UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 108

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 108

SPATIAL SIGN AND RANK BASED SCATTER MATRICES WITH APPLICATIONS

SEIJA SIRKIÄ

JYVÄSKYLÄ
2007

Abstract

This work considers scatter and shape estimators, especially those based on the concepts of spatial sign and rank. A new family of scatter estimators, symmetrized M-estimators of scatter, is introduced. These have the so called independence property which in turn is required in the new solution to the independent component problem also introduced in this work. Some applications of scatter matrices in multivariate hypothesis testing are considered, and some robustness and efficiency properties of the suggested methods are studied.

Acknowledgements

This work has been funded mainly by the COMAS graduate school and also by the Finnish Graduate School of Stochastics and Statistics as well as by the Academy of Finland. Personal grants were granted by Emil Aaltonen Foundation and by Ellen and Artturi Nyyssönen Foundation. The Department of Mathematics and Statistics of the University of Jyväskylä provided the required facilities. All of these institutions are hereby duly aknowledged and thanked.

I am most deeply grateful to my supervisor Prof. Hannu Oja. His never ending enthusiasm towards the scientific work as well as the showing of support and encouragement to all of his students stand as an example to everyone. On top of being a world class scientist he is also a friendly and funny person and so working with him has been both an honor and a pleasure.

My gratitute to Dr. Sara Taskinen, my second supervisor, is equally deep. She has been something for me to look up to in so many ways and I can only hope I am able to continue on the path she showed me. The pieces of advise I got during the years, not only regarding academic issues, were countless and will not be forgotten.

I also need to thank the rest of my co-authors, Dr. Jan Eriksson, Dr. Jaakko Nevalainen and Prof. David E. Tyler. The collaboration was fruitful and enjoyable.

It has certainly not been a lonely job. I have been blessed with the nicest and most inspiring colleagues and co-workers imaginable, both in the Nonparametric and robust multivariate methods research group and in the Department of Mathematics and Statistics. I wish to thank every one of them but I have especially fond feelings for our "lunch seminar". The days I missed it for one reason or other were not complete. Among the people in the department Anni Toivola deserves also special thanks for being the good friend she is and all the help I have got from her.

I thank my parents and my dear sister for the home I grew up in. Without that as my background none of this would ever have been possible.

But the most I thank you, Timo. Your significance to my life in total and this work as a part of it has been enormous and it is something I could never disregard and will never forget. I can only thank you for everything. Your patience and understanding have been incredible.

Heidelberg, October 5th 2007

Seija Sirkiä

List of original publications

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

A Oja, H., Sirkiä, S. and Eriksson, J. (2006), Scatter matrices and independent component analysis, *Austrian Journal of Statistics*, **35**, 175-189.

B Sirkiä, S., Taskinen, S., and Oja, H. (2007), Symmetrised M-estimators of scatter, *Journal of Multivariate Analysis*, **98**, 1611-1629.

C Sirkiä, S., Taskinen, S., Oja, H. and Tyler, D. (2007), Tests and estimates of shape based on spatial signs and ranks, submitted.

D Sirkiä, S., Taskinen, S., Nevalainen, J. and Oja, H. (2007), Multivariate nonparametrical methods based on spatial signs and ranks: The R package SpatialNP, submitted.

The author of this thesis has taken active part in the research for the original publications. In addition, the author has provided one of the examples in Article A, had the main responsibility of writing Articles B and C and parts of Article D, and performed all computation required for Articles B and C. The author is the main author and maintainer of the R package SpatialNP considered in Article D.

Contents

Chapter 1

Introduction

In classical multivariate statistics the covariance matrix plays a key role. Its simple form and good properties make it an extremely powerful concept but it is not completely without flaws. The main shortcoming is that the sample covariance matrix, the regular estimate of covariance, is extremely sensitive to errors in data. Moreover, the covariance matrix does not even exist for very heavy tailed distributions. For these reasons alternative measures and estimators of covariation in the multivariate setting are needed.

This thesis considers scatter and shape matrices and their use in multivariate problems. The methods presented are mainly based on the concepts of spatial signs and ranks which are of nonparametric nature. Most of the theory is derived under the elliptical model, a large family of distributions containing as an important special case the normal distribution but also distributions with naturally occuring extreme values. Robustness properties are therefore also of interest.

In this thesis a new family of multivariate scatter estimators, the symmetrized M-estimators of scatter, is introduced. The estimators of this family have the so called independence property which means that the estimators are diagonal whenever the marginal distributions are independent. These estimators do not need any auxiliary location estimator which also motivates their use. Scatter estimators with the independence property are in turn a prerequisite to a new solution to the independent components problem also presented in this thesis.

Scatter matrices are also considered in the context of hypothesis testing in the multivariate setting as a means to create affine invariant spatial sign and rank tests of location and independence. Further, sphericity tests based on spatial sign and rank scatter matrices are considered. Asymptotic results for the introduced estimators are found and comparisons to existing methods are done via analytic efficiency studies and small sample simulations.

The rest of this introductory part is divided into three parts. In Chapter 2 the

basic definitions for the scatter and shape matrices and the multivariate models considered are given as well as some discussion. Robust and nonparametric methods especially in the scatter estimation and tools to compare estimators in this context are considered in Chapter 3. Finally in Chapter 4 some applications of the scatter and shape matrices, in particular the new method for independent components analysis, are presented.

Chapter 2

Multivariate data and scatter matrices

Multivariate data are more than just several measurements bundled together. Covariations and dependencies between individual, univariate random variables are crucial to any inference based on multivariate data. If they are ignored results become flawed or difficult to obtain. More importantly, the existence and nature of the interactions is often the object of interest. Therefore, describing these interactions is at the heart of multivariate statistics.

2.1 Covariance matrix and the normal model

Mathematically, describing the complete distribution of the random vector describes also the covariances of the marginal random variables completely. By far the most important model for multivariate data is the normal model, described by the gaussian density

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

where $\boldsymbol{\mu} = E[\boldsymbol{x}]$ is the mean vector and

$$\Sigma = E\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\right]$$

is the covariance matrix. For discussion of the reasons for this see for example Anderson (1984). The diagonal elements of Σ give the variances of the marginal random variables and its off-diagonal elements give the covariances of all pairs of marginals.

Under the normal model the mean vector and the covariance matrix describe the distribution completely. Also in the case of a general multivariate distribution the covariance matrix is useful. Although it does not tell everything about the

dependencies between the marginals it is still closely related to the general shape and scale of an observed data cloud. This is visualized in Figure 2.1. The top left plot shows realisations from a bivariate normal distribution, and the top right plot from a distribution with independent Laplace distributions as marginals, both having the identity matrix as covariance matrix. The bottom row shows realisations from distributions that are linear transformations of the ones corresponding to the top row, the same transformation for both sides. The solid line in each plot shows the equal density contour corresponding to the 0.9 quantile, that is, 90 percent of the probability mass is contained inside that contour.

The plotted contours on the left and right hand sides look different but are still similar because of the similar structure behind them. Another way to explain this similarity is that, in fact, the distributions corresponding to the bottom row plots also have a common covariance matrix. This is because the covariance matrix is affine equivariant: if $\boldsymbol{y}$ is an affine transformation of $\boldsymbol{x}$, or formally $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$, then the covariance matrix of $\boldsymbol{y}$ is related to the covariance matrix of $\boldsymbol{x}$ through

$$\Sigma_{\boldsymbol{y}} = A\Sigma_{\boldsymbol{x}}A^T.$$

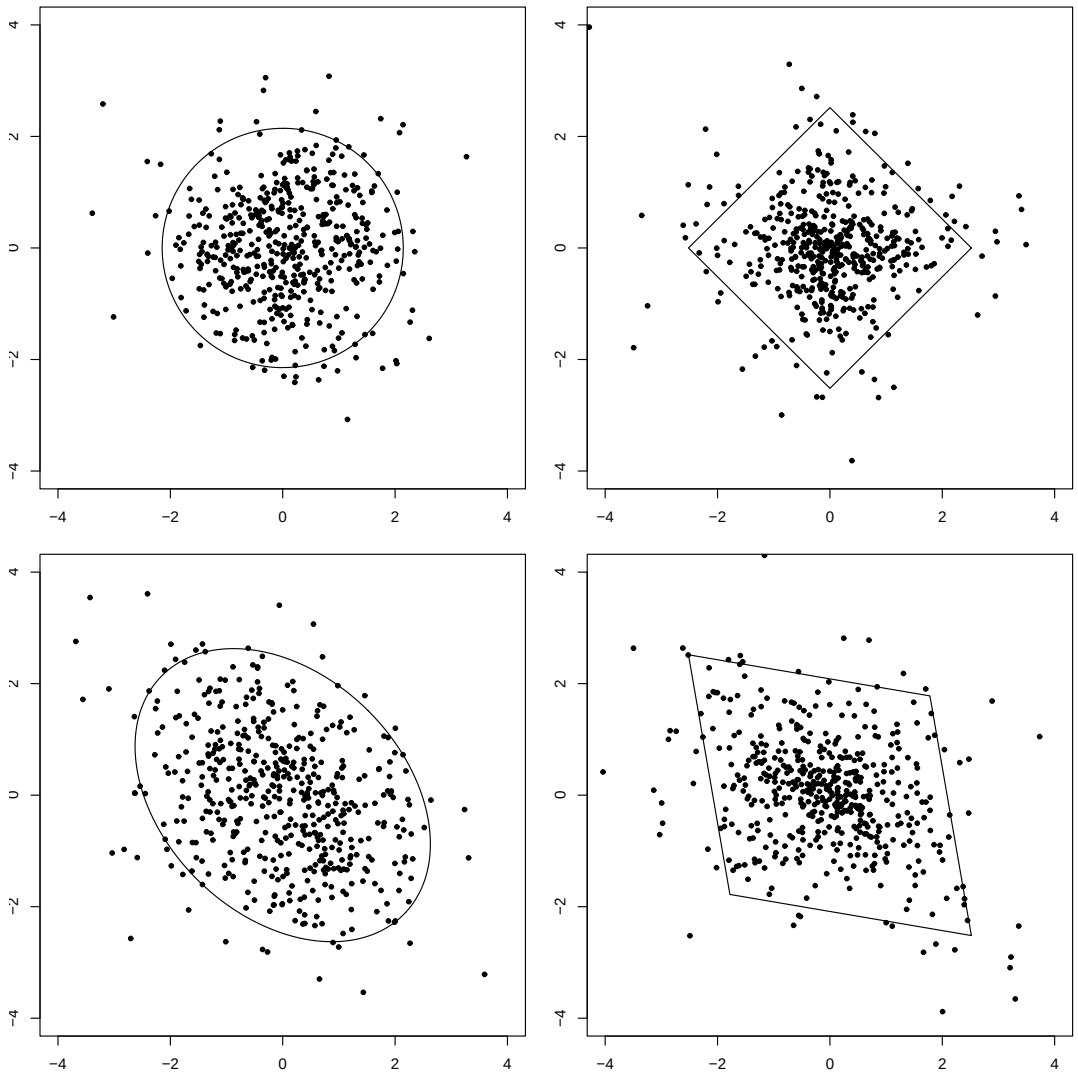
Besides affine equivariance the covariance matrix has other good and strong properties. It is additive: the covariance matrix of the sum of two independent random vectors is simply the sum of the individual covariance matrices. This is a key property in the analysis of variance component models, see for example Searle et al. (1992).

Another property that is interesting is the fact that the covariance matrix is diagonal when the marginals are independent. In the following this is called the independence property. The converse does not hold in general: diagonality of the covariance matrix does not imply the independence of the marginals. This is easy to see by considering the bivariate uniform distribution in the unit square and its rotations which clearly do not have independent marginals. The covariance matrix remains diagonal because of the orthogonal multiplying matrices. The normal model is once again a special case, as the marginals of a gaussian random vector with a diagonal covariance matrix are always independent.

2.2 Scatter and shape matrix functionals

A general scatter matrix functional is now defined as a kind of generalization of the covariance matrix: roughly speaking, a matrix valued functional operating on random vectors is called a scatter functional if it is affine equivariant. Formally, let $\boldsymbol{x}$ be a random p-vector and $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$ its affine transformation where A is

Figure 2.1: Simulated data to illustrate the connection of the covariance matrix and the shape of the observed data cloud



a non-singular $p \times p$ matrix, and let $C(\cdot)$ be a positive definite symmetric matrix valued functional. Then $C(\cdot)$ is a scatter functional if it holds that

$$C(\boldsymbol{y}) = A\ C(\boldsymbol{x})A^T$$

for all such $\boldsymbol{x}$, A and $\boldsymbol{b}$. This requirement alone is enough to ensure that the functional reflects the dependencies between the marginals, or, justifying the name, how observations of the marginal random variables are scattered together. For the covariance matrix the affine equivariance holds for any transformation, even non-singular and non-square matrices, but here it is only required within this restricted group.

A concept closely related to the scatter matrix functional is the shape matrix: the definition of a shape matrix functional is the same as that of a scatter matrix functional but the affine equivariance is required only up to a constant. More precisely, if

$$V(\boldsymbol{y}) \propto A\ V(\boldsymbol{x})A^T$$

holds for any affine transformation as above then $V(\cdot)$ is a shape matrix functional. What may be lost compared to the scatter matrix functional is consistent information on scale. However, for many applications only the shape information is enough. For example, principal components analysis as a method to reduce the number of variables in data only requires information on the order of the principal components and the relative magnitudes of their variances. For this, information about shape is enough, see for example Salibián-Barrera et al. (2006).

Obviously any scatter functional is also a shape functional but some are inherently shape functionals only. Tyler's (1987) M-estimator, for example, is defined as a solution to an estimating equation and this solution is unique only up to a constant. Therefore an additional restiriction is imposed on the scale, namely, that the trace of the matrix is p, equal to the number of dimensions. A similar restriction is common when shape functionals are considered, for example the requirement that the (1,1)-element is equal to one (e.g. Hallin and Paindaveine 2006) or that the determinant of the matrix is equal to one (e.g. Article C of this work). Paindaveine (2007) called the shape defined through the restriction of determinant canonical because only that makes the shape and scale asymptotically independent under the elliptical model (defined in the following).

Sometimes it is reasonable to scale a genuine scatter functional in a way described above (see for example Article B). Clearly, the resulting functional is a shape functional.

2.3 Elliptically symmetric distribution family

The multivariate normal model can also be generalized. The density function of a gaussian random variable can be described by its contours which are cocentric ellipsoids centered at the mean vector and whose shape is given by the covariance matrix. Now, a random vector $\boldsymbol{x}$ with a density function of the form

$$f(\boldsymbol{x}) = \det(\Sigma)^{-1/2} g\left((\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right),$$

where g is a non-negative function on $\mathbb{R}$ such that integral of $g(\boldsymbol{x}^T\boldsymbol{x})$ over $\mathbb{R}^p$ is equal to one, belongs to the elliptically symmetric distribution family. The contours of such a density function are again cocentric ellipsoids centered at $\boldsymbol{\mu}$ and with shape and scale given by Σ which is called here the scatter parameter. For early appearances see Maronna (1976) and Huber (1981, Section 8.4).

The generalization with respect to the normal model is in the function g which gives the relative distances of the equal density contours from each other. If g is such that $\boldsymbol{x}$ has first moments then the location parameter $\boldsymbol{\mu}$ is equal to the mean vector and if $\boldsymbol{x}$ has second moments then the scatter matrix parameter Σ is proportional to the covariance matrix. Because of this Σ can also be called the pseudo-covariance matrix.

Without further restrictions the parameters for a given distribution are ambigious as scale changes of the function g and the scatter matrix Σ can mask each other. This ambiguity can be avoided by restricting the function g in a suitable way. It is also possible to consider only shape, or in other words, to restrict the scale of the scatter parameter for example in one of the three ways explained in Section 2.2. The resulting unique matrix is naturally called a shape parameter.

A special subset of this family consists of the centered and spherically symmetric distributions, that is, the ones with a scatter parameter proportional to the identity matrix, with origin being the symmetry center, i.e. the location parameter. Within this restricted model it is possible to divide the random vector $\boldsymbol{x}$ into two independent parts, its norm $r = \|\boldsymbol{x}\|$ and its direction vector $r^{-1}\boldsymbol{x}$. In fact, the whole elliptically symmetric family is generated by all affine transformations of spherically symmetric distributions. Conversly, it is always possible to "re-transform" an elliptically symmetric random vector into a centered and spherically symmetric random vector by another affine transformation: if $\boldsymbol{x}$ has Σ as its scatter parameter and $\boldsymbol{\mu}$ as its location parameter, then

$$A^{-1}(\boldsymbol{x}-\boldsymbol{\mu}),$$

where A is any matrix for which it holds $AA^T = \Sigma$, is spherically symmetric with respect to the origin. Note that since there are several affine transformations of

a spherically symmetric random variable that have the same distribution, or in particular the same scatter parameter, the standardizing matrix A in the above equation is not unique. In this work the symmetric and positive definite square root is always used. It is denoted and defined by

$$A^{1/2} = U^T L U$$

where L is a diagonal matrix of the square roots of the eigenvalues of A and U is the matrix of corresponding eigenvectors. The Cholesky decomposition is also applicable and used in the literature.

2.4 Independence and IC-model

It was mentioned earlier that the covariance matrix has the independence property. Independence of the marginals does not guarantee the diagonality of a general scatter matrix functional, though. The diagonality of a scatter matrix functional also does not guarantee that the marginals are independent. The covariance matrix under the normal model is an exception: the marginals of a gaussian random vector are independent if and only if its covariance matrix is diagonal. In fact, among the spherically symmetric random vectors the only one with independent marginals is the gaussian random vector.

With these points in mind it is possible to construct another group of multivariate distributions, the one generated from distributions with independent components (marginals) by affine transformations. More formally, these distributions have densities of the form

$$f(\boldsymbol{x}) = |det(A)| \prod_{i=1}^{p} g_i((A\boldsymbol{x})^T \boldsymbol{e}_i),$$

where g_i, $i = 1, \ldots, p$, are univariate density functions, A is a non-singular $p \times p$ matrix and $\boldsymbol{e}_i$ is the ith basis vector ($\boldsymbol{x}^T \boldsymbol{e}_i$ is thus the ith element of $\boldsymbol{x}$). This is called the independent components (IC) -model.

Suppose that a random vector $\boldsymbol{y}$ has independent components and that $\boldsymbol{x} = A^{-1}\boldsymbol{y}$ and further assume that $C(\cdot)$ is a scatter matrix functional. It then holds, by definition, that

$$C(\boldsymbol{x}) = A^{-1}\ C(\boldsymbol{y})(A^{-1})^T =: S.$$

However, an affine transformation $\boldsymbol{z} = B^{-1}\boldsymbol{x}$ where B is such that $BB^T = S$ will not necessarily have the same distribution as $\boldsymbol{y}$ nor even independent components. This means that, unlike in the elliptical family, it is not possible to re-transform

a general member of this family to the generating sub-family using only a scatter matrix functional. The reason for this is that, using the above notation, $C(\boldsymbol{y})$ is not necessarily diagonal while $C(\boldsymbol{z})$ is. Even if $C(\cdot)$ had the independence property the components of $\boldsymbol{z}$ could still be dependent.

Finding the "original" independent components in the above situation is the central question in independent component analysis, or ICA, which is considered in Chapter 4.2 and in Article A. It turns out that under certain assumptions two different scatter matrices can be used to solve this problem, provided that they both have the independence property.

Chapter 3

Robust and nonparametric multivariate methods

Many classical multivariate methods are based on the sample covariance matrix. Examples of these are the generalized T^2-statistic for testing hypotheses about the mean vector, multivariate analysis of variance, principal components analysis and canonical correlation analysis (see Anderson 1984). These methods are usually optimal when the data are from the normal model. The problem is that the sample covariance matrix, the classical estimate of covariance, is very sensitive to errors in the data and miss-specification of the model. This causes also the methods based on the covariance matrix to be unreliable in those cases.

Methods that work correctly even when data contains erroneous and/or extreme observations are called robust. Robust scatter estimation usually involves some kind of downweighting of the most extreme observations. Methods that do not make specific assumptions about the distribution of the data, such as assuming a fixed distribution, up to a finite set of unknown parameters, are called nonparametric. Nonparametric scatter estimation considered in this work uses concepts of spatial (multivariate) signs and ranks. In both cases it is also hoped that the methods do not lose too much of the good properties of the classical methods, for example with respect to efficiency.

3.1 Influence function, breakdown point and efficiency

Influence function (see Hampel et al. 1986) is a way to measure the effect of a single observation on an estimator given in a functional form. It is defined as

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

where $T(F)$ means the value of the estimator functional of interest on a given cdf F and $T(F_\epsilon)$ its value on a contaminated distribution

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

where in turn $\delta_{\boldsymbol{z}}$ is the cdf of a degenerate distribution for which it holds $P(\boldsymbol{z}) = 1$, that is, the Dirac $\delta_{\boldsymbol{z}}$-measure. In other words, the influence function of an estimator is its functional derivative to the direction of a degenerate random vector. It tells the standardized effect on the value of the estimator when a single new observation is introduced at $\boldsymbol{z}$.

A robust estimator is desired to have an influence function that is continuous and bounded. Continuity implies that small shifts in data have only small changes in the estimator. Boundedness implies that no new observation, regardless of how extreme it is, is going to have an arbitrary large effect on the estimator. In this work influence functions are considered in Article B where the influence function of a general symmetrized M-estimator is derived.

Boundedness of the influence function is loosely connected to another measure of robustness, the breakdown point (see again Huber 1981). Roughly speaking, it is the proportion of the data that can be corrupted without completely corrupting the estimator based on such data. More precisely, let T be an estimator and $\boldsymbol{x}_i$, $i = 1, \ldots, n$, a data set. Next, let $\boldsymbol{x}_i^*$ be as the previous data set but with m first vectors changed to arbitrary values. Then the finite sample breakdown point is

$$max \left\{ m : \sup_{\boldsymbol{x}_i^*} \{ d(T(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), T(\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^*)) < \infty \} \right\} / n.$$

where $d(\cdot, \cdot)$ is a suitable metric, relating to the performance of the estimator on the contaminated data with respect to the original data. For vector valued location estimators this metric can be the ordinary euclidian distance. For scatter matrices this metric is usually the largest eigenvalue or the inverse of the smallest eigenvalue, which ever is larger, of the matrix $T^{-1}T^*$, where now T stands for the value of the estimator on the original data and T^* its value on the contaminated data. Breakdown of scatter matrices, though, is not as simple as that of e.g. location, see Davies and Gather (2005).

The breakdown point is then the limit of the finite sample breakdown point as the sample size n tends to infinity. It is now easy to see that if the influence function of an estimator is not bounded its breakdown point is equal to zero. A robust estimator naturally should have a high breakdown point.

Both the influence function and the breakdown point consider the estimator's ability to whitstand errors in the data. Error resistant methods inherently do not "trust" the data completely. So in a situation where there are no errors any inference based on a robust estimator will very likely be less reliable than inference based on a classical estimator, that is, such that assumes a fixed model and that it is the correct one for all data. For hypothesis testing the asymptotic relative

efficiency, or Pitman efficiency, is a common way to measure this difference (see for example Lehmann 1998). The relative efficiency of a test as compared to a reference test is often defined as the ratio of the sample sizes needed to achieve the same size α and the same power $1-\beta$ for a fixed alternative. For chosen values of α and β, the limit of this ratio, as the alternative is approaching the null hypothesis at a correctly chosen rate (often $1/\sqrt{n}$) is then the Pitman efficiency (if it exists). The reference test is often a classical parametric test based on mean vector and covariance matrix, or similar.

Asymptotic relative efficiency of estimators can also be defined in a similar manner, now the equal performance is defined as the estimators having values in a certain neighbourhood of the true value with equal probability. The comparison with large sample sizes requires that the estimators to be compared are consistent with the same convergence rate (often $1/\sqrt{n}$) and have limiting normal distributions. The asymptotic relative efficiency to compare univariate estimators is then simply the ratio of asymptotic variances of the estimators.

In the case where the estimator is inherently multivariate the efficiency of all marginals need not be the same. For example the scatter matrix estimators considered in this work are such that, in the spherically symmetric case, the efficiency of estimation of the diagonal and off-diagonal elements are different. However, the efficiencies of the shape estimators considered can be described by the off-diagonal limiting variances only. This is considered in detail in Article B, where the form of the limiting distribution of scatter estimators in the introduced family of symmetrized M-estimators is found. The efficiencies of certain examples of shape estimators based on those scatter estimators are also computed. Efficiencies of sphericity tests are computed in Article C.

3.2 Spatial signs and ranks

Using the signs and ranks of the observations instead of the original observations is a well known way to implement nonparametric methods in the univariate case, see for example Lehmann (1998) and the references given in its preface. These ideas can be generalized to the multivariate case as well but it is not always very straightforward as there is no natural ordering in the multidimensional space.

The most obvious way to generalize the univariate sign is to use marginal signs to produce vectors with values 1 and -1. These have been used (see Puri and Sen, 1971) but there are problems, most notably the fact that such sign vectors are not affine equivariant, nor even rotation invariant. Using marginals of the observations as univariate observations measured simultaneously also ignores the

idea underlined in the beginning of Chapter 2, that multivariate data is not a collection of the marginal data.

A better way to generalize the univariate signs to the multivariate case becomes obvious if the univariate sign function is written as $sign(x) = |x|^{-1}x$. Interpreting the absolute value as the univariate euclidian norm naturally suggests that the spatial sign function is

$$\boldsymbol{U}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|},$$

with the convention $\mathbf{0}/0 = \mathbf{0}$. Spatial sign vectors are thus unit vectors (cf. the division of a spherically symmetric vector to its length and direction vector, page 12). However, this definition still does not lead to affine equivariant signs although rotation equivariance, that is, equivariance with respect to orthogonal transformations, is obtained. The spatial rank function is now defined under a given distribution F by

$$\boldsymbol{R}(\boldsymbol{x}) = E_{\boldsymbol{y}}[\boldsymbol{U}(\boldsymbol{x} - \boldsymbol{y})],$$

where $\boldsymbol{y}$ is distributed according to F and similarly, the signed rank function

$$\boldsymbol{Q}(\boldsymbol{x}) = \frac{1}{2}E_{\boldsymbol{y}}[\boldsymbol{U}(\boldsymbol{x} - \boldsymbol{y}) + \boldsymbol{U}(\boldsymbol{x} + \boldsymbol{y})].$$

See also Möttönen and Oja (1995). When these functions are applied to a sample of observations and the empirical distribution function the sign, rank and signed rank vectors of the data are obtained. Both rank functions correspond to their univariate counterparts through the fact that the univariate ordinary rank is the sum of signs of pairwise differences (when the sign of zero is taken to be one).

In the sample case a related concept, the set of symmetrized signs is defined by

$$\boldsymbol{U}^s(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{U}(\boldsymbol{x}_i - \boldsymbol{x}_j),\ i, j = 1, \ldots, n.$$

It is then possible to write

$$\begin{aligned} R(\boldsymbol{x}_i) &= ave_j\{\boldsymbol{U}^s(\boldsymbol{x}_i, \boldsymbol{x}_j)\}, \\ Q(\boldsymbol{x}_i) &= (ave_j\{\boldsymbol{U}^s(\boldsymbol{x}_i, \boldsymbol{x}_j)\} + ave_j\{\boldsymbol{U}^s(\boldsymbol{x}_i, -\boldsymbol{x}_j)\})/2 \end{aligned}$$

for the sample spatial ranks and signed ranks. Methods suggested and considered in Article C and Article D rely completely on these four transformations of the data.

3.3 Robust and nonparametric scatter and shape estimation

All of the data transformations introduced in the previous section can be used to create analogies of the covariance matrix. These are the spatial sign covariance matrix

$$SCov(\boldsymbol{x}) = E[\boldsymbol{U}(\boldsymbol{x})\boldsymbol{U}(\boldsymbol{x})^T],$$

the symmetrized spatial sign covariance matrix

$$SSCov(\boldsymbol{x}) = E[\boldsymbol{U}^S(\boldsymbol{x}, \boldsymbol{x}')\boldsymbol{U}^S(\boldsymbol{x}, \boldsymbol{x}')^T],$$

where $\boldsymbol{x}'$ is an independent copy of $\boldsymbol{x}$, the spatial rank covariance matrix

$$RCov(\boldsymbol{x}) = E[\boldsymbol{R}(\boldsymbol{x})\boldsymbol{R}(\boldsymbol{x})^T]$$

and the signed rank covariance matrix

$$SRCov(\boldsymbol{x}) = E[Q(\boldsymbol{x})Q(\boldsymbol{x})^T],$$

with sample versions obtained by replacing the expectation with the average over the sample (and the two copies of the random vector in the case of the symmetrized spatial sign covariance matrix with pairs of observations). All of these are in fact U-statistics as the average is over either the observations themselves or pairs or triplets of them. With the sign covariance matrix and the signed rank covariance matrix it is implicitly assumed that the location is known and equal to the origin, otherwise they have to be combined with a location estimator. See Visuri et al. (2000).

As the sign and rank transformations are not affine equivariant, neither are the covariance matrices based on them and so they are not scatter or shape matrix functionals. However, when the distribution is spherical all of these matrices have expected values proportional to the identity matrix. Therefore they can be applied in the context of sphericity testing. This is thoroughly studied in Article C.

It is possible to suggest shape matrix estimates related to each of the covariance matrices above by an implicit equation. For the sign covariance matrix this is (in the sample form, the function form is obtained by replacing the average by the expectation)

$$ave\left\{\boldsymbol{U}(V^{-1/2}\boldsymbol{x}_i)\boldsymbol{U}(V^{-1/2}\boldsymbol{x}_i)^T\right\} = \frac{1}{p}I_p.$$

This is to say that the estimated shape matrix is the one that standardizes the data so that the signs of the standardized data appear to be "uncorrelated", or in other words the sign covariance matrix of the standardized data is proportional

to the identity matrix. The scale of any possible solution matrix V is ambiguous in this equation. Therefore this equation does not constitute a scatter matrix but it is quite clear that a solution, if any exist, together with a restriction to fix its scale is a shape matrix.

A solution to this particular equation does exist and is known as Tyler's M-estimate (Tyler 1987). It is found by an iterative algorithm suggested by the equation. It is a special case of general M-estimates of scatter (with location known to be the origin, the original form in Maronna (1976) has two simultaneous equations, one for location and another one for scatter) are defined as solutions to equations of the form

$$ave\left\{w(\|V^{-1/2}\boldsymbol{x}_i\|)U(V^{-1/2}\boldsymbol{x}_i)U(V^{-1/2}\boldsymbol{x}_i)^T\right\} \propto I_p,$$

where w is some fixed weight function. Tyler's M is thus obtained by choosing $w(r) = 1$.

The shape estimator corresponding to the symmetrized spatial sign covariance matrix also exists, it is known as Dümbgen's (1998) estimator. It is now possible to suggest a family of estimators such that Dümbgen's estimator is a similar special case of it as Tyler's M is a special case of the M-estimators. This family, symmetrized M-estimators of scatter, is suggested and studied in Article B. They are defined as solutions to equations of a similar form as the regular M-estimators but with signs of the pairwise differences of the observations instead of the signs of the original observations.

The symmetrization makes it unnecessary to know the location of the original observations because the pairwise differences are always located at the origin. What is more important is the fact that symmetrized M-estimators of scatter all have the independence property which is not true for the ordinary M-estimators of scatter and generally overlooked in the literature. The independence property makes it possible to use the symmetrized M-estimators of scatter in independent component analysis, see Section 4.2.

The existence of the scatter estimates corresponding to the rank and signed rank covariance matrices in the general non-elliptic case is uncertain. This is discussed in more detail in Article C. Studying the so called k-step versions of these is on the other hand feasible. It is possible to start with an estimator whose properties are known and apply an iteration step of the form

$$V_k \propto V_{k-1}^{1/2} RCov(V_{k-1}^{-1/2}\boldsymbol{x})V_{k-1}^{1/2},$$

scaled as chosen, which is similar to the one used to find Tyler's or Dümbgen's estimators. The result after a finite number of steps is an estimator whose properties are fairly easy to find. In practice, this sequence appears to converge after a

relatively small number of steps and to a consistent value regardless of the starting point. The properties of this hypothetical limit would most likely be very close to those of a corresponding k-step estimator. The independence property is again of particular interest. If the initial estimator has it, so do all k-step estimators in the iteration sequence of the form given above.

Chapter 4

Applications of sign and rank covariance matrices

In addition to their immediate use in describing the data, the spatial sign and rank covariance matrices and their affine equivariant counterparts can now be applied in several ways. In the following a brief introduction of the methods considered in this work is given. For further reviews on such methods see also Marden (1999a), Möttönen and Oja (1995) and Oja and Randles (2004).

Many of the methods presented here are implemented in the R-package SpatialNP (available at CRAN, `http://cran.r-project.org`). The contents of the package is thoroughly discussed and examples of use are given in Article D.

4.1 Affine invariant nonparametric tests of location and independence and testing of sphericity

A straightforward way to suggest nonparametric tests is to modify an existing classical test by replacing the observations with their signs or ranks. For example, the classical Hotelling's T^2 (Anderson 1984) test statistic for location is constructed by noting that, under the null hypothesis that true location (the mean vector) is the origin,

$$\sqrt{n}\, ave\{\boldsymbol{x}_i\} \rightarrow_d N_p(\mathbf{0}, B),$$

where B is the covariance matrix of $\boldsymbol{x}_i$. Under the null hypothesis the squared norm of the average vector $\bar{\boldsymbol{x}}$ then has a limiting non-central χ^2 distribution and when standardized using the sample covariance matrix

$$\widehat{B} = ave\{\boldsymbol{x}_i \boldsymbol{x}_i^T\}$$

it holds that

$$(\widehat{B}^{-1/2}\bar{\boldsymbol{x}})^T(\widehat{B}^{-1/2}\bar{\boldsymbol{x}}) \rightarrow_d \chi^2(p).$$

This statistic is affine invariant. If now the data is replaced with their spatial signs $\boldsymbol{U}(\boldsymbol{x}_i)$ and the sample covariance matrix by the sample spatial sign covariance matrix $SCov$ a test statistic with the same χ^2 distribution is obtained but it is not affine invariant.

The remedy for the lack of affine invariance is to first standardize the data such that the resulting sign covariance matrix is already, and more importantly, affine invariantly equal to the identity matrix. The test statistic is thus

$$np\|ave\{\boldsymbol{U}(\boldsymbol{z}_i)\}\|^2 \rightarrow_d \chi^2(p)$$

where $\boldsymbol{z}_i = \widehat{V}^{-1/2}\boldsymbol{x}_i$ where in turn $\widehat{V}$ is Tyler's shape estimated on $\boldsymbol{x}_i$. The same kind of location test can be done with signed ranks and the corresponding shape estimator. Further, this method of inner standardisation can be used to find affine invariant nonparametric test of independence similar to the classical Wilks (1935) test, see Taskinen et al. (2003).

As stated above, the spatial sign and rank covariance matrices are proportional to the identity matrix when the data come from a spherical distribution. This fact can be utilized to construct a test for sphericity, or, more generally a test for null hypothesis of the shape parameter being equal to a given matrix. The latter can always be returned to the former by transforming the data according the hypothesized matrix such that under the null the transformed data comes from a spherical distribution. These tests are considered in Article C.

Each of the four sign and rank covariance matrices constitutes a different test. Each test also gives another view to the four shape matrices based on spatial signs and ranks, Tyler's M, Dümbgen's estimator, shape based on spatial ranks and shape based on spatial signed ranks: the shape estimate is the matrix which receives the highest p-value when a corresponding test with that matrix as the null value is performed. In other words the estimate of shape is the one that is least prone to rejection or fits the data best in the light of the test.

4.2 Independent component analysis based on two scatter matrices

Independent component analysis, or ICA, considers a setting where unknown independent random variables, usually called sources in this context, are mixed by an unknown linear mixing, the result of which is observed. The problem is then to find the original sources, or equivalently, the unmixing linear transformation. ICA is then one way to do blind signal separation, that is, to find a set of signals

behind observed data that can be considered "unmixed" in some sense, without much information on what these signals are. Principal components analysis is another example. The difference between the two is that in PCA one wishes to find uncorrelated components that successively explain as much of the total variance as possible. Uncorrelated components can still be dependent, so ICA requires more while on the other hand variance is of no importance.

There are some points that any ICA-method has to take into account. It is never possible, without some technical restrictions, to find exactly the original sources. This is because permutations and rescalings of the sources retain their independence. So, if P is a permutation matrix and D is a diagonal matrix it holds that

$$A\boldsymbol{s} = (AP^{-1}D^{-1})(DP\boldsymbol{s}) = \boldsymbol{x}$$

and there is no way to choose between $\boldsymbol{s}$ and $DP\boldsymbol{s}$ when only $\boldsymbol{x}$ is observed. Another point is that an orthogonal transformation of two independent gaussian variables with equal variances are still independent. This means that if there are more than one gaussian source they cannot be unmixed except up to a rotation.

Many existing ICA-algorithms are based on the following idea: a linear combination of two independent non-gaussian random variables is "more gaussian" than either one of them. This is justified by the central limit theorem. Starting with the observed mixed components the linear combination that is "least normal" then should be one of the original sources. The optimization is then repeated in the remaining orthogonal space, until only one (possibly gaussian) component remains.

This idea requires some measure of gaussianity. FastICA, a popular method proposed by Hyvärinen et al. (2001), chooses entropy, since the normal distribution is the one with maximal entropy when first two moments are fixed. FastICA thus involves a gradient based optimization that minimizes entropy. InfoMax proposed by Adah et al. (2004), in short, aims to maximize the mutual information of separated sources which happens when the sources are independent. Not every algorithm uses optimization, though. For example FOBI (Cardoso, 1989) algorithm is based on fourth order moments, leading to kurtosis as the separating measure.

In this work a new method for ICA is proposed. This method does not rely on optimization of non-gaussianity but on the properties of scatter matrices and the sources themselves. Suppose that S_1 and S_2 are scatter matrices and that both have the independence property. The unmixing matrix based on these matrices and an observed mixing $\boldsymbol{x}$ is

$$B = M_2(S_1(\boldsymbol{x})^{-1/2}\boldsymbol{x})S_1(\boldsymbol{x})^{-1/2},$$

where $M_2(\boldsymbol{z})$ is the matrix of eigenvectors of $S_2(\boldsymbol{z})$. The resulting estimate of the source vector $\boldsymbol{y} = B\boldsymbol{x}$ is such that $S_1(y) = I_p$ and $S_2(y)$ is diagonal, with diagonal elements ordered from highest to lowest. If the original independent sources are such that also for them $S_1(s) = I_p$ and $S_2(s)$ is an ordered diagonal matrix then $\boldsymbol{y}$ equals $\boldsymbol{s}$ up to sign changes, otherwise $\boldsymbol{y}$ equals $\boldsymbol{s}$ up to a rescaling and a permutation.

Since the method uses the eigenvectors of S_2 these should be unambigous to give exactly the correct result. Therefore, in addition to the restriction of there not being more than one gaussian source, no two sources should have the same distribution. If there are, an orthogonal mixing of these sources is found while the other sources are found correctly.

The transformation described above is useful even when the observed data is not produced by a linear mixing of independent sources. The setting is such that the transformed random vector $\boldsymbol{y} = B\boldsymbol{x}$ is ordered with respect to marginal kurtosis defined by the two scatter matrices. In this sense the method is related to the FOBI (Cardoso, 1989) algorithm. Whether this ordering is from high to low or from low to high kurtosis depends on the particular pair of scatter matrices. However, one extremity of kurtosis is a bimodal distribution. This means that this two scatter matrix transformation can be used to find groups in the data: they will be revealed by the first or the last components.

This transformation has yet another useful property: the transformed data set is affine invariant. In other words, the "sources" found are the same regardless of the coordinate system used to describe the original data. Because of this the transformation can be said to describe the data in a data-driven invariant coordinate system, or ICS. It is therefore possible to device affine invariant tests by changing in to these coordinates and performing a corresponding marginal test. The multitude of choices for the two scatter matrices provided for example by the family of symmetrized M-estimators allows for a wide range of tests with for example different robustness or efficiency properties.

4.3 Other methods based on sign and rank covariance matrices

It has been stated above that shape matrices in general and in particular the sign and rank covariance matrices and their affine equivariant counterparts can act as replacements or alternatives of the regular covariance matrix. In principal this carries over to any situation where covariance matrix is used and scale is either of no importance or can be estimated by other means. An example of such

a situation is the principal component analysis as also mentioned before. For a studies of sign and rank covariance matrices in this context see for example Croux et al. (2001) and Marden (1999b).

The main reason for the possibility to use sign and rank covariance matrices in PCA is the fact that principal components are eigenvectors and eigenvectors are of arbitrary length. In canonical correlation analysis the aim is to find linear combinations of two random vectors such that both have uncorrelated marginals and the covariance matrix between the vectors is diagonal with descending diagonal values. This again leads to the eigenvalue setting and therefore to the possibility of using scatter and shape matrices. For a thorough study see Taskinen et al. (2006).

The sign and rank covariance matrices can also be used in multivariate regression setting. An estimate of regression coefficients can be stated as a solution to an estimating equation based on a location estimator, which in turn can be based on spatial signs or ranks. Affine equivariance is then obtained through inner standardization by the corresponding covariance matrix, as with the location and independence tests above. See Article D.

Summary of original publications

In Article A the problem of independent component analysis is stated and a new solution to it based on two scatter matrices with the independence property is given. A proof that this is a solution is given and its performance is illustrated by examples.

Article B considers symmetrized M-estimators of scatter. The family is defined and the influence function and asymptotic distribution of its members are found. Asymptotic relative efficiencies of some members of this group are computed, and a small sample simulation study is performed.

In Article C the covariance matrices and estimates of shape based on spatial signs, symmetrized signs, ranks and signed ranks are treated uniformly. The asymptotic distributions are found and test statistics based on all four transformations for a null hypothesis of sphericity are given. Efficiencies are computed and some power comparisons of the tests in small sample cases are done via a simulation study.

Article D introduces the R package SpatialNP and the methods implemented in it. The use of inner standardization of multivariate tests based on spatial signs and ranks to aquire affine invariance is suggested.

Bibliography

T Adah, T Kim, and V Calhoun. Independent component analysis by complex nonlinearities. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'04)*, 2004.

TW Anderson. *An introduction to Multivariate Statistical Analysis*. Wiley, New York, 2 edition, 1984.

JF Cardoso. Source separation using higher order moments. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'89)*, pages 2109–2112, 1989.

C Croux, E Ollila, and H Oja. Sign and rank covariance matrices: Statistical properties and application to principal component analysis. In Yadolah Dodge, editor, *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pages 257–270. Birkhäuser, Basel, 2001.

PL Davies and U Gather. Breakdown and groups. *Annals of Statistics*, 33: 977–1035, 2005.

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

M Hallin and D Paindaveine. Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Annals of Statistics*, 34: 2707–2756, 2006.

FR Hampel, EM Ronchetti, PJ Rousseeuw, and WJ Stahel. *Robust Statistics: the Approach Based on Influence Functions*. Wiley, New York, 1986.

PJ Huber. *Robust Statistics*. Wiley, New York, 1981.

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

EL Lehmann. *Nonparametrics: Statistical methods based on ranks*. Prentice-Hall, New Yersey, 1998.

JI Marden. Multivariate rank tests. In S Ghosh, editor, *Multivariate Analysis, Design of Experiments, and Survey Sampling*, pages 401–432. CRC, 1999a.

JI Marden. Some robust estimates of principal components. *Statistics & Probability Letters*, 43:349–359, 1999b.

RA Maronna. Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, 4:51–67, 1976.

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

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

D Paindaveine. A canonical definition of shape. *Statistics and Probability Letters (tentatively accepted)*, 2007.

ML Puri and PK Sen. *Nonparametric Methdos in Multivariate Analysis*. Wiley, New York, 1971.

M Salibián-Barrera, S Van Aelst, and Willems G. Principal components analysis based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association*, 101:1198–1211, 2006.

SR Searle, G Casella, and CE McCulloch. *Variance Components*. Wiley, New York, 1992.

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

S Taskinen, C Croux, A Kankainen, E Ollila, and H Oja. Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis*, 97:359–384, 2006.

DE Tyler. A distribution-free M-estimate of multivariate scatter. *Annals of Statistics*, 15:234–251, 1987.

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

SS Wilks. On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3:309–326, 1935.

Original publications