JAAKKO NEVALAINEN

Nonparametric Methods for Multivariate Location Problems with Independent and Cluster Correlated Observations

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Information Sciences of the University of Tampere,
for public discussion in the Paavo Koli Auditorium, Kanslerinrinne 1,
Tampere, on January 12th, 2007, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere
Department of Mathematics, Statistics and Philosophy
Finland

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Juha Siro

Printed dissertation
Acta Universitatis Tamperensis 1200
ISBN 978-951-44-6811-7
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 583
ISBN 978-951-44-6812-4
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2007

Acknowledgements

I am grateful to my supervisor Professor Hannu Oja for his expert guidance, infinite patience, and constant encouragement and enthusiasm throughout the preparation of the thesis.

I warmly thank my co-author Professor Denis Larocque for his valuable advice, and for making the visits to Montréal unforgettable. He sets an excellent example of a modern skilled young researcher in the field of statistics. Dr. Jyrki Möttönen also deserves my appreciation for his accurate input in our joint work.

Professor Thomas P. Hettmansperger and Professor Biman Chakraborty are acknowledged for their constructive remarks on the thesis.

My sincere thanks to Dr. Anne Konu and Dr. Tomi Lintonen for providing insight into well-being in Finnish schools.

It has been my great pleasure to co-operate with the dedicated members of the research group *Nonparametric and Robust Multivariate Methods*, both in matters of science and life.

I wish to thank my parents, my sisters and their families, and my friends for support. My wife Minna and our children Lauri and Kaisa have acted as the underlying backbone of this project, and have kept me smiling happily throughout.

The research was funded by the School of Statistical Information, Inference, and Data Analysis (SIIDA) and the Academy of Finland. The facilities were provided by the Department of Mathematics, Statistics and Philosophy, University of Tampere, and the Tampere School of Public Health.

Tampere, December 2006

Jaakko Nevalainen

Abstract

The aim of this doctoral thesis was to develop efficient nonparametric multivariate methods for independent and identically distributed (i.i.d.) observations and for cluster correlated observations. The first part of the thesis and two of the original articles deal with spatial sign and spatial rank methods, and their affine invariant and equivariant extensions, for the one-sample and the several samples multivariate location problem with i.i.d. observations. The second part and the remaining three original articles focus on the one-sample multivariate location problem with clustered data. Spatial sign methods, with their weighted generalizations and affine invariant and equivariant versions, are considered in this framework. The statistical properties (consistency, limiting distributions, limiting and finite sample efficiencies, robustness, computation) of the procedures are carefully investigated. It is shown that the spatial sign and rank methods have a competitive efficiency relative to the classical techniques, particularly if the data is heavy-tailed or clustered. The efficiencies and other statistical properties of the methods can be improved even further by weighting them in an optimal way. Furthermore, the methods are valid even without moment assumptions, and efficient when the underlying distribution deviates from normality or in the presence of outliers. The proposed procedures are easy to implement on statistical programming languages such as R or SAS/IML.

Key words: affine equivariance; affine invariance; clustered data; multivariate location problem; nonparametric methods; spatial sign and rank.

Contents

Abbreviations

$\sim$	distributed as
^{T}	the transpose of a vector or a matrix
ave$(\cdot)$	arithmetic average of $(\cdot)$
ARE	asymptotic relative efficiency
cov$(\cdot)$	variance-covariance matrix of $(\cdot)$
$E(\cdot)$	expected value of $(\cdot)$
$E_0(\cdot)$	expected value of $(\cdot)$ under the null hypothesis
i.i.d.	independent and identically distributed
MANOVA	multivariate analysis of variance
max$(\cdot)$	maximum of $(\cdot)$
min$(\cdot)$	minimum of $(\cdot)$

List of Original Publications

I. Nevalainen, J., Möttönen, J. & Oja, H. (2006), ”A spatial rank test and corresponding estimators for several samples”, *Submitted*.

II. Nevalainen, J. & Oja, H. (2006), ”SAS/IML macros for a multivariate analysis of variance based on spatial signs”, *Journal of Statistical Software*, 16, 1–17.

III. Nevalainen, J., Larocque, D. & Oja, H. (2007), ”On the multivariate spatial median for clustered data”, *The Canadian Journal of Statistics* (in press).

IV. Larocque, D., Nevalainen, J. & Oja, H. (2007), ”A weighted multivariate sign test for cluster correlated data”, *Biometrika* (in press).

V. Nevalainen, J., Larocque, D. & Oja, H. (2006), ”A weighted spatial median for clustered data”, *Statistical Methods & Applications* (in press).

1 Introduction

Classical multivariate statistical techniques for one-sample location problems and for comparing several treatments assume that the data were from multivariate normal (Gaussian) distributions. The optimal inference methods are then based on the sample mean vectors and the sample covariance matrices. These methods are, however, extremely sensitive to outlying observations, and may be inefficient for heavy-tailed noise distributions. They may even lead to unreliable or invalid results when the underlying distribution strongly deviates from the assumed model. Therefore, some robustness of the statistical procedures is necessary to prevent false conclusions drawn from the data; good robust and nonparametric methods are necessary to prevent avoidable losses in efficiency (Huber, 1980; Hampel, Rochetti, Rousseeuw & Stahel, 1986; Hettmansperger & McKean, 1998). Methods based on spatial signs and ranks to some extent possess these ideal properties.

The outline of the thesis is as follows. In the current Chapter, the multivariate notions of sign and rank and corresponding shape matrices are reviewed. These tools will be needed later on. A description of a school well-being data set is given as well. Competing classical, spatial sign and spatial rank methods for the one-sample and the several samples multivariate location problem, are reviewed in Chapters 2 (independent observations) and 3 (cluster correlated observations). As the theory of multivariate sign and rank methods for independent observations is well studied in the literature, the main contributions of three original papers deal with location problems with clustered data. Two further original papers complete gaps in the development of spatial sign and rank methods. The use of the different methods is illustrated by the means of the example data set.

1.1 Spatial Signs and Ranks

Recall the univariate notions of sign and rank. The sign of an observation x_i is

$$S(x_i) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \tag{1.1}$$

The centered rank of x_i among the data set $x_1, \ldots, x_N$ is

$$R_N(x_i) = \frac{1}{N}\sum_{j=1}^{N} S(x_i - x_j). \tag{1.2}$$

Finally, the signed-rank is the centered rank among the observations $x_1, \ldots, x_N$ and their reflections $-x_1, \ldots, -x_N$:

$$(1.3) \qquad Q_N(x_i) = \frac{1}{2N} \sum_{j=1}^{N} [S(x_i - x_j) + S(x_i + x_j)]$$

The concepts of sign and rank form the basis for univariate nonparametric statistics such as the sign test, the Wilcoxon signed-rank test and the rank-sum test, the median and the Hodges-Lehmann estimator, among others (Lehmann, 1975; Hettmansperger & McKean, 1998; Hollander & Wolfe, 1999, for example).

Several generalizations of the univariate signs and ranks to the multivariate setting have been considered in the literature. *Marginal signs and ranks* are the most obvious ones being just the vectors of the componentwise signs and ranks (Puri & Sen, 1971). Oja (1999) has reviewed methods based on *Oja signs and ranks*. *Interdirections* are yet another concept related to multivariate signs (Randles, 1989), which together with a ranking of the magnitudes of the observations can be employed to generalize signed-rank tests (Peters & Randles, 1990; Hallin & Paindaveine, 2002; Oja & Paindaveine, 2005). Different *data depth* functions provide another way of ranking data points, which can also be used for constructing multivariate nonparametric techniques (Liu, Parelius & Singh, 1999; Zuo & Serfling, 2000). Barnett (1976) and Hettmansperger, Nyblom & Oja (1992) have discussed in detail the difficulties of ordering multivariate data, and have given throughout reviews of the multivariate extensions of sign and rank. Small (1990) and Niinimaa & Oja (1999) have reviewed of the multivariate medians related to the different concepts. From this point on, the terms "sign" and "rank" are used to refer to the spatial signs and the spatial ranks defined in the sequel.

Let $\mathbf{x}_i$ be a vector in $\mathbb{R}^p$. One of the most natural analogues of the univariate sign is the *spatial sign* defined by

$$(1.4) \qquad \mathbf{S}(\mathbf{x}_i) = \begin{cases} \|\mathbf{x}_i\|^{-1}\mathbf{x}_i & \text{if } \mathbf{x}_i \neq \mathbf{0}, \\ \mathbf{0} & \text{if } \mathbf{x}_i = \mathbf{0}, \end{cases}$$

where $\|\mathbf{x}_i\| = \left(\mathbf{x}_i^T \mathbf{x}_i\right)^{1/2}$ is the Euclidian norm. The spatial sign of $\mathbf{x}_i$ is thus a unit vector in $\mathbb{R}^p$ pointing into the direction of $\mathbf{x}_i$.

As in the univariate case, the empirical *spatial centered rank* among the data set $\mathbf{x}_1, \ldots, \mathbf{x}_N$ is defined by

$$(1.5) \qquad \mathbf{R}_N(\mathbf{x}_i) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{S}(\mathbf{x}_i - \mathbf{x}_j).$$

The spatial rank is a vector inside a unit sphere of $\mathbb{R}^p$ pointing from the center of the data cloud approximately into the direction of $\mathbf{x}_i$, retaining both direction and magnitude. The rank function is data dependent, but under general conditions the empirical rank converges uniformly in probability to the theoretical rank function $\mathbf{R}(\mathbf{x}_i) = E\left[\mathbf{S}(\mathbf{x}_i - \mathbf{x})\right]$, where $\mathbf{x} \sim F$ (Möttönen, Oja & Tienari, 1997). Here F is the underlying distribution of the $\mathbf{x}_i$'s. The *spatial signed rank* is just

$$\mathbf{Q}_N(\mathbf{x}_i) \quad = \quad \frac{1}{2N}\sum_{j=1}^{N}\left[\,\mathbf{S}(\mathbf{x}_i-\mathbf{x}_j)+\,\mathbf{S}(\mathbf{x}_i+\mathbf{x}_j)\right]. \tag{1.6}$$

The empirical signed-rank function also converges to its theoretical counterpart $\mathbf{Q}(\mathbf{x}_i)=\frac{1}{2}E\left[\,\mathbf{S}(\mathbf{x}_i-\mathbf{x})+\,\mathbf{S}(\mathbf{x}_i+\mathbf{x})\right]$, $\mathbf{x}\sim F$.

The generalized notions of sign and rank allow for a class of extensions of the univariate statistical concepts and procedures to the multivariate setup. Indeed, the multivariate spatial median (Brown, 1983; Gower, 1974), multivariate spatial quantiles (Chaudhuri, 1996; Koltchinskii, 1997; Chakraborty, 2001, 2003) and other descriptive measures based on them (Serfling, 2004), as well as data depth (Vardi & Zhang, 2000; Serfling, 2002; Gao, 2003), have been defined and studied in the literature. Furthermore, spatial signs and ranks can also be used to construct multivariate nonparametric location tests and estimators (Chaudhuri, 1992; Möttönen & Oja, 1995; Marden, 1999a, for example), multivariate nonparametric tests of independence (Taskinen, 2003), linear models (Bai, Chen, Miao & Rao, 1990; Chakraborty, 2003), and so on.

1.2 Equivariance and Invariance

The next two definitions summarize some important properties for multivariate estimators. Here $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_N)$ is the $p\times N$ data matrix and $\mathbf{1}_N$ is an N-vector of ones.

Definition 1. A vector valued statistic $\widehat{\boldsymbol{\mu}}(\mathbf{X})$ is a *location statistic* if it is *affine equivariant* in the sense that

$$\widehat{\boldsymbol{\mu}}\left(\mathbf{H}\mathbf{X}+\mathbf{g}\mathbf{1}_N^T\right) \quad = \quad \mathbf{H}\widehat{\boldsymbol{\mu}}(\mathbf{X})+\mathbf{g} \tag{1.7}$$

for all nonsingular $p\times p$ matrices $\mathbf{H}$ and for all p-vectors $\mathbf{g}$.

Definition 2. A matrix valued statistic $\widehat{\boldsymbol{\Sigma}}(\mathbf{X})$ is a *scatter statistic* if it is a positive definite, symmetric, $p\times p$ (PDS(p)) and affine equivariant in the sense that

$$\widehat{\boldsymbol{\Sigma}}\left(\mathbf{H}\mathbf{X}+\mathbf{g}\mathbf{1}_N^T\right) \quad = \quad \mathbf{H}\widehat{\boldsymbol{\Sigma}}(\mathbf{X})\mathbf{H}^T \tag{1.8}$$

for all nonsingular $p\times p$ matrices $\mathbf{H}$ and for all p-vectors $\mathbf{g}$.

Under affine transformations, location and scatter statistics adopt to the new coordinate system in a desirable way. A related concept is the *affine invariance* of a statistic, which means that the statistic remains unchanged under the transformation. This will be discussed in the next Chapter.

An orthogonal transformation is obtained if $\mathbf{H}$ is an orthogonal matrix and $\mathbf{g}$ is a vector of zeros. A location or a scatter statistic satisfying (1.7) or (1.8) under such a data transformation is *orthogonally equivariant*. Other interesting special cases are scale transformations ($\mathbf{H}$ is a diagonal matrix and $\mathbf{g}$ is a vector of zeros) and location shifts ($\mathbf{H}$ is an identity matrix). Equivariant statistics under these transformations are said to be *scaling equivariant* and *location equivariant*, respectively. (Hettmansperger & McKean, 1998).

Example 1. *Spatial signs and ranks are orthogonally equivariant but not scaling or location equivariant. Spatial ranks are location invariant.*

1.3 Scatter and Shape Matrices

If a matrix valued estimator $\widehat{\boldsymbol{\Sigma}}(\mathbf{X})$ is PDS(p) and affine equivariant, it is an estimator of *scatter*. The sample covariance matrix is the best known estimator of scatter. However, if a matrix valued estimator $\widehat{\mathbf{V}}(\mathbf{X})$ is PDS(p) and affine equivariant only in that $\widehat{\mathbf{V}}(\mathbf{HX} + \mathbf{g}\mathbf{1}_N^T) \propto \mathbf{H}\widehat{\mathbf{V}}(\mathbf{X})\mathbf{H}^T$, it is an estimator of *shape*. A shape matrix captures information of the shape and orientation of the point cloud, but loses the scale information. However, an estimator of shape often suffices in practice. Shape matrices are particularly useful in the construction of affine invariant spatial sign and rank tests, and corresponding affine equivariant estimators of location.

Shape estimators based on spatial signs and ranks have been studied by Marden (1999b), Visuri, Koivunen & Oja (2000), Croux, Ollila & Oja (2002) and Ollila (2002). These estimators are not affine equivariant, however. An example of an affine equivariant shape matrix estimator is *Tyler's shape matrix* (Tyler, 1987).

Definition 3. Tyler's shape matrix $\mathbf{V}$ with respect to $\boldsymbol{\mu}$ solves the implicit equation

$$E\left[\frac{(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T}{(\mathbf{x}_i - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})}\right] = \frac{1}{p}\mathbf{V}. \tag{1.9}$$

Tyler's matrix is unique (up to a multiplication by a constant) if $N > p(p-1)$. A standardization of a data set by $\widehat{\mathbf{V}}^{-1/2}$ from (1.9) results into signs that appear as if they were from a spherical distribution. The breakdown point of Tyler's shape matrix is $\frac{1}{p}$ (Dümbgen & Tyler, 2005). It is fast and easy to compute. A symmetrized version of Tyler's shape matrix is *Dümbgen's shape matrix* (Dümbgen, 1998).

Definition 4. Dümbgen's shape matrix $\mathbf{V}$ solves the implicit equation

$$E\left[\frac{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{V}^{-1}(\mathbf{x}_i - \mathbf{x}_j)}\right] = \frac{1}{p}\mathbf{V}, \tag{1.10}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are i.i.d.

In the case of elliptical symmetry, the symmetrized version (1.10) estimates the same population quantity as Tyler's matrix (1.9) but its statistical properties are different; e.g. the breakdown point is $1 - \sqrt{1 - \frac{1}{p}}$, which lies in the interval $\left(\frac{1}{2p}, \frac{1}{p}\right)$ (Dümbgen & Tyler, 2005). The main advantage of Dümbgen's matrix over Tyler's matrix is that one does not need a fixed location vector to compute it.

An affine equivariant estimator of scatter based on spatial ranks can be constructed similarly. Write $\mathbf{y}_i = \mathbf{V}^{-1/2}\mathbf{x}_i$. A shape matrix based on signed-ranks may be implicitly defined by

$$\frac{E\left[\mathbf{Q}_N(\mathbf{y}_i)\mathbf{Q}_N(\mathbf{y}_i)^T\right]}{E\left[\mathbf{Q}_N(\mathbf{y}_i)^T\mathbf{Q}_N(\mathbf{y}_i)\right]} = \frac{1}{p}\mathbf{I}_p. \tag{1.11}$$

Oja & Randles (2004) proposed a shape matrix (based on centered ranks) defined via

$$\frac{E\left[\mathbf{R}_N(\mathbf{y}_i)\mathbf{R}_N(\mathbf{y}_i)^T\right]}{E\left[\mathbf{R}_N(\mathbf{y}_i)^T\mathbf{R}_N(\mathbf{y}_i)\right]} = \frac{1}{p}\mathbf{I}_p. \tag{1.12}$$

They also gave an algorithm for its computation but did not prove that it converges. After standardizing by the matrix obtained from (1.11) or (1.12), the signed-ranks or the centered ranks appear as if they were from a spherical distribution inside the unit sphere.

The estimators may in fact be seen as modifications of M-estimators of scatter defined originally by Maronna (1976). Sirkiä, Taskinen & Oja (2006) consider a general approach for symmetrizing M-estimators of scatter. There are many other classes of scatter estimators, with higher breakdown points, like the minimum volume ellipsoid (Rousseeuw, 1985) and the S-estimators (Davies, 1987; Lopuhaä, 1989). It seems, however, the most attractive to use (1.9)–(1.12) together with the spatial sign and rank tests and the corresponding estimators.

1.4 The School Well-Being Data Set

Data set on school well-being illustrates the use of the proposed analysis methods in Chapters 2 and 3.

Konu & Lintonen (2006) collected data on well-being in Finnish schools using an Internet-based tool (Lintonen & Konu, 2006). As a part of a larger study, the pupils in grades 7-9 of the participating schools were asked to fill out a questionnaire on the Internet (`www2.edu.fi/hyvinvointiprofiili`). The pupils were to answer 81 questions on school well-being. The questions were grouped into assessments of school conditions (26 questions), social relationships (19), means for self-fulfillment at school (24) and health status (12). For each question, an ordinal five-point (fully agree, agree, neither agree nor disagree, disagree, fully disagree) answering scale was used. The School Well-Being Profile (Konu & Lintonen, 2006) consists of the means of each group of questions.

In the present example analysis, the pupils' perceptions of their social relationships and of their means for self-fulfillment, and their evolution during the lower secondary school (the Finnish lower secondary school lasts from the 7th grade until the 9th grade), are of interest. Previous studies have

Table 1.1: Number of matched pairs in the example data set for the one sample problem broken down by school.

School ID#	Number of matched pairs	
45	57	(21.9%)
88	5	(1.9%)
171	65	(25.0%)
183	19	(7.3%)
197	18	(6.9%)
224	4	(1.5%)
250	61	(23.5%)
260	31	(11.9%)
Total	260	(100.0%)

shown that the two variables are strongly correlated. Furthermore, they are associated with gender, and related to pupil's health status (health questions are mainly on psychosomatic symptoms).

One sample location problem. The pupils of grades 7 and 9 from the same school were matched for gender and health status. The bivariate differences of the matched pairs were taken as the outcome variables. The question of interest is whether the pupils' well-being changed during the lower secondary school. For purposes of illustration, a subset of eight schools was taken, resulting into a total of $N = 260$ observations for the one-sample analysis. Table 1.1 displays the frequency distribution of the pupils in the schools. The data set was first analyzed by assuming (falsely) that the observations were independent and identically distributed, and secondly, by taking the clustered structure (school memberships of the pupils) into account.

Several samples location problem. Grades 7, 8 and 9 were compared directly as three independent samples. For this reason, only one school was selected for the analysis. The question of interest is whether the three samples have a common location center. The group sizes for the analysis were $n_7 = 58$, $n_8 = 51$ and $n_9 = 52$, where the index refers to the grade.

2 Independent Observations

In this Chapter, the spatial sign and rank methods for one-sample and several samples location problems with i.i.d. observations are reviewed and compared with their classical analogues.

2.1 One Sample

Let $\mathbf{X}_{p\times N} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ be a random sample of i.i.d. observations from an unknown continuous p-variate distribution F symmetric around $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$. By a symmetry of a distribution around $\boldsymbol{\mu}$ it is meant that $\mathbf{x}_i - \boldsymbol{\mu} \sim \boldsymbol{\mu} - \mathbf{x}_i$. Symmetry in turn implies that any location statistic defined as in (1.7) estimates $\boldsymbol{\mu}$. Consider the hypotheses

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \text{ vs. } H_1 : \boldsymbol{\mu} \neq \mathbf{0} \tag{2.1}$$

without loss of generality, and the estimation of the location parameter $\boldsymbol{\mu}$.

The classical test and the estimator. Write

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i \tag{2.2}$$

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{2.3}$$

for the sample mean vector and for the sample covariance matrix, respectively. Hotelling's T^2 test rejects H_0 if

$$T^2 = N\bar{\mathbf{x}}^T\mathbf{S}^{-1}\bar{\mathbf{x}} \geq \frac{(N-1)p}{N-p} F_\alpha(p, N-p), \tag{2.4}$$

where $F_\alpha(p, N-p)$ is the upper αth quantile of an F-distribution with p and $N-p$ degrees of freedom. The test statistic is affine invariant meaning that $T^2(\mathbf{HX}) = T^2(\mathbf{X})$ for any nonsingular $p \times p$ matrix $\mathbf{H}$. The test (2.4) is valid if the observations arise from a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

If the second moments are finite, i.e., the variance-covariance matrix $\boldsymbol{\Sigma}$ exists, the limiting distribution of the sample mean vector is

$$\sqrt{N}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2.5}$$

and under the null hypothesis, the limiting distribution of the test statistic is

$$T^2 \xrightarrow{D} \chi^2_p. \tag{2.6}$$

Thus, Hotelling's T^2 is an asymptotically distribution-free test. The mean serves as an affine equivariant companion estimator to the test. Hotelling's T^2 test and the sample mean are optimal in the presence of underlying normality. The problem is that the they are extremely sensitive to outlying observations—the sample mean has breakdown point of 0% and an unbounded influence function—and they are inefficient for heavy-tailed noise distributions.

The spatial sign test and the spatial median. Write

$$\mathbf{T}_{1N} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{S}(\mathbf{x}_i) \tag{2.7}$$

for the vector valued average of spatial signs. Under the null hypothesis, the limiting distribution of the quadratic form is

$$N\mathbf{T}_{1N}^T\mathbf{B}_1^{-1}\mathbf{T}_{1N} \xrightarrow{D} \chi_p^2, \tag{2.8}$$

where $\mathbf{B}_1 = E_0\left[\mathbf{S}(\mathbf{x}_i)\,\mathbf{S}(\mathbf{x}_i)^T\right]$ is the spatial sign covariance matrix. In the spherically symmetric case, $\mathbf{B}_1 = \frac{1}{p}\mathbf{I}_p$, the test based on the quadratic form $Np\mathbf{T}_{1N}^T\mathbf{T}_{1N}$ is distribution-free. The test is only asymptotically distribution-free, however, if $\mathbf{B}_1$ is replaced by its consistent estimate (Möttönen & Oja, 1995). Asymptotic relative efficiencies of the spatial sign test relative to Hotelling's T^2 have been studied by Brown (1983) and Möttönen et al. (1997). Note that these exceed the efficiencies of a componentwise sign test when $p > 1$, and the efficiencies of Hotelling's T^2 for heavy-tailed distributions.

The estimator corresponding to the spatial sign test is the *spatial median* (Gower, 1974; Brown, 1983). The spatial median minimizes the objective function

$$D(\boldsymbol{\theta}) = E\left(\|\mathbf{x}_i - \boldsymbol{\theta}\| - \|\mathbf{x}_i\|\right). \tag{2.9}$$

(The sample spatial median $\widehat{\boldsymbol{\mu}}$ minimizes the sample counterpart, $D_N(\boldsymbol{\theta}) = \sum_{i=1}^{N}\|\mathbf{x}_i - \boldsymbol{\theta}\|$; the sum of the Euclidian distances.) The spatial median coincides with the univariate median when $p = 1$. It is unique whenever $p \geq 2$ and the observations do not fall on a line (Milasevic & Ducharme, 1987). Unlike the mean, it has the highest possible breakdown point of 50% (Kemperman, 1987; Lopuhaä & Rousseeuw, 1991) and a bounded influence function (Niinimaa & Oja, 1995; Koltchinskii, 1997). However, the performance of the spatial median is adversely affected by inliers, particularly in the bivariate case (Brown, Hall & Young, 1997). If the density $f(\mathbf{x})$ is uniformly bounded,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}\right) \xrightarrow{D} N_p\left(\mathbf{0}, \mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1}\right), \tag{2.10}$$

where $\mathbf{A}_1 = E_0\left[\|\mathbf{x}_i\|^{-1}\left(\mathbf{I}_p - \mathbf{S}(\mathbf{x}_i)\,\mathbf{S}(\mathbf{x}_i)^T\right)\right]$ originates from the Taylor expansion for $D(\boldsymbol{\theta})$ around $\boldsymbol{\mu}$. The proof and the conditions for this result are discussed in Brown (1983), Bai et al. (1990) and Chaudhuri (1992). Most importantly, the conditions do not include the assumption of finite (first or) second moments. For the estimation of $\mathbf{A}_1$ and $\mathbf{B}_1$ matrices, see Bai et al.

(1990), Rao (1988) and Bose & Chaudhuri (1993). Computation of the spatial median is straightforward: a fast and monotonically convergent algorithm has been introduced by Vardi & Zhang (2000, 2001).

The spatial sign test and the spatial median are not, unfortunately, affine invariant/equivariant. Randles (2000) used Tyler's shape matrix (Tyler, 1987) to pretransform the data points $\mathbf{y}_i = \widehat{\mathbf{V}}^{-1/2}\mathbf{x}_i$, and then applied an ordinary spatial sign test on the transformed data points. Randles' test is distribution-free in the elliptic model, and like any orthogonally invariant statistic computed on the $\mathbf{y}_i$'s, it is affine invariant (Randles, 2000; Möttönen, Hüsler & Oja, 2003). It is easy to see that if $\widehat{\mathbf{V}}$ is an affine equivariant estimator of shape e.g. (1.10), the corresponding *transformation retransformation spatial median*

$$\widehat{\mathbf{V}}^{1/2}\widehat{\boldsymbol{\mu}}_1\left(\widehat{\mathbf{V}}^{-1/2}\mathbf{X}\right) \tag{2.11}$$

is affine equivariant (Chakraborty & Chaudhuri, 1996, 1998; Chakraborty, Chaudhuri & Oja, 1998). The transformation retransformation spatial median seems to be as efficient at the spherical model as the spatial median, but more efficient in the general elliptic case. Thorough efficiency studies are still needed. Furthermore, Chakraborty & Chaudhuri (1999) demonstrated that this estimator, combined with a high breakdown point estimator of scatter (shape), can have a breakdown point close to 50%. Hettmansperger & Randles (2002) considered an alternative approach by estimating location and shape simultaneously, based on spatial signs. Their estimators are affine equivariant as well. The procedure appears always to converge but this has not been proven.

The spatial signed-rank test and the spatial Hodges-Lehmann estimator. The multivariate analogue of the Wilcoxon signed-rank test, the spatial signed-rank test, is based on the average of the signed-ranks,

$$\mathbf{T}_{2N} \;=\; \frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_N(\mathbf{x}_i). \tag{2.12}$$

Under the null, its quadratic form

$$N\mathbf{T}_{2N}^T\mathbf{B}_2^{-1}\mathbf{T}_{2N} \;\xrightarrow{D}\; \chi^2_p, \tag{2.13}$$

where $\mathbf{B}_2 = E_0\left[\mathbf{Q}(\mathbf{x}_i)\mathbf{Q}(\mathbf{x}_i)^T\right]$. Chaudhuri (1992), Hössjer & Croux (1995) and Marden (1999a) present other extensions of the signed-rank test statistics. The test is only conditionally distribution-free. This test has superior efficiency to Hotelling's T^2 test for heavy-tailed distributions, and its efficiency at the normal distribution is close to unity (Möttönen et al., 1997; Möttönen & Oja, 2002).

A multivariate extension of the one-sample Hodges-Lehmann estimator (Hodges & Lehmann, 1963) is the *spatial Hodges-Lehmann estimator* (Chaudhuri, 1992). It is defined as the spatial median of the the $\binom{N}{2}$ Walsh averages $(\mathbf{x}_i + \mathbf{x}_j)/2$. Thus, it is the amount of shift that would make the sample

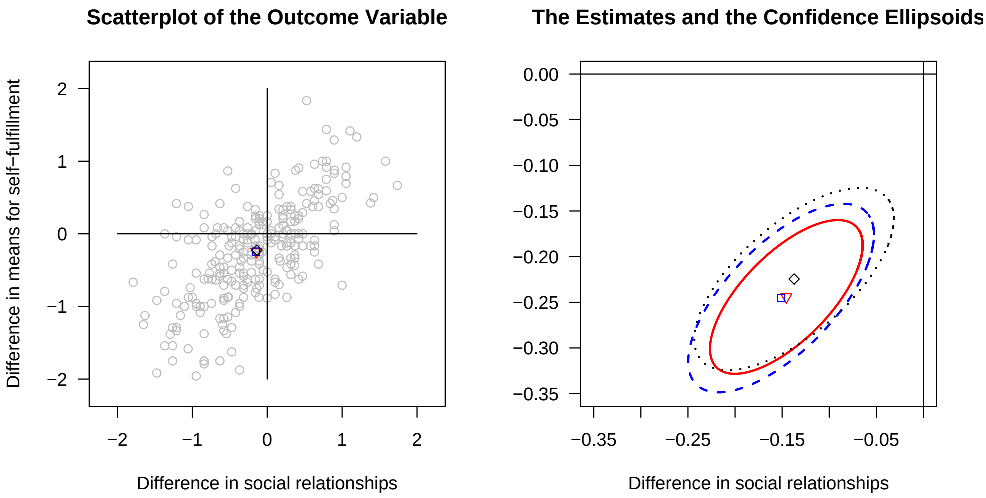

Figure 2.1: A scatterplot of the data and the location estimates with their 95% confidence ellipsoids. The square and the dashed line indicate the sample mean, the diamond and the dotted line the transformation retransformation spatial median, and the triangle and the solid line the transformation retransformation spatial Hodges-Lehmann estimate.

to appear to be centered (in the rank sense) at the origin. Under general assumptions,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}\right) \xrightarrow{D} N_p\left(\mathbf{0}, \mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1}\right), \tag{2.14}$$

where $\mathbf{A}_2 = E_0\left[\|\frac{\mathbf{x}_i+\mathbf{x}_j}{2}\|^{-1}\left(\mathbf{I}_p - \mathbf{S}\left(\frac{\mathbf{x}_i+\mathbf{x}_j}{2}\right)\mathbf{S}\left(\frac{\mathbf{x}_i+\mathbf{x}_j}{2}\right)^T\right)\right]$, $i \neq j$. Its breakdown point is 29.3% and it has a bounded influence function (Möttönen, Oja & Serfling, 2004). Computation proceeds as for the spatial median after finding the pairwise averages (Vardi & Zhang, 2000, 2001).

An affine invariant test is obtained by pretransforming the data with a suitable scatter or shape matrix. In the context of a signed-rank based analysis, a natural choice of the shape matrix would be the one defined in (1.11). An affine equivariant estimator of location can be constructed via the transformation retransformation procedure (Chakraborty & Chaudhuri, 1996, 1998; Chakraborty et al., 1998; Oja & Randles, 2004).

Generalizations of the signed-rank methods have been described by Chaudhuri (1992) and Möttönen et al. (2004).

Example 2. *Consider the one sample example data set described in Section 1.4. Hotelling's T^2, the affine invariant sign test statistic and the affine invariant rank test statistic take values 33.8, 27.7 and 31.5, respectively, which all are highly significant when compared to a χ^2-distribution with 2 degrees of freedom. The estimates of location are plotted in Figure 2.1. The transformation retransformation spatial Hodges-Lehmann estimator provides the most precise estimate for this data set.*

2.2 Several Samples

Let $\mathbf{X}_{p\times N} = (\mathbf{X}_1, \ldots, \mathbf{X}_c)$, where

$$\mathbf{X}_i \quad = \quad (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}),\ i = 1, \ldots, c,$$

be a data matrix of c independent random samples from continuous and symmetric, but otherwise unspecified, p-variate distributions $F(\mathbf{x}-\boldsymbol{\mu}_1), \ldots, F(\mathbf{x}-\boldsymbol{\mu}_c)$, respectively. In the multisample location problem the interest is to confront the hypotheses

$$(2.15) \qquad H_0 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_c \text{ vs. } H_1 : \boldsymbol{\mu}_i\text{'s not all equal.}$$

The estimation of parameters $\boldsymbol{\Delta}_{ij} = \boldsymbol{\mu}_j - \boldsymbol{\mu}_i$, $i, j = 1, \ldots, c$, is of high priority as well.

The classical MANOVA. The Hotelling's trace test statistic for testing the hypothesis (2.15) is

$$(2.16) \qquad T^2 \quad = \quad (N-c)\mathrm{Tr}\left(\mathbf{B}\mathbf{W}^{-1}\right) = \sum_{i=1}^{c} n_i \|\mathbf{S}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\|^2.$$

Here $\bar{\mathbf{x}}_i$ is the mean of the ith sample, $\bar{\mathbf{x}}$ is the global mean, $\mathbf{B}$ is the between-samples sums of squares matrix, $\mathbf{W}$ is the within-samples sums of squares matrix, and $\mathbf{S} = \frac{1}{N-c}\mathbf{W}$ is an unbiased estimator of the variance-covariance matrix $\boldsymbol{\Sigma}$ (if it exists). If the second moments are finite, then under the null hypothesis

$$(2.17) \qquad T^2 \quad \xrightarrow{D} \quad \chi^2_{p(c-1)}.$$

The test statistic is affine invariant in the sense that $T^2(\mathbf{H}\mathbf{X}+\mathbf{g}\mathbf{1}_N^T) = T^2(\mathbf{X})$.

The group differences $\boldsymbol{\Delta}_{ij}$, $i, j = 1, \ldots, c$, can be estimated in an affine equivariant manner by the differences of the sample means $\bar{\boldsymbol{\Delta}}_{ij} = \bar{\mathbf{x}}_j - \bar{\mathbf{x}}_i$ for which

$$(2.18) \qquad \sqrt{N}\left(\bar{\boldsymbol{\Delta}}_{ij} - \boldsymbol{\Delta}_{ij}\right) \quad \xrightarrow{D} \quad N_p\left(\mathbf{0}, \frac{\lambda_i + \lambda_j}{\lambda_i\lambda_j}\boldsymbol{\Sigma}\right),$$

where $\min\{n_i, n_j\} \to \infty$ and $\frac{n_i}{N} \to \lambda_i$, $\frac{n_j}{N} \to \lambda_j$, $0 < \lambda_i, \lambda_j < 1$.

A spatial sign MANOVA. An approach to generalize the one-sample spatial sign test to a several sample case is to replace the sample means, the global mean and the variance-covariance matrix in (2.16) by their nonparametric counterparts. Somorčík (2006) made use of the sample spatial medians, the spatial median of the entire sample, and the estimator of the asymptotic covariance matrix of latter.

An alternative way is to use the spatial sign vectors and the spatial sign covariance matrix directly. Under general assumptions it can be shown that

$$\sum_{i=1}^{c} n_i \|\widehat{\mathbf{B}}_1^{-1/2}\bar{\mathbf{S}}_i\|^2 \quad \xrightarrow{D} \quad \chi^2_{p(c-1)},$$

where $\bar{\mathbf{S}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{S}(\mathbf{x}_{ij} - \widehat{\boldsymbol{\mu}}_{1\cdot})$, $\widehat{\mathbf{B}}_1 = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{n_i} \mathbf{S}(\mathbf{x}_{ij} - \widehat{\boldsymbol{\mu}}_{1\cdot})\,\mathbf{S}(\mathbf{x}_{ij} - \widehat{\boldsymbol{\mu}}_{1\cdot})^T$ and $\widehat{\boldsymbol{\mu}}_{1\cdot}$ is the spatial median of the entire sample. The test statistic is affine invariant when it is computed from standardized data set $\widehat{\mathbf{V}}^{-1/2}\mathbf{X}$. Tyler's matrix with respect to $\widehat{\boldsymbol{\mu}}_1$ can be used as the shape matrix estimator. Yet another to pursue an affine invariant test—as proposed in Nevalainen & Oja (2006, Paper II)—is to apply the simultaneous affine equivariant estimators of location and shape (Hettmansperger & Randles, 2002). Nevalainen & Oja (2006, Paper II) considered a permutation principle for such a multisample sign test, but the proofs of the asymptotic results are still missing from the literature.

The difference of the spatial medians of the jth and the ith sample serves as an orthogonally equivariant estimator $\widehat{\boldsymbol{\Delta}}_{1\cdot ij}$. Under general assumptions,

$$(2.19)\qquad \sqrt{N}\left(\widehat{\boldsymbol{\Delta}}_{1\cdot ij} - \boldsymbol{\Delta}_{ij}\right) \xrightarrow{D} N_p\left(\mathbf{0}, \frac{\lambda_i + \lambda_j}{\lambda_i\lambda_j}\mathbf{A}_1^{-1}\mathbf{B}_1\mathbf{A}_1^{-1}\right)$$

(Möttönen & Oja, 1995). Affine equivariant estimators are constructed by pretransforming the data and retransforming the estimates (by Dümbgen's matrix, for instance), or by using the Hettmansperger-Randles estimators (Hettmansperger & Randles, 2002; Nevalainen & Oja, 2006, Paper II). More theoretical work is needed on the asymptotic behavior of these estimators.

A spatial rank MANOVA. A multivariate Kruskal-Wallis test is based on $\bar{\mathbf{R}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{R}_N(\mathbf{x}_{ij})$, the group average of empirical spatial ranks with respect to the entire sample. Under general assumptions and under the null hypothesis,

$$(2.20)\qquad \sum_{i=1}^{c} n_i \bar{\mathbf{R}}_i^T \widehat{\mathbf{B}}_2^{-1}\bar{\mathbf{R}}_i = \sum_{i=1}^{c} n_i \|\widehat{\mathbf{B}}_2^{-1/2}\bar{\mathbf{R}}_i\|^2 \xrightarrow{D} \chi^2_{p(c-1)}$$

(Möttönen & Oja, 1995; Choi & Marden, 1997, 2002). Note that the average of the spatial ranks over the entire sample is zero. Here

$$(2.21)\qquad \mathbf{B}_2 = E_0\left[\mathbf{R}(\mathbf{x}_{ij})\mathbf{R}(\mathbf{x}_{ij})^T\right]$$

is the spatial rank covariance matrix under the null hypothesis, which can be consistently estimated by the sample counterpart with empirical ranks.

Treatment differences can be directly estimated by the two-sample spatial Hodges-Lehmann estimators. Under general assumptions,

$$(2.22)\qquad \sqrt{N}\left(\widehat{\boldsymbol{\Delta}}_{2\cdot ij} - \boldsymbol{\Delta}_{ij}\right) \xrightarrow{D} N_p\left(\mathbf{0}, \frac{\lambda_i + \lambda_j}{\lambda_i\lambda_j}\mathbf{A}_2^{-1}\mathbf{B}_2\mathbf{A}_2^{-1}\right).$$

The limiting distribution can easily be found heuristically but a solid proof of (2.22) does not seem to appear in the literature. The spatial Hodges-Lehmann estimators suffer, however, from an incompatibility problem: it is not generally true that $\widehat{\boldsymbol{\Delta}}_{2\cdot ij} = \widehat{\boldsymbol{\Delta}}_{2\cdot ik} + \widehat{\boldsymbol{\Delta}}_{2\cdot kj}$. In the univariate case, Lehmann (1963) considered compatible but, in the particular case that $\frac{n_i}{N} \to 0$ for some i,

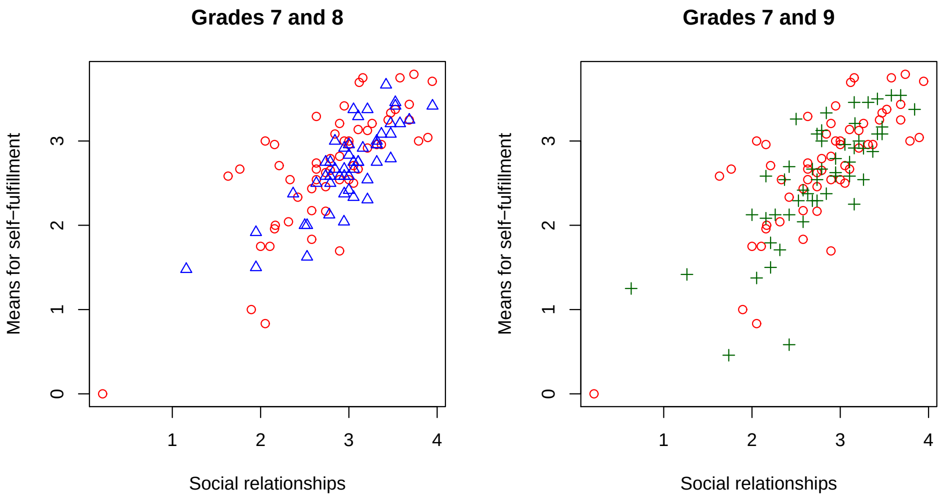

Figure 2.2: Scatterplots of the social relationships and the means for self-fulfillment scores. The circles indicate grade 7, the triangles grade 8 and the pluses grade 9.

inconsistent estimators. Later on, Spjøtvoll (1968) developed weighted estimators that are compatible and consistent under a weaker condition. Their spatial extensions to the multivariate case,

$$\widetilde{\mathbf{\Delta}}_{2\cdot ij} = \frac{1}{N}\sum_{k=1}^{c} n_k \left(\widehat{\mathbf{\Delta}}_{2\cdot ik} + \widehat{\mathbf{\Delta}}_{2\cdot kj}\right), \tag{2.23}$$

possess the properties as well. Asymptotically, the adjusted estimators have the same distribution as the unadjusted ones, but in finite samples the spatial Spjøtvoll (1968) estimators are always at least as efficacious as the spatial Hodges-Lehmann estimators (Nevalainen, Möttönen & Oja, 2006b, Paper I). Affine equivariance can be obtained using similar ideas as before (Nevalainen et al., 2006b, Paper I).

Example 3. *Consider the several samples example data set described in Section 1.4. The outcome variables are plotted in Figure 2.2. Hotelling's* T^2*, the affine invariant several samples sign test statistic and the affine invariant several samples rank test statistic are 9.0* $(p = 0.060)$*, 11.8* $(p = 0.019)$ *and 10.4* $(p = 0.035)$*, correspondingly. Affine equivariant point estimates of location are tabulated in Table 2.1. Note that there is practically no difference in the first three decimal points between the spatial Hodges-Lehmann estimator* $\widehat{\mathbf{\Delta}}_{2\cdot ij}$ *and the adjusted version* $\widetilde{\mathbf{\Delta}}_{2\cdot ij}$ *for a data set of this size with nearly equal group allocation. The different estimates are very similar when comparing grades 7 and 9. Figure 2.2 suggests that grade 8 has a dense region in the middle and a rather symmetric shape whereas the distributions of other two grades seem more dispersed. This could be the explanation why the competing estimates of* $\mathbf{\Delta}_{78}$ *and* $\mathbf{\Delta}_{89}$ *in Table 2.1 differ (but still only slightly).*

Table 2.1: Affine equivariant point estimates of location for Example 3.

	Parameter		
Estimator	$\boldsymbol{\Delta}_{78}$	$\boldsymbol{\Delta}_{79}$	$\boldsymbol{\Delta}_{89}$
$\bar{\boldsymbol{\Delta}}_{ij}$	$(0.189, 0.003)$	$(-0.050, -0.150)$	$(-0.239, -0.154)$
$^*\widehat{\boldsymbol{\Delta}}_{1\cdot ij}$	$(0.173, -0.029)$	$(-0.045, -0.151)$	$(-0.219, -0.122)$
$^*\widehat{\boldsymbol{\Delta}}_{2\cdot ij}$	$(0.152, -0.036)$	$(-0.054, -0.143)$	$(-0.207, -0.110)$
$^*\widetilde{\boldsymbol{\Delta}}_{2\cdot ij}$	$(0.152, -0.035)$	$(-0.054, -0.144)$	$(-0.206, -0.109)$

* indicates that the estimate was computed on a pretransformed data set and transformed back

2.3 Remarks

The methods presented in this section are fairly straightforward to implement in programming languages. For SAS/IML routines, see Möttönen (1997) (rotation invariant/equivariant sign and rank methods) and Nevalainen & Oja (2006, Paper II) (affine invariant/equivariant sign methods). An implementation in R has been done by Sirkiä (2005).

3 Cluster Correlated Observations

Until now, the example data set has been treated as if the schools played no role in the well-being of the pupils. However, it is more likely that pupils within the same school are more similar than pupils across schools. This type of data is called *clustered*. Its special feature—as opposed to the i.i.d. setting covered in Chapter 2—is that two observations from the same cluster may be dependent, but two observations from different clusters are presumed independent. Unless the potential intracluster correlation is taken into account during the course of the analysis, the tests may not maintain their desired level or the estimated standard errors of the estimates may be artificially small.

Recently, several univariate nonparametric tests have been extended to cluster correlated data (Rosner & Grove, 1999; Rosner, Glynn & Ting Lee, 2003; Datta & Satten, 2005; Larocque, 2005). A monograph by Aerts, Geys, Molenberghs & Ryan (2002) provides a complete treatment of clustered data from a parametric perspective. Multivariate approaches to clustered data have been considered by Dueck & Lohr (2005) and Goldstein (2003, chapter 6). Longitudinal data is a potential application of spatial sign and rank methods, but the current literature is only focused on methods based on normality and some other parametric models (Verbeke & Molenberghs, 2000; Diggle, Heagerty, Liang & Zeger, 2002).

3.1 One Sample

Let $\mathbf{X}_{p\times N} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, where

$$\mathbf{X}_i = \begin{pmatrix} x_{i11} & x_{i21} & \cdots & x_{im_i1} \\ x_{i12} & x_{i22} & \cdots & x_{im_i2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1p} & x_{i2p} & \cdots & x_{im_ip} \end{pmatrix},$$

$i = 1, \ldots, n$, $N = m_1 + \ldots + m_n$, be a data matrix such that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are mutually independent. Observations are supposed to be generated from a general "nonparametric" model

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{ij} \tag{3.1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^T$ is the location parameter, and (i) $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ijp})^T$ are random errors assumed to arise from a common continuous and symmetric distribution with $E\left[\mathbf{S}(\boldsymbol{\varepsilon}_{ij})\right] = \mathbf{0}$. Furthermore, assume that

(ii) $(\boldsymbol{\varepsilon}_{i_1 j_1}, \boldsymbol{\varepsilon}_{i_1 j_2}) \sim (\boldsymbol{\varepsilon}_{i_2 k_1}, \boldsymbol{\varepsilon}_{i_2 k_2})$, where i_1, i_2 are two indices in $\{1, \ldots, n\}$, and j_1, j_2 $(j_1 \neq j_2)$ and k_1, k_2 $(k_1 \neq k_2)$ are two indices in $\{1, \ldots, m_{i_1}\}$ and $\{1, \ldots, m_{i_2}\}$, respectively;

(iii) $\boldsymbol{\varepsilon}_{i_1 j_1}$ and $\boldsymbol{\varepsilon}_{i_2 j_2}$ are independent if $i_1 \neq i_2$ but possibly dependent if $i_1 = i_2$;

(iv) As $n \to \infty$,

$$\max\{m_1, \ldots, m_n\} \to M,$$

where M is finite, and

$$\frac{1}{n} \sum_{i=1}^{n} I(m_i = m) \to \alpha_m,$$

$0 \leq \alpha_m \leq 1$ for all $m = 1, \ldots, M$.

The assumptions are explained in more detail in Nevalainen, Larocque & Oja (2007, Paper III).

An adjusted Hotelling's T^2 and the sample mean. It is straightforward to see that

$$\text{(3.2)} \qquad \text{cov}\left(\sqrt{N}\bar{\mathbf{x}}\right) = \text{cov}\left(\boldsymbol{\varepsilon}_{ij}\right) + \frac{\sum_{i=1}^{n} m_i(m_i - 1)}{N} \text{cov}\left(\boldsymbol{\varepsilon}_{ij}, \boldsymbol{\varepsilon}_{ij'}\right)$$

(if the second moments exist). Under general constraints on the sequence of independent random variables $\bar{\mathbf{x}}_{1\cdot}, \bar{\mathbf{x}}_{2\cdot}, \ldots$, where $\bar{\mathbf{x}}_{i\cdot} = \frac{1}{m_i} \sum_{j=1}^{m_i} \mathbf{x}_{ij}$, it follows from the multivariate central limit theorem (Rao, 1965, p. 118) that

$$\text{(3.3)} \qquad \sqrt{N}\left(\bar{\mathbf{x}} - \boldsymbol{\mu}\right) \xrightarrow{D} N_p\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right)$$

where $\boldsymbol{\Sigma}^* = \lim_{n \to \infty} \left[\text{cov}\left(\sqrt{N}\bar{\mathbf{x}}\right)\right]$. The existence of the limit is ensured by assumption (iv). Therefore, under the null hypothesis, the adjusted test statistic

$$\text{(3.4)} \qquad N\bar{\mathbf{x}}^T \left[\widehat{\boldsymbol{\Sigma}}^*\right]^{-1} \bar{\mathbf{x}} \xrightarrow{D} \chi_p^2$$

as $n \to \infty$.

The spatial sign test and the spatial median. Write

$$\text{(3.5)} \qquad \text{cov}\begin{pmatrix} \mathbf{S}(\boldsymbol{\varepsilon}_{ij}) \\ \mathbf{S}(\boldsymbol{\varepsilon}_{ik}) \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{C}_1 \\ \mathbf{C}_1 & \mathbf{B}_1 \end{pmatrix}$$

for all $i = 1, \ldots, n$ and $j, k = 1, \ldots, m_i$ $(j \neq k)$, for the spatial sign covariance structure. Additionally, $\text{cov}\left[\mathbf{S}(\boldsymbol{\varepsilon}_{ij}), \mathbf{S}(\boldsymbol{\varepsilon}_{i'k})\right] = \mathbf{0}$ for $i \neq i'$. Larocque (2003) demonstrates that under general assumptions and under the null hypothesis, as $n \to \infty$,

$$\text{(3.6)} \qquad N\mathbf{T}_{1N}^T \widehat{\mathbf{D}}_1^{-1} \mathbf{T}_{1N} \xrightarrow{D} \chi_p^2,$$

where $\mathbf{T}_{1N}$ is as before and

$$\widehat{\mathbf{D}}_1 = \widehat{\mathbf{B}}_1 + \frac{\sum_{i=1}^n m_i(m_i-1)}{N}\widehat{\mathbf{C}}_1 \qquad (3.7)$$

is a consistent estimator of the asymptotic variance-covariance matrix of $\sqrt{N}\mathbf{T}_{1N}$, denoted by $\mathbf{D}_1$ (also $\widehat{\mathbf{B}}_1 \xrightarrow{P} \mathbf{B}_1$ and $\widehat{\mathbf{C}}_1 \xrightarrow{P} \mathbf{C}_1$). In order to have an affine invariant test, $\mathbf{T}_{1N}$ and $\widehat{\mathbf{D}}_1$ are computed from the pretransformed data matrix $\widehat{\mathbf{V}}^{-1/2}\mathbf{X}$, where $\widehat{\mathbf{V}}$ is any $\sqrt{N}$-consistent estimator of shape e.g. Tyler's matrix (1.9) (Randles, 2000; Larocque, 2003). Interestingly, the sign test suffers less—in efficiency—from intracluster correlation than the adjusted Hotelling's T^2 test (Larocque, 2003; Nevalainen et al., 2007, Paper III).

The corresponding estimator is again the sample spatial median. In Nevalainen et al. (2007, Paper III) it is proven that under general assumptions,

$$\sqrt{N}\left(\widehat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}\right) \xrightarrow{D} N_p\left(\mathbf{0}, \mathbf{A}_1^{-1}\mathbf{D}_1\mathbf{A}_1^{-1}\right), \text{ as } n \to \infty. \qquad (3.8)$$

An affine equivariant spatial median is constructed via the transformation retransformation procedure using a suitable $\sqrt{N}$-consistent estimator of shape e.g. Dümbgen's matrix (Chakraborty & Chaudhuri, 1996, 1998; Chakraborty et al., 1998; Dümbgen, 1998; Nevalainen et al., 2007, Paper III).

Until now, the test statistic and the estimator have been the same as in the i.i.d. case—the proposed variance adjustment ensures that the test and the confidence ellipsoids maintain their nominal level and covering probability. It may, however, be more efficient to make use of the clustered structure also in the construction of the test and the estimator. Larocque, Nevalainen & Oja (2007, Paper IV) and Nevalainen, Larocque & Oja (2006a, Paper V) consider a weighted test based on

$$\mathbf{T}_{1N}^{(w)} = \frac{1}{N}\sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \mathbf{S}(\mathbf{x}_{ij}) \qquad (3.9)$$

and a weighted spatial median minimizing the objective function

$$D_N(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \|\mathbf{x}_{ij} - \boldsymbol{\theta}\|. \qquad (3.10)$$

Under some additional assumptions on the weights, the limiting null distributions of the quadratic form of the weighted test statistic and the limiting distribution of the weighted spatial median are still a central χ_p^2 and a multivariate normal, respectively. Larocque et al. (2007, Paper IV) and Nevalainen et al. (2006a, Paper V) also find optimal weights (in the efficiency sense) under a general family of distributions. The weights can also be used to modify, in a desired way, the breakdown properties of the estimator.

The spatial rank test and the spatial Hodges-Lehmann estimator. Rank procedures for the multivariate location problem with cluster correlated observations still remain to be developed.

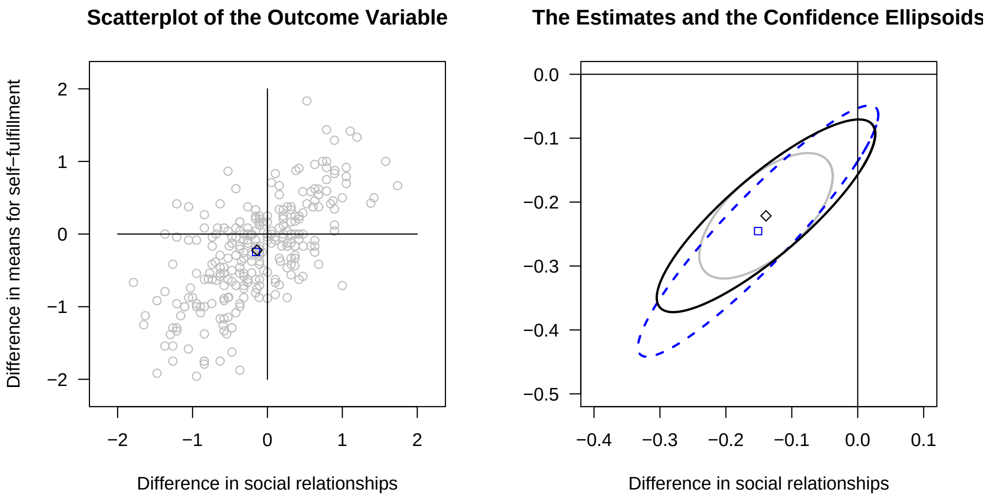

Figure 3.1: A scatterplot of the data and the location estimates with their 95% confidence ellipsoids. The square and the dashed line indicate the adjusted sample mean and the diamond and the solid line the transformation retransformation spatial median. The grey line indicates a confidence ellipsoid for the transformation retransformation spatial median without the variance correction due to clustering.

Example 4. *Consider the one sample example data set described in Section 1.4. The adjusted Hotelling's T^2 and the affine invariant sign test statistic are 13.8 (p=0.001) and 3.9 (p=0.143), respectively. The estimates and their confidence ellipsoids are plotted in Figure 3.1. Note the different estimation strategy of $\mathbf{B}_1$ under the null hypothesis (with respect to the null point) and under the alternative (with respect to the point estimate of location). The classical approach traditionally uses the same estimate of $\mathbf{\Sigma}$ in the test and in the construction of confidence bounds, so a comparison of test results is not entirely fair. The confidence ellipsoids are of a similar magnitude.*

In this example, the clustered structure of the data cannot be neglected: clustering has a remarkable impact on the values of the test statistics (as compared to the previous i.i.d. analysis) and on the magnitude of the confidence ellipsoid. In this particular example data set, the estimated intracluster correlation is only around 0.05. The explanation is the huge number of covariance terms due to large clusters (Table 1.1).

3.2 Several Samples

A nonparametric treatment of the multivariate clustered several samples problem does not seem to appear in the literature. Univariate nonparametric procedures have been studied, for instance, by Rosner & Grove (1999), Rosner et al. (2003) and Datta & Satten (2005).

3.3 Remarks

While the assumption that the number of clusters goes to infinity is quite standard, it may sometimes be unrealistic to make this assumption. A study can have a fixed set of clusters, and collect a large number of subunits within these clusters. In such cases it is questionable whether the approximations based on asymptotical results assuming $n \to \infty$ are accurate. Instead, suppose that $\min\{m_1, \ldots, m_n\} \to \infty$. A sufficient condition for the standardized variance of the test statistic (3.7) to converge is to assume that $\mathbf{C}_1 = O\left(\frac{1}{N}\right)$. In other words, the entire intracluster correlation structure should asymptotically vanish. This type of assumption could be reasonable if there is a natural way to measure the distance between the observations within the same cluster, like in the context of time series, for example. Otherwise, the assumption seems rather unattractive. Note that the finite sample estimator of $\mathbf{D}_1$ appears to be the same.

Summaries of Original Publications

I. In the multivariate several samples location problem, it is usually of interest to present estimates of treatment effects along with the test. The two-sample spatial Hodges-Lehmann estimators are apparent companions to a multivariate Kruskal-Wallis test. However, these estimators generally fail to satisfy the compatibility property $\widehat{\mathbf{\Delta}}_{2\cdot ij} = \widehat{\mathbf{\Delta}}_{2\cdot ik} + \widehat{\mathbf{\Delta}}_{2\cdot kj}$, which makes them difficult to interpret. In this paper adjusted estimators possessing this property are considered. A simulation study is carried out in order to study their finite sample efficiencies. Limiting distributions and efficiencies are presented as well.

II. The affine invariant multivariate sign test due to Randles (2000) and the companion estimator due to Hettmansperger & Randles (2002) are reviewed in this paper. Extensions to a multisample case are proposed and discussed from a practical perspective. The methods are compared with their classical analogues. A new SAS/IML tool for performing a spatial sign based multivariate analysis of variance is introduced.

III. The multivariate one sample location problem with clustered data is considered from a nonparametric viewpoint. The spatial median and its affine equivariant version are proposed as companion estimators to the affine invariant sign test of Larocque (2003). The asymptotics of the proposed estimators are extended to cluster dependent data, and their limiting as well as finite sample efficiencies for multivariate t-distributions are explored. As an important result, it is found that the efficiency of the spatial median suffers less from intracluster correlation than the mean vector. An application of the new method, using data on well-being in Finnish schools, is given.

IV. The multivariate one sample location problem with clustered data is considered. A family of multivariate weighted sign tests is introduced. Under weak assumptions, the test statistic is asymptotically distributed as a chi-squared random variable as the number of clusters goes to infinity. The asymptotic distribution of the test statistic is also given for a local alternative model under multivariate normality. Optimal weights maximizing Pitman asymptotic efficiency are provided. These weights depend on the cluster sizes and on the intracluster correlation. Several approaches for estimating these weights are presented. Using Pitman asymptotic efficiency, it is shown that appropriate weighting can substantially increase the efficiency compared to a test that gives the same weight to each cluster. A multivariate weighted t-test is also introduced. The finite-sample performance of the weighted sign test is explored through a simulation study which shows that the proposed approach is very competitive. A real data example illustrates the practical application of the methodology.

V. A weighted spatial median is proposed for the multivariate one sample location problem with cluster correlated data. Its limiting distribution is derived under mild conditions and it is shown to be multivariate normal. Asymptotic as well as finite sample efficiencies and breakdown properties are considered, and supplied with illustrative examples. It turns out that there are potentially meaningful gains in estimation efficiency. An affine equivariant weighted spatial median is developed in parallel. This paper provides companion estimates to the weighted affine invariant sign test proposed by Larocque et al. (2007, Paper IV).

References

Aerts, M., Geys, H., Molenberghs, G. & Ryan, L. M. (eds.) (2002), *Topics in Modelling of Clustered Data*, Boca Raton, Florida, USA: Chapman & Hall/CRC.

Bai, Z. D., Chen, R., Miao, B. Q. & Rao, C. R. (1990), "Asymptotic theory of least distances estimate in multivariate linear models", *Statistics*, 4, 503–519.

Barnett, V. (1976), "The ordering of multivariate data", *Journal of the Royal Statistical Society, Series A*, 139, 318–355.

Bose, A. & Chaudhuri, P. (1993), "On the dispersion of multivariate median", *Annals of the Institute of Statistical Mathematics*, 45, 541–548.

Brown, B. M. (1983), "Statistical uses of the spatial median", *Journal of the Royal Statistical Society, Series B*, 45, 25–30.

Brown, B. M., Hall, P. & Young, G. A. (1997), "On the effect of inliers on the spatial median", *Journal of Multivariate Analysis*, 63, 88–104.

Chakraborty, B. (2001), "On affine equivariant multivariate quantiles", *Annals of the Institute of Statistical Mathematics*, 53, 380–403.

——— (2003), "On multivariate quantile regression", *Journal of Statistical Planning and Inference*, 110, 109–132.

Chakraborty, B. & Chaudhuri, P. (1996), "On a transformation and retransformation technique for constructing affine equivariant multivariate median", in *Proceedings of the American Mathematical Society* (Vol. 124), Vol. 124, pp. 2539–2547.

——— (1998), "On an adaptive transformation and retransformation estimate of multivariate location", *Journal of the Royal Statistical Society, Series B*, 60, 145–157.

——— (1999), "A note on the robustness of multivariate medians", *Statistics & Probability Letters*, 45, 269–276.

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

Chaudhuri, P. (1992), ”Multivariate location estimation using extension of R-estimates through U-statistics type approach”, *The Annals of Statistics*, 20, 897–916.

——— (1996), ”On a geometric notion of quantiles for multivariate data”, *Journal of the American Statistical Association*, 91, 862–868.

Choi, K. & Marden, J. (1997), ”An approach to multivariate rank tests in multivariate analysis of variance”, *Journal of the American Statistical Association*, 92, 1581–1590.

——— (2002), ”Multivariate analysis of variance using spatial ranks”, *Sociological Methods & Research*, 30, 341–366.

Croux, C., Ollila, E. & Oja, H. (2002), ”Sign and rank covariance matrices: statistical properties and application to principal component analysis”, in *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, ed. Y. Dodge, Basel, Switzerland: Birkäuser Verlag, pp. 257–269.

Datta, S. & Satten, G. A. (2005), ”Rank-sum tests for clustered data”, *Journal of the American Statistical Association*, 100, 908–915.

Davies, P. L. (1987), ”Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices”, *The Annals of Statistics*, 15, 1269–1292.

Diggle, P. J., Heagerty, P., Liang, K.-Y. & Zeger, S. L. (2002), *Analysis of Longitudinal Data* (2nd ed.), Oxford, UK: Oxford University Press.

Dueck, A. & Lohr, S. (2005), ”Robust estimation of multivariate covariance components”, *Biometrics*, 61, 162–169.

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

Dümbgen, L. & Tyler, D. E. (2005), ”On the breakdown properties of some multivariate M-functionals”, *Scandinavian Journal of Statistics*, 32, 247–264.

Gao, Y. (2003), ”Data depth based on spatial rank”, *Statistics & Probability Letters*, 65, 217–225.

Goldstein, H. (2003), *Multilevel Statistical Models* (3rd ed.), London, UK: Arnold.

Gower, J. C. (1974), ”The mediancentre”, *Applied Statistics*, 23, 466–470.

Hallin, M. & Paindaveine, D. (2002), ”Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks”, *The Annals of Statistics*, 30, 1103–1133.

Hampel, F. R., Rochetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York, USA: John Wiley & Sons.

Hettmansperger, T. P. & McKean, J. W. (1998), *Robust Nonparametric Statistical Methods*, London, UK: Arnold.

Hettmansperger, T. P., Nyblom, J. & Oja, H. (1992), ”On multivariate notions of sign and rank”, in L_1-*Statistical Analysis and Related Methods*, ed. Y. Dodge, Amsterdam: North Holland: Elsevier, pp. 267–278.

Hettmansperger, T. P. & Randles, R. H. (2002), ”A practical affine equivariant multivariate median”, *Biometrika*, 89, 851–860.

Hodges, J. L. & Lehmann, E. L. (1963), ”Estimates of location based on rank tests”, *The Annals of Mathematical Statistics*, 34, 598–611.

Hollander, M. & Wolfe, D. A. (1999), *Nonparametric Statistical Methods* (2nd ed.), New York, USA: John Wiley & Sons.

Hössjer, O. & Croux, C. (1995), ”Generalizing univariate signed rank statistics for testing and estimating multivariate location parameter”, *Nonparametric Statistics*, 4, 293–308.

Huber, P. J. (1980), *Robust Statistics*, New York, USA: John Wiley & Sons.

Kemperman, J. H. B. (1987), ”The median of a finite measure on a Banach space”, in *Statistical data analysis based on the* L_1-*norm and related methods*, ed. Y. Dodge, North-Holland, Amsterdam: Elsevier Science Publishers B. V., pp. 217–230.

Koltchinskii, V. I. (1997), ”M-estimation, convexity and quantiles”, *The Annals of Statistics*, 25, 435–477.

Konu, A. & Lintonen, T. (2006), ”School well-being in grades four to twelve”, *Health Education Research*, 21, 633–642.

Larocque, D. (2003), ”An affine-invariant multivariate sign test for cluster correlated data”, *The Canadian Journal of Statistics*, 31, 437–455.

——— (2005), ”The Wilcoxon signed-rank test for cluster correlated data”, in *Statistical Modeling and Analysis for Complex Data Problems*, eds. P. Duchesne & B. Rémillard, New York, USA: Springer, pp. 309–323.

Larocque, D., Nevalainen, J. & Oja, H. (2007), ”A weighted multivariate sign test for cluster correlated data”, *Biometrika (in press).*

Lehmann, E. L. (1963), ”Robust estimation in analysis of variance”, *The Annals of Mathematical Statistics*, 34, 957–966.

——— (1975), *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco, California, USA: Holden-Day.

Lintonen, T. & Konu, A. (2006), ”The Well-Being Profile—an Internet tool for school health promotion”, *Promotion & Education (in press).*

Liu, R. Y., Parelius, J. M. & Singh, K. (1999), ”Multivariate analysis by data depth: descriptive statistics, graphics and inference”, *The Annals of Statistics*, 27, 783–858.

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.

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

Marden, J. I. (1999a), ”Multivariate rank tests”, in *Multivariate Analysis, Design of Experiments, and Survey Sampling*, ed. S. Ghosh, New York, USA: M. Dekker, pp. 401–432.

——— (1999b), ”Some robust estimates of principal components”, *Statistics & Probability Letters*, 43, 349–359.

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

Milasevic, P. & Ducharme, G. R. (1987), ”Uniqueness of the spatial median”, *The Annals of Statistics*, 15, 1332–1333.

Möttönen, J. (1997), ”SAS/IML macros for spatial sign and rank tests”, *Mathematics*. University of Oulu.

Möttönen, J., Hüsler, J. & Oja, H. (2003), ”Multivariate nonparametric tests in a randomized complete block design”, *Journal of Multivariate Analysis*, 85, 106–129.

Möttönen, J. & Oja, H. (1995), ”Multivariate spatial sign and rank methods”, *Nonparametric Statistics*, 5, 201–213.

——— (2002), ”Efficiency of the spatial rank test”, *Unpublished manuscript.*

Möttönen, J., Oja, H. & Serfling, R. J. (2004), ”Multivariate generalized spatial signed-rank methods”, *Journal of Statistical Research*, 39, 19–42.

Möttönen, J., Oja, H. & Tienari, J. (1997), ”On the efficiency of multivariate spatial sign and rank tests”, *The Annals of Statistics*, 25, 542–552.

Nevalainen, J., Larocque, D. & Oja, H. (2006a), ”A weighted spatial median for clustered data”, *Statistical Methods & Applications (in press).*

——— (2007), ”On the multivariate spatial median for clustered data”, *The Canadian Journal of Statistics (in press).*

Nevalainen, J., Möttönen, J. & Oja, H. (2006b), ”A spatial rank test and corresponding estimators for several samples”, *Submitted.*

Nevalainen, J. & Oja, H. (2006), ”SAS/IML macros for a multivariate analysis of variance based on spatial signs”, *Journal of Statistical Software*, 16, 1–17.

Niinimaa, A. & Oja, H. (1995), ”On the influence functions of certain bivariate medians”, *Journal of the Royal Statistical Society, Series B*, 57, 565–574.

——— (1999), ”Multivariate median”, in *Encyclopedia of Statistical Sciences* (Vol. 3), eds. S. Kotz, C. B. Read & D. Banks, New York, USA: John Wiley & Sons, Vol. 3, pp. 497–505.

Oja, H. (1999), ”Affine invariant multivariate sign and rank tests and corresponding estimates: a review”, *Scandinavian Journal of Statistics*, 26, 319–343.

Oja, H. & Paindaveine, D. (2005), ”Optimal signed-rank tests based on hyperplanes”, *Journal of Statistical Planning and Inference*, 135, 300–323.

Oja, H. & Randles, R. H. (2004), ”Multivariate nonparametric tests”, *Statistical Science*, 19, 598–605.

Ollila, E. (2002), ”Sign and rank covariance matrices with applications to multivariate analysis”, Ph.D. thesis, University of Jyväskylä.

Peters, D. & Randles, R. H. (1990), ”A multivariate signed-rank test for the one-sample location problem”, *Journal of the American Statistical Association*, 85, 552–557.

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

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

——— (2000), ”A simpler, affine-invariant, multivariate, distribution-free sign test”, *Journal of the American Statistical Association*, 95, 1263–1268.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*, New York, USA: John Wiley & Sons.

——— (1988), ”Methodology based on the L_1-norm in statistical inference”, *Sankhyā*, 50, 289–313.

Rosner, B., Glynn, R. J. & Ting Lee, M.-L. (2003), ”Incorporation of clustering effects for the Wilcoxon rank sum test: a large sample approach”, *Biometrics*, 59, 1089–1098.

Rosner, B. & Grove, D. (1999), ”Use of the Mann-Whitney U-test for clustered data”, *Statistics in Medicine*, 18, 1387–1400.

Rousseeuw, P. J. (1985), "Multivariate estimation with high breakdown point", in *Mathematical Statistics and Applications* (Vol. B), eds. W. Grossmann, G. Pug, I. Vincze & W. Wertz, Dordrecht: Reidel, pp. 283–297.

Serfling, R. J. (2002), "A depth function and a scale curve based on spatial quantiles", in *Statistical data analysis based on the L_1-norm and related methods*, ed. Y. Dodge, Basel, Switzerland: Birkhäuser Verlag, pp. 25–38.

——— (2004), "Nonparametric multivariate descriptive measures based on spatial quantiles", *Journal of Statistical Planning and Inference*, 123, 259–278.

Sirkiä, S. (2005), *R-functions for multivariate sign and rank tests.* Available from Internet: `http://www.maths.jyu.fi/~ssirkia/signrank/`.

Sirkiä, S., Taskinen, S. & Oja, H. (2006), "Symmetrised M-estimators of multivariate scatter", *Journal of Multivariate Analysis (conditionally accepted).*

Small, C. G. (1990), "A survey of multidimensional medians", *International Statistical Review*, 58, 263–277.

Somorčík, J. (2006), "Tests using spatial median", *Austrian Journal of Statistics*, 35, 331–338.

Spjøtvoll, E. (1968), "A note on robust estimation in analysis of variance", *The Annals of Mathematical Statistics*, 39, 1486–1492.

Taskinen, S. (2003), "On nonparametric tests of independence and robust canonical correlation analysis", Ph.D. thesis, University of Jyväskylä.

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

Vardi, Y. & Zhang, C.-H. (2000), "The multivariate L_1-median and associated data depth", in *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 97), Vol. 97, pp. 1423–1426.

——— (2001), "A modified Weiszfeld algorithm for the Fermat-Weber location problem", *Mathematical Programming, Series A*, 90, 559–566.

Verbeke, G. & Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York, USA: Springer.

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

Zuo, Y. & Serfling, R. J. (2000), "General notions of statistical depth function", *The Annals of Statistics*, 28, 461–482.

A Spatial Rank Test and Corresponding Estimators for Several Samples*

by Jaakko Nevalainen[1], Jyrki Möttönen[2] and Hannu Oja[3]

Version: September 29, 2006

* *Published from Statistics & Probability Letters, with permission from Elsevier*

[1]*Department of Mathematics, Statistics and Philosophy*
33014 University of Tampere, Finland
e-mail: `jaakko.nevalainen@uta.fi`

[2]*Department of Mathematics and Statistics, University of Kuopio, Finland*
[3]*Tampere School of Public Health, University of Tampere, Finland*

Abstract

In the several samples location problem, it is usually of interest to present estimates of treatment effects along with the test. The spatial Hodges-Lehmann estimators $\widehat{\mathbf{\Delta}}_{ij}$ of the differences between treatments i and j are apparent companions to a multivariate Kruskal-Wallis test. However, these estimators generally fail to satisfy the property $\widehat{\mathbf{\Delta}}_{ij} = \widehat{\mathbf{\Delta}}_{ik} + \widehat{\mathbf{\Delta}}_{kj}$, making them incompatible with each other. In this paper we consider adjusted estimators possessing this property. A simulation study is carried out in order to study their finite sample efficiencies. Limiting distributions and efficiencies are presented as well.

Key words: Kruskal-Wallis test; Multivariate several samples rank test; Spatial Hodges-Lehmann estimator; Spatial rank.

1 Introduction

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_c)$ be a $p \times N$ data matrix consisting of c independent random samples

$$\begin{aligned}
\mathbf{X}_1 &= (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1}), \\
\mathbf{X}_2 &= (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n_2}), \\
&\vdots \\
\mathbf{X}_c &= (\mathbf{x}_{c1}, \ldots, \mathbf{x}_{cn_c}),
\end{aligned}$$

from p-variate continuous distributions $F(\mathbf{x}-\boldsymbol{\mu}_1)$, $F(\mathbf{x}-\boldsymbol{\mu}_2)$, ..., $F(\mathbf{x}-\boldsymbol{\mu}_c)$, respectively, where $N = n_1 + \ldots + n_c$. Knowing little of the underlying distribution, we are interested in finding out whether there are differences in location between the samples, and if so, estimating those differences.

First recall the notions of spatial sign and rank. A multivariate extension of univariate sign, the spatial sign of vector $\mathbf{x}_{ij}$, is defined as

$$\mathbf{S}(\mathbf{x}_{ij}) = \begin{cases} \|\mathbf{x}_{ij}\|^{-1}\mathbf{x}_{ij}, & \text{if } \mathbf{x}_{ij} \neq \mathbf{0}; \\ \mathbf{0}, & \text{if } \mathbf{x}_{ij} = \mathbf{0}, \end{cases}$$

where $\|\cdot\|$ denotes the Euclidian length. Thus, the spatial sign is a p-variate unit vector. The empirical spatial centered rank of $\mathbf{x}_{ij}$ among the data set $\mathbf{X}$ is defined as

$$\mathbf{R}_N(\mathbf{x}_{ij}) = \frac{1}{N}\sum_{k=1}^{c}\sum_{l=1}^{n_k}\mathbf{S}(\mathbf{x}_{ij}-\mathbf{x}_{kl}).$$

This gives a vector inside the unit sphere pointing from the center of the data cloud $\mathbf{X}$ approximately to the direction of $\mathbf{x}_{ij}$. Spatial ranks are data dependent, but they converge uniformly in probability to their theoretical values $\mathbf{R}(\mathbf{x}_{ij}) = E\left[\mathbf{S}\left(\mathbf{x}_{ij}-\mathbf{x}\right)\right]$, $\mathbf{x}\sim F$.

2 A Multivariate Kruskal-Wallis Test

The hypotheses of interest are

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_c \text{ not all equal.}$$

A classical univariate nonparametric test for the problem is the Kruskal-Wallis test. Hettmansperger et al. (1998) gave a multivariate extension of the Kruskal-Wallis test, identical to the classical test in the univariate case, based on affine invariant ranks. Similar approach can be taken with spatial ranks as outlined next.

Write $\overline{\mathbf{R}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{R}_N(\mathbf{x}_{ij})$ for the group average of empirical spatial ranks with respect to the entire sample. Under the null hypothesis,

$$Q^2 = \sum_{i=1}^{c} n_i \overline{\mathbf{R}}_i^T \widehat{\mathbf{B}}^{-1}\overline{\mathbf{R}}_i \xrightarrow{D} \chi^2_{p(c-1)}$$

(Möttönen and Oja, 1995; Choi and Marden, 1997), where

$$\widehat{\mathbf{B}} = \text{ave}\left\{\mathbf{R}_N(\mathbf{x}_{ij})\mathbf{R}_N(\mathbf{x}_{ij})^T\right\}$$

is a consistent estimator (under H_0) of the spatial rank covariance matrix

$$\mathbf{B} = E_0\left[\mathbf{R}(\mathbf{x}_{ij})\mathbf{R}(\mathbf{x}_{ij})^T\right].$$

3 Estimation of Treatment Effects

Denote the difference between treatments i and j by $\boldsymbol{\Delta}_{ij} = \boldsymbol{\mu}_j - \boldsymbol{\mu}_i$. Natural companion estimator $\widehat{\boldsymbol{\Delta}}_{ij}$ to the test is the multivariate two-sample spatial Hodges-Lehmann estimator, that is, the sample spatial median of the n_in_j pairwise differences $\mathbf{x}_{jk}-\mathbf{x}_{il}$ ($k = 1,\ldots,n_j$; $l = 1,\ldots,n_i$). Note that $\widehat{\boldsymbol{\Delta}}_{ij} = -\widehat{\boldsymbol{\Delta}}_{ji}$ and $\widehat{\boldsymbol{\Delta}}_{ii} = \mathbf{0}$. Möttönen and Oja (1995) state but do not prove the following asymptotic result concerning this estimator:

Theorem 1 *Assume that min* $\{n_i, n_j\} \to \infty$ *and that* $\frac{n_i}{N} \to \lambda_i$ *and* $\frac{n_j}{N} \to \lambda_j$, $0 < \lambda_i, \lambda_j < 1$. *Under general assumptions,*

$$\sqrt{N}(\widehat{\boldsymbol{\Delta}}_{ij} - \boldsymbol{\Delta}_{ij}) \xrightarrow{D} N_p\left(\mathbf{0}, \frac{\lambda_i+\lambda_j}{\lambda_i\lambda_j}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right).$$

Here

$$\mathbf{A} \;=\; E\left[\frac{1}{\|\mathbf{x}_{jk}-\mathbf{x}_{il}-\boldsymbol{\Delta}_{ij}\|}\left(\mathbf{I}_p - \mathbf{S}(\mathbf{x}_{jk}-\mathbf{x}_{il}-\boldsymbol{\Delta}_{ij})\mathbf{S}(\mathbf{x}_{jk}-\mathbf{x}_{il}-\boldsymbol{\Delta}_{ij})^T\right)\right].$$

To find a covariance matrix estimate for $\widehat{\boldsymbol{\Delta}}_{ij}$, the $\mathbf{A}$ matrix can be estimated by

$$\widehat{\mathbf{A}} \;=\; \text{ave}\left\{\frac{1}{\|\mathbf{x}_{jk}-\mathbf{x}_{il}-\widehat{\boldsymbol{\Delta}}_{ij}\|}\left(\mathbf{I}_p - \mathbf{S}(\mathbf{x}_{jk}-\mathbf{x}_{il}-\widehat{\boldsymbol{\Delta}}_{ij})\mathbf{S}(\mathbf{x}_{jk}-\mathbf{x}_{il}-\widehat{\boldsymbol{\Delta}}_{ij})^T\right)\right\},$$

where the average is taken over all possible pairs $(\mathbf{x}_{ij}, \mathbf{x}_{kl})$ from all the samples. An estimate of the $\mathbf{B}$ matrix can be obtained via

$$\widehat{\mathbf{B}} \;=\; \text{ave}\left\{\mathbf{R}_N(\mathbf{x}_{ij}-\widehat{\boldsymbol{\Delta}}_{1i})\mathbf{R}_N(\mathbf{x}_{ij}-\widehat{\boldsymbol{\Delta}}_{1i})^T\right\}.$$

For completeness of this paper, a heuristic proof of the limiting normality is presented in the Appendix. Chaudhuri (1992) considers the spatial one-sample Hodges-Lehmann estimator and Hodges and Lehmann (1963) the univariate two-sample estimation problem.

The inconvenience with the above estimators—just like in the univariate case, or when using multivariate marginal ranks—is that the obtained estimates are not generally compatible in the sense that $\widehat{\boldsymbol{\Delta}}_{ij} = \widehat{\boldsymbol{\Delta}}_{ik} + \widehat{\boldsymbol{\Delta}}_{kj}$.

To overcome this problem, consider competing estimators of treatment effects. An estimator of the difference between the ith and the jth treatment using the kth treatment as a reference is

$$\widetilde{\boldsymbol{\Delta}}_{ij\cdot k} \;=\; \widehat{\boldsymbol{\Delta}}_{ik} + \widehat{\boldsymbol{\Delta}}_{kj}. \tag{1}$$

This type of estimator can be useful in a situation where the treatment effect of interest cannot be estimated directly, but only via a third treatment. Note that $\widetilde{\boldsymbol{\Delta}}_{ij\cdot i} = \widetilde{\boldsymbol{\Delta}}_{ij\cdot j} = \widehat{\boldsymbol{\Delta}}_{ij}$. Taking the average over the treatments groups

$$\widetilde{\boldsymbol{\Delta}}_{ij} \;=\; \frac{1}{c}\sum_{k=1}^{c}\widetilde{\boldsymbol{\Delta}}_{ij\cdot k} \tag{2}$$

yields a generalization of the univariate estimator proposed by Lehmann (1963). The adjusted estimators (1) and (2) are consistent only if $\frac{n_i}{N} \to \lambda_i$, $0 < \lambda_i < 1$, for all $i = 1, \ldots, c$. Estimators that are consistent under a weaker condition that $\frac{n_i}{N} \to \lambda_i$ and $\frac{n_j}{N} \to \lambda_j$ are obtained by weighting the estimators by the relative group size of the reference sample (Spjøtvoll, 1968):

$$\bar{\boldsymbol{\Delta}}_{ij} \;=\; \frac{1}{N}\sum_{k=1}^{c} n_k\widetilde{\boldsymbol{\Delta}}_{ij\cdot k} \tag{3}$$

When $n_1 = \ldots = n_c$ the spatial Spjøtvoll's estimators (3) reduce to the spatial Lehmann's estimators (2).

Theorem 2 *Assume that min* $\{n_i, n_j, n_k\} \to \infty$ *and that* $\frac{n_i}{N} \to \lambda_i$, $\frac{n_j}{N} \to \lambda_j$ *and* $\frac{n_k}{N} \to \lambda_k$, $0 < \lambda_i, \lambda_j, \lambda_k < 1$. *Then, under general assumptions,*

$$\sqrt{N}\left(\widetilde{\boldsymbol{\Delta}}_{ij\cdot k} - \widehat{\boldsymbol{\Delta}}_{ij}\right) \;\xrightarrow{P}\; \mathbf{0}$$

for all i, j, k.

Corollary 1 *Under the assumptions of Theorem 2,*

$$\sqrt{N}\left(\widetilde{\mathbf{\Delta}}_{ij} - \widehat{\mathbf{\Delta}}_{ij}\right) \xrightarrow{P} \mathbf{0} \text{ and}$$
$$\sqrt{N}\left(\bar{\mathbf{\Delta}}_{ij} - \widehat{\mathbf{\Delta}}_{ij}\right) \xrightarrow{P} \mathbf{0}$$

for all i, j, k.

Theorem 2 and Corollary 1 imply that the alignment of the estimates (with respect to the other $c - 2$ treatments) does not alter their limiting distributions. However, it is unclear what happens to the efficiency of the adjusted estimators in finite samples, particularly if the sample sizes are widely disparate. This question will be addressed in the next section.

4 Efficiencies

Recall that the spatial Hodges-Lehmann estimator is much more efficient than the mean difference vector for heavy-tailed distributions, and nearly as efficient at the normal model (Table 1). As the adjusted estimators share the limiting distribution of the spatial Hodges-Lehmann estimator, their asymptotic relative efficiencies are identical as well.

The finite sample efficiencies of the adjusted estimators $\widetilde{\mathbf{\Delta}}_{12\cdot 3}$, $\widetilde{\mathbf{\Delta}}_{12}$ and $\bar{\mathbf{\Delta}}_{12}$ relative to the unadjusted estimator $\widehat{\mathbf{\Delta}}_{12}$ shown in Figures 1, 2 and 3, respectively, are based on simulations from a univariate normal distribution (10000 repetitions) and a bivariate spherical normal distribution (1000 repetitions) for three groups. Efficiencies for spherical distributions in general are likely to be approximately the same. The simulations were conducted in R (R Development Core Team, 2004).

The efficiency of the estimator $\widetilde{\mathbf{\Delta}}_{12\cdot 3}$, based on a reference sample of size 1 ($n_3 = 1$), is approximately 0.7 ($p = 1$) and 0.8 ($p = 2$). At this point, the value is merely the observed relative efficiency of the the difference of the (spatial) medians related to the (spatial) Hodges-Lehmann estimator. Adding a few observations to the third group quickly improves the performance of the estimator $\widetilde{\mathbf{\Delta}}_{12\cdot 3}$. At $n_1 = n_2 = n_3$, the observed relative efficiency is close to unity (Figure 1). Our further simulation studies suggest that, as $n_1 = n_2$ remain fixed and n_3 increases, the finite sample efficiency increases even beyond 1, but only very slightly.

The behavior of the spatial Lehmann's estimator $\widetilde{\mathbf{\Delta}}_{12}$, being the average of the $\widetilde{\mathbf{\Delta}}_{12\cdot k}$ estimators, is very similar (Figure 2). It is superior to the estimator $\widetilde{\mathbf{\Delta}}_{12\cdot 3}$, because it is never worse than 90% efficient, and because it reaches the efficiency of the spatial Hodges-Lehmann estimator much faster. The reasons are easy to see: as $\widetilde{\mathbf{\Delta}}_{12} = \frac{1}{3}(2\widehat{\mathbf{\Delta}}_{12} + \widetilde{\mathbf{\Delta}}_{12\cdot 3})$, the (spatial) Hodges-Lehmann estimator $\widehat{\mathbf{\Delta}}_{12}$ receives the most weight in the computation.

Finally, the spatial Spjøtvoll's estimator $\bar{\mathbf{\Delta}}_{12}$ seems to have the same efficiency as the spatial Hodges-Lehmann estimator $\widehat{\mathbf{\Delta}}_{12}$ when n_3 is small (Figure 3). As n_3 increases and n_1, n_2 remain fixed, the estimator tends to $\widetilde{\mathbf{\Delta}}_{12\cdot 3}$. Thus, for large n_3 it can be slightly better than $\widehat{\mathbf{\Delta}}_{12}$. The weighting procedure enables the spatial Spjøtvoll's estimator $\bar{\mathbf{\Delta}}_{12}$ to protect itself against efficiency losses due to extreme group allocations in both directions, thus making it a superior estimator. If the sample sizes are approximately the same, it does not make a difference which estimator is used.

5 Affine Invariant/Equivariant Versions

The test and the estimators based on spatial ranks are orthogonally invariant and equivariant but not affine invariant and equivariant. However, if $\widehat{\mathbf{V}}$ is an affine equivariant estimator of shape in the sense that $\widehat{\mathbf{V}}(\mathbf{HX}+\mathbf{g1}_N^T) \propto \mathbf{H}\widehat{\mathbf{V}}\mathbf{H}^T$, then

- any orthogonally invariant test computed on the transformed data set $\mathbf{Y} = \widehat{\mathbf{V}}^{-1/2}\mathbf{X}$ is affine invariant (Randles, 2000; Möttönen et al., 2003), and
- any orthogonally equivariant estimator computed on the transformed data set, and retranformed, $\widehat{\mathbf{V}}^{1/2}\widehat{\boldsymbol{\Delta}}(\mathbf{Y})$, is affine equivariant. This procedure is widely known as the transformation retransformation technique (Chakraborty and Chaudhuri, 1996; Chakraborty et al., 1998; Chakraborty and Chaudhuri, 1998).

Therefore, a solution is to perform the multivariate spatial rank test on a transformed data set $\mathbf{Y}$, and to retransform the treatment difference estimates obtained from $\mathbf{Y}$ back to the original scale by $\widehat{\mathbf{V}}^{1/2}$. Here the most natural approach is to use spatial ranks in the estimation of the shape matrix $\mathbf{V}$ (Oja and Randles, 2004). One possibility is to apply a shape matrix defined by the implicit equation

$$p\sum_{i=1}^{c}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{R}_{n_i}(\mathbf{y}_{ij})\mathbf{R}_{n_i}(\mathbf{y}_{ij})^T\right] = \sum_{i=1}^{c}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{R}_{n_i}(\mathbf{y}_{ij})^T\mathbf{R}_{n_i}(\mathbf{y}_{ij})\right]\mathbf{I}_p, \tag{4}$$

where $\mathbf{y}_{ij} = \widehat{\mathbf{V}}^{-1/2}\mathbf{x}_{ij}$ and $\mathbf{R}_{n_i}(\mathbf{y}_{ij})$ is the spatial rank of $\mathbf{y}_{ij}$ among $\mathbf{y}_{i1},\ldots,\mathbf{y}_{in_i}$. After standardization by $\widehat{\mathbf{V}}^{-1/2}$ obtained from (4) the spatial ranks appear as if they were from a spherical distribution. For similar definitions of shape matrices based on spatial signs, see earlier work of Tyler (1987) and Dümbgen (1998). Oja and Randles (2004) also gave an algorithm for the computation of a shape matrix similar to (4). The algorithm always seems to converge, but the actual proof is missing.

Acknowledgements

The research work of Jaakko Nevalainen and Hannu Oja was partially supported by research grants from the Academy of Finland and by the doctoral program of School of Statistical Information, Inference, and Data Analysis (SIIDA).

Appendix

Proof of Theorem 1. Let $\mathbf{x}_1,\ldots,\mathbf{x}_m$ and $\mathbf{y}_1,\ldots,\mathbf{y}_n$ be two independent random samples from $F(\mathbf{x}-\boldsymbol{\mu}_x)$ and $F(\mathbf{y}-\boldsymbol{\mu}_y)$. It is not a restriction to assume that $\boldsymbol{\Delta}_{xy} = \boldsymbol{\mu}_y - \boldsymbol{\mu}_x = \mathbf{0}$. The estimator $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}_{xy}$ satisfies

$$\frac{\sqrt{N}}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{S}(\mathbf{y}_j - \mathbf{x}_i - \widehat{\boldsymbol{\Delta}}) = \mathbf{0}.$$

Suppose that $\widehat{\boldsymbol{\Delta}}$ is $\sqrt{N}$-consistent for $\boldsymbol{\Delta}$, where $N = m+n$. This can be shown in the multivariate case as in Nevalainen et al. (2006). Write $\widehat{\boldsymbol{\Delta}}^* = \sqrt{N}\widehat{\boldsymbol{\Delta}}$. Then the Taylor expansion around $\widehat{\mathbf{\Delta}}^* = \mathbf{0}$ gives

$$\mathbf{0} \;=\; \frac{\sqrt{N}}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{S}(\mathbf{y}_j - \mathbf{x}_i) - \widehat{\mathbf{A}}\widehat{\mathbf{\Delta}}^* + o_P(1),$$

where $\widehat{\mathbf{A}} \xrightarrow{P} \mathbf{A}$ and

$$\sqrt{N}\left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{S}(\mathbf{y}_j - \mathbf{x}_i) - \frac{1}{n}\sum_{j=1}^{n}\mathbf{R}(\mathbf{y}_j) + \frac{1}{m}\sum_{i=1}^{m}\mathbf{R}(\mathbf{x}_i)\right] \xrightarrow{P} \mathbf{0}.$$

Therefore

$$\sqrt{N}\left[\widehat{\mathbf{\Delta}} + \mathbf{A}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}\mathbf{R}(\mathbf{y}_j) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{R}(\mathbf{x}_i)\right)\right] \xrightarrow{P} \mathbf{0}$$

and the result follows.

Proof of Theorem 2. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m$, $\mathbf{y}_1, \ldots, \mathbf{y}_n$ and $\mathbf{z}_1, \ldots, \mathbf{z}_l$ be three independent random samples from $F(\mathbf{x})$. Let $N = m + n + l$. By Theorem 1

$$\sqrt{N}\left[\widehat{\mathbf{\Delta}}_{xz} + \mathbf{A}^{-1}\left(\frac{1}{l}\sum_{k=1}^{l}\mathbf{R}(\mathbf{z}_k) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{R}(\mathbf{x}_i)\right)\right] \xrightarrow{P} \mathbf{0},$$

and

$$\sqrt{N}\left[\widehat{\mathbf{\Delta}}_{zy} + \mathbf{A}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}\mathbf{R}(\mathbf{y}_j) - \frac{1}{l}\sum_{k=1}^{l}\mathbf{R}(\mathbf{z}_k)\right)\right] \xrightarrow{P} \mathbf{0}.$$

But then simply

$$\sqrt{N}\left[\widehat{\mathbf{\Delta}}_{xy\cdot z} + \mathbf{A}^{-1}\left(\frac{1}{n}\sum_{j=1}^{n}\mathbf{R}(\mathbf{y}_j) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{R}(\mathbf{x}_i)\right)\right] \xrightarrow{P} \mathbf{0}$$

and therefore $\sqrt{N}\widehat{\mathbf{\Delta}}_{xy\cdot z}$ and $\sqrt{N}\widehat{\mathbf{\Delta}}_{xy}$ have the same limiting distribution.

References

B. Chakraborty and P. Chaudhuri. On a transformation and re-transformation technique for constructing affine equivariant multivariate median. In *Proceedings of the American Mathematical Society*, volume 124, pages 2539–2547, 1996.

B. Chakraborty and P. Chaudhuri. On an adaptive transformation and retransformation estimate of multivariate location. *Journal of the Royal Statistical Society, Series B*, 60: 145–157, 1998.

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

P. Chaudhuri. Multivariate location estimation using extension of R-estimates through U-statistics type approach. *The Annals of Statistics*, 20:897–916, 1992.

K. Choi and J. Marden. An approach to multivariate rank tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 92:1581–1590, 1997.

L. Dümbgen. On Tyler's M-functional of scatter in high dimension. *Annals of the Institute of Statistical Mathematics*, 50:471–491, 1998.

T. P. Hettmansperger, J. Möttönen, and H. Oja. Affine invariant multivariate rank tests for several samples. *Statistica Sinica*, 8:785–800, 1998.

J. L. Hodges and E. L. Lehmann. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34:598–611, 1963.

E. L. Lehmann. Robust estimation in analysis of variance. *The Annals of Mathematical Statistics*, 34:957–966, 1963.

J. Möttönen and H. Oja. Multivariate spatial sign and rank methods. *Nonparametric Statistics*, 5:201–213, 1995.

J. Möttönen, H. Oja, and J. Tienari. On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, 25:542–552, 1997.

J. Möttönen, J. Hüsler, and H. Oja. Multivariate nonparametric tests in a randomized complete block design. *Journal of Multivariate Analysis*, 85:106–129, 2003.

J. Nevalainen, D. Larocque, and H. Oja. The multivariate spatial median for clustered data. *Conditionally accepted to the Canadian Journal of Statistics*, 2006.

H. Oja and R. H. Randles. Multivariate nonparametric tests. *Statistical Science*, 19: 598–605, 2004.

R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2004. URL `http://www.R-project.org`. ISBN 3-900051-00-3.

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

E. Spjøtvoll. A note on robust estimation in analysis of variance. *The Annals of Mathematical Statistics*, 39:1486–1492, 1968.

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

Table 1: Asymptotic relative efficiency of the spatial Hodges-Lehmann estimator relative to the mean difference vector under multivariate t-distributions (Möttönen et al., 1997).

	Degrees of freedom			
Dimension	3	6	10	∞
1	1.900	1.164	1.054	0.955
2	1.953	1.187	1.071	0.967
3	1.994	1.200	1.081	0.973
6	2.050	1.219	1.095	0.984
10	2.093	1.229	1.103	0.989

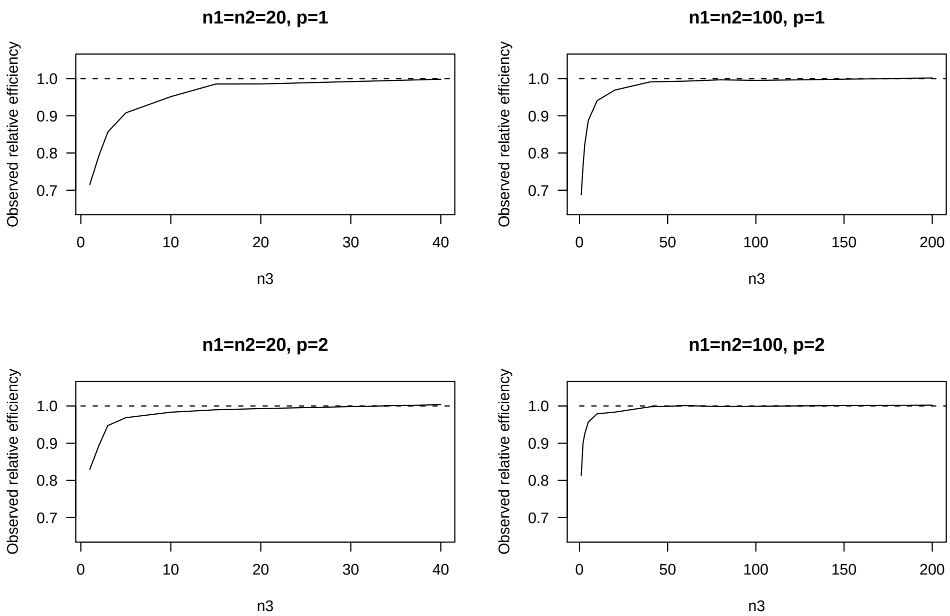

Figure 1: The observed relative efficiency of $\widetilde{\boldsymbol{\Delta}}_{12\cdot 3}$ relative to $\widehat{\boldsymbol{\Delta}}_{12}$.

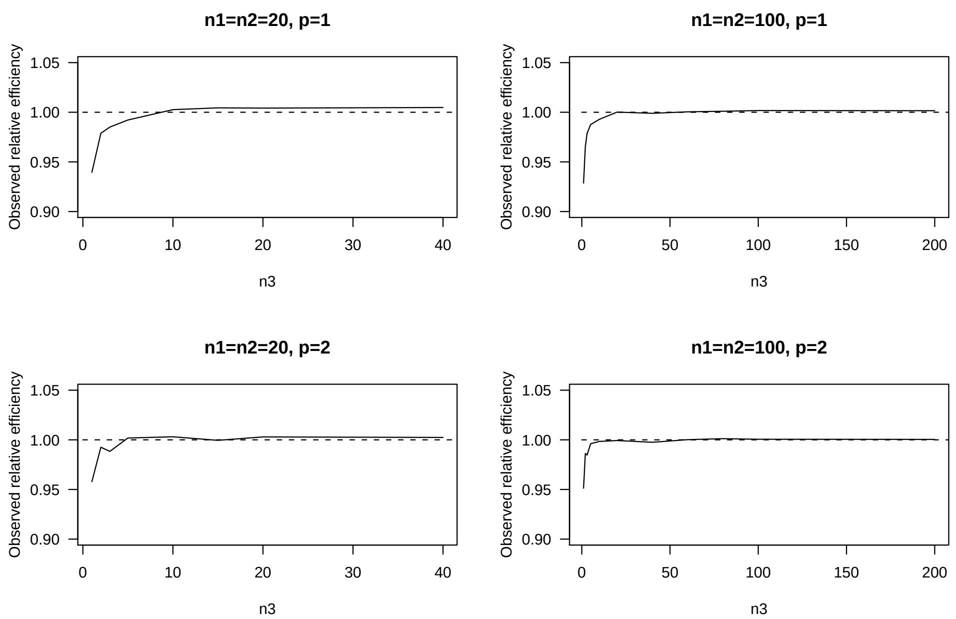

Figure 2: The observed relative efficiency of $\widetilde{\mathbf{\Delta}}_{12}$ relative to $\widehat{\mathbf{\Delta}}_{12}$.

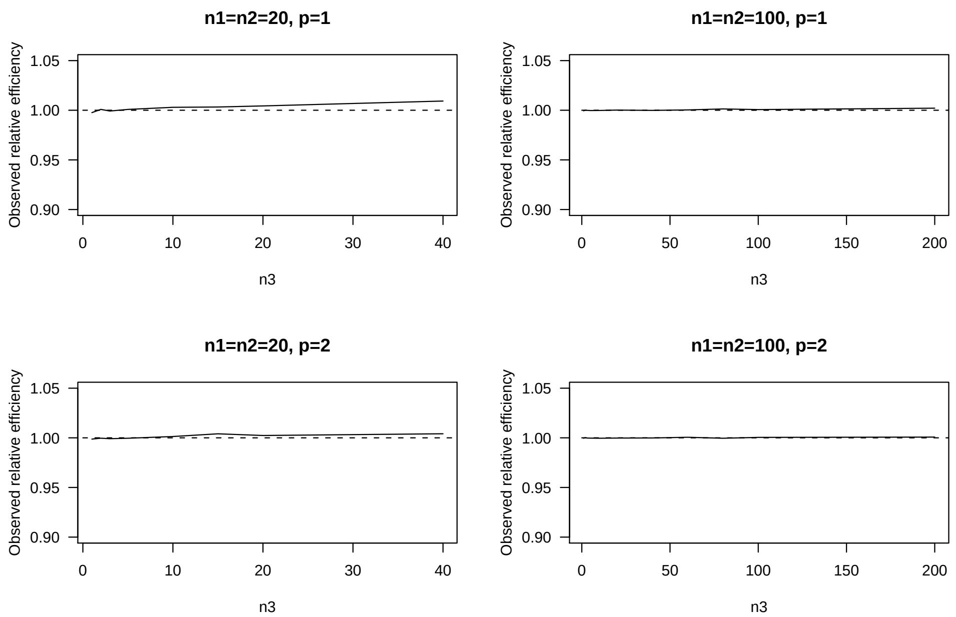

Figure 3: The observed relative efficiency of $\bar{\mathbf{\Delta}}_{12}$ relative to $\widehat{\mathbf{\Delta}}_{12}$.

Journal of Statistical Software

May 2006, Volume 16, Issue 5. http://www.jstatsoft.org/

SAS/IML Macros for a Multivariate Analysis of Variance Based on Spatial Signs

Jaakko Nevalainen
University of Tampere

Hannu Oja
University of Tampere

Abstract

Recently, new nonparametric multivariate extensions of the univariate sign methods have been proposed. Randles (2000) introduced an affine invariant multivariate sign test for the multivariate location problem. Later on, Hettmansperger and Randles (2002) considered an affine equivariant multivariate median corresponding to this test. The new methods have promising efficiency and robustness properties. In this paper, we review these developments and compare them with the classical multivariate analysis of variance model. A new SAS/IML tool for performing a spatial sign based multivariate analysis of variance is introduced.

Keywords: affine invariance/equivariance, spatial sign, multivariate analysis of variance, multivariate sign test, multivariate median, SAS.

1. Introduction

Classical statistical techniques for multivariate location problems such as Hotelling's T^2 tests, multivariate analysis of variance (MANOVA) and multivariate multiple regression analysis rely on the assumption that the data were from a multivariate normal distribution. The inference methods are then based on the assumption of multivariate normality, the sample mean vector and the sample covariance matrix. However, these methods are extremely sensitive to outlying observations and they are inefficient for heavy tailed noise distributions.

Möttönen and Oja (1995) reviewed multivariate sign and rank tests and the corresponding estimates based on the L_1-type objective function. The tests and estimates were rotation invariant and equivariant, but not affine invariant/equivariant. Recently, new nonparametric multivariate extensions of the univariate sign methods have been proposed. Randles (2000) developed an affine invariant one-sample multivariate sign test. Hettmansperger and Randles (2002) considered an affine equivariant multivariate median corresponding to this test. Their approach combines the simultaneous use of the spatial median (Brown 1983), Tyler's M-estimate of scatter (Tyler 1987) and the transformation-retransformation technique (Chakraborty, Chaudhuri, and Oja 1998). Chakraborty *et al.* (1998) used a similar idea as Hettmansperger and Randles (2002), but not Tyler's scatter matrix. Like Randles' test, the Hettmansperger and Randles (2002) estimate is fairly easy to compute.

In this paper, we will first recall the classical MANOVA model. In Section 3 we review the multivariate spatial sign tests and estimators analoguous to their classical alternatives. In addition, we outline some ideas how to approximate the precision of the estimates. Section 4 introduces new SAS macros written in Interactive Matrix Language (IML) for performing the analysis. As far as the authors are aware, these procedures are not currently available in standard software packages. Finally, the use of the SAS/IML tools is illustrated by an example. The complete SAS/IML code is available at http://www.jstatsoft.org/v16/i05/.

In the following sections we will assume that there are c independent random samples of p-dimensional observations. Let

$$\mathbf{X} = (\mathbf{x}_{11} \cdots \mathbf{x}_{1n_1}\ \mathbf{x}_{21} \cdots \mathbf{x}_{2n_2}\ \cdots\ \mathbf{x}_{c1} \cdots \mathbf{x}_{cn_c})$$

denote the $p \times N$ data matrix, where $\mathbf{x}_{ij} = (x_{ij1}\ x_{ij2} \cdots x_{ijp})^\top$ represents the jth observation of the ith sample. Further write $N = n_1 + \cdots + n_c$ for the total number of observations. In practice, the data matrix is often given as a transpose of $\mathbf{X}$. We are interested in drawing conclusions on the parameter set $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}$, where $\boldsymbol{\mu}_i$ denotes the center of symmetry of the ith sample, and $\boldsymbol{\Sigma}$ the covariance (or scatter) matrix (assumed to be common for all the samples). Alternatively, one may wish to parametrize the model by $\boldsymbol{\mu}_1, \boldsymbol{\Delta}_{12}, \ldots, \boldsymbol{\Delta}_{1c}, \boldsymbol{\Sigma}$, where $\boldsymbol{\Delta}_{1i} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_1$ represents the difference between sample i and the first sample used as a reference (e.g. placebo). In general, we wish to estimate both sets of parameters, and construct the associated location tests.

Let $\mathbf{B}$ denote a nonsingular $p \times p$ matrix and $\mathbf{b}$ a $p \times 1$ vector. A location estimate $\hat{\boldsymbol{\mu}}_i(\mathbf{X})$ and a scatter matrix estimate $\hat{\boldsymbol{\Sigma}}(\mathbf{X})$ are affine equivariant if

$$\begin{aligned} \hat{\boldsymbol{\mu}}_i\left(\mathbf{B}\mathbf{X} + \mathbf{b}\mathbf{1}_N^\top\right) &= \mathbf{B}\hat{\boldsymbol{\mu}}_i(\mathbf{X}) + \mathbf{b} \text{ and} \\ \hat{\boldsymbol{\Sigma}}\left(\mathbf{B}\mathbf{X} + \mathbf{b}\mathbf{1}_N^\top\right) &= \mathbf{B}\hat{\boldsymbol{\Sigma}}(\mathbf{X})\mathbf{B}^\top. \end{aligned}$$

A test statistic $\mathbf{T}(\mathbf{X})$ is affine invariant if

$$\mathbf{T}\left(\mathbf{B}\mathbf{X} + \mathbf{b}\mathbf{1}_N^\top\right) = \mathbf{T}(\mathbf{X}).$$

These definitions simply mean that a rescaling, a rotation or a shift of the data should results into corresponding transformation in the estimates, but the value of the test statistic should remain unchanged.

2. Classical MANOVA

When more than one attribute is measured per observational unit and the observational units arise from independent populations, the design is typically analyzed by multivariate analysis of variance techniques. Classical MANOVA assumption is that the outcome vectors $\mathbf{x}_{ij}$ $(p \times 1)$ are generated from the model

$$\mathbf{x}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{ij},$$

where $\boldsymbol{\mu}_i = (\mu_{i1}\,\mu_{i2}\,\cdots\,\mu_{ip})^\top$ is the location center for the ith sample (population), and $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}\,\varepsilon_{ij2}\,\cdots\,\varepsilon_{ijp})^\top$ are independent and identically distributed random errors from a multivariate normal distribution $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. In a one-sample case, the classical test for the location problem is Hotelling's T^2 test.

Lemma 1 *Hotelling's T^2 statistic for testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is*

$$T^2 = N\,\bar{\mathbf{x}}^\top \mathbf{S}^{-1}\bar{\mathbf{x}},$$

where $\bar{\mathbf{x}}$ is the sample mean vector and $\mathbf{S}$ is the sample covariance matrix. Furthermore,

$$\frac{N-p}{(N-1)p}\,T^2 \text{ has an } F_{p,N-p} \text{ distribution.}$$

In a multisample case, the interest is to test the null hypothesis of no difference in location between the samples

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c$$

assuming a common covariance matrix $\boldsymbol{\Sigma}$. Under the null hypothesis, the maximum likelihood estimator of a joint $\boldsymbol{\mu}$ is the sample mean vector over the combined sample, and the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is the pooled sample covariance matrix. For hypothesis testing, we may use the two-sample Hotelling's T^2 statistic, or in a more general c-sample case, the Hotelling's trace statistic:

Lemma 2 *Hotelling's trace statistic for testing $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c$ is*

$$T^2 = (N-c)\,\mathrm{Tr}\left(\mathbf{B}\mathbf{W}^{-1}\right),$$

where $\mathbf{B}$ is the between-samples sums of squares matrix and $\mathbf{W}$ the within-samples sums of squares matrix. Under the null hypothesis, the test statistic is asymptotically $\chi^2_{p(c-1)}$ distributed.

Write

$$\mathbf{z}_{ij} = \left(\frac{1}{N-c}\mathbf{W}\right)^{-1/2}(\mathbf{x}_{ij} - \bar{\mathbf{x}})$$

for standardized observations with the sample mean vector zero and the sample covariance matrix $\mathbf{I}_p$, and

$$\bar{\mathbf{z}}_i = \left(\frac{1}{N-c}\mathbf{W}\right)^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})$$

for their sample means. We can write

$$\begin{aligned}
(N-c)\,\mathrm{Tr}\left(\mathbf{B}\mathbf{W}^{-1}\right) &= (N-c)\sum_{i=1}^{c} n_i \mathrm{Tr}\left((\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^\top \mathbf{W}^{-1}\right) \\
&= (N-c)\sum_{i=1}^{c} n_i \mathrm{Tr}\left(\mathbf{W}^{-1/2}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^\top \mathbf{W}^{-1/2}\right) \\
&= \sum_{i=1}^{c} n_i \mathrm{Tr}\left(\bar{\mathbf{z}}_i \bar{\mathbf{z}}_i^\top\right) = \sum_{i=1}^{c} n_i \|\bar{\mathbf{z}}_i\|^2. \qquad (1)
\end{aligned}$$

Note that the limiting distribution is still $\chi^2_{p(c-1)}$ if the covariance matrix estimate $(N-c)^{-1}\mathbf{W}$ is replaced by the regular pooled sample covariance matrix $\mathbf{S}$. We will observe similarities between the trace statistic and a multivariate spatial sign test statistic later on.

3. Spatial sign MANOVA

In this section we present sign based competitors to the Hotelling's T^2 statistic and to the sample mean vector. Estimation and test constructions are based on the spatial signs of suitably standardized outcome vectors.

Multivariate extension of the sign concept, the spatial sign vector, is defined as

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|^{-1}\mathbf{x} & \text{if } \mathbf{x} \neq \mathbf{0} \\ \mathbf{0} & \text{if } \mathbf{x} = \mathbf{0} \end{cases}$$

where $\|\mathbf{x}\| = (\mathbf{x}^\top \mathbf{x})^{1/2}$ is the Euclidean length of vector $\mathbf{x}$. Spatial signs are clearly rotation equivariant but not affine equivariant.

Let $\mathbf{V}$ denote the scatter matrix defined by Tyler (1987), which is the solution to

$$\mathsf{E}\left(\frac{\mathbf{V}^{1/2}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{V}^{1/2}}{(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})}\right) = \frac{1}{p}\mathbf{I}_p.$$

Tyler's scatter matrix is affine equivariant, but unique only up to a multiplication by a constant; we will choose the symmetric version with $\mathrm{Tr}(\mathbf{V}) = p$. For a sign based analysis, it suffices to standardize by $\mathbf{z}_{ij} = \hat{\mathbf{V}}^{-1/2}(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)$. A location estimate is needed as well, and its selection will be discussed in the subsequent sections. The standardization is an analogue to the Mahalanobis transformation in the classical multivariate analysis, but instead of standardizing the sample variance-covariance matrix of the original data, this standardization produces a standardized variance-covariance matrix for the spatial sign vectors. For standardized data, the sign vectors then tend to lie uniformly on the unit sphere (see Figures 1, 2 and 3). Denote the direction vectors by $\mathbf{u}_{ij} = \mathbf{S}(\mathbf{z}_{ij})$ and the radius by $r_{ij} = \|\mathbf{z}_{ij}\|$.

Again assume that the outcome vectors are generated from

$$\mathbf{x}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{ij},$$

where the residuals can be decomposed as $\boldsymbol{\varepsilon}_{ij} = \boldsymbol{\Sigma}^{1/2} r_{ij}\mathbf{u}_{ij}$. Moving roughly from strong to minimal conditions, different model assumptions of the underlying distribution in terms of the direction vector $\mathbf{U}_{ij}$ and the radius $R_{ij} \geq 0$ can be listed as follows (Randles 2000):

1. Multivariate normal
   - $\mathbf{U}_{ij}$ is uniformly distributed on a p-dimensional unit sphere,
   - $R^2_{ij} \sim \chi^2_p$, and
   - $\mathbf{U}_{ij}$ and R_{ij} are independent.
2. Elliptical symmetry
   - $\mathbf{U}_{ij}$ is uniformly distributed on a p-dimensional unit sphere and

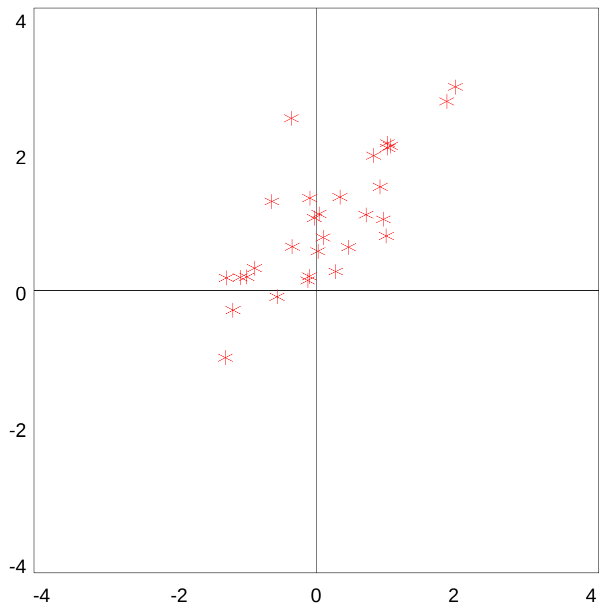

Figure 1: Observations $\mathbf{x}_j$ from a bivariate normal distribution

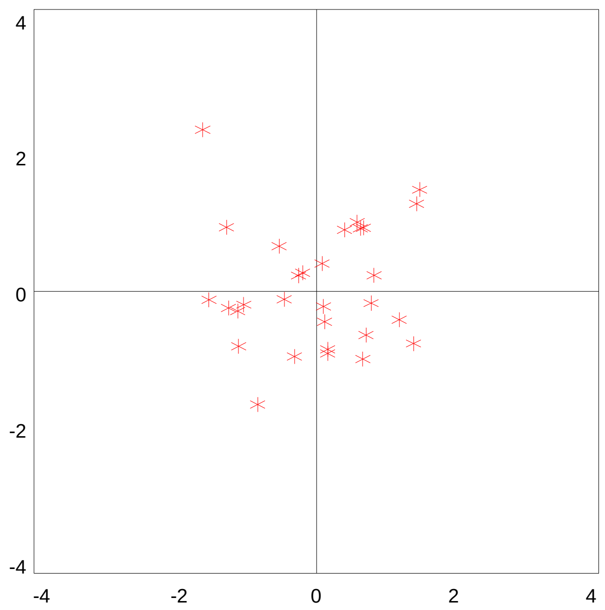

Figure 2: Standardized observations $\mathbf{z}_j$

- $\mathbf{U}_{ij}$ and R_{ij} are independent.

3. Elliptical directions

- $\mathbf{U}_{ij}$ is uniformly distributed on a p-dimensional unit sphere.

4. Symmetry

- $R_{ij}\mathbf{U}_{ij}$ has the same distribution as $-R_{ij}\mathbf{U}_{ij}$.

5. Directional symmetry

- $\mathbf{U}_{ij}$ has the same distribution as $-\mathbf{U}_{ij}$.

The families are not subsets of each other: for a hierarchy between these symmetry assumptions see Randles (2000). Multivariate spatial sign methods are typically distribution-free in the family of elliptical directions. If the underlying distribution is skewed, the location parameter $\boldsymbol{\mu}$ is the population median vector rather than the mean vector (symmetry center in models 1, 2 and 4). Similarly, $\boldsymbol{\Sigma}$ is the covariance matrix in the multivariate normal model, and proportional to the covariance matrix (if it exists) in the elliptical symmetry model.

3.1. One-sample case

Consider testing the null hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against the alternative hypothesis $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ (without loss of generality). For standardized signs $\mathbf{u}_j = \mathbf{S}\left(\hat{\mathbf{V}}^{-1/2}(\mathbf{x}_j - \mathbf{0})\right)$, seek an estimate of Tyler's scatter matrix as the solution to the implicit equation

$$\text{ave}\left\{\mathbf{u}_j\mathbf{u}_j^\top\right\} = \frac{1}{p}\mathbf{I}_p.$$

Obviously, the estimate is not influenced by r_j. Hence, a distribution-free test in the family of elliptical directions is given by

Lemma 3 *Under the null hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$, the limiting distribution of the multivariate spatial sign test statistic*

$$Q^2 = Np\|\text{ave}\{\mathbf{u}_j\}\|^2$$

is χ^2 with p degrees of freedom.

The development was given by Randles (2000). The test statistic Q^2 is affine invariant.

For small samples, Randles (2000) proposes the use of a sign change test. For the family of directionally symmetric distributions, it leads into a conditionally distribution-free test. Let $\mathbf{U}$ denote a $p \times N$ matrix with $\mathbf{u}_j$ as the jth column. Furthermore, let $\mathbf{S}_1, \ldots, \mathbf{S}_M$, denote independent random $N \times N$ diagonal sign change matrices with 2^N equiprobable values of $\text{diag}(\pm 1, \ldots, \pm 1)$. Since $\hat{\mathbf{V}}$ is sign change invariant, the p-value can be estimated by

$$\hat{p} = \frac{\#\{Q^2(\mathbf{U}\mathbf{S}_m) \geq Q^2(\mathbf{U})\}}{M},$$

that is, by the proportion of cases where $Q^2(\mathbf{U}\mathbf{S}_m) \geq Q^2(\mathbf{U})$, $m = 1, ..., M$.

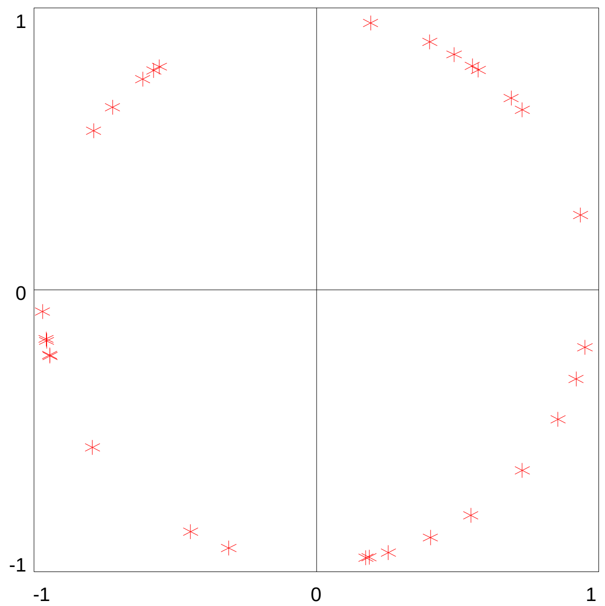

Figure 3: Direction vectors $\mathbf{u}_j$

Möttönen, Oja, and Tienari (1997) studied the limiting efficiency of multivariate sign tests for multivariate t-distributions. They show that the efficiency relative to Hotelling's test is 0.785 even in a bivariate normal case (∞ degrees of freedom). In four dimensions, they obtained relative efficiencies of 0.884, 1.051 and 2.250 for ∞, 10 and 4 degrees of freedom, respectively. In dimension 10, the same figures were 0.951, 1.131 and 2.422. See also Randles (1989) for the family of elliptically symmetric power family distributions.

Hettmansperger and Randles (2002) introduced the simultaneous estimation of location and scatter in the one-sample case. They computed a multivariate location estimate and a scatter matrix estimate to satisfy

$$\text{ave}\{\mathbf{u}_j\} = \mathbf{0} \text{ and } \text{ave}\{\mathbf{u}_j\mathbf{u}_j^\top\} = \frac{1}{p}\mathbf{I}_p \tag{2}$$

for standardized signs $\mathbf{u}_j = \mathbf{S}\left(\hat{\mathbf{V}}^{-1/2}(\mathbf{x}_j - \hat{\boldsymbol{\mu}})\right)$. The solutions to the equations are the transformation-retransformation spatial median and Tyler's scatter matrix, respectively. Standardization by the resulting location and scatter estimates distributes direction vectors uniformly into a unit sphere centered at $\mathbf{0}$ (Figure 3). Another important property of the estimates is that they are affine equivariant. The property is reached by the above utilization of the transformation-retransformation procedure (Chakraborty *et al.* 1998).

3.2. Several samples case

Next consider c independent random samples with cumulative distribution functions $F(\mathbf{x} - \boldsymbol{\mu}_1), F(\mathbf{x} - \boldsymbol{\mu}_2), \ldots, F(\mathbf{x} - \boldsymbol{\mu}_c)$, i.e. it is assumed that the underlying distributions have a joint scatter matrix and differ only in location. Our interest is to test the null hypothesis of no treatment difference

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c$$

or, equivalently,

$$H_0 : \mathbf{\Delta}_{12} = \cdots = \mathbf{\Delta}_{1c} = \mathbf{0}.$$

Furthermore, we wish to estimate the centers of symmetry $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_c$ for each sample, and the treatment differences $\mathbf{\Delta}_{12}, \ldots, \mathbf{\Delta}_{1c}$ with respect to a reference location $\boldsymbol{\mu}_1$.

Start by constructing the standardized sign vectors $\mathbf{u}_{ij} = \mathbf{S}\left(\hat{\mathbf{V}}^{-1/2}(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}})\right)$, where $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{V}}$ are the null case estimates (obtained as in the one-sample estimation case). Then, if $\hat{\mathbf{V}}$ is a $\sqrt{N}$-consistent estimate and $\hat{\boldsymbol{\mu}}$ the corresponding transformation-retransformation spatial median, we have for elliptical F that

Lemma 4 *Under the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_c$, the multisample multivariate spatial sign test statistic*

$$Q^2 = p \sum_{i=1}^{c} n_i \|\text{ave}_j\{\mathbf{u}_{ij}\}\|^2 \tag{3}$$

has a limiting χ^2 distribution with $p\,(c-1)$ degrees of freedom.

("ave$_j$" means the average taken over j.) A conditionally distribution-free test can be obtained by permuting (Oja and Randles 2004): Let $\mathbf{P}_1, \ldots, \mathbf{P}_M$ denote random $N \times N$ permutation matrices with $N!$ equiprobable values obtained by permuting the rows of an identity matrix $(N \times N)$. As $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{V}}$ are permutation invariant, p-value can be estimated as

$$\hat{p} = \frac{\#\{Q^2(\mathbf{UP}_m) \geq Q^2(\mathbf{U})\}}{M}$$

where $\mathbf{U}$ is the data set consisting of standardized signs.

The test statistic resembles the Hotelling's trace test statistic (1) in a classical MANOVA setting. But (3) is based on the directions only. For limiting efficiencies, see Randles (1989) and Möttönen *et al.* (1997).

Figure 4 displays an illustration of a bivariate non-null case. The direction vectors of the two samples are concentrated on different parts of the unit circle.

Estimation is extended to a c-sample case as follows. Choose $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_c$ and $\hat{\mathbf{V}}$ so that they satisfy

$$\text{ave}\{\mathbf{u}_{1j}\} = \cdots = \text{ave}\{\mathbf{u}_{cj}\} = \mathbf{0} \text{ and } \text{ave}\left\{\mathbf{u}_{ij}\mathbf{u}_{ij}^\top\right\} = \frac{1}{p}\mathbf{I}_p$$

for standardized signs $\mathbf{u}_{ij} = \mathbf{S}\left(\hat{\mathbf{V}}^{-1/2}(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)\right)$. The resulting estimates are the sample transformation-retransformation spatial medians utilizing a joint Tyler's scatter matrix. Due to the affine equivariance property, the differences $\mathbf{\Delta}_{12}, \ldots, \mathbf{\Delta}_{1c}$ can be constructed as the differences of the transformation-retransformation spatial medians.

3.3. Estimation of accuracy

The following asymptotic result gives a way to approximate the precision of the estimates.

Lemma 5 *In the elliptically symmetric case*

$$\sqrt{N}\,(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \longrightarrow_D N_p\left(\mathbf{0}, \frac{p}{(p-1)^2}\;\left[\mathsf{E}\left[r^{-1}\right]\right]^{-2}\mathbf{V}\right).$$

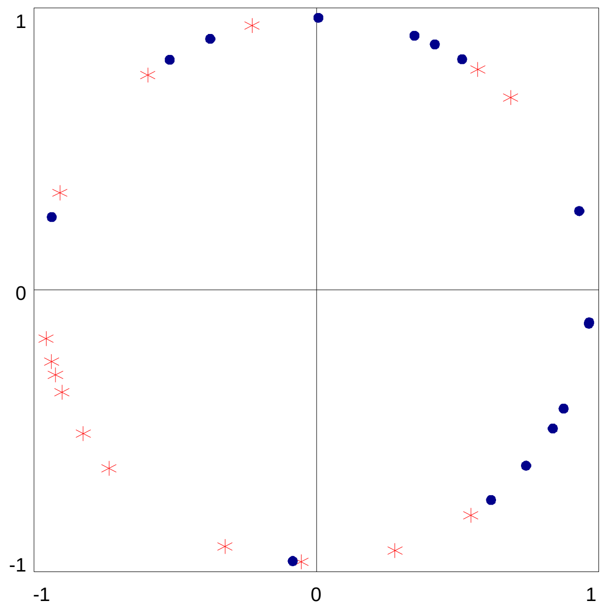

Figure 4: Direction vectors $\mathbf{u}_{ij}$ from two bivariate normal distributions

Therefore, an estimate of the covariance matrix is achieved by

$$\widehat{\mathsf{COV}}(\hat{\boldsymbol{\mu}}) = \frac{p}{N(p-1)^2} \left[\mathrm{ave}\{r_j^{-1}\}\right]^{-2} \hat{\mathbf{V}}.$$

See for example Ollila, Oja, and Croux (2003b), Ollila, Hettmansperger, and Oja (2003a), and Hettmansperger and Randles (2002). In the case of several samples, an estimate $\widehat{\mathsf{COV}}(\hat{\boldsymbol{\mu}}_i)$ is obtained by replacing N by n_i in Lemma 5. The covariance matrix estimate of $\hat{\boldsymbol{\Delta}}_{1j}$ is easily obtained as $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_c$ are asymptotically independent.

Another possibility to estimate precision is to use distribution-free methods such as bootstrapping and delete-1 jackknife. These methods are quite attractive since they require no assumption of the underlying distribution or, assuming that some basic prerequisites are fulfilled, a large sample size.

To get a bootstrap covariance matrix estimate, generate bootstrap samples $\mathbf{X}_1^*, \ldots, \mathbf{X}_B^*$ by sampling (with replacement) from the observed sample $\mathbf{X}$, keeping sample size fixed. Then compute the desired estimate from each bootstrap sample.

Lemma 6 *A bootstrap estimator of the covariance matrix of* $\hat{\boldsymbol{\mu}}$ *is*

$$\widehat{\mathsf{COV}}(\hat{\boldsymbol{\mu}}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\boldsymbol{\mu}}_b^* - \ \mathrm{ave}\{\hat{\boldsymbol{\mu}}^*\}\right) \left(\hat{\boldsymbol{\mu}}_b^* - \ \mathrm{ave}\{\hat{\boldsymbol{\mu}}^*\}\right)^\top$$

where $\hat{\boldsymbol{\mu}}_b^* = \hat{\boldsymbol{\mu}}(\mathbf{X}_b^*)$ *is the location estimate from the bth bootstrap sample.*

In case of more than one sample, we wish to make use of the model assumption of a common scatter matrix $\mathbf{V}$. After standardization by the estimates $\hat{\boldsymbol{\mu}}_i$ and $\hat{\mathbf{V}}$, $\mathbf{z}_{ij}$ vectors are approximately ”independent and identically distributed”. The idea is to sample (with replacement) from the data set (as if it were one sample)

$$\mathbf{Z} = \{\mathbf{z}_{11}, \ldots, \mathbf{z}_{1n_1}, \cdots, \mathbf{z}_{c1}, \ldots, \mathbf{z}_{cn_c}\}.$$

Then transform each $\mathbf{z}^*_{ij}$ back to obtain $\mathbf{x}^*_{ij} = \hat{\mathbf{V}}^{1/2}\mathbf{z}^*_{ij} + \hat{\boldsymbol{\mu}}_i$. These vectors then constitute the bootstrap sample $\mathbf{X}^* = \left(\mathbf{x}^*_{ij}\right)$. Then we can proceed as usual.

Note that some healthy caution is needed with bootstrapping. As pointed out by Stromberg (1997), there is in fact a "high" probability of generating a single bootstrap sample with an unusually large proportion of outlying observations. This proportion might even exceed the breakdown point of the estimator. Thus, even for robust methods, bootstrap estimation may sometimes fail in the presence of outliers. Yet another problem could arise when the sample size is small compared to the dimension of the data.

To overcome possible limitations of bootstrapping, an alternative approach is a delete-1 jackknife estimator.

Lemma 7 *The delete-1 jackknife estimator of covariance matrix of $\hat{\boldsymbol{\mu}}$ in the one-sample case is*

$$\widehat{\mathsf{COV}}(\hat{\boldsymbol{\mu}}) = \frac{N-1}{N} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\mu}}^{(i)} - \hat{\boldsymbol{\mu}}\right) \left(\hat{\boldsymbol{\mu}}^{(i)} - \hat{\boldsymbol{\mu}}\right)^\top$$

where $\hat{\boldsymbol{\mu}}^{(i)}$ is the location estimate from a sample without the ith observation.

We have not used jackknife methods in a case of several samples.

Delete-1 jackknife does not always work well, for example, in conjunction with a nonsmooth estimator such as the vector of marginal medians (Shao and Wu 1989). However, delete-1 jackknife appears to perform nicely with the transformation-retransformation spatial median.

4. SAS/IML modules

The programs are organized as macros, which consist of frequently used modules and the master code. This section outlines the functionality of the modules, so that an advanced user can modify and make further use of them. The SAS/IML programs (`sgnmanova_1.sas` and `sgnmanova_c.sas`) are available at http://www.jstatsoft.org/v16/i05/.

4.1. Modules for estimation of location and scatter

Modules `estimate_1` (one-sample case) and `estimate_c` (c-sample case) perform the estimation procedure. The estimation algorithm uses the steps

1. Compute the direction vectors $\mathbf{u}_{ij}$ by the current estimate values.
2. Update $\hat{\mathbf{V}}$.
3. Update $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_c$.
4. Return to 1 and continue until convergence.

Vector of the componentwise medians and a $p \times p$ identity matrix are used as starting values. Iteration steps are given by

$$\begin{aligned}\hat{\mathbf{V}} &\leftarrow p\,\hat{\mathbf{V}}^{1/2}\,\text{ave}_{ij}\left\{\mathbf{u}_{ij}\mathbf{u}_{ij}^{\top}\right\}\hat{\mathbf{V}}^{1/2} \text{ and}\\ \hat{\boldsymbol{\mu}}_i &\leftarrow \hat{\boldsymbol{\mu}}_i + \left[\text{ave}_j\left\{r_{ij}^{-1}\right\}\right]^{-1}\hat{\mathbf{V}}^{1/2}\text{ave}_j\{\mathbf{u}_{ij}\}.\end{aligned}$$

(Hettmansperger and Randles 2002; Vardi and Zhang 2001; Oja and Randles 2004). The symmetric transformation matrix $\hat{\mathbf{V}}^{-1/2}$ is found via the spectral decomposition of the matrix $\hat{\mathbf{V}}$.

We have also implemented a protection against landing iteration on a data point (Vardi and Zhang 2001). Their modification ensures that iteration moves on even then. The need for such protection is rare, but it does have practical value in bootstrapping, since—for some bootstrap samples—the iteration might encounter a large mass of data on a single point.

`estimate_1` Input for the module are the data matrix and the desired level of precision. The module returns a $(p+1) \times p$ matrix where the first row is the location estimate and the remaining rows consist of the scatter matrix estimate.

`estimate_c` Input for the module are the data matrix, the desired level of precision and the number of samples. The module returns a $(p+c) \times p$ matrix where the first c rows are the location estimates and the remaining rows consist of the scatter matrix estimate.

4.2. Modules for hypothesis testing

Modules for testing the null hypothesis are named as `test_1` (one-sample case) and `test_c` (multisample case).

`test_1` Input for the module are the data matrix, the desired level of precision and the number of sign change permutations. Module estimates the scatter matrix under H_0 (fixed location), and returns a 1×4 vector with value of the test statistic, a p-value based on the limiting distribution, a p-value based on a sign change permutation distribution and its standard error (from a binomial distribution) as elements.

`test_c` Input for the module are the data matrix, the desired level of precision and the number of permutations. Calls the `estimate_1` module. The module returns a 1×4 vector with value of the test statistic, a p-value based on the limiting distribution, a p-value based on a permutation distribution and its standard error (from binomial distribution) as elements.

Small number of permutations guarantees a reasonable computation time.

4.3. Modules for estimation of accuracy

Module `asymptotic` estimates the covariance matrix of $\hat{\boldsymbol{\mu}}$ based on the limiting distribution. Similarly, module `bootstrap` estimates the covariance matrix by bootstrapping, and module `jackknife` estimates it by the delete-1 jacknife. Note that `jackknife` module is available only for the one-sample case.

`asymptotic` Input for the module consist of the data matrix, the estimated parameters values and the desired level of precision. In a multisample case the number of samples has to be given as well. The module returns a $p \times p$ covariance matrix estimate in a one-sample case, and a $cp \times (p+1)$ matrix in a multisample case, where the first column identifies the rows which contain the covariance matrix estimate of $\hat{\boldsymbol{\mu}}_i$.

`bootstrap` Input for the module consist of the data matrix, the desired level of precision and the number of bootstrap samples. In a multisample case the number of samples has to be given as well. Calls the respective `estimate` modules. The module returns a $p \times p$ covariance matrix estimate in a one-sample case, and a $cp \times (p+1)$ matrix in a multisample case, where the first column identifies the rows which contain the covariance matrix of $\hat{\boldsymbol{\mu}}_i$.

`jackknife` Input for the module consist of the data matrix, location estimate and the desired level of precision. Calls the `estimate_1` module. The module returns a $p \times p$ covariance matrix estimate.

It is a good idea to start with a small number of bootstrap samples.

5. Examples

5.1. Multivariate normal distribution

We simulated a two-sample case ($n_1 = n_2 = 50$)

$$\mathbf{x}_{1j} \sim N_3(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) \text{ and } \mathbf{x}_{2j} \sim N_3(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 & 1 \\ & 3 & 1 \\ & & 3 \end{pmatrix}.$$

One-sample analysis. We start by analysing the first sample data as a one-sample problem. The null hypothesis of interest is $H_0 : \boldsymbol{\mu}_1 = \mathbf{0}$. The SAS statements

```
%INCLUDE '<full path>\sgnmanova_1.sas';
sgnmanova_1(y3onesam, eps=1E-9, nperm=1000, nboot=500);
```

produce the output

```
                                    Q2
Value of the test statistic:   0.6305067


                                  P_AS
p-value (large sample appr.):  0.8894144


                             P_PERM      SE_P
p-value (sign change test):     0.898 0.0095706 ( 1000 permutations)
```

The null hypothesis is thus not rejected; p-values based on the limiting χ^2_3-distribution and on a permutation distribution were 0.889 and 0.898, respectively. The estimate of $\boldsymbol{\mu}_1$ and covariance matrix estimates of $\hat{\boldsymbol{\mu}}_1$ are obtained from the output as well (the output is reduced to fit it on the page):

```
   MU
-0.054617 0.0340526 0.0930009

   COV_AS
0.0236718 0.0211768 0.0165879
0.0211768 0.0594587 0.0227896
0.0165879 0.0227896 0.049833

   COV_BS
0.025025  0.0262765 0.0181631
0.0262765 0.0674611 0.0308509
0.0181631 0.0308509 0.0511991

   COV_JK
0.0238118 0.0238157 0.0178145
0.0238157 0.061081  0.0289153
0.0178145 0.0289153 0.0480224
```

The subindices "AS", "BS" and "JK" refer to the approximation method by asymptotics, bootstrapping and jackknife, respectively (see Section 3.3). The estimates are very similar. For comparison, the sample mean vector is $\bar{\mathbf{x}}_1 = (-0.02\ 0.09\ 0.06)^\top$ and the estimated covariance matrix of the sample mean is

$$\begin{pmatrix} 0.020 & 0.020 & 0.014 \\ & 0.054 & 0.018 \\ & & 0.048 \end{pmatrix}.$$

The mean is slightly more accurate in the normal case. But, if just one observation of the data set is contaminated (by adding, say, 10 to all its components), the covariance matrix estimates are:

$$\widehat{\mathsf{COV}}_{\mathrm{AS}}(\hat{\boldsymbol{\mu}}_1) = \begin{pmatrix} 0.027 & 0.024 & 0.019 \\ & 0.062 & 0.025 \\ & & 0.048 \end{pmatrix}, \ \widehat{\mathsf{COV}}_{\mathrm{BS}}(\hat{\boldsymbol{\mu}}_1) = \begin{pmatrix} 0.028 & 0.027 & 0.021 \\ & 0.068 & 0.031 \\ & & 0.051 \end{pmatrix},$$

$$\widehat{\mathsf{COV}}_{\mathrm{JK}}(\hat{\boldsymbol{\mu}}_1) = \begin{pmatrix} 0.026 & 0.024 & 0.019 \\ & 0.062 & 0.030 \\ & & 0.044 \end{pmatrix} \text{ and } \widehat{\mathsf{COV}}(\bar{\mathbf{x}}_1) = \begin{pmatrix} 0.059 & 0.059 & 0.042 \\ & 0.093 & 0.046 \\ & & 0.065 \end{pmatrix}..$$

The covariance matrix of $\hat{\boldsymbol{\mu}}_1$ is almost unaffected, but the covariance matrix of the sample mean nearly doubles in size. This reflects the robustness of the spatial median against outliers. Despite of a single outlier, bootstrapping worked well. We will return to the robustness studies in the two-sample case.

Two-sample analysis. Now we move on to the sample comparisons. The interest is to test for differences in location, and to estimate the location, shift and scatter. Analysis for the two-sample data set was performed by the SAS statements

MANOVA	Parameter	Estimate	Standard error
Spatial sign	$\boldsymbol{\mu}_1$	$(-0.05\ 0.03\ 0.09)^\top$	$(0.15\ 0.28\ 0.24)^\top$
	$\boldsymbol{\mu}_2$	$(0.71\ 0.55\ 0.93)^\top$	$(0.15\ 0.28\ 0.24)^\top$
	$\boldsymbol{\Delta}_{12}$	$(0.77\ 0.52\ 0.85)^\top$	$(0.21\ 0.39\ 0.34)^\top$
Classical	$\boldsymbol{\mu}_1$	$(-0.02\ 0.09\ 0.06)^\top$	$(0.14\ 0.26\ 0.22)^\top$
	$\boldsymbol{\mu}_2$	$(0.71\ 0.55\ 0.91)^\top$	$(0.14\ 0.26\ 0.22)^\top$
	$\boldsymbol{\Delta}_{12}$	$(0.73\ 0.46\ 0.86)^\top$	$(0.20\ 0.37\ 0.32)^\top$

Table 1: Estimates for location and shift. Standard errors are based on large sample approximations.

```
%INCLUDE '<full path>\sgnmanova_c.sas';
sgnmanova_c(y3, 2, 1E-9, 1000, 500);
```

Resulting location and shift estimates are shown in Table 1. The sample covariance matrix and Tyler's scatter matrix (used to transform the data), both standardized to $\mathrm{Tr}(\cdot) = 3$, are very much alike:

$$\frac{3}{\mathrm{Tr}(\mathbf{S})}\mathbf{S} = \begin{pmatrix} 0.42 & 0.51 & 0.38 \\ & 1.48 & 0.42 \\ & & 1.10 \end{pmatrix} \text{ and } \hat{\mathbf{V}} = \begin{pmatrix} 0.42 & 0.47 & 0.43 \\ & 1.48 & 0.50 \\ & & 1.10 \end{pmatrix}$$

Test results are presented in Table 2. To demonstrate the robustness of the multivariate spatial sign test we contaminated the elements of a single observation in the first sample by adding a positive constant to all its elements. The effect on the multivariate spatial test is small, but Hotelling's trace test fails completely for large contamination values.

Contamination factor	Hotelling's trace	Multivariate spatial sign test χ^2_3	1000 permutations
none	0.002	0.006	0.003 (0.002)
1	0.003	0.008	0.007 (0.003)
10	0.155	0.013	0.011 (0.003)
100	0.726	0.014	0.012 (0.003)

Table 2: p-values for testing $H_0 : \boldsymbol{\Delta}_{12} = \mathbf{0}$. The standard error of the p-value estimate is given in parentheses.

Naturally, the same phenomenon is reflected in the corresponding estimates. For a contamination factor of 100,

$$\begin{aligned} \hat{\boldsymbol{\mu}}_1 &= (\ -0.02 \quad 0.07 \quad 0.17\ )^\top, \text{ and} \\ \bar{\mathbf{x}}_1 &= (\ 1.98 \quad 2.09 \quad 2.06\ )^\top. \end{aligned}$$

The Hettmansperger-Randles estimate is still close to the true value, but the sample mean vector is totally destroyed.

5.2. Multivariate Cauchy distribution

In this section we study the behavior of the estimates for a heavy-tailed error distribution. We simulated a data set ($N = 50$) from a multivariate Cauchy distribution using the model

$$\mathbf{x}_j = \frac{\mathbf{y}_j}{z_j},$$

where $\mathbf{y}_j \sim N_3(\mathbf{0}, \mathbf{I}_3)$, $z_j^2 \sim \chi_1^2$, and $\mathbf{y}_j$ and z_j are independent. Then $\mathbf{x}_j$ has a spherical multivariate Cauchy distribution. The distribution does not possess finite moments, and it has very heavy tails.

An analysis by `sgnmanova_1` macro gives

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 0.01 & -0.06 & -0.19 \end{pmatrix}^\top$$
$$\hat{\mathbf{V}} = \begin{pmatrix} 0.902 & 0.107 & 0.166 \\ & 1.124 & -0.081 \\ & & 0.974 \end{pmatrix}$$

i.e. natural estimates of the center of symmetry and the spatial sign covariance. Different covariance matrix estimates of the location estimate are

$$\widehat{\mathsf{COV}}_{\mathrm{AS}}(\hat{\boldsymbol{\mu}}) = \begin{pmatrix} 0.034 & 0.004 & 0.006 \\ & 0.043 & -0.003 \\ & & 0.037 \end{pmatrix},$$
$$\widehat{\mathsf{COV}}_{\mathrm{BS}}(\hat{\boldsymbol{\mu}}) = \begin{pmatrix} 0.048 & 0.011 & 0.012 \\ & 0.050 & -0.002 \\ & & 0.046 \end{pmatrix}, \text{ and}$$
$$\widehat{\mathsf{COV}}_{\mathrm{JK}}(\hat{\boldsymbol{\mu}}) = \begin{pmatrix} 0.039 & 0.011 & 0.007 \\ & 0.050 & -0.001 \\ & & 0.035 \end{pmatrix},$$

giving results mainly in the same direction. Due to the extreme values generated by the underlying Cauchy distribution, bootstrapping appears to slightly overestimate the elements of the variance-covariance matrix.

Due to the lack of finite moments of the noise distribution, a classical analysis is not helpful at all:

$$\bar{\mathbf{x}} = \begin{pmatrix} 0.36 & -1.72 & 0.04 \end{pmatrix}^\top$$
$$\mathbf{S} = \begin{pmatrix} 17.720 & -38.555 & -2.000 \\ & 142.475 & 3.274 \\ & & 6.156 \end{pmatrix}$$

By coincidence, the p-values were close to each other: $p = 0.749$ and $p = 0.776$ for the spatial sign test and the Hotelling's T^2 test, respectively.

6. Concluding remarks

Hettmansperger and Randles (2002) recognized that the conditions for the existence and the uniqueness of simultaneous solutions to the estimating equations have not been established. In authors' experience, however, the algorithm appears always to converge.

Lopuhaä and Rousseeuw (1991) showed that the spatial median has a 50% breakdown point. The breakdown point of Tyler's scatter matrix is positive, and generally within the interval $[1/(p+1), 1/p]$. Both the location estimator and the scatter estimator have bounded influence functions. Given these robustness qualities, the minimal model assumptions and the good efficiency properties, multivariate spatial sign methods are attractive alternatives to the classical procedures particularly for skewed or heavy-tailed distributions, or in the presence of outliers.

Acknowledgements

Authors wish to thank the Academy of Finland and the School of Statistical Information, Inference and Data Analysis (SIIDA) for supporting this research.

References

Brown BM (1983). "Statistical Uses of the Spatial Median." *Journal of the Royal Statistical Society B*, **45**, 25–30.

Chakraborty B, Chaudhuri P, Oja H (1998). "Operating Transformation Retransformation on Spatial Median and Angle Test." *Statistica Sinica*, **8**, 767–784.

Hettmansperger TP, Randles RH (2002). "A Practical Affine Equivariant Multivariate Median." *Biometrika*, **89**, 851–860.

Lopuhaä HP, Rousseeuw PJ (1991). "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices." *The Annals of Statistics*, **19**, 229–248.

Möttönen J, Oja H (1995). "Multivariate Spatial Sign and Rank Methods." *Nonparametric Statistics*, **5**, 201–213.

Möttönen J, Oja H, Tienari J (1997). "On the Efficiency of Multivariate Spatial Sign and Rank Tests." *The Annals of Statistics*, **25**, 542–552.

Oja H, Randles RH (2004). "Multivariate Nonparametric Tests." *Statistical Science*, **19**, 598–605.

Ollila E, Hettmansperger TP, Oja H (2003a). "Affine Equivariant Multivariate Sign Methods." Under revision.

Ollila E, Oja H, Croux C (2003b). "The Affine Equivariant Sign Covariance Matrix: Asymptotic Behavior and Efficiencies,." *Journal of Multivariate Analysis*, **87**, 328–355.

Randles RH (1989). "A Distribution-free Multivariate Sign Test Based on Interdirections." *Journal of the American Statistical Association*, **84**, 1045–1050.

Randles RH (2000). "A Simpler, Affine-invariant, Multivariate, Distribution-free Sign Test." *Journal of the American Statistical Association*, **95**, 1263–1268.

Shao J, Wu CFJ (1989). "A General Theory for Jackknife Variance Estimation." *The Annals of Statistics*, **17**, 1176–1197.

Stromberg AJ (1997). "Robust Covariance Estimates Based on Resampling,." *Journal of Statistical Planning and Inference*, **57**, 321–334.

Tyler DE (1987). "A Distribution-free M-estimator of Multivariate Scatter." *The Annals of Statistics*, **15**, 234–251.

Vardi Y, Zhang CH (2001). "A Modified Weiszfeld Algorithm for the Fermat-Weber Location Problem." *Mathematical Programming A*, **90**, 559–566.

Affiliation:

Jaakko Nevalainen
Department of Mathematics, Statistics and Philosophy
33014 University of Tampere, Finland
E-mail: jaakko.nevalainen@uta.fi
URL: http://www.uta.fi/~jaakko.nevalainen/

Hannu Oja
Tampere School of Public Health
33014 University of Tampere, Finland
E-mail: hannu.oja@uta.fi
URL: http://www.uta.fi/~hannu.oja/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 16, Issue 5 *Submitted:* 2004-11-23
May 2006 *Accepted:* 2006-05-17

On the multivariate spatial median for clustered data

Jaakko NEVALAINEN, Denis LAROCQUE and Hannu OJA

Key words and phrases: Affine equivariance; Clustered data; Intracluster correlation; Multivariate location problem; Spatial median.
MSC 2000: Primary 62H15, 62G10; secondary 62E20.

Abstract: The authors consider the multivariate one-sample location problem with clustered data from a nonparametric viewpoint. They propose the spatial median and its affine equivariant version as companion estimators to the affine invariant sign test of Larocque (2003). The authors extend the asymptotics of the proposed estimators to cluster dependent data, and explore the limiting as well as finite sample efficiencies for multivariate t-distributions. They demonstrate that the efficiency of the spatial median suffers less from intracluster correlation than the mean vector. They give an application of the new method using data on well-being in Finnish schools.

La médiane spatiale pour données corrélées en grappes

Résumé : Les auteurs considèrent le problème de position multivarié avec des données corrélées en grappes d'un point de vue non-paramétrique. Ils proposent la médiane spatiale et sa version affine équivariante comme complément au test du signe affine invariant de Larocque (2003). Les auteurs généralisent les résultats asymptotiques des estimateurs proposés au cas de données corrélés en grappes et étudient l'efficacité asymptotique pour les lois t-multivariées. De plus, ils explorent l'efficacité pour des tailles d'échantillons finies à l'aide de simulations. Ils démontrent que la médiane spatiale est moins affectée par la corrélation intra-grappe que le vecteur des moyennes. Finalement, les auteurs illustrent la nouvelle méthode en utilisant des données recueillies dans les écoles finlandaises.

1. INTRODUCTION

Hierarchical structures are common in a variety of data. Hierarchies can be present due to genetic or circumstantial similarities of the observational units, or may arise because of the design of the study. Think of questionnaires on economical welfare posed on all family members in a sample of families. The answers of the members of the same family tend to be more alike than answers chosen at random from the whole sample. The patients from the same hospital in a multinational longitudinal study are subject to the same treating practices, and may have similar diets and genetic backgrounds. Therefore, they may also respond to treatment similarly. A common feature of these two examples is that the data can be arranged, in a natural way, into presumably independent groups, called clusters. However, the outcomes of the members of the same cluster may be correlated. This intracluster correlation must not be overlooked. Ignorance of this characteristic of the data might lead to artificially small p-values and estimated standard errors of the estimates (Figure 3).

In recent work, generalized estimating equations have been popular for the analysis of clustered data (Stoner & Leroux 2002; Williamson, Datta & Satten 2003). Aerts, Geys, Molenberghs & Ryan

(2002) devote a whole book to the topics of modelling of clustered data, mainly by parametric methods. Several extensions of univariate nonparametric tests to cluster correlated data have also been proposed (Rosner & Grove 1999; Rosner, Glynn & Ting Lee 2003; Datta & Satten 2005; Larocque 2005). A multivariate analysis of clustered data based on the normality assumption appears in Goldstein (2003, chapter 6). The use of multivariate nonparametric techniques in the clustered data area, however, has received attention only very lately. In a recent paper, Larocque (2003) developed an affine invariant multivariate sign test for a one-sample location problem with cluster correlated data. Dueck & Lohr (2005) considered robust estimation of variance components and developed a robust estimator of multivariate location in parallel. However, they did not present any asymptotic results for the location estimator.

The goal of this paper is to construct a multivariate location estimator, which can be used as a companion estimate to the test described by Larocque (2003). The spatial median is a natural candidate in the context of spatial sign tests. Its affine equivariant version will be treated as well. These estimators are examples of multivariate medians, which require only mild assumptions of the underlying distribution, are suitable for symmetric as well as skewed distributions, are more efficient than the vector of componentwise medians, are more efficient than the mean vector for heavy-tailed noise distributions, are robust against outliers, and finally, are easy to compute. In the present paper, we will show the consistency of the proposed estimators and derive their limiting distributions with clustered data. Furthermore, we will study the asymptotic and finite sample efficiencies. An example data set will be carefully analyzed to illustrate the performance of the estimators in practice.

2. NOTATION AND ASSUMPTIONS

Let $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \cdots, x_{ijp})^T$ denote the p-dimensional continuous outcome vector of the jth individual ($j = 1, \ldots, m_i$) in the ith cluster ($i = 1, \ldots, n$). Write $N = \sum_{i=1}^n m_i$ for the total number of observations. For clustered data, the data matrix $\mathbf{X}$ ($p \times N$) can be partitioned as

$$\mathbf{X} = (\mathbf{X}_1 \;\; \mathbf{X}_2 \;\; \cdots \;\; \mathbf{X}_n),$$

where each partition represents a cluster consisting of

$$\mathbf{X}_i = (\mathbf{x}_{i1} \; \mathbf{x}_{i2} \;\; \cdots \;\; \mathbf{x}_{im_i}) = \begin{pmatrix} x_{i11} & x_{i21} & \cdots & x_{im_i1} \\ x_{i12} & x_{i22} & \cdots & x_{im_i2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1p} & x_{i2p} & \cdots & x_{im_ip} \end{pmatrix}.$$

The cluster memberships are assumed to be known and the clusters are assumed to be independent of each other throughout this paper. Suppose that the observations arise from a general multivariate location model

$$\mathbf{x}_{ij} = \boldsymbol{\mu}_0 + \boldsymbol{\varepsilon}_{ij},$$

where $\boldsymbol{\mu}_0 = (\mu_{01}, \ldots, \mu_{0p})^T$ is the population spatial median minimizing the objective function $D(\boldsymbol{\mu}) = E\left(\|\mathbf{x}_{ij} - \boldsymbol{\mu}\| - \|\mathbf{x}_{ij}\|\right)$ ($\|\cdot\|$ denotes the Euclidean norm). Define that $\|\boldsymbol{\varepsilon}_{ij}\|^{-1}\boldsymbol{\varepsilon}_{ij} = \mathbf{0}$, if $\boldsymbol{\varepsilon}_{ij} = \mathbf{0}$. Then $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ijp})^T$ is a random error from an unknown continuous distribution, same for all i, j, with $E\left(\|\boldsymbol{\varepsilon}_{ij}\|^{-1}\boldsymbol{\varepsilon}_{ij}\right) = \mathbf{0}$.

The rest of this section lists the necessary assumptions for the construction of the asymptotic results. In asymptotic considerations it is assumed that the number of clusters goes to infinity. As $n \to \infty$, the sequence of cluster sizes $\{m_n\}$ is treated as a sequence of constants such that

(A1) $\max\{m_1, \ldots, m_n\} \to M$ and
(A2) $n^{-1}\sum_{i=1}^n I(m_i = m) \to \lambda_m$,

where M is finite and $0 \leq \lambda_m \leq 1$ for all $m = 1, \ldots, M$. This means that the cluster sizes are assumed to be bounded in a uniform manner. Secondly, the underlying marginal distribution of $\boldsymbol{\varepsilon}_{ij}$ is assumed to fulfill the following conditions.

(B1) $E(\|\boldsymbol{\varepsilon}_{ij} - \boldsymbol{\mu}\| - \|\boldsymbol{\varepsilon}_{ij}\|) > 0, \forall \boldsymbol{\mu} \neq \mathbf{0}$.
(B2) The density function $f(\boldsymbol{\varepsilon})$ is bounded and continuous, and $f(0) > 0$.

The theoretical objective function $D(\boldsymbol{\mu}) = E\left(\|\mathbf{x}_{ij} - \boldsymbol{\mu}\| - \|\mathbf{x}_{ij}\|\right)$ is formulated in this way to guarantee that it is finite everywhere, and (B1) says that the unique solution lies at $\boldsymbol{\mu}_0$. If $p \geq 2$, the condition (B2) implies the existence of $E\left(\|\boldsymbol{\varepsilon}_{ij}\|^{-1}\right)$, and ensures that the observations lie in a genuinely p-dimensional space. The final assumption deals with the joint distribution of $\boldsymbol{\varepsilon}_{ij}$ and $\boldsymbol{\varepsilon}_{i'k}$. In the clustered data setting, $\boldsymbol{\varepsilon}_{ij}$ is independent of $\boldsymbol{\varepsilon}_{i'k}$ $(i \neq i')$, but it can depend on $\boldsymbol{\varepsilon}_{ij'}$ $(j \neq j')$. Suppose that

(B3) $(\boldsymbol{\varepsilon}_{i_1 j_1}, \boldsymbol{\varepsilon}_{i_1 j_2}) \sim (\boldsymbol{\varepsilon}_{i_2 k_1}, \boldsymbol{\varepsilon}_{i_2 k_2})$,

where i_1, i_2 are any two indices in $\{1, \ldots, n\}$, and j_1, j_2 $(j_1 \neq j_2)$ and k_1, k_2 $(k_1 \neq k_2)$ are two indices chosen in $\{1, \ldots, m_{i_1}\}$ and $\{1, \ldots, m_{i_2}\}$, respectively. Most importantly, this condition implies that the intracluster correlation is the same for every cluster.

3. ESTIMATION THROUGH THE SPATIAL MEDIAN

The vector valued statistic $\mathbf{T}_N = N^{-1} \sum_i \sum_j \|\mathbf{x}_{ij}\|^{-1} \mathbf{x}_{ij}$ is the spatial sign test statistic for testing the null hypothesis $H_0 : \boldsymbol{\mu}_0 = \mathbf{0}$ (without loss of generality). The test statistic is based only on the spatial signs (or directions) of the observations and ignores their distances. See Brown (1983) and Möttönen & Oja (1995), for example.

Define

$$\begin{aligned}
\mathbf{A} &= \begin{cases} 2f(0), & \text{if } p = 1; \\ E_0\left\{\frac{1}{\|\mathbf{x}_{ij}\|}\left(\mathbf{I}_p - \frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^T}{\|\mathbf{x}_{ij}\|^2}\right)\right\}, & \text{if } p \geq 2, \end{cases} \\
\mathbf{B} &= E_0\left\{\frac{\mathbf{x}_{ij}\mathbf{x}_{ij}^T}{\|\mathbf{x}_{ij}\|^2}\right\}, \text{ and} \\
\mathbf{C} &= E_0\left\{\frac{\mathbf{x}_{ij}\mathbf{x}_{ij'}^T}{\|\mathbf{x}_{ij}\|\,\|\mathbf{x}_{ij'}\|}\right\} \; (j \neq j').
\end{aligned}$$

The notation $E_0(\cdot)$ means that the expectation is taken under the null hypothesis. These expectations exist as it is assumed that (B2) is true. The matrix $\mathbf{A}$ originates from the linear approximation of the spatial median, and the matrices $\mathbf{B}$ and $\mathbf{C}$ from the covariance matrix of the spatial sign test statistic. If intracluster dependency is present (which often implies that $\mathbf{C} \neq \mathbf{0}$), straightforward calculation gives

LEMMA 1. *Under the null hypothesis, the test statistic $\sqrt{N}\mathbf{T}_N$ has the covariance matrix*

$$\mathbf{D}_N = \mathbf{B} + \frac{1}{N}\sum_{i=1}^{n} m_i(m_i - 1)\,\mathbf{C}.$$

The conditions (A1) and (A2) imply that $\lim_{n\to\infty}\left[N^{-1}\sum_i m_i(m_i - 1)\right]$ exists. Write $\mathbf{D} = \lim_{n\to\infty} \mathbf{D}_N$.

LEMMA 2. *Under the null hypothesis, the limiting distribution of the test statistic* $\sqrt{N}\mathbf{T}_N$ *is multivariate normal* $N_p(\mathbf{0}, \mathbf{D})$.

Clustering has an effect only on the variance-covariance matrix of the test statistic. In the case of independent observations, the test statistic has an asymptotic $N_p(\mathbf{0}, \mathbf{B})$ distribution.

The estimator associated with the spatial sign test is the spatial median (Gower, 1974; Brown, 1983). The sample spatial median minimizes

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} \|\mathbf{x}_{ij} - \boldsymbol{\mu}\|,$$

that is, the sum of the Euclidian distances to the data points. It is found as the solution to the estimating equation

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{\mathbf{x}_{ij} - \boldsymbol{\mu}}{\|\mathbf{x}_{ij} - \boldsymbol{\mu}\|} = \mathbf{0}.$$

Vardi & Zhang (2001) give a fast, monotonically convergent algorithm for the computation of the spatial median. The solution is unique when the observations do not fall on a straight line. This is true almost surely by the condition (B2). Furthermore, the estimator $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}(\mathbf{X})$ is strongly consistent for $\boldsymbol{\mu}_0$. Its limiting distribution is given by the next theorem.

THEOREM 1. *Under the general assumptions (A1)–(A2) and (B1)–(B3), the limiting distribution of* $\sqrt{N}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)$ *is multivariate normal* $N_p(\mathbf{0}, \mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1})$.

In the case of independent observations the limiting distribution is simply $N_p(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$. Again, only the covariance matrix of the estimator is corrected for the intracluster correlation. Another potentially efficient approach would be to find a weighted estimator minimizing the objective function $\sum_i \sum_j w_{ij}\|\mathbf{x}_{ij} - \boldsymbol{\mu}\|$. The weights could be chosen in an optimal way; for example, in such a way that the determinant of the covariance matrix of the estimator is minimized. A detailed study of such weighted methods will appear in Nevalainen, Larocque & Oja (2006a) and Larocque, Nevalainen & Oja (2007).

Write $\mathbf{Z} = \mathbf{X} - \widehat{\boldsymbol{\mu}}\mathbf{1}_N^T$ for the matrix of estimated residuals and $\mathbf{z}_{ij}$ for its elements. In order to obtain an estimate of the covariance matrix of $\widehat{\boldsymbol{\mu}}$ for $p \geq 2$, obvious candidates for estimators of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are

$$\widehat{\mathbf{A}} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{\frac{1}{\|\mathbf{z}_{ij}\|}\left(\mathbf{I}_p - \frac{\mathbf{z}_{ij}\mathbf{z}_{ij}^T}{\|\mathbf{z}_{ij}\|^2}\right)\right\}, \quad (1)$$

$$\widehat{\mathbf{B}} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\mathbf{z}_{ij}\mathbf{z}_{ij}^T}{\|\mathbf{z}_{ij}\|^2}, \quad (2)$$

$$\widehat{\mathbf{C}} = \left\{\sum_{i=1}^{n} m_i(m_i - 1)\right\}^{-1}\sum_{i=1}^{n}\sum_{j}\sum_{j'}\frac{\mathbf{z}_{ij}\mathbf{z}_{ij'}^T}{\|\mathbf{z}_{ij}\|\,\|\mathbf{z}_{ij'}\|}, (1 \leq j, j' \leq m_i, j \neq j'). \quad (3)$$

A natural estimator of $\mathbf{D}$ is then

$$\widehat{\mathbf{D}} = \widehat{\mathbf{B}} + \frac{1}{N}\sum_{i=1}^{n} m_i(m_i - 1)\,\widehat{\mathbf{C}}. \quad (4)$$

THEOREM 2. *Under the general assumptions (A1)–(A2) and (B1)–(B3),* $\widehat{\mathbf{A}}$ *and* $\widehat{\mathbf{D}}$ *are weakly consistent estimators of* $\mathbf{A}$ *and* $\mathbf{D}$*, respectively.*

The desired covariance matrix estimator of the sample spatial median is then

$$\widehat{\text{cov}}(\widehat{\boldsymbol{\mu}}) = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{D}}\widehat{\mathbf{A}}^{-1}.$$

4. ESTIMATION THROUGH THE TRANSFORMATION RETRANSFORMATION SPATIAL MEDIAN

Consider affine transformations of the data matrix by an arbitrary nonsingular $p \times p$ matrix $\mathbf{H}$ and a p-vector $\mathbf{g}$ $(p \geq 2)$

$$\mathbf{X} \longmapsto \mathbf{HX} + \mathbf{g}\mathbf{1}_N^T,$$

where $\mathbf{1}_N$ is an N-vector of ones. Then $\widehat{\boldsymbol{\mu}}(\mathbf{HX} + \mathbf{g}\mathbf{1}_N^T) = \mathbf{H}\,\widehat{\boldsymbol{\mu}}(\mathbf{X}) + \mathbf{g}$ is not generally true; it holds only if $\mathbf{H}$ is an orthogonal matrix. Thus, the spatial median is not affine equivariant: it is location and rotation equivariant, but not equivariant under marginal rescaling. A well known way to achieve the affine equivariance property for the spatial median is to utilize the transformation retransformation procedure (Chakraborty & Chaudhuri 1996; Chakraborty, Chaudhuri & Oja 1998). If $\widehat{\mathbf{V}}$ is an affine equivariant estimator of scatter (or shape) and $\widehat{\boldsymbol{\mu}}$ is the spatial median, then the transformation retransformation spatial median

$$\widetilde{\boldsymbol{\mu}}(\mathbf{X}) = \widehat{\mathbf{V}}^{1/2}\widehat{\boldsymbol{\mu}}(\widehat{\mathbf{V}}^{-1/2}\mathbf{X})$$

has the affine equivariance property. The next theorem shows that in the spherical case $\widehat{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\mu}}$ share the same limiting distribution and thus possess the same limiting efficiencies.

THEOREM 3. *Assume* $\sqrt{N}\left(\widehat{\mathbf{V}} - \mathbf{I}_p\right) = O_P(1)$ *. Then*

$$\sqrt{N}\,(\widetilde{\boldsymbol{\mu}}(\mathbf{X}) - \widehat{\boldsymbol{\mu}}(\mathbf{X})) \xrightarrow{P} \mathbf{0}.$$

More generally, for elliptical distributions, $\widetilde{\boldsymbol{\mu}}$ is a consistent estimator of $\boldsymbol{\mu}_0$, and

$$\sqrt{N}\,(\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) \xrightarrow{D} N_p\left(\mathbf{0}, \mathbf{V}^{1/2}\mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1}\mathbf{V}^{1/2}\right),$$

where $\mathbf{A}$ and $\mathbf{D}$ come from the distribution of the standardized data. The result holds for a slightly more general class of distributions called the elliptical directions class. See Randles (2000) for the different notions of symmetry.

As the spatial median is orthogonally equivariant, it does not matter (asymptotically) which $\sqrt{N}$-consistent affine equivariant scatter matrix (or shape matrix) is used in the transformation. Shape matrices based on the spatial signs are appealing candidates, since they are estimated in the same spirit as the spatial median itself. In the construction of an affine invariant test, Randles (2000) and Larocque (2003) used Tyler's shape matrix (Tyler 1987) with respect to the origin (the null point). The use of Tyler's shape matrix necessitates a fixed location vector and it is therefore not directly suitable for the estimation purpose. Hettmansperger & Randles (2002) studied the simultaneous estimation of location and shape using Tyler's shape matrix and the spatial median, but the existence or the uniqueness of such simultaneous solutions has not been proven. Dümbgen (1998) proposed a symmetrized version of the Tyler's shape matrix, which in the clustered data setting can be defined by the implicit equation

$$E\left\{\frac{(\mathbf{x}_{ij} - \mathbf{x}_{i'k})(\mathbf{x}_{ij} - \mathbf{x}_{i'k})^T}{(\mathbf{x}_{ij} - \mathbf{x}_{i'k})^T\mathbf{V}^{-1}(\mathbf{x}_{ij} - \mathbf{x}_{i'k})}\right\} = \frac{1}{p}\mathbf{V},$$

where $\mathbf{x}_{ij}$ and $\mathbf{x}_{i'k}$ are two independent copies of $\mathbf{x}$. Dümbgen's shape matrix can be computed just like Tyler's shape matrix after constructing the pairwise differences. Computation algorithms are described by Randles (2000) and Oja & Randles (2004).

Since the asymptotic theory of shape matrices for clustered data is still unknown, our strategy is to choose one observation at random from each cluster, and to estimate Dümbgen's shape matrix based on these n independent observations. This primitive solution suffices, because the $\widehat{\mathbf{V}}$ matrix serves the sole purpose of obtaining affine equivariance. All the data points should of course be used when computing any other estimates. In a pursuit of maximal efficiency, one may take K samples of size n, compute the corresponding shape estimates $\widehat{\mathbf{V}}_1, \ldots, \widehat{\mathbf{V}}_K$, and take the average $K^{-1}\sum_{k=1}^{K}\widehat{\mathbf{V}}_k$ as the final estimate of $\mathbf{V}$.

5. ASYMPTOTIC RELATIVE EFFICIENCY

Efficiency calculations are based on the multivariate model

$$\mathbf{x}_{ij} = \frac{\mathbf{a}_i + \boldsymbol{\varepsilon}_{ij}}{(s_i^2/\nu)^{1/2}}, \tag{5}$$

where the cluster effect $\mathbf{a}_i \sim N_p(\mathbf{0}, \tau^2\mathbf{I}_p)$, the residual $\boldsymbol{\varepsilon}_{ij} \sim N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$, and the scale $s_i^2 \sim \chi^2_\nu$ are mutually independent. The model can be interpreted as a multivariate random effect model allowing for a cluster-specific location and a cluster-specific scale. The formulation implies that $\mathbf{x}_{ij}$ has a multivariate t-distribution with ν degrees of freedom.

Consider the balanced case with $m_1 = m_2 = \ldots = m$. Without loss of generality, suppose that $\tau^2 + \sigma^2 = 1$. In this particular model, $\|\mathbf{x}_{ij}\|$ and $\|\mathbf{x}_{ij}\|^{-1}\mathbf{x}_{ij}$ are independent. Moreover, $\|\mathbf{a}_i + \boldsymbol{\varepsilon}_{ij}\|^2 \sim \chi^2_p$. Let ϱ denote the intracluster correlation coefficient $\tau^2(\tau^2+\sigma^2)^{-1} = \tau^2$, which is the correlation between the kth components of two observations from the same cluster. Then one obtains

$$\begin{aligned}
\mathbf{A} &= \frac{\sqrt{2}\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu}\Gamma(\frac{\nu}{2})}\frac{\sqrt{2}\Gamma(\frac{p+1}{2})}{p\Gamma(\frac{p}{2})}\mathbf{I}_p \\
\mathbf{D} &= \left(1 + (m-1)\varrho\frac{F(1/2, 1/2; p/2+1; \varrho^2)}{F(1/2, 1/2; p/2+1; 1)}\right)\frac{1}{p}\mathbf{I}_p
\end{aligned}$$

where $F(\cdot)$ denotes Gauss' hypergeometric function (Abramowitz & Stegun 1970; Saw 1983; Larocque 2003). The covariance matrix of the mean vector in this model simplifies to

$$\operatorname{cov}\left(\sqrt{N}\bar{\mathbf{x}}\right) = \frac{\nu\Gamma(\frac{\nu-2}{2})}{2\Gamma(\frac{\nu}{2})}\{1 + (m-1)\varrho\}\mathbf{I}_p. \tag{6}$$

In the normal case ($\nu = \infty$), the expression reduces to $\{1 + (m-1)\varrho\}\mathbf{I}_p$. Based on the p-root of the ratio of Wilks' generalized variances of the two estimators, the asymptotic relative efficiency of the spatial median relative to the mean vector is given by

$$\begin{aligned}
\operatorname{ARE}(\widehat{\boldsymbol{\mu}}, \bar{\mathbf{x}}) = \left\{\frac{\det(\operatorname{cov}(\bar{\mathbf{x}}))}{\det(\mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1})}\right\}^{1/p} &= \frac{\Gamma(\frac{\nu-2}{2})\Gamma^2(\frac{\nu+1}{2})}{\Gamma^3(\frac{\nu}{2})}\frac{2\Gamma^2(\frac{p+1}{2})}{p\Gamma^2(\frac{p}{2})}\{1 + (m-1)\varrho\} \\
&\quad \times \left\{1 + (m-1)\varrho\frac{F(1/2, 1/2; p/2+1; \varrho^2)}{F(1/2, 1/2; p/2+1; 1)}\right\}^{-1}.
\end{aligned}$$

Figure 1 shows the asymptotic relative efficiencies for different degrees of freedom, dimensions and cluster sizes as a function of the intracluster correlation coefficient. As in the case of independent observations, the heavier the tails of the underlying distribution, the better the relative

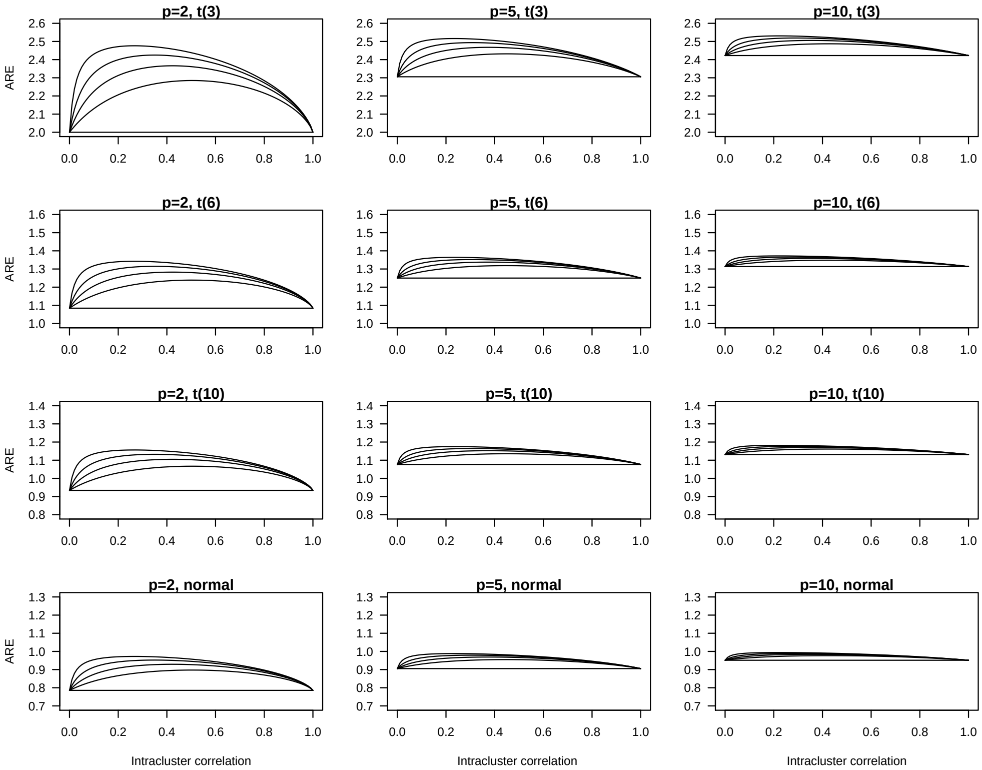

Figure 1: ARE of the spatial median versus the mean for multivariate t-distribution models as a function of the intraclass correlation coefficient at cluster sizes $m = 1$ (the straight line), 5, 10, 20 and 50 (the uppermost curve). Note the different scaling of the vertical axes.

efficiency of the spatial median. The ARE increases as the dimension increases. Interestingly, the plots also suggest that the spatial median suffers less from intracluster correlation than the mean vector. Larocque (2003) obtained similar results for the affine invariant sign test in the multivariate normal case. A possible explanation is that the covariance matrix of the spatial median depends only on the correlation between the spatial signs of the data points, and not on the correlation between the distances of the data points. The mean vector, however, depends on both of them. The shapes of the curves bend farther away from a straight line as the cluster size increases. High dimension reduces the phenomenon: as $p \to \infty$, the shapes of the curves approach that of a straight line regardless of the cluster size.

Note that the limiting covariance matrix of the spatial median does not depend on the choice of the scale parameter in the model (5). To confirm this, see formulas (1), (2), (3) and (4). However, the same is not true for the sample mean and for the intracluster correlation coefficient. Thus, the efficiencies at models with different restrictions on the scale parameter would be different from the case considered.

More limiting efficiency results for spatial sign methods can be found in Möttönen, Oja & Tienari (1997) and in Larocque (2003). The efficiencies of the tests and the corresponding estimators coincide.

6. SIMULATION STUDY

Simulations were based on model (5) and were conducted in R (R Development Core Team 2004).

The spatial median was computed from 10000 generated samples, and its covariance matrix was estimated from the sample distribution. The known covariance matrix (6) was used for the sample mean. The finite sample efficiency comparison of the estimators is based on the ratio of the determinants of the estimated covariance matrices to the power of $1/p$.

Simulation results are presented in Figure 2 for the multivariate normal distribution and the multivariate t_6 distribution, for different sample sizes and for two dimensions. The simulated

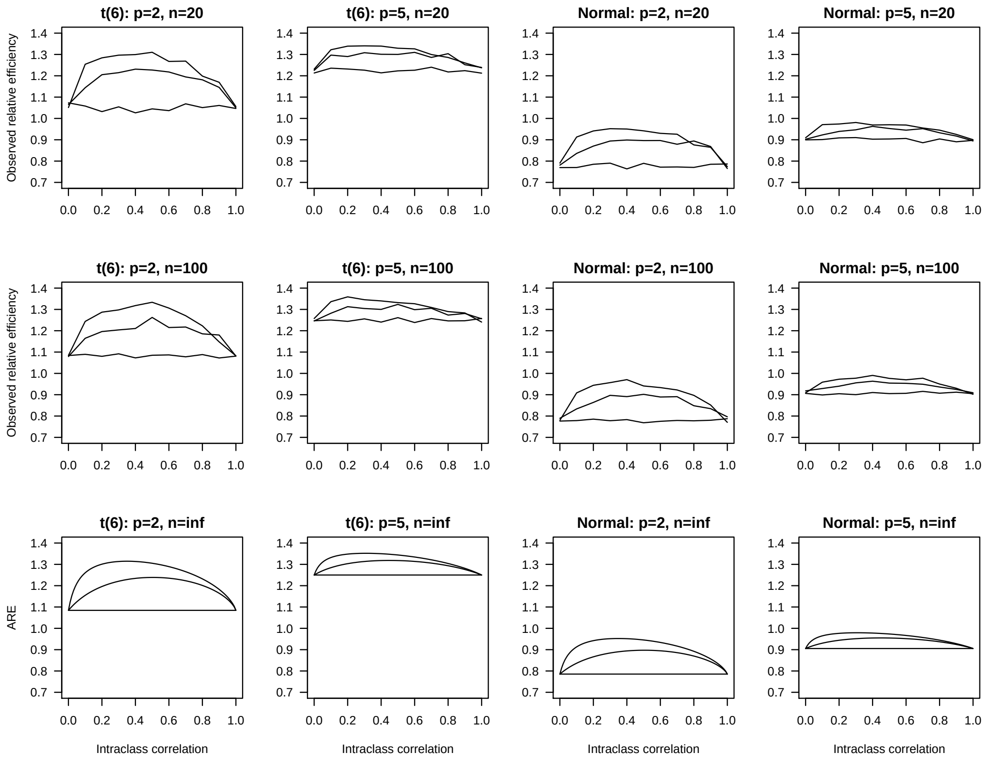

Figure 2: Relative efficiencies of the spatial median versus the mean for the multivariate t_6-distribution and for the multivariate normal distribution as a function of the intraclass correlation coefficient at cluster sizes $m = 1$ (the nearly straight line), 5 and 20 (the uppermost curve).

curves resemble the asymptotic ones to a high extent in all of the cases considered. Thus, the approximations based on the asymptotic results appear to be accurate, and one can trust the efficiencies for finite samples as well.

Figure 3 illustrates what might happen if the clustered structure of the data is ignored in the analysis. The data points were generated from a spherical bivariate normal distribution centred at the origin with an intraclass correlation of $\varrho = 0.5$. The figure shows the sample spatial median and two 95% confidence ellipsoids for it: one which was estimated with a correction for the intracluster correlation, and another one, which was computed falsely assuming i.i.d. observations. As the number of clusters increases and the effective sample size grows, there is a true gain in precision; both of the ellipsoids clearly reflect this. Increasing cluster size has only a minor effect on the size of the corrected ellipsoid. However, the uncorrected ellipsoid reacts strongly and gets artificially small—it does not even include the true value in five of the nine cases. An analyst unaware of the clustered data structure could easily report too optimistic confidence limits or too small p-values, or even draw false conclusions from the data.

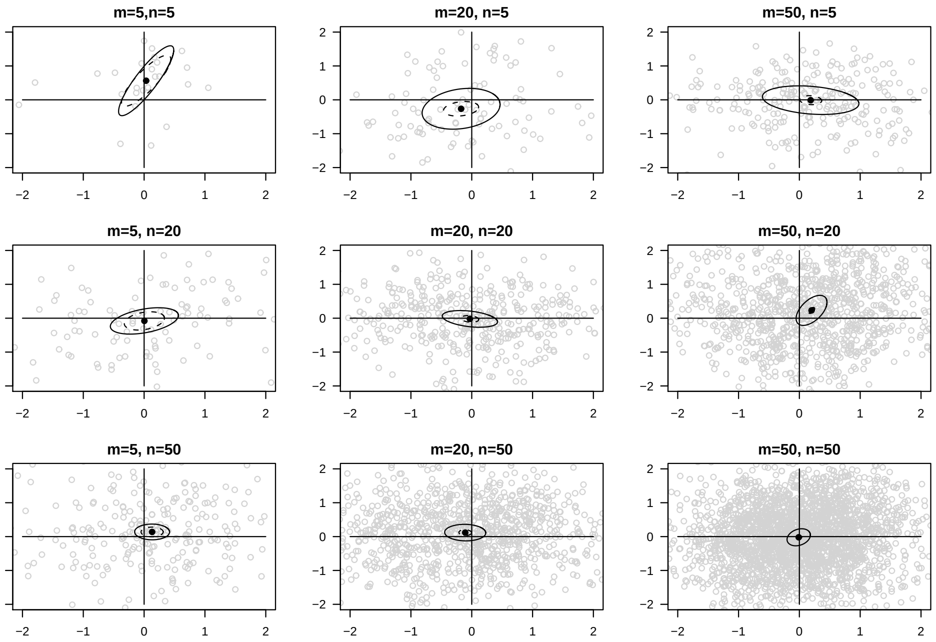

Figure 3: The confidence ellipsoids for the spatial medians of generated random samples from a spherical bivariate normal distribution with an intracluster correlation $\varrho = 0.5$. The true spatial median lies at the origin. The dashed line indicates the confidence ellipsoid estimated without taking the intracluster correlation into account, and the solid line indicates the corrected confidence ellipsoid.

7. EXAMPLE

Konu & Lintonen (2006) collected data on well-being in Finnish schools using an Internet-based tool (Lintonen & Konu 2006). As a part of a larger study, the pupils in grades 7-9 of the participating schools were asked to fill out a questionnaire on the Internet (`www2.edu.fi/hyvinvointi-profiili`). The data were gathered during the 2004-2005 school year. Participation of the schools was strictly voluntary. In addition to basic background information (gender, age, grade), the pupils were to answer 81 questions on school well-being. The questions were grouped into assessments of school conditions (26 questions), social relationships (19), means for self-fulfillment at school (24) and health status (12). For each question, an ordinal five-point (fully agree, agree, neither agree nor disagree, disagree, fully disagree) answering scale was used. The School Well-Being Profile (Konu & Lintonen 2006) consists of the means of each group of questions.

In the present example, we are interested in the pupils' perceptions of their social relationships (SR) and of their means for self-fulfillment (MSF), and their evolution during the lower secondary school (the Finnish lower secondary school lasts from the 7th grade until the 9th grade). Previous studies have shown that the two variables are strongly correlated; the means for self-fulfillment score is partly associated with the social relationships between the pupils and the teachers. Furthermore, they are associated with gender, and related to pupil's health status (health questions are mainly on psychosomatic symptoms). For comparison of grades 7 and 9, we matched the pupils from the same school for gender and health status. The bivariate differences of the matched pairs were taken as the outcomes for comparison. The analyzed data set is then a subset of the original data set, and it consists of 2438 pupils from 19 schools ($N = 1219$ for each analysis).

The boxplots of the two scores at grade 7 are shown in Figure 4. There is some heterogeneity between the schools, but a little bit less than initially expected at the onset of the study. Other studies have indicated that the relative influence of school is stronger on the performance of the pupils than on their well-being. There is a great variation in the number of questionnaires per school: in the smallest schools only four pupils answered the questionnaire, but in some schools as many as 166 pupils answered the questionnaire. The average number of questionnaires per school was 64. The data also seems to be skewed (Figures 4 and 5).

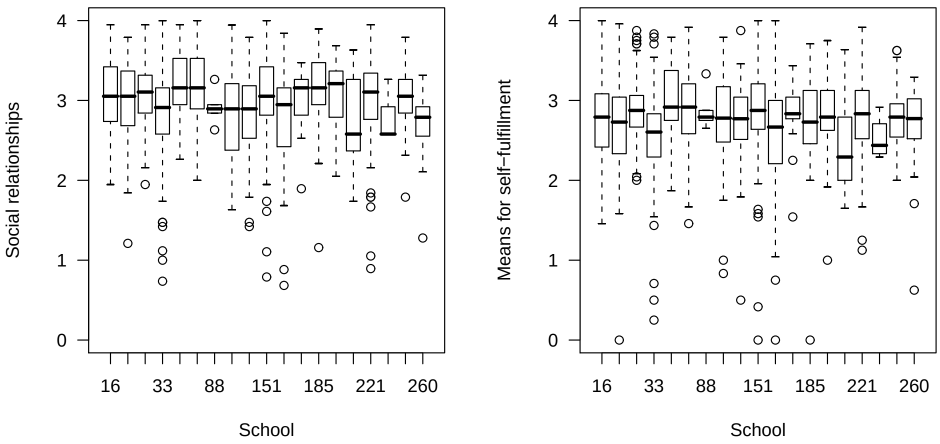

Figure 4: Social relationship and means for self-fulfillment scores at grade 7 broken down by school.

The sample mean vector $\overline{\mathbf{x}}$, the sample spatial median $\widehat{\boldsymbol{\mu}}$, and the sample transformation retransformation spatial median $\widetilde{\boldsymbol{\mu}}$ for the bivariate outcome $(\text{SR}, \text{MSF})^T$ are summarized in Table 1. The transformation matrix estimate $\widehat{\mathbf{V}}^{-1/2}$ was computed $K = 100$ times from n randomly chosen independent observations (one per cluster at each repetition). The final estimate was taken to be the arithmetic average of the repetitions.

Table 1: The estimates of location for the outcome at the 7th grade and the 9th grade, and for the difference outcome of the matched pairs.

	Grade 7th	Grade 9th	Matched-pair difference
$\overline{\mathbf{x}}$	$(3.00, 2.76)^T$	$(2.80, 2.50)^T$	$(-0.20, -0.26)^T$
$\widehat{\boldsymbol{\mu}}$	$(3.03, 2.78)^T$	$(2.84, 2.55)^T$	$(-0.19, -0.23)^T$
$\widetilde{\boldsymbol{\mu}}$	$(3.04, 2.79)^T$	$(2.85, 2.55)^T$	$(-0.19, -0.23)^T$

The estimates are quite similar to each other, but the mean vector is slightly away from the others, apparently due to the asymmetry of the data cloud (Figure 5). The covariance matrix estimates at grade 7 for the three competing location estimators are:

$$\widehat{\text{cov}}\,(\bar{\mathbf{x}}) = \begin{pmatrix} 0.00114 & 0.00108 \\ 0.00108 & 0.00125 \end{pmatrix},$$
$$\widehat{\text{cov}}\,(\widehat{\boldsymbol{\mu}}) = \begin{pmatrix} 0.00102 & 0.00091 \\ 0.00091 & 0.00102 \end{pmatrix},$$
$$\widehat{\text{cov}}\,(\widetilde{\boldsymbol{\mu}}) = \begin{pmatrix} 0.00094 & 0.00082 \\ 0.00082 & 0.00094 \end{pmatrix}.$$

The estimated covariance matrix of the transformation retransformation spatial median has the smallest determinant perhaps suggesting the best estimation efficiency. The corresponding confidence ellipsoids are presented in Figure 5.

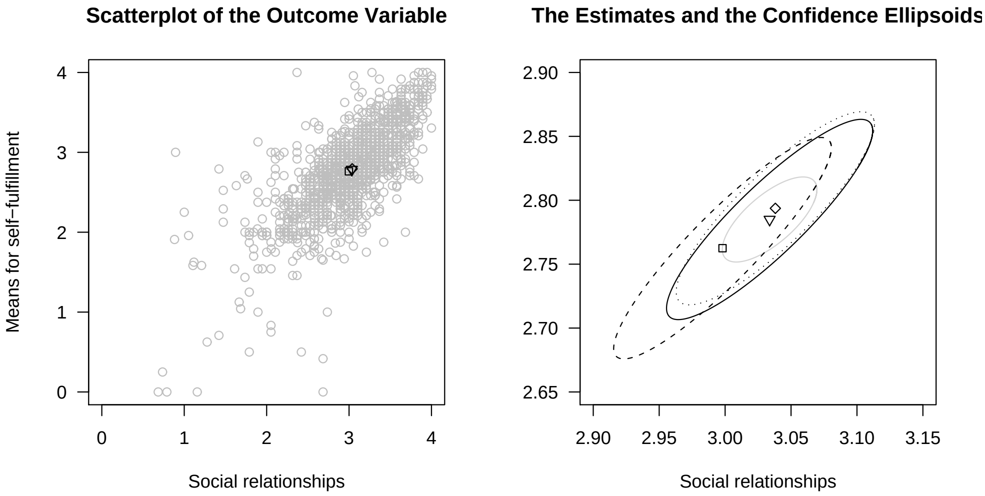

Figure 5: Scores at the 7th grade and the location estimates with their 95% confidence ellipsoids. The triangle and the solid line correspond to the spatial median, the diamond and the dotted line to the transformation retransformation spatial median, and the square and the dashed line to the mean vector. The grey line indicates a confidence ellipsoid for the spatial median without the correction for clustering.

The intracluster correlation estimates for the 7th grade are $\hat{\varrho}_{\text{SR}} = 0.045$ and $\hat{\varrho}_{\text{MSF}} = 0.049$, indicating small intracluster dependency. Even though there is only little positive intracluster correlation, the correction for it has a surprisingly big impact on the confidence ellipsoid. As a result of the large average cluster size and the heterogeneity in the cluster sizes, there is an enormous number of covariance terms. Therefore, the multiplying constant $N^{-1}\sum_i m_i(m_i-1)$ of equation (4) takes a large value (here ≈ 97) allowing the matrix $\widehat{\mathbf{C}}$ to have a meaningful impact on the covariance matrix of the estimator. The estimates of $\mathbf{B}$ and $\mathbf{C}$ from the untransformed data are

$$\widehat{\mathbf{B}} = \begin{pmatrix} 0.49288 & 0.21497 \\ 0.21497 & 0.50712 \end{pmatrix}$$
$$\widehat{\mathbf{C}} = \begin{pmatrix} 0.01603 & 0.01575 \\ 0.01575 & 0.02162 \end{pmatrix}$$

This result emphasizes how important it is to adjust, even for small intracluster correlations.

When considering the difference between the 9th and the 7th grade, part of the intracluster correlation vanishes ($\hat{\varrho}_{\mathrm{SR}} = 0.017$ and $\hat{\varrho}_{\mathrm{MSF}} = 0.021$). Substraction of the outcomes of the matched pairs eliminates the correlation deriving from level differences between the schools. In this case it appears that the schools have very little or no influence on the changes of the social relationships and the means for self-fulfillment during lower secondary school. The confidence ellipsoids do not include the origin so there is a statistically significant ($p < 0.05$) overall decrease in the scores from the 7th to the 9th grade (Figure 6).

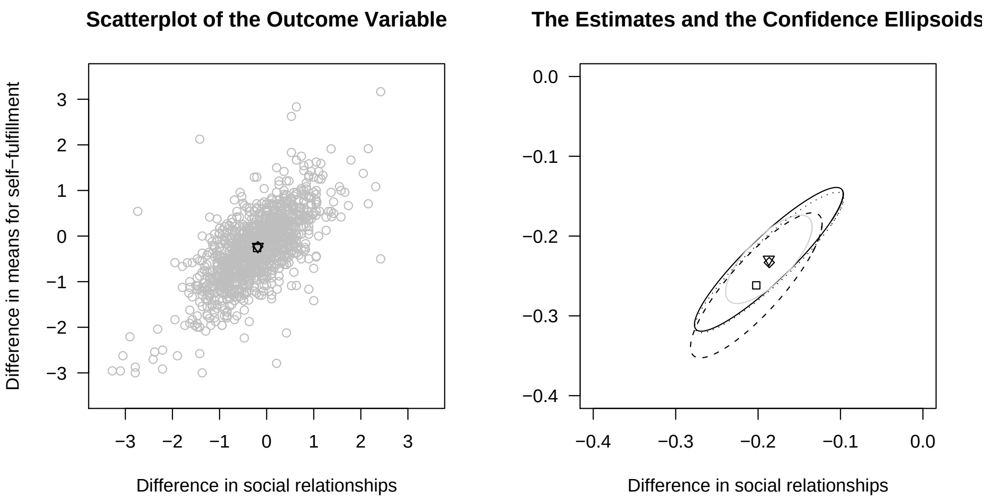

Figure 6: Score differences (9th - 7th grade) and the location estimates with their 95% confidence ellipsoids. The triangle and the solid line correspond to the spatial median, the diamond and the dotted line to the transformation retransformation spatial median, and the square and the dashed line to the mean vector. The grey line indicates a confidence ellipsoid for the spatial median without the correction for clustering.

APPENDIX

Without loss of generality, the true spatial median is assumed to lie at the origin throughout this section. For notational convenience, the subindices are often dropped.

Proof of Lemma 1. Follows by straightforward calculation.

Proof of Lemma 2. Follows from the multivariate extension of the central limit theorem (Serfling 1980, p. 30).

Proof of Theorem 1. Write

$$\begin{aligned} D(\boldsymbol{\mu}) &= E\left(\|\mathbf{x}-\boldsymbol{\mu}\| - \|\mathbf{x}\|\right) \text{ for the theoretical objective function, and} \\ D_N(\boldsymbol{\mu}) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\|\mathbf{x}_{ij}-\boldsymbol{\mu}\| - \|\mathbf{x}_{ij}\|\right) \text{ for its sample counterpart.} \end{aligned}$$

The spatial median which minimizes the theoretical objective function is then $\mathbf{0}$.

Lemma 3.

$$D(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} + o(\|\boldsymbol{\mu}\|^2).$$

Proof of Lemma 3. Lemma 19 in Arcones (1998).

I. Almost sure convergence

Lemma 4. $\widehat{\boldsymbol{\mu}} \to \mathbf{0}$ *almost surely.*

Proof of Lemma 4.

1. First write

$$D_N(\boldsymbol{\mu}) = \frac{1}{N}\sum_{i=1}^{n}\left[\sum_{j=1}^{m_i}\left(\|\mathbf{x}_{ij}-\boldsymbol{\mu}\| - \|\mathbf{x}_{ij}\|\right)\right] = \frac{1}{N}\sum_{i=1}^{n} S_i.$$

The sums $S_1, S_2, \ldots, S_n$ are independent random variables with $E(S_i) = m_i\, D(\boldsymbol{\mu})$ and $\text{var}(S_i) = \sigma_i^2$. Since $|\,\|\mathbf{a}-\mathbf{b}\| - \|\mathbf{a}\|\,| \le \|\mathbf{b}\|$, we have that $\sigma_i^2 \le m_i^2\|\boldsymbol{\mu}\|^2$. Furthermore, as

$$\sum_{i=1}^{\infty}\frac{\sigma_i^2}{i^2} \le \|\boldsymbol{\mu}\|^2\sum_{i=1}^{\infty}\frac{m_i^2}{i^2} \le \|\boldsymbol{\mu}\|^2\sum_{i=1}^{\infty}\frac{\max\{m_1^2,\ldots,m_i^2\}/\sqrt{i}}{i^{3/2}} \le C\sum_{i=1}^{\infty}\frac{1}{i^{3/2}}$$

for some $C > 0$, the series converges to a finite constant (by the integral test). Kolmogorov's strong law of large numbers (Serfling 1980, p. 27) gives

$$\frac{1}{n}\sum_{i=1}^{n} S_i - \frac{1}{n}\sum_{i=1}^{n} m_i D(\boldsymbol{\mu}) \xrightarrow{a.s.} 0.$$

Hence, we conclude that $D_N(\boldsymbol{\mu}) \longrightarrow D(\boldsymbol{\mu})$ almost surely.

2. As $D_N(\boldsymbol{\mu})$ and $D(\boldsymbol{\mu})$ are finite and convex, also

$$\sup_{\|\boldsymbol{\mu}\|\leq C} \|D_N(\boldsymbol{\mu}) - D(\boldsymbol{\mu})\| \xrightarrow{a.s.} 0$$

   for all $C > 0$ (Theorem 10.8 in Rockafellar 1970).

3. Write

$$\widehat{\boldsymbol{\mu}}^* = \arg\min_{\|\boldsymbol{\mu}\|\leq C} D_N(\boldsymbol{\mu}).$$

   Then $D_N(\widehat{\boldsymbol{\mu}}^*) \to 0$ and $\widehat{\boldsymbol{\mu}}^* \to \mathbf{0}$ almost surely, because

$$D_N(\widehat{\boldsymbol{\mu}}^*) \leq D_N(\mathbf{0}) \xrightarrow{a.s.} D(\mathbf{0}) \leq D(\widehat{\boldsymbol{\mu}}^*).$$

   This can be seen as $D(\mathbf{0}) = 0$ and $\|D_N(\widehat{\boldsymbol{\mu}}^*) - D(\widehat{\boldsymbol{\mu}}^*)\| \xrightarrow{a.s.} 0$.

4. Let $E > C$. If $\|\boldsymbol{\mu}\| \in [C, E]$, then $D_N(\boldsymbol{\mu})$ converges uniformly to $D(\boldsymbol{\mu})$ almost surely (Theorem 10.8 in Rockafellar 1970). As $[C, E]$ is a compact set and due to (B1), almost surely

$$\inf_{\|\boldsymbol{\mu}\|\in[C,E]} D_N(\boldsymbol{\mu}) \to \inf_{\|\boldsymbol{\mu}\|\in[C,E]} D(\boldsymbol{\mu}) = D(\boldsymbol{\mu}^*) = \delta > 0 \text{ and } D_N(\mathbf{0}) \to \mathbf{0}$$

   for some $\|\boldsymbol{\mu}^*\| \in [C, E]$. Convexity of $D_N(\boldsymbol{\mu})$ then shows that $\|\widehat{\boldsymbol{\mu}}\| \leq C$ almost surely.

II. Convergence rate

DEFINITION 1. Write

$$g(\mathbf{x}; \boldsymbol{\mu}) = \|\boldsymbol{\mu}\|^{-1} \left(\|\mathbf{x} - \boldsymbol{\mu}\| - \|\mathbf{x}\| - D(\boldsymbol{\mu}) + \boldsymbol{\mu}^T \|\mathbf{x}\|^{-1}\mathbf{x}\right), \text{for } \boldsymbol{\mu} \neq \mathbf{0},$$

and $g(\mathbf{x}; \mathbf{0}) = 0$. Further write $G_N(\boldsymbol{\mu}) = N^{-1} \sum_i \sum_j g(\mathbf{x}_{ij}; \boldsymbol{\mu})$.

LEMMA 5. *If* $\mathbf{S}_N \xrightarrow{P} \mathbf{0}$ *then* $\sqrt{N} G_N(\mathbf{S}_N) \xrightarrow{P} 0$.

Proof of Lemma 5. First note that

$$\begin{aligned} E\left\{\sqrt{N} G_N(\boldsymbol{\mu})\right\} &= 0 \text{ and that} \\ \operatorname{var}\left\{\sqrt{N} G_N(\boldsymbol{\mu})\right\} &\leq E\left\{g^2(\mathbf{x}; \boldsymbol{\mu})\right\} + \frac{\sum_i m_i(m_i - 1)}{N} E\left\{g^2(\mathbf{x}; \boldsymbol{\mu})\right\}. \end{aligned}$$

Moreover, $\|g(\mathbf{x}; \boldsymbol{\mu})\| \leq 3$ and

$$\frac{1}{\|\boldsymbol{\mu}\|} \left\| \|\mathbf{x} - \boldsymbol{\mu}\| - \|\mathbf{x}\| - D(\boldsymbol{\mu}) + \boldsymbol{\mu}^T \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| \leq 2 \min\left(1, \frac{\|\boldsymbol{\mu}\|}{\|\mathbf{x}\|}\right) + \frac{D(\boldsymbol{\mu})}{\|\boldsymbol{\mu}\|}.$$

(Lemma 19 in Arcones 1998.) By the bounded convergence theorem (Cramér 1945, p. 67) the integral and the limit for continuous and bounded functions can be interchanged. We take the limit $\|\boldsymbol{\mu}\| \xrightarrow{a.s.} \mathbf{0}$ first and immediately see that also

$$\begin{aligned} E\left\{4 \min\left(1, \frac{\|\boldsymbol{\mu}\|^2}{\|\mathbf{x}\|^2}\right)\right\} &\to 0, \\ E\left\{\left[\frac{D(\boldsymbol{\mu})}{\|\boldsymbol{\mu}\|}\right]^2\right\} &\to 0 \text{ and} \\ E\left\{2 \min\left(1, \frac{\|\boldsymbol{\mu}\|}{\|\mathbf{x}\|}\right) \frac{D(\boldsymbol{\mu})}{\|\boldsymbol{\mu}\|}\right\} &\to 0. \end{aligned}$$

Check Lemma 3 to see that $D(\boldsymbol{\mu})/\|\boldsymbol{\mu}\|$ is continuous and bounded.

Thus, var $\left\{\sqrt{N}G_N(\boldsymbol{\mu})\right\}$ goes to zero (see also (A1)). The result then follows as $\mathbf{S}_N$ is bounded in probability.

LEMMA 6. $\sqrt{N}\|\widehat{\boldsymbol{\mu}}\| = O_P(1)$.

Proof of Lemma 6.

1. There exist $C > 0$ and $\delta > 0$ such that if $\|\boldsymbol{\mu}\| < C$ then $D(\boldsymbol{\mu}) \geq \delta\|\boldsymbol{\mu}\|^2$. Use Lemma 3.
2. Almost surely

$$\begin{aligned}\delta N\|\widehat{\boldsymbol{\mu}}\|^2 &\leq ND(\widehat{\boldsymbol{\mu}}) \\ &= ND_N(\widehat{\boldsymbol{\mu}}) - N\|\widehat{\boldsymbol{\mu}}\|G_N(\widehat{\boldsymbol{\mu}}) + N\widehat{\boldsymbol{\mu}}^T\mathbf{T}_N \\ &\leq -N\|\widehat{\boldsymbol{\mu}}\|G_N(\widehat{\boldsymbol{\mu}}) + N\widehat{\boldsymbol{\mu}}^T\mathbf{T}_N.\end{aligned}$$

   The last inequality follows from $D_N(\widehat{\boldsymbol{\mu}}) \leq D_N(\mathbf{0}) = 0$.
3. Therefore

$$\delta N\|\widehat{\boldsymbol{\mu}}\|^2 \leq \sqrt{N}\widehat{\boldsymbol{\mu}}^T\mathbf{O}_N$$

   almost surely where $\mathbf{O}_N$ is bounded in probability. Divide this by $\sqrt{N}\|\widehat{\boldsymbol{\mu}}\|$ and the proof then follows.

III. Asymptotic multinormality

Now we are ready to prove Theorem 1.

1. If $\sqrt{N}\mathbf{S}_N$ is bounded in probability, then $\sqrt{N}\|\mathbf{S}_N\|\ \sqrt{N}G_N(\mathbf{S}_N) \xrightarrow{P} 0$.
2. As $\sqrt{N}\|\widehat{\boldsymbol{\mu}}\|\ \sqrt{N}G_N(\widehat{\boldsymbol{\mu}}) \xrightarrow{P} 0$,

$$\begin{aligned}ND_N(\widehat{\boldsymbol{\mu}}) &= ND(\widehat{\boldsymbol{\mu}}) - \sqrt{N}\widehat{\boldsymbol{\mu}}^T\ \sqrt{N}\mathbf{T}_N + o_P(1) \\ &= \frac{N}{2}\widehat{\boldsymbol{\mu}}^T\mathbf{A}\widehat{\boldsymbol{\mu}} - \sqrt{N}\widehat{\boldsymbol{\mu}}^T\ \sqrt{N}\mathbf{T}_N + o_P(1).\end{aligned}$$

3. As $\sqrt{N}\|\mathbf{A}^{-1}\mathbf{T}_N\|\ \sqrt{N}G_N(\mathbf{A}^{-1}\mathbf{T}_N) \xrightarrow{P} 0$,

$$ND_N(\mathbf{A}^{-1}\mathbf{T}_N) = -\frac{N}{2}\mathbf{T}_N^T\mathbf{A}^{-1}\mathbf{T}_N + o_P(1).$$

4. Then

$$\begin{aligned}N\|\mathbf{A}^{1/2}\widehat{\boldsymbol{\mu}} - \mathbf{A}^{-1/2}\mathbf{T}_N\|^2 &= N\mathbf{T}_N^T\mathbf{A}^{-1}\mathbf{T}_N + N\widehat{\boldsymbol{\mu}}^T\mathbf{A}\widehat{\boldsymbol{\mu}} - 2N\mathbf{T}_N^T\widehat{\boldsymbol{\mu}} \\ &= 2N\left(D_N(\widehat{\boldsymbol{\mu}}) - D_N(\mathbf{A}^{-1}\mathbf{T}_N)\right) + o_P(1) \\ &= o_P(1).\end{aligned}$$

   The last equality follows from $D_N(\widehat{\boldsymbol{\mu}}) \leq D_N(\mathbf{A}^{-1}\mathbf{T}_N)$ and from the fact that

$$2N\left(D_N(\widehat{\boldsymbol{\mu}}) - D_N(\mathbf{A}^{-1}\mathbf{T}_N)\right) + o_P(1) \geq 0.$$

5. Thus $\sqrt{N}\widehat{\boldsymbol{\mu}} - \mathbf{A}^{-1}\sqrt{N}\mathbf{T}_N \xrightarrow{P} 0$. □

Proof of Theorem 2. Nevalainen, Larocque & Oja (2006b).

Proof of Theorem 3. Write $\widehat{\mathbf{\Gamma}} = \widehat{\mathbf{V}}^{-1/2}$ and

$$\widetilde{\boldsymbol{\mu}}(\mathbf{X}) \quad = \quad \widehat{\mathbf{\Gamma}}^{-1}\widehat{\boldsymbol{\mu}}\left(\widehat{\mathbf{\Gamma}}\mathbf{X}\right).$$

for the corresponding affine equivariant transformation retransformation spatial median. Furthermore, write

$$\mathbf{T}_N\left(\mathbf{\Gamma}\right) \quad = \quad \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\mathbf{\Gamma}\mathbf{x}_{ij}}{\|\mathbf{\Gamma}\mathbf{x}_{ij}\|}.$$

As in Randles (2000), it can be proven that $\sqrt{N}\left(\mathbf{T}_N\left(\widehat{\mathbf{\Gamma}}\right) - \mathbf{T}_N\left(\mathbf{I}_p\right)\right) = o_P(1)$ as $\widehat{\mathbf{\Gamma}}$ is $\sqrt{N}$-consistent. Then (using contiguity) also

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\widehat{\mathbf{\Gamma}}(\mathbf{x}_{ij}-\delta/\sqrt{N})}{\|\widehat{\mathbf{\Gamma}}(\mathbf{x}_{ij}-\delta/\sqrt{N})\|} \text{ and } \frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\mathbf{x}_{ij}-\delta/\sqrt{N}}{\|\mathbf{x}_{ij}-\delta/\sqrt{N}\|}$$

have the same limiting distribution $N_p(\mathbf{A}\delta, \mathbf{D})$. As each coordinate of $\|\mathbf{x}_{ij} - \boldsymbol{\mu}\|^{-1}(\mathbf{x}_{ij} - \boldsymbol{\mu})$ is monotone with respect to each coordinate of $\boldsymbol{\mu}$, this implies that $\widetilde{\boldsymbol{\mu}}$, which solves

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\widehat{\mathbf{\Gamma}}\left(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}}\right)}{\|\widehat{\mathbf{\Gamma}}\left(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}}\right)\|} \quad = \quad \mathbf{0},$$

must be $\sqrt{N}$-consistent as well.

Write $\mathbf{\Gamma}^* = \sqrt{N}\left(\widehat{\mathbf{\Gamma}} - \mathbf{I}_p\right)$ and $\boldsymbol{\mu}^* = \sqrt{N}\widetilde{\boldsymbol{\mu}}$. ($\mathbf{\Gamma}^*$ and $\boldsymbol{\mu}^*$ are both bounded in probability.) Using Taylor's expansion around $\mathbf{\Gamma}^* = \mathbf{0}$ and $\boldsymbol{\mu}^* = \mathbf{0}$, it can be shown that

$$\mathbf{0} = \frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\widehat{\mathbf{\Gamma}}\left(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}}\right)}{\|\widehat{\mathbf{\Gamma}}\left(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}}\right)\|} = \frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\mathbf{x}_{ij}}{\|\mathbf{x}_{ij}\|} - \mathbf{A}\boldsymbol{\mu}^* + \mathbf{R}\left(\boldsymbol{\mu}^*, \mathbf{\Gamma}^*\right)$$

where $\mathbf{R}\left(\boldsymbol{\mu}^*, \mathbf{\Gamma}^*\right) \xrightarrow{P} \mathbf{0}$. This gives the desired result. □

Remark. There is no loss in generality in assuming $\sqrt{N}(\widehat{\mathbf{V}} - \mathbf{I}_p) = O_P(1)$. Write $\mathbf{Y} = \mathbf{\Gamma}\mathbf{X}$. By Theorem 3, $\mathbf{\Gamma}\widehat{\mathbf{\Gamma}}^{-1}\widehat{\boldsymbol{\mu}}\left(\widehat{\mathbf{\Gamma}}\mathbf{\Gamma}^{-1}\mathbf{Y}\right)$ and $\widehat{\boldsymbol{\mu}}(\mathbf{Y})$ have the same limiting distribution. Thus, the limiting distribution of $\widehat{\mathbf{\Gamma}}^{-1}\widehat{\boldsymbol{\mu}}\left(\widehat{\mathbf{\Gamma}}\mathbf{X}\right)$ is the same as the limiting distribution of $\mathbf{\Gamma}^{-1}\widehat{\boldsymbol{\mu}}(\mathbf{Y})$. Affine equivariance gives the result.

ACKNOWLEDGEMENTS

The research work of Jaakko Nevalainen and Hannu Oja was partially supported by research grants from the Academy of Finland and by the doctoral program of the School of Statistical Information, Inference, and Data Analysis (SIIDA). The work of Denis Larocque was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Fonds québécois de la recherche sur la nature et les technologies (FQRNT). The authors also wish to thank Anne Konu and Tomi Lintonen for providing the example data set and insight into it.

M. Abramowitz & I. E. Stegun (1970), editors. *Handbook of Mathematical Functions*. Dover, New York.

M. Aerts, H. Geys, G. Molenberghs & L. M. Ryan (2002), editors. *Topics in Modelling of Clustered Data*. Chapman & Hall/CRC, Boca Raton, Florida.

M. A. Arcones (1998). Asymptotic theory for M-estimators over a convex kernel. *Econometric Theory*, 14, 387–422.

B. M. Brown (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society Series B*, 45, 25–30.

B. Chakraborty & P. Chaudhuri (1996). On a transformation and re-transformation technique for constructing affine equivariant multivariate median. In *Proceedings of the American Mathematical Society*, volume 124, pages 2539–2547.

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

H. Cramér (1945). *Mathematical Methods of Statistics*. Almqvist & Wiksells, Uppsala, Sweden.

S. Datta & G. A. Satten (2005). Rank-sum tests for clustered data. *Journal of the American Statistical Association*, 100, 908–915.

A. Dueck & S. Lohr (2005). Robust estimation of multivariate covariance components. *Biometrics*, 61, 162–169.

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

H. Goldstein (2003). *Multilevel Statistical Models*. Third Edition. Arnold, London.

J. C. Gower (1974). The mediancentre. *Applied Statistics*, 23, 466–470.

T. P. Hettmansperger & R. H. Randles (2002). A practical affine equivariant multivariate median. *Biometrika*, 89, 851–860.

A. I. Konu & T. P. Lintonen (2006). School well-being in grades four to twelve. *Health Education Research*, 21, 633–642.

D. Larocque (2003). An affine-invariant multivariate sign test for cluster correlated data. *The Canadian Journal of Statistics*, 31, 437–455.

D. Larocque (2005). The Wilcoxon signed-rank test for cluster correlated data. *Statistical Modeling and Analysis for Complex Data Problems*. P. Duchesne & B. Rémillard, Editors. Springer, 309–323.

D. Larocque, J. Nevalainen & H. Oja. (2007). A weighted multivariate sign test for cluster-correlated data. *Biometrika*, in press.

T. P. Lintonen & A. I. Konu (2006). The Well-Being Profile—an Internet tool for school health promotion. *Promotion & Education*, in press.

J. Möttönen & H. Oja (1995). Multivariate spatial sign and rank methods. *Nonparametric Statistics*, 5, 201–213.

J. Möttönen, H. Oja & J. Tienari (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, 25, 542–552.

J. Nevalainen, D. Larocque & H. Oja (2006a). A weighted spatial median for clustered data. *Statistical Methods & Applications: Special Issue on Robust Statistics*, in press.

J. Nevalainen, D. Larocque & H. Oja (2006b). On the multivariate spatial median for clustered data (full version). Available from Internet: `http://www.uta.fi/~jn67732/files/SpatMedClust_Full.pdf`.

H. Oja & R. H. Randles (2004). Multivariate nonparametric tests. *Statistical Science*, 19, 598–605.

R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2004. URL `http://www.R-project.org`. ISBN 3-900051-00-3.

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

R. T. Rockafellar (1970). *Convex Analysis.* Princeton University Press, Princeton, New Jersey.

B. Rosner, R. J. Glynn & M.-L. Ting Lee (2003). Incorporation of clustering effects for the Wilcoxon rank sum test: a large sample approach. *Biometrics*, 59, 1089–1098.

B. Rosner & D. Grove (1999). Use of the Mann-Whitney U-test for clustered data. *Statistics in Medicine*, 18, 1387–1400.

J. G. Saw (1983). Dependent unit vectors. *Biometrika*, 70, 665–671.

R. J. Serfling (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

J. A. Stoner & B. G. Leroux (2002). Analysis of clustered data: a combined estimating equations approach. *Biometrika*, 89, 567–578.

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

Y. Vardi & C.-H. Zhang (2001). A modified Weiszfeld algorithm for the Fermat-Weber location problem. *Mathematical Programming, Series A*, 90, 559–566.

J. M. Williamson, S. Datta & G. A. Satten (2003). Marginal analyses of clustered data when cluster size is informative. *Biometrics*, 59, 36–42.

Received ???
Accepted ???

Jaakko NEVALAINEN: `jaakko.nevalainen@uta.fi`
Department of Mathematics, Statistics and Philosophy
33014 University of Tampere, Finland

Denis LAROCQUE: `denis.larocque@hec.ca`
Department of Quantitative Methods
HEC Montréal, Québec
Canada H3T 2A7

Hannu OJA: `hannu.oja@uta.fi`
Tampere School of Public Health
33014 University of Tampere, Finland