for the delightful collaboration. I am also grateful to my thesis reviewers Professor Jana Jurečková and Professor Davy Paindaveine for their constructive comments on my work.

The research work for this thesis was carried out at the Department of Mathematics and Statistics, University of Jyväskylä during the years 2000-2003. The research was funded by the COMAS graduate school and the Academy of Finland. Additional financial support was provided by the Jenny and Antti Wihuri foundation and the Ellen and Artturi Nyyssönen foundation. These organizations and foundations are greatly acknowledged.

Finally, I would like to thank my dearest ones. My parents Pirkko and Pertti and brothers Jero and Riku deserve warm thanks for their support from the very beginning. My deepest gratitude is to my husband Lasse for his patience, love and support.

Jyväskylä, November 2003

Sara Taskinen

List of Original Publications

This thesis consists of an introductory part and publications listed below. In the introductory part they will be referred to Article A, Article B, etc.

A Taskinen, S., Kankainen, A. and Oja, H. (2003), "Tests of independence based on sign and rank covariances", in: *Developments in Robust Statistics. International Conference on Robust Statistics 2001* (eds. R. Dutter, P. Filzmoser, U. Gather and P.J. Rousseeuw), pp. 387–403, Springer-Verlag, Heidelberg.

B Taskinen, S., Kankainen, A. and Oja, H. (2003), "Sign test of independence between two random vectors", *Statistics and Probability Letters*, **62**, 9–21.

C Taskinen, S., Oja, H. and Randles, R.H. (2003), "Multivariate nonparametric tests of independence", Submitted.

D Taskinen, S., Kankainen, A. and Oja, H. (2003), "Rank scores tests of multivariate independence", To appear (shortened version) in: *Theory and Applications of Recent Robust Methods. Series: Statistics for Industry and Technology* (eds. M. Hubert, G. Pison, A. Struyf and S. Van Aelst), Birkhauser, Basel.

E Taskinen, S., Croux, C., Kankainen, A., Ollila, E. and Oja H. (2003), "Canonical analysis based on scatter matrices", Submitted.

Contents

Abstract **3**

Acknowledgements **4**

List of Original Publications **5**

1 Introduction **8**

2 Tests of Independence and Canonical Analysis in Normal Population **10**

2.1 Classical Tests of Independence 10

2.1.1 Independence and Second Moments 10

2.2 Test Statistics 11

2.3 Canonical Correlation Analysis 13

2.3.1 Introduction 13

2.3.2 Population Canonical Correlation Analysis 13

2.3.3 Estimation of Canonical Correlations and Vectors 14

2.3.4 Tests for Canonical Correlations 15

3 Nonparametric and Robust Methods **17**

3.1 Estimation of Location and Scatter 17

3.1.1 Elliptical Distributions 17

3.1.2 Location, Scatter and Shape Functionals 18

3.1.3 Influence Functions and Limiting Distributions 19

3.1.4 Estimation of Multivariate Location and Scatter 21

3.1.5 Estimators Based on Signs and Ranks 23

3.2 Nonparametric Tests of Independence 24

3.2.1 Bivariate Tests of Independence 24

3.2.2 Componentwise Quadrant Statistics 26
3.2.3 Tests Based on Interdirections 27
3.2.4 Tests Based on Standardized Spatial Signs and Ranks 27
3.2.5 Asymptotic Relative Efficiencies 28
3.2.6 Finite-sample Efficiencies 30
3.3 Canonical Correlation Analysis 34
3.3.1 Robust Methods in Canonical Analysis 34
3.3.2 Canonical Analysis Based on Robust Scatter Matrices 34
3.3.3 Asymptotic Relative Efficiencies 36
3.3.4 Tests for Canonical Correlations 36

4 Summary of Original Publications **38**

Bibliography **40**

Original Publications **45**

Chapter 1

Introduction

Classical tests of independence and canonical correlation analysis assume that the observations come from a multivariate normal distribution. The tests and estimators are then derived using the sample mean and sample covariance matrix, which are the maximum likelihood estimators of unknown mean vector and covariance matrix. Under the assumption of multivariate normality, the classical methods are optimal. It is, however, well known that, if the data are spoiled with outlying observations, that is, observations far away from the bulk of the data, or if the underlying distribution has heavier tails than multivariate normal distribution, the techniques based on the sample mean vector and sample covariance matrix perform poorly.

In this thesis, nonparametric and robust alternatives to normal theory methods are considered and their statistical properties (limiting distributions, efficiencies, robustness) are studied. The tests of independence are based on the multivariate sign and rank concepts. When testing independence between two multivariate vectors, standardized spatial signs and ranks are shown to be very useful. These affine equivariant signs and ranks are based on the approach launched in Randles (2000) and Hettmansperger and Randles (2002). The resulting tests are highly resistant to outliers. They are also more efficient than the normal theory based tests when the underlying distribution is heavy-tailed.

A natural way to robustify canonical correlation analysis, is to use robust scatter or shape matrices when estimating the canonical correlations and vectors. This approach is examined in the second part of this thesis. It turns out that the estimates based on affine equivariant sign covariance matrix (Visuri et al., 2000) are highly efficient but not very resistant to outliers. More robust estimates are obtained using for example minimum covariance determinant (Rousseeuw, 1985) estimators. The resulting estimates, however, suffer from poor efficiency properties.

The outline of this introductory part is as follows. In Chapter 2, the normal theory based methods used in testing independence and canonical correlation analysis are shortly reviewed. In Chapter 3, robust and nonparametric alternatives for normal theory methods are considered. At first, different tools needed in deriving robust procedures are presented. These include robust estimators of location and scatter and concepts of sign and rank. Classical as well as recently proposed tests of independence are then presented and discussed, and some of them are compared through limiting and finite-sample efficiencies. Finally, the robust canonical analysis is shortly reviewed, and it is shown how any scatter matrix may be used in estimating canonical correlations and vectors.

Chapter 2

Tests of Independence and Canonical Analysis in Normal Population

2.1 Classical Tests of Independence

2.1.1 Independence and Second Moments

Assume that $\boldsymbol{x}$ is a random vector from a k-variate continuous distribution with cumulative distribution function (cdf) F, and let $\boldsymbol{x}$ be partitioned into p- and q-dimensional subvectors $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ respectively, that is,

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}^{(1)} \\ \boldsymbol{x}^{(2)} \end{pmatrix}.$$

In the following, we wish to test the hypothesis

$$H_0: \quad \boldsymbol{x}^{(1)} \text{ and } \boldsymbol{x}^{(2)} \text{ are independent,}$$

which means that the joint distribution $F(\boldsymbol{x})$ may be written as

$$F(\boldsymbol{x}) = F_1(\boldsymbol{x}^{(1)})F_2(\boldsymbol{x}^{(2)}),$$

where F_1 is the marginal cdf of $\boldsymbol{x}^{(1)}$ and F_2 is the marginal cdf of $\boldsymbol{x}^{(2)}$.

Multivariate distributions are often described with first and second moments (if they exists). Then the partitioned mean vector of $\boldsymbol{x}$ is

$$\boldsymbol{\mu} = E\begin{pmatrix} \boldsymbol{x}^{(1)} \\ \boldsymbol{x}^{(2)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}$$

and the partitioned covariance matrix is

$$\Sigma = E\left[\begin{pmatrix} \boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)} \\ \boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{(1)} - \boldsymbol{\mu}^{(1)} \\ \boldsymbol{x}^{(2)} - \boldsymbol{\mu}^{(2)} \end{pmatrix}^T\right] = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with nonsingular $(p+q) \times (p+q)$, $p \times p$ and $q \times q$ matrices Σ, Σ_{11} and Σ_{22}. It is easy to see that the hypothesis of independence implies the hypothesis

$$H_0' : \quad \Sigma_{12} = 0,$$

that is, each variable in the first set is uncorrelated with each variable in the other set. Note that in the bivariate case ($p = q = 1$), the covariance matrix may be written as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

where σ_1^2 and σ_2^2 are the variances of $x^{(1)}$ and $x^{(2)}$, and ρ is the correlation between $x^{(1)}$ and $x^{(2)}$. In this case, $H_0' :\ \rho = 0$.

When the classical tests of independence are derived, one often assumes that $\boldsymbol{x}$ comes from a k-variate normal distribution with density function

$$f(\boldsymbol{x}) = |2\pi\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}.$$

Under this model, the hypothesis H_0 is equivalent to the hypothesis H_0'. In general this is, however, not true.

2.2 Test Statistics

Assume now that $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ is a random sample from a k-variate normal distribution and that each $\boldsymbol{x}_i$ is partitioned into p- and q-dimensional subvectors $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$. The classical test statistics for testing the null hypothesis of independence may be derived using the maximum likelihood estimators of $\boldsymbol{\mu}$ and Σ, that is, the sample mean vector and sample covariance matrix

$$\bar{\boldsymbol{x}} = \text{ave}\{\boldsymbol{x}_i\} \quad \text{and} \quad S = \text{ave}\{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T\},$$

where "ave" denotes the average taken over $i = 1, \dots, n$. The partitioned $\bar{\boldsymbol{x}}$ and S are

$$\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{\boldsymbol{x}}^{(1)} \\ \bar{\boldsymbol{x}}^{(2)} \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \tag{2.1}$$

respectively.

In the bivariate case ($p = q = 1$), a natural test statistic for testing $H_0' :\ \rho = 0$ is the Pearson product moment correlation coefficient (1896)

$$\widehat{r} = \frac{\sum_{i=1}^n (x_i^{(1)} - \bar{x}^{(1)})(x_i^{(2)} - \bar{x}^{(2)})}{\sqrt{\sum_{i=1}^n (x_i^{(1)} - \bar{x}^{(1)})^2}\sqrt{\sum_{i=1}^n (x_i^{(2)} - \bar{x}^{(2)})^2}},$$

that is, the maximum likelihood estimator of ρ. Under H_0', $\sqrt{n}\,\widehat{r} \to_d N(0,1)$.

Assume next that $p \geq 2$ and $q \geq 2$. Wilks (1935) showed that the likelihood ratio test statistic for testing $H_0' :\ \Sigma_{12} = 0$ is

$$W = \frac{|S|}{|S_{11}||S_{22}|}. \tag{2.2}$$

Under H_0', $-n \log W \to_d \chi^2_{pq}$. Note that when W is used in testing independence in nonnormal populations, one needs to assume that the fourth moments of underlying distribution are finite. Other classical test of independence is the Pillai's trace (1955)

$$P = Tr(S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}), \tag{2.3}$$

which is asymptotically equivalent with Wilks' test, that is, $n(P + \log W) \to_P 0$.

Muirhead (1982) showed that Wilks' test W is affine invariant, that is, its value is not changed under the group of transformations

$$\mathcal{G} = \{\boldsymbol{x} \to D\boldsymbol{x} + \boldsymbol{d}\}, \tag{2.4}$$

where $\boldsymbol{d}$ is a $p + q$ vector and

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

is an arbitrary matrix with nonsingular $p \times p$ and $q \times q$ matrices D_1 and D_2. See also Section 2.3.4. Invariance implies that the value of W does not depend on the chosen marginal coordinate systems and the performance of W is consistent under different covariance structures of $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$. Also Pillai's trace is affine invariant under $\mathcal{G}$.

2.3 Canonical Correlation Analysis

2.3.1 Introduction

The purpose of canonical correlation analysis is to describe the linear interrelations between two random vectors $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$. The method was proposed by Hotelling (1935, 1936), who applied the technique to study the relationship between a set of mental test variables and a set of physical variables.

In canonical analysis one forms new separate coordinate systems for $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$. In both systems, the variables are defined as linear combinations of original vectors, so that, the marginal variables are uncorrelated and have unit variances. In addition, the first variables are linear combinations with maximum correlation. The second variables are such combinations that have maximum correlation and are uncorrelated with the first ones, and so on. The linear combinations are then called canonical variates and corresponding maximum correlations are canonical correlations.

Canonical analysis reduces the correlation structure implicit in Σ to a form involving only few canonical correlations and therefore provides an useful method to reduce the dimensionality of a problem. If one wishes to study the interrelations between two large sets of variates, by canonical analysis one can consider only those linear combinations of each set that are most highly correlated.

2.3.2 Population Canonical Correlation Analysis

Let now k-dimensional random vector $\boldsymbol{x}$ be partitioned as in Section 2.1.1 and assume without loss of generality that $p \leq q$. In canonical analysis one finds a $p \times p$ matrix $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$, a $q \times q$ matrix $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_q)$ and a $p \times p$ diagonal matrix $R = \mathrm{diag}(\rho_1, \ldots, \rho_p)$ with $\rho_1 \geq \cdots \geq \rho_p$ so that

$$\begin{pmatrix} A^T & 0 \\ 0 & B^T \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} I_p & (R, 0) \\ (R, 0)^T & I_q \end{pmatrix}. \tag{2.5}$$

The diagonal elements of R are then called **canonical correlations**, the columns of A and B are **canonical vectors** and the linear combinations

$$\boldsymbol{u} = A^T \boldsymbol{x}^{(1)} \quad \text{and} \quad \boldsymbol{v} = B^T \boldsymbol{x}^{(2)}$$

are **canonical variates**.

By means of canonical variates, the entire relationship between $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ can be expressed using only p canonical correlations. From (2.5) one finds that for $i = 1, \ldots, p$ and $j = 1, \ldots, q$

$$\mathrm{Cov}(u_i, v_j) = \begin{cases} \rho_i, & i = j \\ 0, & i \neq j \end{cases}$$

and for $i, j = 1, \ldots, p$ and $k, l = 1, \ldots, q$

$$\mathrm{Cov}(u_i, u_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad \text{and} \quad \mathrm{Cov}(v_k, v_l) = \begin{cases} 1, & k = l \\ 0, & k \neq l. \end{cases}$$

In other words, the first canonical variates u_1 and v_1 are only correlated with each other, the second variates u_2 and v_2 with each other, and so on. Corresponding correlations are $\rho_1, \ldots, \rho_p$. Note that, if $\rho_1 > \cdots > \rho_p$, then the canonical vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ and $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ are uniquely defined up to a sign. Remaining $q-p$ canonical vectors $\boldsymbol{b}_{p+1}, \ldots, \boldsymbol{b}_q$ are uniquely defined expect for multiplication on the right by an orthogonal $(q-p) \times (q-p)$ matrix (see Anderson, 1984, pp. 493).

Finally note that from (2.5) one has that

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}A = A(R, 0)(R, 0)^T$$

and

$$\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}B = B(R, 0)^T(R, 0).$$

Thus the squared canonical correlations and canonical vectors can be found explicitely as eigenvalues and eigenvectors of

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad \text{and} \quad \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12},$$

so that, the eigenvectors are chosen to satisfy $A^T\Sigma_{11}A = I_p$ and $B^T\Sigma_{22}B = I_q$.

2.3.3 Estimation of Canonical Correlations and Vectors

The sample canonical correlations and vectors can be found as in the previous section by replacing Σ by its sample counterpart. If the data come from a multivariate normal distribution, then the maximum likelihood estimators for canonical correlations $\widehat{R} = \mathrm{diag}(\widehat{r}_1, \ldots, \widehat{r}_p)$ and vectors $\widehat{A} = (\widehat{\boldsymbol{a}}_1, \ldots, \widehat{\boldsymbol{a}}_p)$ and $\widehat{B} = (\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_q)$ are obtained as eigenvalues and eigenvectors of

$$S_{11}^{-1}S_{12}S_{22}^{-1}S_{21} \quad \text{and} \quad S_{22}^{-1}S_{21}S_{11}^{-1}S_{12},$$

with eigenvectors chosen to satisfy $A^T S_{11} A = I_p$ and $B^T S_{22} B = I_q$.

The limiting distribution of the canonical correlations in the multivariate normal case was derived already in Hsu (1941). His result is valid under very general assumptions on the population canonical correlations. The limiting distributions of the canonical vectors have been considered in several papers. See Anderson (1999) for a detailed discussion on these. Anderson also gave the complete limiting distributions of the canonical correlations and vectors assuming that the nonzero population correlations are distinct. He showed among other things that if $p = q$, then for $i = 1, \ldots, p$, the marginal distributions of $\sqrt{n}(\widehat{r}_i - \rho_i)$, $\sqrt{n}(\widehat{\boldsymbol{a}}_i - \boldsymbol{a}_i)$ and $\sqrt{n}(\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i)$ are asymptotically normal with zero mean and asymptotic variances

$$\mathrm{ASV}(\widehat{r}_i) = (1 - \rho_i)^2,$$

$$\mathrm{ASV}(\widehat{\boldsymbol{a}}_i) = \frac{1}{2}\boldsymbol{a}_i\boldsymbol{a}_i^T + \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{(\rho_j^2 + \rho_i^2 - 2\rho_j^2\rho_i^2)(1 - \rho_i^2)}{(\rho_i^2 - \rho_j^2)^2}\boldsymbol{a}_j\boldsymbol{a}_j^T$$

and

$$\mathrm{ASV}(\widehat{\boldsymbol{b}}_i) = \frac{1}{2}\boldsymbol{b}_i\boldsymbol{b}_i^T + \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{(\rho_j^2 + \rho_i^2 - 2\rho_j^2\rho_i^2)(1 - \rho_i^2)}{(\rho_i^2 - \rho_j^2)^2}\boldsymbol{b}_j\boldsymbol{b}_j^T.$$

2.3.4 Tests for Canonical Correlations

In Section 2.2, we considered tests for the null hypothesis $H_0' : \Sigma_{12} = 0$. Since from (2.5) one has that $A^T \Sigma_{12} B = (R, 0)$, the hypothesis is equivalent to

$$H_0' : \ \rho_1 = \cdots = \rho_p = 0,$$

and the test statistics given in (2.2) and (2.3) can be used for testing the canonical correlations also. Note that Wilks' test statistic can be expressed in terms of sample canonical correlations as

$$W = \frac{|S|}{|S_{11}||S_{22}|} = |I_p - S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}| = \prod_{i=1}^{p} (1 - \widehat{r}_i^2)$$

and Pillai's trace becomes

$$P = Tr(S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}) = \sum_{i=1}^{p} \widehat{r}_i^2.$$

Muirhead (1982) showed that under the group of transformations $\mathcal{G}$ given in (2.4), any invariant test is a function of squared sample canonical correlations. The invariance of the test statistics thus follows.

If $H_0' : \rho_1 = \cdots = \rho_p = 0$ is rejected, it may be of interest to study how many population canonical correlations differ from 0, that is, how many canonical correlations are needed to describe the relationships between $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$. Fujikoshi (1974) showed that the likelihood ratio test for testing $H_0' : \rho_{k+1} = \cdots = \rho_p = 0$ is based on

$$W' = \prod_{i=k+1}^{p} (1 - \widehat{r}_i^2).$$

Under H_0', $-n \log W' \to_d \chi^2_{(p-k)(q-k)}$.

Chapter 3

Nonparametric and Robust Methods

3.1 Estimation of Location and Scatter

3.1.1 Elliptical Distributions

In the multivariate normal case, the methods based on sample mean and sample covariance matrix are optimal. It is, however, well known that the sample mean and covariance matrix and methods based on them are highly sensitive to outlying observations. In this chapter we consider some nonparametric and robust alternatives for tests of independence and canonical correlation analysis and start by defining the tools needed in derivations.

Since the elliptical distributions are often used in the robustness studies, we first recall the definition of corresponding distributions. A k-dimensional random vector $\boldsymbol{x}$ has a continuos elliptically symmetric distribution if the density function is of the form

$$f(\boldsymbol{x}) = |\Sigma|^{-1/2} f_0(\Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})), \tag{3.1}$$

where Σ is a positive definite symmetric $k \times k$ matrix (PDS(k)) and $f_0(\boldsymbol{z}) = \exp\{-\rho(||\boldsymbol{z}||)\}$ with $\boldsymbol{z} = \Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$. Note that if $\boldsymbol{x}$ comes from an elliptical distribution, then the standardized variable $\boldsymbol{z}$ has a spherical distribution, and $\boldsymbol{z}$ can be decomposed as $\boldsymbol{z} = r\boldsymbol{u}$, where $r = ||\boldsymbol{z}||$ and $\boldsymbol{u} = ||\boldsymbol{z}||^{-1}\boldsymbol{z}$ are independent with $\boldsymbol{u}$ being uniformly distributed on the unit sphere. If $\rho(r) = r^2/2$, the multivariate normal distribution is obtained and

$$\rho(r) = \frac{k+\nu}{2} \log\left(1 + \frac{r^2}{\nu}\right)$$

gives the k-variate t distribution with ν degrees of freedom. Finally,

$$\rho(r) = -\log\left\{(1-\epsilon)\exp\left(-\frac{r^2}{2}\right) + \epsilon c^{-k}\exp\left(-\frac{r^2}{2c^2}\right)\right\}$$

yields the k-variate contaminated normal distribution, that is, the distribution with density function $f(\boldsymbol{x}) = (1-\epsilon)\phi(\boldsymbol{x}) + \epsilon c^{-k}\phi(\boldsymbol{x}/c)$, where ϕ denotes the density of standard normal distribution, $\epsilon \in [0,1]$ and $c > 1$.

3.1.2 Location, Scatter and Shape Functionals

In this section we define the statistical functionals of multivariate location, scatter and shape. Let $T(F)$ and $C(F)$ denote the location vector and scatter matrix functionals. Alternatively one can write $T(\boldsymbol{x})$ and $C(\boldsymbol{x})$, if $\boldsymbol{x}$ is a random vector with cumulative distribution function F. We say that a k-variate vector functional is a **location vector** if it is affine equivariant, that is, if

$$T(A\boldsymbol{x} + \boldsymbol{b}) = AT(\boldsymbol{x}) + \boldsymbol{b}$$

for any nonsingular $k \times k$ matrix A and k-vector $\boldsymbol{b}$. Further a $k \times k$ matrix valued functional $C(F)$ is a **scatter matrix** if it is PDS(k) and affine equivariant in the sense that

$$C(A\boldsymbol{x} + \boldsymbol{b}) = AC(\boldsymbol{x})A^T.$$

In several applications, like in canonical correlation analysis, it is enough to estimate the covariance matrix up to a constant. The so called shape matrix can be seen as a "standardized" scatter matrix. The functional $V(F)$ is a **shape matrix** if it is PDS(k) with $Det(V) = 1$ and affine equivariant in the sense that

$$V(A\boldsymbol{x} + \boldsymbol{b}) = \{Det[AV(\boldsymbol{x})A^T]\}^{-1/k} AV(\boldsymbol{x})A^T.$$

The condition $Det(V) = 1$ is sometimes replaced by the condition $Tr(V) = k$. See Ollila et al. (2003b). Note that if $C(F)$ is a scatter matrix, then

$$V(F) = \{Det[C(F)]\}^{-1/k} C(F).$$

The shape matrix can, however, be defined without any reference to a scatter matrix. Note also that if $\boldsymbol{x}$ is a random vector from an elliptically symmetric distribution with cdf F, then the affine equivariance properties of the functionals imply that $T(\boldsymbol{x}) = \boldsymbol{\mu}$, $C(\boldsymbol{x}) = c_0\Sigma$ and $V(\boldsymbol{x}) = [Det(\Sigma)]^{-1/k}\Sigma$. Thus, in the elliptic case, the location vectors and shape matrices estimate the same population quantities. The scatter matrices are not directly comparable, and a correction factor $(1/c_0)$ depending both on C and the spherical distribution corresponding to F is needed to quarantee the Fisher-consistency to Σ.

3.1.3 Influence Functions and Limiting Distributions

Influence function measures the robustness of a functional against a single outlier, that is, the effect of an infinitesimal contamination located at point $\boldsymbol{z}$. For robust estimators the influence functions are bounded and continuous. Consider now the contaminated distribution

$$F_\epsilon = (1-\epsilon)F + \epsilon\Delta_{\boldsymbol{z}},$$

where $\Delta_{\boldsymbol{z}}$ denotes the cdf of a distribution putting all its mass at $\boldsymbol{z}$. The influence function (Hampel et al., 1986) of a functional T at F is then defined as

$$\mathrm{IF}(\boldsymbol{z}, T, F) = \lim_{\epsilon\to 0}\frac{T(F_\epsilon) - T(F)}{\epsilon}.$$

Let now F be a spherical distribution and write $\boldsymbol{z} = r\boldsymbol{u}$ where $r = ||\boldsymbol{z}||$ is the length and $\boldsymbol{u} = ||\boldsymbol{z}||^{-1}\boldsymbol{z}$ is the direction of the contamination vector $\boldsymbol{z}$. The influence functions of location, scatter and shape functionals are derived in Hampel et al. (1986), Croux and Haesbrock (2000) and Ollila et al. (2003b). The influence function of location functional T is given by

$$\mathrm{IF}(\boldsymbol{z}, T, F) = \gamma_T(r)\boldsymbol{u},$$

where $\gamma_T(r)$ is a real valued function depending on T and F. Further, the influence functions of scatter and shape matrix functionals are

$$\mathrm{IF}(\boldsymbol{z}; C, F) = \alpha_C(r)\boldsymbol{u}\boldsymbol{u}^T - \beta_C(r)I_k$$

and

$$\mathrm{IF}(\boldsymbol{z}; V, F) = \alpha_V(r)\left[\boldsymbol{u}\boldsymbol{u}^T - \frac{1}{k}I_k\right]$$

for some real valued functions $\alpha_C(r)$, $\beta_C(r)$ and $\alpha_V(r)$.

As an example, consider the population mean vector and covariance matrix functionals

$$T(F) = E_F(\boldsymbol{x}) \quad \text{and} \quad C(F) = E_F[(\boldsymbol{x} - E_F(\boldsymbol{x}))(\boldsymbol{x} - E_F(\boldsymbol{x}))^T].$$

At spherical F, the corresponding influence functions are

$$\mathrm{IF}(\boldsymbol{z}, T, F) = r\boldsymbol{u}, \quad \text{and} \quad \mathrm{IF}(\boldsymbol{z}; C, F) = r^2\boldsymbol{u}\boldsymbol{u}^T - \frac{E_F(r^2)}{k} I_k.$$

Since the functions are linear and quadratic in r, the estimates are not robust against outliers.

Influence functions can also be used when computing the asymptotic variances of the estimates. If $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is a random sample from a spherical distribution and F_n is the corresponding empirical cdf, then natural location, scatter and shape estimates are $\widehat{\boldsymbol{\mu}} = T(F_n)$, $\widehat{C} = C(F_n)$ and $\widehat{V} = V(F_n)$. Under general assumptions, the limiting distributions of $\sqrt{n}\widehat{\boldsymbol{\mu}}$, $\sqrt{n}\mathrm{vec}(\widehat{C} - I_k)$ and $\sqrt{n}\mathrm{vec}(\widehat{V} - I_k)$ are multivariate normal distributions with mean vectors zero and covariance matrices

$$\mathrm{ASV}(\widehat{\boldsymbol{\mu}}; F) = E_F[\mathrm{IF}(\boldsymbol{x}, T, F)\mathrm{IF}(\boldsymbol{x}, T, F)^T],$$

$$\mathrm{ASV}(\widehat{C}; F) = E_F[\mathrm{vec}\{\mathrm{IF}(\boldsymbol{x}, C, F)\}\mathrm{vec}\{\mathrm{IF}(\boldsymbol{x}, C, F)\}^T]$$

and

$$\mathrm{ASV}(\widehat{V}; F) = E_F[\mathrm{vec}\{\mathrm{IF}(\boldsymbol{x}, V, F)\}\mathrm{vec}\{\mathrm{IF}(\boldsymbol{x}, V, F)\}^T].$$

Here "vec" is a matrix operator that stacks the columns of the matrix on top of each other. As shown in Ollila et al. (2003a, 2003b), the limiting covariance of location estimate may be written using the marginal variance of an element of $\sqrt{n}\widehat{\boldsymbol{\mu}}$, and the covariances of scatter and shape estimates are characterized by the asymptotic variances of diagonal and off-diagonal elements of scatter and shape matrices as follows.

$$\mathrm{ASV}(\widehat{\boldsymbol{\mu}}; F) = \mathrm{ASV}(\widehat{\mu}_1; F) I_k,$$

$$\begin{aligned}\mathrm{ASV}(\widehat{C}; F) = {} & \mathrm{ASV}(\widehat{C}_{12}; F) \left[I_{k^2} + I_{k,k} - 2\mathrm{vec}(I_k)\mathrm{vec}(I_k)^T\right] \\ & + \mathrm{ASV}(\widehat{C}_{11}; F)\mathrm{vec}(I_k)\mathrm{vec}(I_k)^T\end{aligned}$$

and

$$\mathrm{ASV}(\widehat{V};F) = \mathrm{ASV}(\widehat{V}_{12};F)\left[I_{k^2} + I_{k,k} - \frac{2}{k}\mathrm{vec}(I_k)\mathrm{vec}(I_k)^T\right],$$

where $I_{k,k}$ is the so called commutation matrix, that is, for any $k \times k$ matrix A, $\mathrm{vec}(A^T) = I_{k,k}\mathrm{vec}(A)$, and

$$\mathrm{ASV}(\widehat{\mu}_1;F) = \frac{E_F[\gamma_T^2(r)]}{k}, \quad \mathrm{ASV}(\widehat{V}_{12};F) = \frac{E_F[\alpha_V^2(r)]}{k(k+2)},$$

$$\mathrm{ASV}(\widehat{C}_{12};F) = \frac{E_F[\alpha_C^2(r)]}{k(k+2)} \tag{3.2}$$

and

$$\mathrm{ASV}(\widehat{C}_{11};F) = \frac{3E_F[\alpha_C^2(r)]}{k(k+2)} - \frac{2E_F[\alpha_C(r)\beta_C(r)]}{k} + E_F[\beta_C^2(r)]. \tag{3.3}$$

At elliptical distributions, the expressions for influence functions and limiting variances can be derived using the affine equivariance properties of functionals. See Ollila et al. (2003a, 2003b), for example.

3.1.4 Estimation of Multivariate Location and Scatter

Several robust techniques for estimating multivariate location and scatter have been proposed in the literature. In the following we briefly review some of the techniques in more detail.

Consider first the regular **maximum likelihood (ML) estimators** of location and scatter at elliptical model given in (3.1). Write $\boldsymbol{z}_i = C^{-1/2}(\boldsymbol{x}_i - \boldsymbol{\mu})$ for the standardized observations. Then $r_i = ||\boldsymbol{z}_i||$ and $\boldsymbol{u}_i = ||\boldsymbol{z}_i||^{-1}\boldsymbol{z}_i$, and simultaneous ML-estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{C}$ solve

$$\mathrm{ave}\{w_1(r_i)\boldsymbol{u}_i\} = \boldsymbol{0} \quad \text{and} \quad \mathrm{ave}\{w_2(r_i)\boldsymbol{u}_i\boldsymbol{u}_i^T\} = I_k,$$

with the weight functions $w_1(r) = \psi(r)$ and $w_2(r) = r\psi(r)$, where $\psi(r) = \rho'(r)$ is the optimal location score function. Note that in the multivariate normal case, $\psi(r) = r$, and the sample mean vector and covariance matrix are obtained.

M-estimators of location and scatter were first proposed by Maronna (1976). Huber (1981) extended Maronna's definition by defining M-estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{C}$ as solutions of

$$\mathrm{ave}\{v_1(r_i)(\boldsymbol{x}_i - \boldsymbol{\mu})\} = \boldsymbol{0} \quad \text{and} \quad \mathrm{ave}\{v_2(r_i)(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T - v_3(r_i)C\} = 0,$$

where v_1, v_2 and v_3 are some real-valued functions on $[0, \infty)$. Note that M-estimates include ML-estimates as a particular case. For the existence and uniqueness, influence functions and asymptotic normality of M-estimators, see Maronna (1976), Huber (1981) and Kent and Tyler (1991), for example. As an example of M-estimators, consider Huber's M-estimator that use $v_3(r) = 1$ and $v_i(r) = \psi_i(r)/r$, for $i = 1, 2$, where $\psi_1(r) = \psi_H(r, c)$ and $\psi_2(r) = \psi_H(r, c^2)$ and

$$\psi_H(r, c) = \min\{r, \max\{r, -c\}\}$$

is known as Huber's psi-function. The tuning constant c is chosen so that $q = Pr(\chi_k^2 \leq c^2)$. According to Tyler (1986), the asymptotic breakdown point of Huber's M-estimator is then $\epsilon^* = \min\{1/c^2, 1 - k/c^2\}$ for $c^2 > k$. Thus the breakdown point decreases with the dimension. Note that the breakdown point is another tool to measure the robustness of an estimator. Loosely speaking it is the maximum fraction of outliers in the sample that an estimator can tolerate.

S-estimators were introduced by Rousseeuw and Leroy (1987) and Davies (1987). They are defined as solutions $\widehat{\boldsymbol{\mu}}$ and $\widehat{C}$ to the problem of minimizing $det(C)$ subject to

$$\text{ave}\{\rho(r_i)\} = b,$$

where $\rho : \mathbb{R} \to [0, \infty)$ is bounded, nondecreasing and sufficiently smooth, and $0 < b < \sup \rho$. For the general theory and properties of S-estimators, see Davies (1987), Lopuhaä (1989) and Lopuhaä and Rousseeuw (1991). An example of function ρ is Tukey's biweight function

$$\rho(r, c) = \min\left\{\frac{r^2}{2} - \frac{r^4}{2c^2} + \frac{r^6}{6c^4}, \frac{c^2}{6}\right\},$$

that yields to the biweight S-estimator. To attain ϵ^* asymptotic breakdown point, the constant c can be chosen so that $b = \epsilon^* \rho(c)$.

Other affine equivariant estimators of location and scatter include for example the **minimum volume ellipsoid (MVE) estimators** and **minimum covariance determinant (MCD) estimators** introduced by Rousseeuw (1985). MVE-estimators are computed using the smallest regular ellipsoid containing at least half of the observations. Location estimate is then the center of the ellipsoid and scatter matrix is defined by the shape of the ellipsoid. MCD-estimators are defined using such subset h observations whose covariance matrix has smallest determinant. The location and scatter estimates are then given by the average and covariance matrix computed over this optimal subset. Typically, the size of the subset equals $h = [n(1 - \epsilon^*)]$. The properties of MVE- and MCD-estimators are studied in Davies (1992), Butler et al. (1993) and Croux and Haesbroeck (1999) among others.

3.1.5 Estimators Based on Signs and Ranks

The methods based on signs and ranks are widely used in estimation and testing problems. The reason for their popularity is that they are usually robust, efficient and valid under very weak assumptions about the underlying population. In the following, we will consider location and shape matrix estimators based on signs and ranks.

Recall first the definition of univariate sign and rank. Let $x_1, \ldots, x_n$ be a univariate data set. The univariate sign function is

$$S(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ +1, & x > 0. \end{cases}$$

The centered rank function is defined as

$$R(x) = \text{ave}\{S(x - x_i)\}$$

and the univariate median $\widehat{\mu}$ of the x_i's satisfies $\text{ave}\{S(\widehat{\mu} - x_i)\} = 0$. The multivariate extensions of the univariate sign and rank are now easily constructed. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a k-variate data set. The spatial sign function is then defined as

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \|\boldsymbol{x}\|^{-1}\boldsymbol{x}, & \boldsymbol{x} \neq \boldsymbol{0} \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0}, \end{cases}$$

where $\|\boldsymbol{x}\| = (\boldsymbol{x}^T\boldsymbol{x})^{1/2}$ is the (Euclidean) length of the vector $\boldsymbol{x}$, that is, $\boldsymbol{S}(\boldsymbol{x})$ is a unit vector in the direction of $\boldsymbol{x}$. The spatial rank function is defined as

$$\boldsymbol{R}(\boldsymbol{x}) = \text{ave}\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{x}_i)\}$$

and the spatial median $\widehat{\boldsymbol{\mu}}$ (Brown, 1983) of the $\boldsymbol{x}_i$'s satisfies $\text{ave}\{\boldsymbol{S}(\widehat{\boldsymbol{\mu}} - \boldsymbol{x}_i)\} = \boldsymbol{0}$. See also Möttönen and Oja (1995).

Hettmansperger and Randles (2002) proposed affine equivariant location and shape matrix estimators based on the spatial sign concept. Write $\boldsymbol{z}_i = V^{-1/2}(\boldsymbol{x}_i - \boldsymbol{\mu})$ for the standardized observations. Then the location and shape estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ solve

$$\text{ave}\{\boldsymbol{S}(\boldsymbol{z}_i)\} = \boldsymbol{0} \quad \text{and} \quad k\,\text{ave}\{\boldsymbol{S}(\boldsymbol{z}_i)\boldsymbol{S}^T(\boldsymbol{z}_i)\} = I_k.$$

The resulting shape matrix estimate is Tyler's M-estimate (Tyler, 1987) and the location estimate is the transformation retransformation spatial median. For the transformation retransformation technique, see Chakraborty et al. (1998).

The shape matrix estimates may also be based on the spatial rank concept. Dümbgen (1998) considered the estimate similar to Tyler's M-estimate only computed on differences $\boldsymbol{z}_i - \boldsymbol{z}_j$. The shape matrix estimate $\widehat{V}$ is then chosen to satisy

$$k\,\text{ave}_{i<j}\{\boldsymbol{S}(\boldsymbol{z}_i - \boldsymbol{z}_j)\boldsymbol{S}^T(\boldsymbol{z}_i - \boldsymbol{z}_j)\} = I_k.$$

Moreover, $\widehat{V}$ may also be chosen so that

$$k\,\text{ave}\{\boldsymbol{R}(\boldsymbol{z}_i)\boldsymbol{R}^T(\boldsymbol{z}_i)\} = \text{ave}\{\boldsymbol{R}^T(\boldsymbol{z}_i)\boldsymbol{R}(\boldsymbol{z}_i)\}I_k.$$

Note that, since the above approaches are based on differences of standardized observations, they avoid estimation of location parameter.

If $\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ are some location and shape matrix estimates and the standardized observations are defined as $\boldsymbol{z}_i = \widehat{V}^{-1/2}(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})$, then the vectors $\boldsymbol{S}(\boldsymbol{z}_i)$ are called the **standardized spatial sign vectors** and corresponding rank vectors $\boldsymbol{R}(\boldsymbol{z}_i) = \text{ave}_j\{\boldsymbol{S}(\boldsymbol{z}_i - \boldsymbol{z}_j)\}$ are the **standardized spatial rank vectors**. Unlike the spatial sign and rank vectors of original observations, these standardized sign and rank vectors are affine equivariant. See Randles (2000) and Article C, for example. Note that, in the standardization, any $\sqrt{n}$-consistent shape or scatter matrix estimate may be used.

Besides spatial signs and ranks, several other multivariate extensions of the univariate signs and ranks are also found in the literature. For the concepts of marginal signs and ranks and affine equivariant (Oja) signs and ranks, see for example Hettmansperger and McKean (1998) and references therein. In Visuri et al. (2000), sign and rank covariance matrices based on marginal, spatial and affine equivariant (Oja) signs and ranks are defined and their usefulness in scatter matrix estimation is discussed. The statistical properties of spatial sign and rank covariance matrices are studied in Marden (1999), Visuri et al. (2000) and Croux et al. (2002). For the properties of affine equivariant sign and rank covariance matrices, see Ollila et al. (2003a, 2003c) and Visuri et al. (2003).

3.2 Nonparametric Tests of Independence

3.2.1 Bivariate Tests of Independence

Several nonparametric tests of bivariate independence have been proposed in the literature. See for example Hájek and Šidák (1967), for a review of those. Classical nonparametric competitors to Pearson product moment correlation coefficient are

Blomqvist's quadrant test (1950), Kendall's tau (1938) and Spearman's rho (1904). The tests are based on the univariate sign and rank concepts and are defined as follows. Let $\widehat{\mu}^{(1)}$ be a univariate median of the $x_i^{(1)}$'s and write $\widehat{S}_i^{(1)} = S(x_i^{(1)} - \widehat{\mu}^{(1)})$, $S_{ij}^{(1)} = S(x_i^{(1)} - x_j^{(1)})$ and $R_i^{(1)} = \text{ave}_j\{S(x_i^{(1)} - x_j^{(1)})\}$ for the centered signs, signs of pairwise differences and centered ranks. If $\widehat{S}_i^{(2)}$, $S_{ij}^{(2)}$ and $R_i^{(2)}$ are corresponding signs and ranks based on $x_1^{(2)}, \ldots, x_n^{(2)}$, then the Blomqvist's quadrant statistic is

$$Q = \text{ave}\{\widehat{S}_i^{(1)} \widehat{S}_i^{(2)}\},$$

Kendall's tau is

$$\tau = \text{ave}_{i<j}\{S_{ij}^{(1)} S_{ij}^{(2)}\}$$

and Spearman's rho is

$$\rho = \text{ave}\{R_i^{(1)} R_i^{(2)}\}.$$

The test statistics are thus covariances between centered signs, signs of the pairwise differences, and centered ranks, respectively. Also several rank scores tests of bivariate independence are found in the literature. If $\widehat{D}_i^{(1)}$ and $\widehat{D}_i^{(2)}$ are regular ranks of $|x_i^{(1)} - \widehat{\mu}^{(1)}|$ and $|x_i^{(2)} - \widehat{\mu}^{(2)}|$, and $a(u)$ and $b(u)$ are some score functions, then the test statistics may be written as

$$T = \text{ave}\{a(\widehat{D}_i^{(1)}) b(\widehat{D}_i^{(2)}) \widehat{S}_i^{(1)} \widehat{S}_i^{(2)}\}.$$

For asymptotically equivalent rank scores tests, see Hájek and Šidák (1967).

The properties of Q, τ and ρ have been widely studied in the literature. For example, their asymptotic relative efficiencies have been considered in several papers. To compute asymptotic efficiencies, a model of dependence have to be chosen to serve as an alternative to the null hypothesis of independence. Loosely speaking, the asymptotic relative efficiency of the test relative to the competing test is then the ratio of sample sizes such that the tests achieve equal power against equal alternatives. Note that if T_n is the test statistic such that under the alternative hypothesis H_n, $\sqrt{n}T_n \rightarrow_d N(\mu, \sigma^2)$, then μ^2/σ^2 is the so called efficacy of the test and asymptotic relative efficiencies are obtained as ratios of efficacies.

In the normal distribution case, natural alternatives include bivariate normal distributions with nonzero correlation. The efficiency of Q relative to the Pearson product moment correlation coefficient $\widehat{r}$ is then $(2/\pi)^2$ and the efficiencies of τ and ρ relative to $\widehat{r}$ are $(3/\pi)^2$. Other classes of alternatives are considered in Konijn (1954), Farlie (1960), Bhuchongkul (1964) and Hájek and Šidák (1967) among others. Konijn (1954) defined his alternatives as $x_i^{(1)} = \lambda_1 y_i + \lambda_2 z_i$ and $x_i^{(2)} = \lambda_3 y_i + \lambda_4 z_i$, where y_i and z_i are independent random variables and $\lambda_1, \ldots, \lambda_4 \in \mathbb{R}$. He compared the efficiencies of Q, τ and ρ using several choices of underlying distribution. Hájek and Šidák (1967) considered alternatives of type $x_i^{(1)} = y_i^{(1)} + \Delta z_i$ and $x_i^{(2)} = y_i^{(2)} + \Delta z_i$, where $\Delta > 0$, $y_i^{(1)}$, $y_i^{(2)}$ and z_i are mutually independent, $y_i^{(1)}$ and $y_i^{(2)}$ are distributed according to f and g and $0 < Var(z_i) < \infty$. They showed that when f and g are of logistic type, the test based on ρ is the locally most powerful test and when f and g are of double-exponential type, Q yields the locally most powerful test.

In Article A, the applicability of different sign and rank covariances (Visuri et al., 2000) in testing independence is studied. Note that when marginal sign and rank covariances are used, the resulting statistics are Q, τ and ρ. The tests are compared with the classical sample covariance through asymptotic efficiencies using the dependence model similar to a model proposed by Gieser and Randles (1997). Also the robustness properties of statistics are studied.

3.2.2 Componentwise Quadrant Statistics

Puri and Sen (1971) proposed nonparametric analogues to Wilks' likelihood ratio test based on componentwise rankings. Let T be a $(p+q) \times (p+q)$ matrix with elements

$$T_{st} = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{D_{si}}{n+1}\right) J\left(\frac{D_{ti}}{n+1}\right),$$

where $D_{si} = \sum_j I(x_{sj} \le x_{si})$ and $J(u)$ is an arbitrary score function. If T is partitioned as the sample covariance matrix in (2.1), then the test statistic of Puri and Sen is of the form

$$S^J = \frac{|T|}{|T_{11}||T_{22}|}.$$

Note that with special choices of score functions, S^J yields to multivariate extensions of Blomqvist's quadrant statistic and Spearman's rho. Puri and Sen (1971) showed that under H_0, $-n \log S^J \to_d \chi^2_{pq}$, so their test is a natural competitor of Wilks' likelihood ratio test W. The test, however, is not invariant under the group $\mathcal{G}$ given in (2.4). Thus its performance depends on the variance-covariance structure of the underlying data.

3.2.3 Tests Based on Interdirections

Gieser and Randles (1997) introduced a nonparametric test of independence based on interdirections (Randles, 1989). Their test statistic is defined as

$$Q_1^2 = \text{ave}\{\cos(\pi\widehat{p}^{(1)}(i;i'))\cos(\pi\widehat{p}^{(2)}(i;i'))\},$$

where, for example, $\widehat{p}^{(1)}(i;i')$ is the fraction of hyperplanes formed by the origin and $p-1$ vectors $\boldsymbol{x}_{i^*}^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)}$ with $i^* \neq i$ and $i^* \neq i'$ such that $\boldsymbol{x}_i^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)}$ and $\boldsymbol{x}_{i'}^{(1)} - \widehat{\boldsymbol{\mu}}^{(1)}$ are on the opposite sides of the hyperplane. The observations are centered using some affine equivariant location estimate $\widehat{\boldsymbol{\mu}}^{(1)}$ based on $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$. Note that in the bivariate case, Q_1^2 reduces to Blomqvist's quadrant statistic. Gieser and Randles showed that under H_0 and for elliptically symmetric $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, $npq\,Q_1^2 \to_d \chi_{pq}^2$. They also showed that their test is affine invariant under the group $\mathcal{G}$.

As shown in Article C, interdirections can be used in deriving multivariate analogues to Kendall's tau and Spearman's rho, respectively. The statistics are then based on

$$\tau_1^2 = \text{ave}\{\cos(\pi p^{(1)}(i,j;i',j'))\cos(\pi p^{(2)}(i,j;i',j'))\}$$

and

$$\rho_1^2 = \text{ave}\{\cos(\pi p^{(1)}(i,j;i',j'))\cos(\pi p^{(2)}(i,k;i',k'))\},$$

where, for example, $p^{(1)}(i,j;i',j')$ represents the fraction of hyperplanes formed by the origin and differences $\boldsymbol{x}_{i^*}^{(1)} - \boldsymbol{x}_{j^*}^{(1)}$ such that $\boldsymbol{x}_i^{(1)} - \boldsymbol{x}_j^{(1)}$ and $\boldsymbol{x}_{i'}^{(1)} - \boldsymbol{x}_{j'}^{(1)}$ are on the opposite sides of the hyperplane. (Here i^* and j^* are different from any of $\{i, j, i'$ or $j'\}$). Note that the Kendall and Spearman analogues do not require centering on a location estimator. It is also remarkable that no scatter or shape matrix estimate is needed in the interdirection approach.

3.2.4 Tests Based on Standardized Spatial Signs and Ranks

In Article B, more practical extension of Blomqvist's quadrant test is introduced. The statistic is created by forming p-dimensional standardized sign vectors $\widehat{\boldsymbol{S}}_i^{(1)} = \boldsymbol{S}(\boldsymbol{z}_i^{(1)})$ based on $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$ and q-dimensional sign vectors $\widehat{\boldsymbol{S}}_i^{(2)} = \boldsymbol{S}(\boldsymbol{z}_i^{(2)})$ based on $\boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_n^{(2)}$. The test is then based on

$$Q_2^2 = ||\text{ave}\{\widehat{\boldsymbol{S}}_i^{(1)}\widehat{\boldsymbol{S}}_i^{(2)^T}\}||^2,$$

where $||A||^2 = Tr(A^TA)$. In Article B, the data points are standardized with transformation retransformation spatial medians and Tyler's M-estimates defined in Section 3.1.5. However, in the standardization, any affine equivariant $\sqrt{n}$-consistent estimates can be used. In the elliptic case, Q_2^2 is asymptotically equivalent with the interdirection test Q_1^2, but much easier to compute in practise.

As shown in Article C, the multivariate extensions of Kendall's tau and Spearmans's rho can be derived similarly. If p-dimensional standardized sign vectors $\widehat{\boldsymbol{S}}_{ij}^{(1)} = \boldsymbol{S}(\boldsymbol{z}_i^{(1)} - \boldsymbol{z}_j^{(1)})$ and rank vectors $\widehat{\boldsymbol{R}}_i^{(1)} = \text{ave}_j\{\boldsymbol{S}(\boldsymbol{z}_i^{(1)} - \boldsymbol{z}_j^{(1)})\}$ are formed based on $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_n^{(1)}$ and q-dimensional vectors $\widehat{\boldsymbol{S}}_{ij}^{(2)}$ and $\widehat{\boldsymbol{R}}_i^{(2)}$ are formed based on $\boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_n^{(2)}$, then the multivariate analogue to Kendall's tau is

$$\tau_2^2 = ||\text{ave}_{i<j}\{\widehat{\boldsymbol{S}}_{ij}^{(1)} \widehat{\boldsymbol{S}}_{ij}^{(2)^T}\}||^2$$

and the multivariate analogue to Spearman's rho is

$$\rho_2^2 = ||\text{ave}\{\widehat{\boldsymbol{R}}_i^{(1)} \widehat{\boldsymbol{R}}_i^{(2)^T}\}||^2.$$

Note that, since the signs are based on differences, only the shape estimate is needed for standardizing the observations. In Article C, the standardization is done using shape estimates similar to Tyler's M-estimate only computed on differences of observations. See Section 3.1.5, for corresponding definitions. Again, in the elliptic case, τ_2^2 and ρ_2^2 are asymptotically equivalent with their interdirection counterparts τ_1^2 and ρ_1^2, but much easier to compute in practise. For the asymptotic properties of τ_2^2 and ρ_2^2, see Article C.

In Article D, rank scores tests of multivariate independence are introduced. The test statistics are of the form

$$T^2 = \left|\left|\text{ave}\left\{a\left(\frac{\widehat{D}_i^{(1)}}{n+1}\right) b\left(\frac{\widehat{D}_i^{(2)}}{n+1}\right) \widehat{\boldsymbol{S}}_i^{(1)} \widehat{\boldsymbol{S}}_i^{(2)^T}\right\}\right|\right|^2,$$

where $a(u)$ and $b(u)$ are continuous, monotone and square integrable score functions, and $\widehat{D}_i^{(1)}$ and $\widehat{D}_i^{(2)}$ denote the regular ranks of $||\boldsymbol{z}_i^{(1)}||$ and $||\boldsymbol{z}_i^{(2)}||$. In Article D, three different choices of the score functions, namely the sign scores, the Wilcoxon scores and the van der Waerden scores, are discussed in more detail.

3.2.5 Asymptotic Relative Efficiencies

In the multivariate case, only a few models of dependence are found in the literature. Puri and Sen (1971) considered some interesting sequences of alternatives and computed the efficiencies of S^J relative to the likelihood ratio test under multinormal distribution using some special choices of score functions.

Gieser and Randles (1997) used in their efficiency comparisons a generalization of the model introduced by Konijn (1954). Their model of dependence is given by

$$\begin{pmatrix} \boldsymbol{y}_i^{(1)} \\ \boldsymbol{y}_i^{(2)} \end{pmatrix} = \begin{pmatrix} (1-\Delta)I_p & \Delta M_1 \\ \Delta M_2 & (1-\Delta)I_q \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix} = A_\Delta \begin{pmatrix} \boldsymbol{x}_i^{(1)} \\ \boldsymbol{x}_i^{(2)} \end{pmatrix}, \tag{3.4}$$

where $\Delta = \delta/\sqrt{n}$, $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$ are independent p- and q-variate random vectors, and M_1 and M_2 are arbitrary $p \times q$ and $q \times p$ matrices chosen so that A_Δ^{-1} exists. Gieser (1993) derived the sufficient conditions under which the sequence of alternatives $H_n : \Delta = \delta/\sqrt{n}$ is contiguous to the null hypothesis $H_0 : \Delta = 0$. For the motivation of using models of type (3.4), see Konijn (1954) and Gieser and Randles (1997).

Hannan (1956) showed that when comparing test statistics that have under H_n limiting noncentral chi-squared distributions, the asymptotic relative efficiencies are obtained as ratios of noncentrality parameters. Gieser and Randles (1997) computed the asymptotic efficiencies of Q_1^2 relative to Wilks' test W assuming that the distributions of $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are elliptically symmetric. Using the exponential power family, they showed that Q_1^2 is more efficient than W when the underlying distributions are heavy-tailed. They also considered the efficiencies of Q_1^2 relative to quadrant test analogue S^J and showed that Q_1^2 performs better when $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are spherically symmetric. In Gieser (1993), the comparisons are also made in t distribution case.

In Articles B and C, the efficiency comparisons of Q_2^2, τ_2^2 and ρ_2^2 relative to Wilks' test are made using the model given in (3.4) and assuming that the underlying distributions are elliptically symmetric. The efficiencies are computed in the multivariate normal distribution, t distribution and contaminated normal distribution cases. In the multivariate normal case, the efficiency of Q_2^2 relative to W is low but as the underlying distribution becomes very heavy-tailed, Q_2^2 outperforms W. Kendall and Spearman analogues τ_2^2 and ρ_2^2 appear to be equally efficient. Their asymptotic efficiencies relative to W are very high even in the multinormal case. In Article D, the asymptotic efficiencies of different rank scores tests are considered. The Wilcoxon scores test seems to be highly efficient when $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are low-dimensional vectors. In the multinormal case, the van der Waerden scores test is as efficient as W and in the considered heavy-tailed cases, more efficient than other rank scores tests. The asymptotic efficiency of van der Waerden scores test relative to W is also considered in Paindaveine (2003). He showed that under elliptical distributions, the efficiency is always larger or equal to 1.

As an example, we consider in the following the efficiencies of Q_2^2, ρ_2^2 and van der Waerden scores test, denoted by T_2^2, in different multivariate normal and contaminated normal distribution cases. In Table 3.1, the efficiencies of Q_2^2 and ρ_2^2 are listed in the case when the distributions of $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are p- and q-variate normal. Note that $\mathrm{ARE}(T_2^2, W) = 1$ for all dimensions. In Table 3.2, the efficiencies of Q_2^2 and ρ_2^2 are listed in the case of contaminated normal distributions with $\epsilon = 0.1$ and $c = 3$ and 6. Again, the efficiencies of van der Waerden test T_2^2 do not depend on the dimensions at all. For $c = 3$, $\mathrm{ARE}(T_2^2, W) = 1.254$ and for $c = 6$, $\mathrm{ARE}(T_2^2, W) = 1.891$. Thus, according to asymptotic efficiencies, the van der Waerden scores test seems to be the best one for light-tailed distributions and in heavy-tailed cases, Spearman analogue performs better than the other tests.

Table 3.1: $\mathrm{ARE}(Q_2^2, W)$ and $\mathrm{ARE}(\rho_2^2, W)$ (between parentheses) at different p- and q-variate normal distributions.

	p				
q	2	3	5	8	10
2	0.617	0.667	0.711	0.738	0.747
	(0.934)	(0.941)	(0.948)	(0.954)	(0.956)
3		0.721	0.769	0.798	0.807
		(0.948)	(0.955)	(0.961)	(0.963)
5			0.820	0.851	0.861
			(0.963)	(0.969)	(0.971)
8				0.883	0.894
				(0.974)	(0.976)
10					0.905
					(0.979)

3.2.6 Finite-sample Efficiencies

Besides comparing asymptotic relative efficiencies, it may be of interest to make efficiency comparisons for small sample sizes. Gieser and Randles (1997) compared the finite-sample powers of Q_1^1, S^J and W under exponential power family by a simple simulation study. In Articles B, C and D, the finite-sample powers of tests based on standardized signs and ranks are compared in the multivariate normal distribution, t distribution and contaminated normal distribution cases.

Table 3.2: ARE(Q_2^2, W) and ARE(ρ_2^2, W) (between parentheses) at different p- and q-variate contaminated normal distributions for $\epsilon = 0.1$ and for selected values of c.

		p				
	q	2	3	5	8	10
	2	0.773	0.836	0.892	0.926	0.937
		(1.174)	(1.188)	(1.203)	(1.213)	(1.217)
	3		0.904	0.964	1.000	1.013
			(1.202)	(1.217)	(1.228)	(1.232)
$c = 3$	5			1.028	1.067	1.080
				(1.233)	(1.243)	(1.247)
	8				1.107	1.121
					(1.254)	(1.258)
	10					1.135
						(1.262)
	2	1.166	1.260	1.344	1.395	1.413
		(1.917)	(1.949)	(1.981)	(2.002)	(2.010)
	3		1.362	1.453	1.508	1.527
			(1.982)	(2.014)	(2.035)	(2.043)
$c = 6$	5			1.550	1.608	1.628
				(2.047)	(2.069)	(2.076)
	8				1.669	1.690
					(2.091)	(2.099)
	10					1.711
						(2.106)

In the following, we compare the empirical powers of W, Q_2^2, ρ_2^2 and T_2^2 using samples of sizes $n = 30$ and 60 from a multivariate normal distribution and from a contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$. Independent $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ samples were generated from the considered distributions and the transformation in (3.4) with $M_1 = M_2^T = I_p$ was used for chosen values of δ to introduce dependence into the model. The test statistics were then computed and corresponding p-values were obtained using chi-square approximations to the null distributions. The process was replicated 1500 times. Empirical powers using the multivariate normal distribution are illustrated in Figure 3.1 and using the contaminated normal distribution in Figure 3.2.

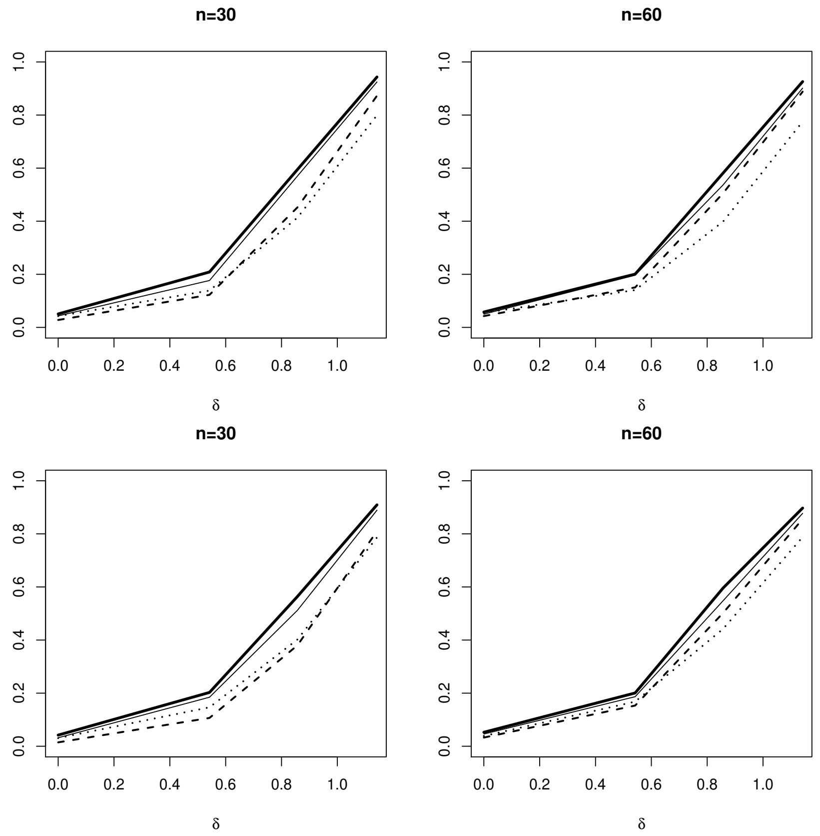

Figure 3.1: Empirical powers for $p = q = 3$ (first row) and $p = q = 5$ (second row) using the multivariate normal distribution The thick solid line denotes W, the thin solid line ρ_2^2, the thick dotted line T_2^2 and the thin dotted line Q_2^2.

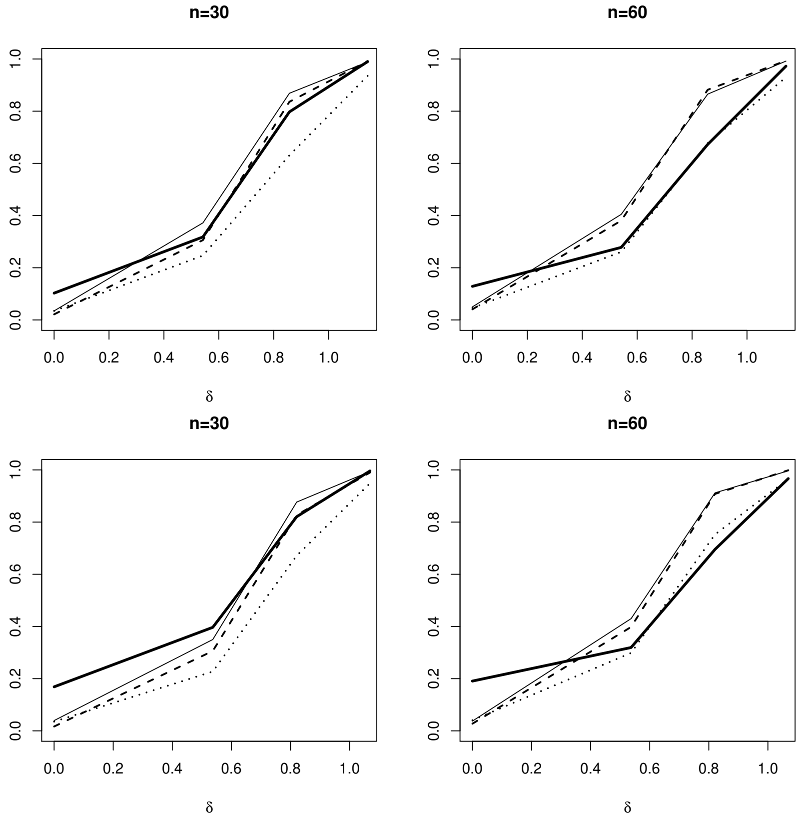

Figure 3.2: Empirical powers for $p = q = 3$ (first row) and $p = q = 5$ (second row) using the contaminated normal distribution with $\epsilon = 0.1$ and $c = 6$. The thick solid line denotes W, the thin solid line ρ_2^2, the thick dotted line T_2^2 and the thin dotted line Q_2^2.

The simulation results show that the finite-sample performance of ρ_2^2 is very good. In the multinormal case, Wilks' test is the most efficient one, but ρ_2^2 is very competitive with it. According to asymptotic results, T_2^2 should be equally powerful with Wilks' test, but especially for $n = 30$, it performs poorly. The empirical powers of Q_2^2 are low in all considered cases, as was expected. The sizes of tests are very close to the designated size 0.05. In the contaminated normal distribution case, the size of W varies widely above 0.05. It is therefore difficult to compare W with other tests. For $n = 30$, ρ_2^2 seems to be slightly more powerful than T_2^2, but as n increases, no significant differences can be seen between tests. The empirical powers of Q_2^2 are again lower than those of ρ_2^2 and T_2^2.

3.3 Canonical Correlation Analysis

3.3.1 Robust Methods in Canonical Analysis

In Section 2.3.3, sample covariance matrix was used in estimating the canonical correlations and vectors. A natural way to robustify canonical correlation analysis is to use robust scatter or shape matrix in estimation. Campbell (1982) used M-estimators in canonical analysis by estimating scatter matrices of each group separately. Karnel (1991) estimated correlations and vectors using partitioned M-estimator instead of sample covariance matrix. He also studied the robustness properties of his procedure using empirical influence function of canonical correlation. Recently, Croux and Dehon (2002) used robust scatter matrices in canonical analysis. They gave expressions for influence functions of canonical correlations and vectors based on any affine equivariant scatter matrix and studied the MCD-based methods in more detail. In Article E, robust canonical analysis based on any affine equivariant scatter or shape matrices is considered. Influence functions and the limiting distributions of correlations and vectors are derived under elliptical distribution. Several different scatter matrices are compared.

Alternative ways to robustify canonical analysis include projection pursuit approach and robust alternating regression. See Oliveira and Branco (2000) and Dehon et al. (2000), for example.

3.3.2 Canonical Analysis Based on Robust Scatter Matrices

Assume now that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is a random sample from a k-variate elliptically symmetric distribution F and that each $\boldsymbol{x}_i$ is partitioned into p- and q-dimensional subvectors $\boldsymbol{x}_i^{(1)}$ and $\boldsymbol{x}_i^{(2)}$, respectively. If C is any partitioned affine equivariant scatter matrix estimator, then the canonical correlations $R = \mathrm{diag}(\rho_1, \ldots, \rho_p)$ and vectors $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$ and $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_q)$ can be defined implicitely as in Section 2.3.2 as a solution of

$$\begin{pmatrix} A^T & 0 \\ 0 & B^T \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} I_p & (R, 0) \\ (R, 0)^T & I_q \end{pmatrix}.$$

Note that the canonical correlations based on different scatter matrices estimate the same population quantities and are directly comparable, but for canonical vectors, a correction factor is needed. As shown in Article E, in estimating canonical correlations, a shape matrix estimator V can be used instead of C. Again, the same canonical correlations are obtained, but the vectors are unique up to a constant.

Write next $\widehat{R} = \text{diag}(\widehat{r}_1, \ldots, \widehat{r}_p)$, $\widehat{A} = (\widehat{\boldsymbol{a}}_1, \ldots, \widehat{\boldsymbol{a}}_p)$ and $\widehat{B} = (\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_q)$ for canonical correlation and vector estimates based on scatter matrix $\widehat{C}$. The limiting distributions of canonical correlations based on sample covariance matrix at elliptical F are derived in Muirhead and Waternaux (1980) and Eaton and Tyler (1994). The first authors assumed that all population canonical correlations are distinct. The latter authors derived the distribution in more gereral case and state that their result is valid when any affine equivariant scatter matrix is used. In Croux and Dehon (2002), the variances and covariances of canonical correlations and vectors are estimated using influence functions.

In Article E, the limiting distributions of correlations and vectors are found under elliptical distributions. If $\rho_1 > \cdots > \rho_p > 0$ and the limiting distributions of $\sqrt{n}\text{vec}(\widehat{C} - C)$ is multivariate normal with zero mean and covariance matrix $\text{ASV}(\widehat{C}; F)$, then the marginal distributions of $\sqrt{n}(\widehat{r}_i - \rho_i)$, $\sqrt{n}(\widehat{\boldsymbol{a}}_i - \boldsymbol{a}_i)$ and $\sqrt{n}(\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i)$ are asymptotically normal with zero mean and asyptotic variances

$$\text{ASV}(\widehat{r}_i; F) = (1 - \rho_i^2)^2 \text{ASV}(\widehat{C}_{12}; F_0),$$

$$\begin{aligned}\text{ASV}(\widehat{\boldsymbol{a}}_i; F) = &\frac{1}{4}\text{ASV}(\widehat{C}_{11}; F_0)\boldsymbol{a}_i\boldsymbol{a}_i^T \\ &+ \text{ASV}(\widehat{C}_{12}; F_0)\sum_{\substack{j=1 \\ j\neq i}}^{p} \frac{(\rho_j^2 + \rho_i^2 - 2\rho_j^2\rho_i^2)(1-\rho_i^2)}{(\rho_i^2 - \rho_j^2)^2}\boldsymbol{a}_j\boldsymbol{a}_j^T\end{aligned}$$

and

$$\begin{aligned}\text{ASV}(\widehat{\boldsymbol{b}}_i; F) = &\frac{1}{4}\text{ASV}(\widehat{C}_{11}; F_0)\boldsymbol{b}_i\boldsymbol{b}_i^T \\ &+ \text{ASV}(\widehat{C}_{12}; F_0)\sum_{\substack{j=1 \\ j\neq i}}^{q} \frac{(\rho_j^2 + \rho_i^2 - 2\rho_j^2\rho_i^2)(1-\rho_i^2)}{(\rho_i^2 - \rho_j^2)^2}\boldsymbol{b}_j\boldsymbol{b}_j^T,\end{aligned}$$

where $\rho_j = 0$ for $j > p$, and the asymptotic variances of diagonal and off-diagonal elements of $\widehat{C}$ at spherical distribution F_0 corresponding to F are given in (3.3) and (3.2). Note that if $\widehat{C}$ is the sample covariance matrix, then at normal distribution $\text{ASV}(\widehat{C}_{11}; F_0) = 2$ and $\text{ASV}(\widehat{C}_{12}; F_0) = 1$. Thus the asymptotic variances given by Anderson (1999) are obtained.

3.3.3 Asymptotic Relative Efficiencies

Having general formulas for asymptotic variances and covariances of canonical correlations and vectors allows us to compare the estimates based on different scatter and shape matrices by means of asymptotic relative efficiencies. Note that, in the case of estimators, asymptotic relative efficiency describes the accuracy of an estimator as compared to the other estimator and is obtained as ratio of asymptotic variances. Write now $\widehat{r}_{i,C}$ and $\widehat{\boldsymbol{a}}_{i,C}$ for the canonical correlation and vector estimates based on $\widehat{C}$. As shown in Article E, at elliptical distribution F, the efficiencies of estimates based on $\widehat{C}$ relative to those based on $\widehat{C}'$ are given by the following ratios

$$\mathrm{ARE}(\widehat{a}_{ii,C}, \widehat{a}_{ii,C'}) = \frac{\mathrm{ASV}(\widehat{C}_{11}; F_0)}{\mathrm{ASV}(\widehat{C}'_{11}; F_0)}$$

and

$$\mathrm{ARE}(\widehat{a}_{ij,C}, \widehat{a}_{ij,C'}) = \mathrm{ARE}(\widehat{r}_{i,C}, \widehat{r}_{i,C'}) = \frac{\mathrm{ASV}(\widehat{C}_{12}; F_0)}{\mathrm{ASV}(\widehat{C}'_{12}; F_0)},$$

where F_0 denotes the spherical distribution corresponding to F.

Table 3.3 lists the efficiencies of estimates based on some robust scatter matrices $\widehat{C}$ relative to those based on the sample covariance matrix $\widehat{C}'$ at different multivariate normal distributions. The scatter matrices considered are Huber's M-estimator with $q = 0.9$, 25% breakdown biweight S-estimator and 25% breakdown MCD-estimator. See Section 3.1.4, for corresponding definitions. Moreover, the scatter matrix estimator (SCM) based on affine equivariant sign covariance matrix was included in comparisons (Ollila et al., 2003a).

In the considered multivariate normal cases, the estimates based on sample covariance matrix are only slightly more efficient than those based on the sign covariance matrix. The estimates based on Huber's M-estimator and biweight S-estimator are also highly efficient, but MCD-based estimates perform poorly especially in low dimensions.

3.3.4 Tests for Canonical Correlations

Muirhead and Waternaux (1980) considered the likelihood ratio test (2.2) for testing the null hypothesis $H_0' : \ \rho_{k+1} = \cdots = \rho_p = 0$. They showed that at elliptical F with finite fourth moments, $-n \log W \to_d (1+\kappa)\chi^2_{(p-k)(q-k)}$, where κ denotes the kurtosis of the underlying distribution. Thus a robust test for canonical correlations is obtained by dividing $-n \log W$ by a consistent estimate of $(1+\kappa)$.

Table 3.3: AREs of canonical correlation and vector estimates based on robust scatter matrices relative to the estimates based on sample covariance matrix at different $(p+q)$-variate normal distributions. The efficiencies $\text{ARE}(\widehat{a}_{ii,C}, \widehat{a}_{ii,C'})$ are listed in left column and $\text{ARE}(\widehat{a}_{ij,C}, \widehat{a}_{ij,C'}) = \text{ARE}(\widehat{r}_{i,C}, \widehat{r}_{i,C'})$ in right column.

$p=q$	SCM	S	M	MCD	SCM	S	M	MCD
2	0.974	0.960	0.942	0.336	0.982	0.953	0.961	0.284
3	0.986	0.978	0.963	0.391	0.991	0.975	0.975	0.356
5	0.994	0.989	0.979	0.459	0.988	0.996	0.986	0.438
8	0.998	0.994	0.988	0.515	0.998	0.993	0.992	0.502
10	0.998	0.991	0.991	0.538	0.999	0.995	0.994	0.529

Since the canonical correlations based on different scatter or shape matrices estimate the same population quantities, any scatter or shape matrix can be used in testing the null hypothesis $H_0' : \ \rho_1 = \cdots = \rho_p = 0$. In Article B, a test statistic analogous to Pillai's test statistic (2.3) is derived. If a shape matrix estimator is partitioned as

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

then the test statistic is based on

$$P' = Tr(V_{11}^{-1} V_{12} V_{22}^{-1} V_{21}).$$

The tests are naturally affine invariant under the group $\mathcal{G}$ in (2.4), and have under H_0' a limiting chi-squared distribution with pq degrees of freedom. Note also that in the multinormal case, the tests based on shape (or scatter) matrices may be used in testing independence.

Chapter 4

Summary of Original Publications

In Article A, the applicability of marginal, spatial and affine equivariant sign and rank concepts in constructing bivariate tests of independence is studied. The influence functions of the statistics are given for robustness considerations. Limiting distributions are derived under the null hypothesis as well as under interesting sequences of alternatives. Asymptotic relative efficiencies with respect to the classical correlation test are computed in bivariate normal and t distribution cases. The theory is illustrated by an example.

In Article B, a new affine invariant extension of Blomqvist's quadrant test based on standardized spatial signs is proposed. The limiting distributions are derived in the elliptic case and the asymptotic relative efficiencies with respect to the classical Wilks' test are computed in the multivariate normal, t distribution and contaminated normal distribution cases. Simulations are used to compare finite-sample powers and an example is used to illustrate the robustness of test.

In Article C, new affine invariant extensions of Kendall's tau and Spearman's rho are introduced. In this article, we focus on tests based on standardized spatial signs and ranks, but also their interdirection counterparts are considered. Asymptotic theory is developed and the tests are compared using asymptotic and finite-sample efficiencies. The theory is illustrated by an example.

In Article D, new rank scores tests for testing multivariate independence are proposed. The tests are constructed by combining standardized spatial signs and univariate ranks of the Mahalanobis-type distances of observations from the origin. Three different choices of score functions are discussed in more detail. The limiting distributions are derived and asymptotic and finite-sample efficiencies are compared. Also the robustness properties of tests are studied.

In Article E, robust canonical correlation analysis is considered. The influence functions as well as limiting variances and covariances of canonical correlations and vectors based on affine equivariant scatter matrices are derived in the elliptic case. Also the shape matrix based canonical analysis is considered. Limiting and finite-sample efficiencies of estimators based on different scatter matrices are compared through theoretical and simulation studies. The theory is illustrated by an example.

Bibliography

[1] Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, (2nd ed.), Wiley, New York.

[2] Anderson, T.W. (1999), "Asymptotic theory for canonical correlation analysis", *Journal of Multivariate Analysis*, **70**, 1–29.

[3] Bhuchongkul, S. (1964), "A class of nonparametric tests for independence in bivariate populations", *Annals of Mathematical Statistics*, **35**, 138–149.

[4] Blomqvist, N. (1950), "On a measure of dependence between two random variables", *Annals of Mathematical Statistics*, **21**, 593–600.

[5] Brown, B.M. (1983), "Statistical uses of the spatial median", *J*ournal of the Royal Statistical Society, Series B, **45**, 25–30.

[6] Butler, R.W., Davies, P.L. and Jhun, M. (1993), "Asymptotics of the minimum covariance determinant estimator", *The Annals of Statistics*, **21**, 1385–1400.

[7] Campbell, N.A. (1982), "Robust procedures in multivariate analysis II. Robust canonical variate analysis", *Applied Statistics*, **31**, 1–8.

[8] Chakraborty, B., Chaudhuri, P. and Oja, H. (1998), "Operating transformation retransformation on spatial median and angle test", *Statistica Sinica*, **8**, 767–784.

[9] Croux, C. and Haesbroeck, G. (1999), "Influence function and efficiency of the minimum covariance determinant scatter matrix estimator", *Journal of Multivariate Analysis*, **71**, 161–190.

[10] Croux, C. and Haesbroeck, G. (2000), "Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies", *Biometrika*, **87**, 603–618.

[11] Croux, C. and Dehon, C. (2002), "Analyse canonique basée sur des estimateurs robustes de la matrice de covariance", *La Revue de Statistique Appliquée*, **2**, 5–26.

[12] Croux, C., Ollila, E., and Oja, H. (2002), "Sign and rank covariance matrices: statistical properties and application to principal components analysis", in: *Statistical Data Analysis Based on the L1-Norm and Related Methods* (ed. Y. Dodge), pp. 257–271, Birkhauser, Basel.

[13] Davies, P.L. (1987), "Asymptotic behavior of S-Estimates of multivariate location parameters and dispersion matrices", *The Annals of Statistics*, **15**, 1269-1292.

[14] Davies, P.L. (1992), "The asymptotics of Rousseeuw's minimum volume ellipsoid", *The Annals of Statistics*, **20**, 1828–1843.

[15] Dehon, C., Filzmoser, P. and Croux, C. (2000), "Robust methods for canonical correlation analysis", in: *Data Analysis, Classification, and Related Methods* (eds. H.A.L. Kiers, J.P. Rasson, P.J.F. Groenen and M. Schrader), pp. 321–326, Springer-Verlag, Berlin.

[16] Dümbgen, L. (1998), "On Tyler's M-functional of scatter in high dimension", *Annals of the Institute of Statistical Mathematics*, **50**, 471–491.

[17] Eaton, M.L. and Tyler, D. (1994), "The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis", *Journal of Multivariate Analysis*, **34**, 439–446.

[18] Farlie, D.J.G. (1960), "The performance of some correlation coefficients for a general bivariate distributions", *Biometrika*, **47**, 307–323.

[19] Fujikoshi, Y. (1974), "The likelihood ratio tests for the dimensionality of regression coefficients", *Journal of Multivariate Analysis*, **4**, 327–340.

[20] Gieser, P.W. (1993), "A new nonparametric test for independence between two sets of variates", unpublished Ph.D. thesis, University of Florida, Gainesville.

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

[22] Hájek, J. and Šidák, Z. (1967), *Theory of Rank Tests*, Academic Press, New York.

[23] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.J. (1986), *Robust Statistics: the Approach Based on Influence Functions*, Wiley, New York.

[24] Hannan, E.J. (1956), "The asymptotic powers of certain tests based on multiple correlations", *Journal of the Royal Statistical Society, Series B*, **18**, 227–233.

[25] Hettmansperger, T.P. and McKean, J.W. (1998), *Robust Nonparametric Statistical Methods*, Arnold, London.

[26] Hettmansperger, T.P. and Randles, R.H. (2002), "A practical affine equivariant multivariate median", *Biometrika*, **89**, 851–860.

[27] Hotelling, H. (1935), "The most predictable criterion", *Journal of Educational Psychology*, **26**, 139–142.

[28] Hotelling, H. (1936), "Relations between two sets of variables", *Biometrika*, **28**, 321–377.

[29] Hsu, P.L. (1941), "On the limiting distribution of the canonical correlations", *Biometrika*, **32**, 38–45.

[30] Huber, P.J. (1981), *Robust Statistics*, Wiley, New York.

[31] Karnel, G. (1991), "Robust canonical correlation and correspondence analysis", *The Frontiers of Statistical Scientific Theory and Industrial Applications*, 335–354.

[32] Kendall, M.G. (1938), "A new measurement of rank correlation", *Biometrika*, **30**, 81–93.

[33] Kent, J.T. and Tyler, D.E. (1991), "Redescending M-estimates of multivariate location and scatter", *The Annals of Statistics*, **19**, 2102–2119.

[34] Konijn, H.S. (1954), "On the power of certain tests for independence in bivariate populations", *Annals of Mathematical Statistics*, **25**, 300–323.

[35] Lopuhaä, H.P. (1989), "On the relation between S-estimators and M-estimators of multivariate location and covariance", *The Annals of Statistics*, **17**, 1662–1683.

[36] Lopuhaä, H.P. and Rousseeuw, P.J. (1991), "Breakdown points of affine equivariant estimators of multivariate location and covariance matrices", *The Annals of Statistics*, **19**, 229–248.

[37] Marden, J.I. (1999), "Some robust estimates of principal components", *Statistics and Probability Letters*, **43**, 349–359.

[38] Maronna, R.A. (1976), "Robust M-estimators of multivariate location and scatter", *The Annals of Statistics*, **4**, 51–67.

[39] Muirhead, R.J. and Waternaux, C.M. (1980), "Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations", *Biometrika*, **67**, 31–43.

[40] Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.

[41] Möttönen J. and Oja, H. (1995), "Multivariate spatial sign and rank methods", *Journal of Nonparametric Statistics*, **5**, 201–213.

[42] Oliveira, M.R. and Branco, J.A. (2000), "Projection pursuit approach to robust canonical correlation analysis", in: *Proceedings in Computational Statistics 2000* (eds. J.G. Bethlehem and P.G.M. van der Heijden), pp. 415–420, Physica Verlag, Heidelberg.

[43] Ollila, E., Oja, H. and Croux, C. (2003a), "The affine equivariant sign covariance matrix: Asymptotic behavior and efficiencies", *Journal of Multivariate Analysis*, in print.

[44] Ollila, E., Hettmansperger, T.P. and Oja, H. (2003b), "Affine equivariant multivariate sign methods", Under revision.

[45] Ollila, E., Croux, C., and Oja, H. (2003c), "Influence function and asymptotic efficiency of the affine equivariant rank covariance matrix", Conditionally accepted in *Statistica Sinica.*

[46] Paindaveine, D. (2003), "Elementary derivations of multivariate non-admissibility results", Submitted.

[47] Pearson, K. (1896), "Mathematical contributions to the theory of evolution-III. Regression, heredity and panmixia", *Philosophical Transactions of the Royal Society of London, Series A*, **187**, 253–318.

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

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

[50] Randles, R.H. (1989), "A distribution-free multivariate sign test based on interdirections", *Journal of the American Statistical Association*, **84**, 1045–1050.

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

[52] Rousseeuw, P.J. (1985), "Multivariate estimation with high breakdown point", in: *Mathematical Statistics and Applications*, Vol. B, (eds. W. Grossmann, G. Pflug, I. Vincze, and W. Werty), pp. 283–297, Reider, Dodrecht.

[53] Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, Wiley, New York.

[54] Spearman, C. (1904), "The proof and measurement of association between two things", *American Journal of Psychology*, **15**, 72–101.

[55] Tyler, D.E. (1986), "Breakdown properties of the M-estimators of multivariate scatter", Report, Department of Statistics, Rutgers University.

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

[57] Visuri, S., Oja, H. and Koivunen, V. (2000), "Sign and rank covariane matrices", *Journal of Statistical Planning and Inference*, **91**, 557–575.

[58] Visuri, S., Ollila, E., Koivunen, V., Möttönen, J. and Oja, H. (2003), "Affine equivariant multivariate rank methods", *Journal of Statistical Planning and Inference*, **114**, 161–185.

[59] Wilks, S.S. (1935), "On the independence of k sets of normally distributed statistical variables", *Econometrica*, **3**, 309–326.