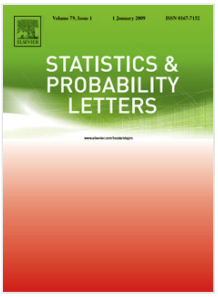
Characteristics of multivariate distributions and the invariant coordinate system

Pauliina Ilmonen, Jaakko Nevalainen, Hannu Oja

Accepted Manuscript



Characteristics of multivariate distributions and the invariant coordinate system

Pauliina Ilmonen, Jaakko Nevalainen, Hannu Oja

PII: S0167-7152(10)00233-6
DOI: 10.1016/j.spl.2010.08.010
Reference: STAPRO 5769

To appear in: *Statistics and Probability Letters*

Received date: 12 April 2010
Revised date: 13 August 2010
Accepted date: 16 August 2010

Please cite this article as: Ilmonen, P., Nevalainen, J., Oja, H., Characteristics of multivariate distributions and the invariant coordinate system. *Statistics and Probability Letters* (2010), doi:10.1016/j.spl.2010.08.010

Characteristics of multivariate distributions and the invariant coordinate system

Pauliina Ilmonen[a,*], Jaakko Nevalainen[b], Hannu Oja[a]

[a] *Tampere School of Public Health, FI-33014 University of Tampere, Finland.*
[b] *Statistics/Department of Social Research, FI-20014 University of Turku, Finland.*

Abstract

We consider a semiparametric multivariate location-scatter model where the standardized random vector of the model is fixed using simultaneously two location vectors and two scatter matrices. The approach using location and scatter functionals based on the first four moments serves as our main example. The four functionals yield in a natural way the corresponding skewness, kurtosis and unmixing matrix functionals. Affine transformation based on the unmixing matrix transforms the variable to an invariant coordinate system. The limiting properties of the skewness, kurtosis, and unmixing matrix estimates are derived under general conditions. We discuss related statistical inference problems, the role of the sample statistics in testing for normality and ellipticity, and connections to invariant coordinate selection and independent component analysis.

Keywords: asymptotic normality, independent component analysis, invariant coordinate selection, multivariate kurtosis, multivariate skewness, scatter matrix
2000 MSC: 62H10, 62H12, 62G05, 62G20, 62F12

1. Introduction

Consider the $n \times p$ data matrix $X = (x_1, ..., x_n)^T$, where $x_1, ..., x_n$ is a random sample from a p-variate distribution. Different *location-scatter*

*Corresponding author: Pauliina Ilmonen.
Email addresses: pauliina.ilmonen@uta.fi (Pauliina Ilmonen), jaakko.nevalainen@utu.fi (Jaakko Nevalainen), hannu.oja@uta.fi (Hannu Oja)

models are obtained if one assumes that

$$x_i = \Omega z_i + \mu, \quad i = 1, ..., n,$$

where $Z = (z_1, ..., z_n)^T$ is an unobservable random sample from a "standardized" distribution, μ is a *location vector* and Ω is a full-rank $p \times p$ matrix, termed the *mixing matrix* in the independent component analysis (ICA) literature. The inverse of Ω, $\Gamma = \Omega^{-1}$, is the *unmixing matrix*, and $\Sigma = \Omega\Omega^T$ is the *scatter matrix*. Posing various assumptions on the distribution of the z_i yields different parametric or semiparametric multivariate models with parametrized by μ and Σ, or by μ and Ω (Nordhausen et al. (2010)).

The following location-scatter models arise from this general structure and are often considered and discussed in the literature.

1. The classical multivariate methods rely on the assumption of multivariate normality, that is, $z_i \sim N_p(0, I_p)$, $i = 1, ..., n$. The location parameter μ is the mean vector and the scatter parameter Σ the covariance matrix. As $Oz_i \sim N_p(0, I_p)$ for all orthogonal matrices O, the mixing matrix Ω or the unmixing matrix Γ are defined only up to an orthogonal transformation in the *multivariate normal model*.
2. In the *multivariate elliptical model* it is assumed that $z_i \sim Oz_i$ for all orthogonal matrices O. (Notation $x \sim y$ means that random variables x and y are similarly distributed.) To fix Σ it is often assumed that $E(\|z_i\|^2) = p$ or that $Med(\|z_i\|^2) = \chi^2_{p,1/2}$. (The first configuration naturally requires that finite second moments exists, but the second allows to avoid any moment assumptions.) As in the multivariate normal model, Ω and Γ are again defined only up to an orthogonal transformation. Elliptical distributions are thus symmetric in the sense that $z_i \sim Oz_i$ for all O, but they may vary in their kurtosis properties. The model permits for heavier (or lighter) tails than the multivariate normal model, and therefore elliptical models are commonly seen as a more realistic alternative to the multivariate normal model. Robust testing and estimation procedures, for example, often assume ellipticity.
3. Another type of model family is obtained if one presumes that the observations arise from a *parametric independent component (IC) model*. Here the z_i are assumed to have independent and standardized components and the density function $f(z_i) = \prod_{j=1}^{p} f_j(z_{ij})$ with some known standardized marginal densities $f_1, ..., f_p$. Matrix Γ is unique for distinct standardized densities $f_1, ..., f_p$.

4. In a generalization of the parametric IC model, the *semiparametric independent component model*, z_i is assumed to consist of independent and standardized components such that $E(z_i) = 0$ and $E(z_i z_i^T) = I_p$, or $Med(z_{ij}) = 0$ and $Med(z_{ij}^2) = \chi^2_{1,1/2}$. But then Ω and Γ are defined only up to permutations and sign changes of the columns and rows, respectively. In ICA the goal is to estimate Γ (up to a permutation, rescaling and sign changes of the rows). In parametric and semiparametric IC models, skewness and kurtosis properties are characteristics of the marginal distributions.

Instead of location-scatter models we will work under a general *semiparametric location-scatter-skewness-kurtosis model* (shortly: semiparametric model), which includes all (continuous as well as discrete) multivariate distributions with finite fourth moments. Thus, many of the more conventional models listed above overlap with the semiparametric model, which was first introduced by Nordhausen et al. (2010).

The paper is organized as follows. The semiparametric model is defined in Section 2 along with a discussion of related unmixing matrix, skewness and kurtosis functionals, G, d and L. Section 3 gives useful asymptotic results for the corresponding estimates $\hat{\Gamma}$, $\hat{\delta}$ and $\hat{\Lambda}$ even outside the semiparametric model. More detailed results for the moment-based estimates are given in Section 4. Statistical properties of the fourth-order blind identification FOBI estimate are obtained as a side-product. Section 5 discusses the uses of sample statistics in testing and estimation problems.

2. Definitions and preliminary results

2.1. Semiparametric model

A multivariate semiparametric model can be defined with natural parameters for multivariate location, scatter, skewness and kurtosis, respectively. The z_i needs to be standardized in a special way using two moment-based location functionals, T_1 and T_2, and two moment-based scatter matrix functionals S_1 and S_2. Next we establish the model, the moment-based location and scatter functionals, and their connection to the model parameters.

Semiparametric model. *Assume that $X = (x_1, ..., x_n)^T$ is random sample from a p-variate distribution such that*

$$x_i = \Omega z_i + \mu, \quad i = 1, ..., n,$$

where the z_i are standardized so that

$$\begin{aligned} E(z_i) &= 0, \\ E(z_i z_i^T) &= I_p, \\ E(z_i z_i^T z_i) &= p \cdot \delta \quad \textit{and} \\ E(z_i z_i^T z_i z_i^T) &= (p+2) \cdot \Lambda \end{aligned}$$

where δ is a p-vector with all components $\delta_i \geq 0$, $i = 1, ..., p$, and Λ is a diagonal matrix with diagonal elements $\lambda_1 \geq ... \geq \lambda_p > 0$.

The semiparametric model was first introduced in Nordhausen et al. (2010). The parameters in the model are the mean vector μ, the covariance matrix $\Sigma = \Omega\Omega^T$, the skewness vector δ based on third moments, and the kurtosis matrix Λ. We will return to these concepts shortly. The model is general in the sense that it includes all p-variate distributions with finite fourth moments, but of course rules out heavy-tailed distributions. The mixing and unmixing matrices, Ω and $\Gamma = \Omega^{-1}$, are uniquely defined if $\delta_i > 0$, $i = 1, ..., p$, and $\lambda_1 > ... > \lambda_p > 0$. When the model parameters are fixed in this way, the unmixing matrix can be used to transform the random vector to an invariant coordinate system (Tyler et al. (2009, ICS)). If the components of z_i are independent, the unmixing matrix Γ is the fourth-order blind identification (FOBI) functional by Cardoso (1989), a solution in ICA. Alternative models and multivariate skewness and kurtosis measures are obtained if the moment-based location and scatter measures are replaced by some other, *e.g.* robust, multivariate location measures and scatter measures.

The model obviously includes the multivariate normal model with $\delta = 0$ and $\Lambda = I_p$. For elliptical distribution $\delta = 0$ and $\Lambda = \lambda I_p$, where λ is a kurtosis parameter, which may not be finite. In the elliptical model Ω is thus defined only up to an orthogonal transformation. However, $\Sigma = \Omega\Omega^T$ is uniquely defined. IC models are included when the marginal densities possess finite fourth-order moments. Recall that the target parameter of ICA is Γ.

2.2. Location and scatter functionals

In robust and nonparametric communities the characteristics of a distribution are often described by functionals. Let F_x be the cumulative distribution function (cdf) of a p-variate random variable x. A *location functional* $T(F_x)$ is a vector-valued $(p \times 1)$ functional, which is affine equivariant in the sense that

$$T(F_{Ax+b}) = AT(F_x) + b$$

for all nonsingular $p \times p$ matrices A and for all p-vectors b. A *scatter functional* $S(F_x)$ is a $p \times p$ -matrix-valued functional which is positive definite and affine equivariant in the sense that

$$S(F_{Ax+b}) = AS(F_x)A^T$$

for all nonsingular $p \times p$ matrices A and for all p-vectors b. A scatter functional S is said to possess the *independence property* if $S(F_x)$ is a diagonal matrix for all x with independent components—a property which not all scatter matrices enjoy, but which is essential in independent component analysis. The first examples of location and scatter functionals are the mean vector and covariance matrix:

$$T_1(F_x) = E(x) \quad \text{and} \quad S_1(F_x) = E\left((x - E(x))(x - E(x))^T\right).$$

The covariance matrix S_1 has the independence property. In the semiparametric model $T_1(F_x) = \mu$ and $S_1(F_x) = \Sigma$.

Location and scatter functionals can be based on the third and fourth moments as well. A location functional based on third moments is

$$T_2(F_x) = \frac{1}{p} E\left((x - E(x))^T S_1(F_x)^{-1}(x - E(x))x\right).$$

Last, a scatter matrix based on fourth moments is

$$S_2(F_x) = \frac{1}{p+2} E\left((x - E(x))(x - E(x))^T S_1(F_x)^{-1}(x - E(x))(x - E(x))^T\right),$$

which has the independence property (Oja et al. (2006)). Note now that these functionals can be used to standardize the random vectors in the semiparametric model as clearly

$$T_1(F_{z_i}) = 0, \quad T_2(F_{z_i}) = \delta, \quad S_1(F_{z_i}) = I_p, \quad \text{and} \quad S_2(F_{z_i}) = \Lambda.$$

2.3. Skewness, kurtosis, and unmixing matrix functionals

Without fixing any particular location and scatter functionals, like the moment-based functionals in the above, the *unmixing matrix functional* G $(p \times p)$, *skewness functional* d $(p \times 1)$ and *kurtosis functional* L $(p \times p)$, based on two pairs of some location and scatter functionals, (T_1, S_1) and (T_2, S_2), can be defined as follows.

Definition 2.1. *Let a matrix-valued functional G, a vector-valued functional d, and a diagonal matrix-valued functional L be defined so that, if $z = G(F_x)(x - T_1(F_x))$, then*

$$T_1(F_z) = 0, \quad S_1(F_z) = I_p, \quad T_2(F_z) = d, \quad \textit{and} \quad S_2(F_z) = L,$$

where $d \geq 0$ and the diagonal elements of L are in a decreasing order.

Note that G and L are solutions of the eigenvector and eigenvalue problem

$$S_1^{-1} S_2 G^T = G^T L.$$

A solution G is then unique up to a permutation, rescaling and sign changes of the rows. Among these, Definition 2.1 then picks up the solution G for which $S_1(F_z) = I_p$, $T_2(F_z) \geq 0$ and $S_2(F_z)$ is a diagonal matrix with decreasing diagonal elements. The first condition fixes the scales, the second one the signs, and the third one the order of the rows of G. The solution G then also satisfies

$$GS_1G^T = I_p \quad \text{and} \quad GS_2G^T = L,$$

where, as before, L is a diagonal matrix consisting of the eigenvalues of $S_1^{-1}S_2$. The solution is unique if $d > 0$ and L has distinct diagonal elements.

For functionals G, d, and L we then have the following lemma:

Lemma 2.1. *Assume the semiparametric model $x = \Omega z + \mu$, where for some location functionals T_1 and T_2 and for some scatter functionals S_1 and S_2, $T_1(F_z) = 0$, $S_1(F_z) = I_p$, $T_2(F_z) = \delta > 0$ and $S_2(F_z) = \Lambda$ is a diagonal matrix with diagonal elements $\lambda_1 > ... > \lambda_p > 0$. If G is based on (T_1, S_1) and (T_2, S_2), then*

$$G(F_x) = \Gamma, \quad d(F_x) = \delta, \quad \textit{and} \quad L(F_x) = \Lambda.$$

The functionals are affine equivariant and invariant in the sense that

$$G(F_{Ax+b}) = G(F_x)A^{-1}, \quad d(F_{Ax+b}) = d(F_x), \quad \textit{and} \quad L(F_{Ax+b}) = L(F_x).$$

for all nonsingular $p \times p$ matrices A and all p-vectors b.

The values of the functionals at the empirical distribution F_n yield natural Fisher consistent estimates of the corresponding population quantities. For

an estimate of $T(F_x)$ we then write $T(F_n)$ or $T(X)$ where X is an $n \times p$ data matrix. To simplify notation, we write

$$T_1 = T_1(F_x), \quad S_1 = S_1(F_x), \quad T_2 = T_2(F_x), \quad \text{and} \quad S_2 = S_2(F_x),$$

and

$$\hat{T}_1 = T_1(F_n), \quad \hat{S}_1 = S_1(F_n), \quad \hat{T}_2 = T_2(F_n), \quad \text{and} \quad \hat{S}_2 = S_2(F_n).$$

Most of the time, the interest lies in the population parameters

$$\Gamma = G(F_x), \quad \delta = d(F_x), \quad \text{and} \quad \Lambda = L(F_x)$$

and their Fisher consistent estimates

$$\hat{\Gamma} = G(F_n), \quad \hat{\delta} = d(F_n), \quad \text{and} \quad \hat{\Lambda} = L(F_n),$$

respectively. Of course, the estimates $\hat{\delta}$, $\hat{\Gamma}$ and $\hat{\Lambda}$ adopt the same equivariance and invariance properties as the corresponding functionals meaning that

$$\begin{aligned} G(XA^T + 1_n b^T) &= G(X)A^{-1}, \\ d(XA^T + 1_n b^T) &= d(X) \quad \text{and} \\ L(XA^T + 1_n b^T) &= L(X). \end{aligned}$$

2.4. Connections to ICA, ICS and classical skewness and kurtosis measures

If T_1, T_2, S_1 and S_2 are the moment-based functionals we can say more, and find similarities in the literature. First, if the observations come from a continuous distribution, the estimates exist and are unique with probability one. Second, the estimate $\hat{\Gamma}$ is the well-known FOBI estimate in the IC model. Third, in the univariate case ($p = 1$), $||\hat{\delta}||^2$ and $\hat{\Lambda}$ reduce to the classical univariate skewness and kurtosis measures

$$\frac{[E(x - E(x))^3]^2}{[E(x - E(x))^2]^3} \quad \text{and} \quad \frac{E(x - E(x))^4}{3\,[E(x - E(x))^2]^2}.$$

In the multivariate case Mardia (1970) defined different moment-based measures of skewness and kurtosis for a sample $X = (x_1, ..., x_n)'$ as

$$b_1 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} ((x_i - \hat{T}_1)^T \hat{S}_1^{-1} (x_j - \hat{T}_1))^3$$

and

$$b_2 = \frac{1}{n}\sum_{i=1}^{n}((x_i - T_1)^T \hat{S}_1^{-1}(x_i - T_1))^2$$

which use the moments of elements of a maximal invariant statistic, the $n \times n$ matrix

$$(X - 1_n\hat{T}_1^T)\hat{S}_1^{-1}(X - 1_n\hat{T}_1^T)^T.$$

Our skewness and kurtosis statistics are based on third and fourth moments of another maximal invariant statistic, namely, the $n \times p$ matrix

$$\hat{Z} = (X - 1_n\hat{T}_1^T)\hat{\Gamma}^T.$$

Matrix $\hat{Z}$ gives the observations in an invariant coordinate system (Tyler et al. (2009)). Still another invariant coordinate system (maximal invariant statistic) based on $p + 1$ observations was proposed by Chakraborty and Chaudhuri (1999). Their approach is known as the transformation-retransformation approach. See also Serfling (2010) for a general discussion on standardization, weak covariance, transformation-retransformation, and strong invariant coordinate system functionals. Bera and John (1983) use the third and fourth moments of the "scaled residuals" in $(X-1_n\hat{T}_1^T)\hat{S}_1^{-1/2}$ (with a symmetric square root matrix). Unlike Mardia's statistics and our skewness and kurtosis statistics, their statistics are not affine invariant, however.

3. Asymptotical properties

We are interested in the limiting behavior of the estimates $\hat{\delta}$, $\hat{\Gamma}$ and $\hat{\Lambda}$. As the estimates are affine equivariant and invariant, it is not a restriction to assume that

$$T_1 = 0, \quad S_1 = I_p, \quad T_2 = \delta \quad \text{and} \quad S_2 = \Lambda \quad \text{, and therefore} \quad \Gamma = I_p.$$

For uniqueness, we assume that $\delta_i > 0$, $i = 1, ..., p$, and the diagonal elements of Λ are strictly ordered so that $\lambda_1 > ... > \lambda_p > 0$.

We assume that the location and scatter estimates, not necessarily moment-based yet, are root-n consistent, that is,

$$\sqrt{n}\hat{T}_1 = O_p(1) \quad \text{and} \quad \sqrt{n}(\hat{S}_1 - I_p) = O_p(1)$$

as well as

$$\sqrt{n}(\hat{T}_2 - \delta) = O_p(1) \quad \text{and} \quad \sqrt{n}(\hat{S}_2 - \Lambda) = O_p(1).$$

Then we have the following result.

Theorem 3.1. *If $\hat{T}_1$, $\hat{S}_1$, $\hat{T}_2$ and $\hat{S}_2$ are root-n consistent, then so are $\hat{\delta}$, $\hat{\Gamma}$ and $\hat{\Lambda}$ and*

$$\sqrt{n}(\hat{\delta} - \delta) = \sqrt{n}(\hat{T}_2 - \delta) - \sqrt{n}\hat{T}_1 + \sqrt{n}(\hat{\Gamma} - I_p)\delta + o_p(1)$$

and

$$\begin{aligned}
\sqrt{n}(\hat{\Gamma}_{ii} - 1) &= -\frac{1}{2}\sqrt{n}((\hat{S}_1)_{ii} - 1) + o_p(1), \\
(\lambda_i - \lambda_j)\sqrt{n}\hat{\Gamma}_{ij} &= \sqrt{n}(\hat{S}_2)_{ij} - \lambda_i\sqrt{n}(\hat{S}_1)_{ij} + o_p(1),\ i \neq j, \quad \textit{and} \\
\sqrt{n}(\hat{\Lambda}_{ii} - \lambda_i) &= \sqrt{n}((\hat{S}_2)_{ii} - \lambda_i) - \lambda_i\sqrt{n}((\hat{S}_1)_{ii} - 1) + o_p(1).
\end{aligned}$$

Proof. First note that since the transformation $(T_1, S_1, T_2, S_2) \to (d, G, L)$ is continuous in a neighborhood of $(0, I_p, \delta, \Lambda)$, and $(\hat{T}_1, \hat{S}_1, \hat{T}_2, \hat{S}_2) \to_P (0, I_p, \delta, \Lambda)$, also $(\hat{\delta}, \hat{\Gamma}, \hat{\Lambda}) \to_P (\delta, I_p, \Lambda)$. As

$$\hat{\delta} = T_2\left((X - 1_n\hat{T}_1^T)\hat{\Gamma}^T\right),$$

it follows by affine equivariance that

$$\sqrt{n}\left(T_2\left((X - 1_n\hat{T}_1^T)\hat{\Gamma}^T\right) - \delta\right) = \sqrt{n}\hat{\Gamma}(\hat{T}_2 - \hat{T}_1 - \delta) + \sqrt{n}(\hat{\Gamma} - I_p)\delta.$$

Then by Slutsky's theorem $\sqrt{n}\hat{\Gamma}(\hat{T}_2 - \hat{T}_1 - \delta) - \sqrt{n}(\hat{T}_2 - \delta) + \sqrt{n}\hat{T}_1$ converges to 0 in distribution and therefore in probability as well. Thus

$$\sqrt{n}\hat{\Gamma}(\hat{T}_2 - \hat{T}_1 - \delta) = \sqrt{n}(\hat{T}_2 - \delta) - \sqrt{n}\hat{T}_1 + o_p(1),$$

and the first part of the theorem follows. For $\hat{\Gamma}$ and $\hat{\Lambda}$ we utilize the estimating equations

$$\hat{\Gamma}\hat{S}_1\hat{\Gamma}^T = I_p \ \text{ and } \ \hat{\Gamma}\hat{S}_2\hat{\Gamma}^T = \hat{\Lambda}.$$

Then

$$\begin{aligned}
(\hat{\Gamma} - I_p)\hat{S}_1\hat{\Gamma}^T + (\hat{S}_1 - I_p)\hat{\Gamma}^T + (\hat{\Gamma} - I_p)^T &= 0 \ \text{ and} \\
(\hat{\Gamma} - I_p)\hat{S}_2\hat{\Gamma}^T + (\hat{S}_2 - \Lambda)\hat{\Gamma}^T + \Lambda(\hat{\Gamma} - I_p)^T &= \hat{\Lambda} - \Lambda
\end{aligned}$$

and Slutsky's theorem gives

$$\begin{aligned}
\sqrt{n}(\hat{S}_1 - I_p) &= -\sqrt{n}(\hat{\Gamma} - I_p) - \sqrt{n}(\hat{\Gamma} - I_p)^T + o_p(1) \ \text{ and} \\
\sqrt{n}(\hat{S}_2 - \Lambda) &= -\sqrt{n}(\hat{\Gamma} - I_p)\Lambda - \sqrt{n}\Lambda(\hat{\Gamma} - I_p)^T + \sqrt{n}(\hat{\Lambda} - \Lambda) + o_p(1).
\end{aligned}$$

These equations yield the desired results for $\hat{\Gamma}$ and $\hat{\Lambda}$. □

In the matrix form, we can write

$$\begin{aligned}
\sqrt{n}\ diag(\hat{\Gamma} - I_p) &= -\frac{1}{2}\sqrt{n}\ diag(S_1 - I_p) + o_p(1), \\
\sqrt{n}\ (\hat{\Gamma} - diag(\hat{\Gamma})) &= \sqrt{n}\ H \odot \Big((\hat{S}_2 - \Lambda) - (\hat{S}_1 - I_p)\Lambda\Big) + o_p(1), \quad \text{and} \\
\sqrt{n}\ (\hat{\Lambda} - \Lambda) &= \sqrt{n}\ diag\Big((\hat{S}_2 - \Lambda) - (\hat{S}_1 - I_p)\Lambda\Big) + o_p(1),
\end{aligned}$$

where H is a $p \times p$ matrix with elements

$$H_{ij} = 0, \text{ if } i = j, \quad \text{and} \quad H_{ij} = (\lambda_i - \lambda_j)^{-1}, \text{ if } i \neq j,$$

$diag(\Gamma)$ for example is a diagonal matrix with the same diagonal elements as Γ, and $\odot$ means the Hadamard (entrywise) product.

Note that the principal component analysis is a special case here: if one takes $S_1 = I_p$ (constant) and $S_2 = S$, the theorem gives the limiting behavior of the eigenvectors and eigenvalues of S.

4. Limiting distributions of the moment-based estimates

We next establish the limiting distributions of the estimates $\hat{\delta}$, $\hat{\Gamma}$ and $\hat{\Lambda}$ obtained by using the moment-based location and scatter statistics

$$\hat{T}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \text{and} \quad \hat{S}_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{T}_1)(x_i - \hat{T}_1)^T$$

and

$$\hat{T}_2 = \frac{1}{np}\sum_{i=1}^{n}(x_i - \hat{T}_1)^T \hat{S}_1^{-1}(x_i - \hat{T}_1)x_i$$

and

$$\hat{S}_2 = \frac{1}{n(p+2)}\sum_{i=1}^{n}(x_i - \hat{T}_1)(x_i - \hat{T}_1)^T \hat{S}_1^{-1}(x_i - \hat{T}_1)(x_i - \hat{T}_1)^T.$$

Again, as the estimates are affine equivariant and invariant, we may assume that the population values are

$$T_1 = 0, \quad S_1 = I_p, \quad T_2 = \delta \quad \text{and} \quad S_2 = \Lambda \quad \text{, and therefore} \quad \Gamma = I_p.$$

Thus $x_i = z_i$, $i = 1, ..., n$, and we write

$$\tilde{T}_1 = \frac{1}{n}\sum_{i=1}^{n} z_i, \quad \text{and} \quad \tilde{S}_1 = \frac{1}{n}\sum_{i=1}^{n} z_i z_i^T$$

and

$$\tilde{T}_2 = \frac{1}{np}\sum_{i=1}^{n} z_i z_i^T z_i \quad \text{and} \quad \tilde{S}_2 = \frac{1}{n(p+2)}\sum_{i=1}^{n} z_i z_i^T z_i z_i^T.$$

If the first eight moments of z_i exist, the joint distribution of

$$\sqrt{n}\begin{pmatrix} \tilde{T}_1 \\ vec(\tilde{S}_1 - I_p) \\ \tilde{T}_2 - \delta \\ vec(\tilde{S}_2 - \Lambda) \end{pmatrix}$$

is, by the central limit theorem, a $2(p + p^2)$ variate (singular) normal distribution with mean zero and covariance matrix given by

$$D = E\left(\begin{pmatrix} z_i \\ vec(z_i z_i^T - I_p) \\ \frac{1}{p} z_i z_i^T z_i - \delta \\ vec\left(\frac{1}{p+2} z_i z_i^T z_i z_i^T - \Lambda\right) \end{pmatrix}\begin{pmatrix} z_i \\ vec(z_i z_i^T - I_p) \\ \frac{1}{p} z_i z_i^T z_i - \delta \\ vec\left(\frac{1}{p+2} z_i z_i^T z_i z_i^T - \Lambda\right) \end{pmatrix}^T\right).$$

One can show that

$$\sqrt{n}\begin{pmatrix} \hat{T}_1 \\ vec(\hat{S}_1 - I_p) \\ \hat{T}_2 - \delta \\ vec(\hat{S}_2 - \Lambda) \end{pmatrix} = C\sqrt{n}\begin{pmatrix} \tilde{T}_1 \\ vec(\tilde{S}_1 - I_p) \\ \tilde{T}_2 - \delta \\ vec(\tilde{S}_2 - \Lambda) \end{pmatrix} + o_p(1),$$

where

$$C = \begin{pmatrix} I_p & 0 & 0 & 0 \\ 0 & I_{p^2} & 0 & 0 \\ C_{31} & C_{32} & I_p & 0 \\ C_{41} & C_{42} & 0 & I_{p^2} \end{pmatrix}$$

with

$$C_{31} = -\frac{2}{p} I_p \quad \text{and} \quad C_{32} = -\frac{1}{p} E\left(z_i^T \otimes (z_i z_i^T)\right)$$

and

$$C_{41} = -\frac{1}{p+2}\left[E\left((z_i^T z_i)(I_p \otimes z_i)\right) + E\left((z_i^T z_i)(z_i \otimes I_p)\right) + 2 \cdot E\left(z_i \otimes (z_i z_i^T)\right)\right]$$

and

$$C_{42} = -\frac{1}{p+2}E\left((z_i z_i^T) \otimes (z_i z_i^T)\right).$$

The asymptotic normality of

$$\sqrt{n}\begin{pmatrix} \hat{T}_1 \\ vec(\hat{S}_1 - I_p) \\ \hat{T}_2 - \delta \\ vec(\hat{S}_2 - \Lambda) \end{pmatrix}$$

then follows.

Finally, if $\delta_i > 0$, $i = 1, ..., p$, and the diagonal elements of Λ are strictly ordered, we get

$$\sqrt{n}\begin{pmatrix} \hat{\delta} - \delta \\ vec\left(\hat{\Lambda} - \Lambda\right) \\ vec\left(\hat{\Gamma} - \Gamma\right) \end{pmatrix} = B\sqrt{n}\begin{pmatrix} \hat{T}_1 \\ vec\left(\hat{S}_1 - I_p\right) \\ \hat{T}_2 - \delta \\ vec\left(\hat{S}_2 - \Lambda\right) \end{pmatrix},$$

where

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ 0 & B_{22} & 0 & B_{24} \\ 0 & B_{32} & 0 & B_{34} \end{pmatrix}$$

with

$$B_{11} = -I_p, \text{ and } B_{12} = [\delta^T \otimes I_p]\left[-\frac{1}{2}diag(vec(I_p)) - diag(vec(H))(\Lambda \otimes I_p)\right]$$

and

$$B_{13} = I_p \text{ and } B_{14} = [\delta^T \otimes I_p]diag(vec(H))$$

and

$$B_{22} = -diag(vec(I_p))(\Lambda \otimes I_p) \text{ and } B_{24} = diag(vec(I_p))$$

and

$$B_{32} = -\frac{1}{2} diag(vec(I_p)) - diag(vec(H))(\Lambda \otimes I_p) \quad \text{and} \quad B_{34} = diag(vec(H)).$$

(The matrix H is given in Section 3.)

We have thus proved the following theorem.

Theorem 4.1. *Assume that X is a random sample from the semiparametric model with $\Omega = I_p$, $\mu = 0$, $\delta_i > 0$, $i = 1, ..., p$, and $\lambda_1 > ... > \lambda_p > 0$. Assume also that the first eight moments of z_i are finite. Then*

$$\sqrt{n} \begin{pmatrix} \hat{\delta} - \delta \\ vec\left(\hat{\Lambda} - \Lambda\right) \\ vec\left(\hat{\Gamma} - I_p\right) \end{pmatrix}$$

has the limiting (singular) $p + 2p^2$-variate normal distribution with mean value zero and covariance matrix $BCDC^T B^T$.

By affine equivariance and invariance properties of the estimates this generalizes to

Corollary 4.1. *Assume that X is a random sample from the semiparametric model with $\delta_i > 0$, $i = 1, ..., p$, and $\lambda_1 > ... > \lambda_p > 0$. Assume also that the first eight moments of z_i are finite. Then*

$$\sqrt{n} \begin{pmatrix} \hat{\delta} - \delta \\ vec\left(\hat{\Lambda} - \Lambda\right) \\ vec\left(\hat{\Gamma} - \Gamma\right) \end{pmatrix}$$

has the limiting (singular) $p + 2p^2$-variate normal distribution with mean value zero and covariance matrix $ABCDC^T B^T A^T$ where

$$A = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{p^2} & 0 \\ 0 & 0 & \Gamma^T \otimes I_p \end{pmatrix}.$$

Remark 4.1. *Under the assumption of central symmetry (i.e. $z_i \sim -z_i$) the value of $\delta = 0$, but even then the unmixing matrix functional satisfying $\hat{\Gamma}\hat{S}_1\hat{\Gamma}^T = I_p$ is affine equivariant up to the signs of its row vectors. In order to fix the signs of the unmixing matrix functional and the estimate of it, we can require for example that $\Gamma 1_p > 0$ (and set $\hat{\Gamma}1_p > 0$). Now, even under the assumption of central symmetry, the limiting joint distribution of $\hat{\Gamma}$ and $\hat{\Lambda}$ is the one given in Theorem 4.1 and Corollary 4.1. In the elliptic case $\delta = 0$ and $\Lambda = \lambda I_p$ and the limiting behavior of $\hat{\delta}$, $\hat{\Lambda}$, and $\hat{\Gamma}$ is unknown. However, the limiting properties of $\|\hat{\delta}\|^2$ and the mean and variance of the elements of $\hat{\Lambda}$ are known, see Section 5.2.*

Remark 4.2. *In this section we derived the asymptotic joint distribution of the skewness, kurtosis and unmixing matrix estimators for moment-based functionals. Since Theorem 3.1 is not restricted to moment-based functionals, the same methodology can be used when other location and scatter functionals are considered, as long as the joint limiting distribution of the corresponding location and scatter estimates is known.*

5. Applications and concluding remarks

5.1. Statistical inference

The results in Section 4 can be used to find estimates of the limiting covariance matrices of the estimates $\hat{\Gamma}$, $\hat{\delta}$ and $\hat{\Lambda}$. These could then in turn be used in the construction of confidence ellipsoids for the parameters. The results can also be employed in the development and the conduct of interesting testing procedures, which we are currently working on.

To estimate the limiting distribution in practice, one can proceed as follows.

1. Calculate $\hat{T}_1$, $\hat{T}_2$, $\hat{S}_1$ and $\hat{S}_2$.
2. Find estimates $\hat{\Gamma}$, $\hat{\delta}$ and $\hat{\Lambda}$.
3. Transform observations to the invariant coordinate system:
$$\hat{Z} = (X - 1_n\hat{T}_1^T)\hat{\Gamma}^T.$$
4. Find estimates $\hat{D}$, $\hat{C}$, $\hat{B}$ and $\hat{A}$: in the formulas for D and C replace the expectations by averages and the (unknown) z_i by $\hat{z}_i$, $i = 1, ..., n$.

5. Then, approximately,

$$\begin{pmatrix} \hat{\delta} - \delta \\ vec\left(\hat{\Lambda} - \Lambda\right) \\ vec\left(\hat{\Gamma} - \Gamma\right) \end{pmatrix} \sim N_{p+2p^2}\left(0, \frac{1}{n}\hat{A}\hat{B}\hat{C}\hat{D}\hat{C}^T\hat{B}^T\hat{A}^T\right).$$

A bootstrap technique can also be used to estimate the distributions of the sample statistics: Let U be a random $n \times n$ matrix such that the rows are independent and the row vectors have $Multin(1; (1/n, ..., 1/n))$ distribution. Then U is called a bootstrap matrix and UX is a bootstrap sample. The bootstrap estimates of the covariance matrices of $\hat{\Gamma}$, $\hat{\delta}$ and $\hat{\Lambda}$, for example, can be found as follows.

1. Calculate $\hat{T}_1$, $\hat{T}_2$, $\hat{S}_1$ and $\hat{S}_2$.
2. Find estimates $\hat{\Gamma}$, $\hat{\delta}$ and $\hat{\Lambda}$.
3. Choose M independent bootstrap matrices $U_1, ..., U_M$.
4. Calculate M bootstrap samples $X_i^* = U_i X$, $i = 1, ..., M$.
5. Calculate M bootstrap estimates

$$\hat{\delta}_i^* = d(X_i^*), \quad \hat{\Gamma}_i^* = G(X_i^*) \text{ and } \hat{\Lambda}_i^* = L(X_i^*), \quad i = 1, ..., M.$$

6. Calculate the sample covariance matrix of

$$\begin{pmatrix} \hat{\delta}_i^* - \hat{\delta} \\ vec\left(\hat{\Lambda}_i^* - \hat{\Lambda}\right) \\ vec\left(\hat{\Gamma}_i^* - \hat{\Gamma}\right) \end{pmatrix}, \quad i = 1, ..., M.$$

5.2. Tests for normality and ellipticity

Skewness and kurtosis statistics can be used to test for normality and/or ellipticity. In the elliptic case $\delta = 0$ and $\Lambda = \lambda I_p$. In the multivariate normal case $\lambda = 1$. Our assumptions for δ and Λ stated in Theorem 3.1 are thus not true, and the limiting behavior of $\hat{\delta}$, $\hat{\Lambda}$, and $\hat{\Gamma}$ is unknown. However, the limiting properties of $\|\hat{\delta}\|^2$ and the mean and variance of the elements of $\hat{\Lambda}$ are known, and may be used in the following way.

As skewness and kurtosis are affine invariant, it is not a restriction to assume that the distribution is spherical. Then

$$T_1 = T_2 = 0, \quad S_1 = I_p \text{ and } S_2 = \lambda I_p,$$

for some $\lambda > 0$. It is well known that in the spherical case the location statistics often satisfy

$$\sqrt{n}\hat{T}_i = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\gamma_i(r_j)u_j + o_p(1), \quad i = 1, 2,$$

and the scatter statistics satisfy

$$\sqrt{n}(\hat{S}_i - S_i) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\alpha_i(r_j)u_j u_j^T - \beta_i(r_j)S_i\right) + o_p(1), \quad i = 1, 2,$$

where $r_i = \|x_i\|$ and $u_i = \|x_i\|^{-1}x_i$, $i = 1, ..., n$. Functions γ_i, α_i and β_i then give the influence functions for T_i and S_i, respectively.

Kankainen et al. (2007) proposed the use of the skewness and kurtosis statistics

$$U = (\hat{T}_2 - \hat{T}_1)^T \hat{S}_1^{-1}(\hat{T}_2 - \hat{T}_1) \quad \text{and} \quad W = ||\hat{S}_1^{-1}\hat{S}_2 - I_p||^2$$

for testing multivariate normality. It is then straightforward to see that

$$U = \|\hat{\delta}\|^2 \quad \text{and} \quad W = \|\hat{\Lambda} - I_p\|^2 = tr\left((\hat{\Lambda} - I_p)^2\right).$$

Kankainen et al. (2007) proved that

Theorem 5.1. *In the multivariate normal model,*

(i) the limiting distribution of nU is that of $\eta_1 U_1$, where $U_1 \sim \chi^2_p$ and $\eta_1 = (1/p)E[(\gamma_1(r) - \gamma_2(r))^2]$ with $r^2 \sim \chi^2_k$;

(ii) the limiting distribution of nW is that of

$$\eta_2\, W_1 \;+\; \eta_3\, W_2,$$

where $W_1 \sim \chi^2_{p(p+1)/2-1}$ and $W_2 \sim \chi^2_1$ are independent,

$$\eta_2 = \frac{2}{p(p+2)}E[(\alpha_2(r) - \alpha_1(r))^2]$$

and

$$\eta_3 = \frac{1}{p}E[(\alpha_2(r) - \alpha_1(r))^2] - 2E[(\alpha_2(r) - \alpha_1(r))(\beta_2(r) - \beta_1(r))] + pE[(\beta_2(r) - \beta_1(r))^2].$$

The expected values are calculated for $r^2 \sim \chi^2_p$.

The statistics U and W_1 can be used to test for ellipticity as well (but with different limiting distributions). Mardia (1970) advocated using his skewness and kurtosis statistics to test for multivariate normality. Under the null hypothesis of multivariate normality

$$\frac{nb_1}{6} \quad \text{and} \quad \frac{n(b_2 - p(p+1))^2}{8p(p+2)}$$

have limiting chi-square distributions with $p(p+1)(p+2)/6$ and 1 degrees of freedom, correspondingly. Kankainen et al. (2007) obtained the limiting Pitman efficiencies of U and W with respect to Mardia's statistics for contiguous sequences of contaminated normal distributions.

Nordhausen et al. (2010) discuss the general idea of using of $\hat{\delta}$ and $\hat{\Lambda}$ in the selection of an appropriate model for the data. Our results now provide the basic elements to convert their ideas into formal inference tools.

5.3. Invariant coordinate selection

Tyler et al. (2009) introduced a general method for exploring multivariate data called the *invariant coordinate selection*. In their approach, they used two shape (not scatter) matrices to find invariant coordinate system; the resulting coordinate system is invariant up to coordinatewise location, sign, and scale. Here, by associating two scatter statistics together with two location statistics, we also fix the location, sign, and the scale, and obtain a fully invariant coordinate system. The invariant coordinate system is useful in many ways. Plotting the observations in the new coordinate system

$$\hat{Z} = (X - 1_n \hat{T}_1^T)\hat{\Gamma}^T,$$

helps in finding outlying observations and clusters in the data. In the case of mixtures of elliptical distributions, a subset of invariant coordinates corresponds to Fisher's linear discriminant subspace (Tyler et al. (2009)). Invariant coordinate selection may thus be seen as a tool for dimension reduction as well. Note that $\hat{Z}$ is a maximal invariant statistic under the group of affine transformations.

5.4. Independent component model

In the semiparametric independent component model, matrix $\hat{\Gamma}$ is a solution to the ICA problem. If T_1, S_1, T_2 and S_2 are the moment-based estimates, then $\hat{\Gamma}$ is the well-known FOBI estimate. Our approach thus gives

a whole family of unmixing matrix estimates for the ICA problem. Furthermore, the limiting properties of the estimates can be considered and compared in our approach. Nordhausen et al. (2009) and Oja et al. (2010) found optimal signed-rank tests for location and independence in the independent component model.

Acknowledgements

The authors wish to thank the editor and the anonymous referee for their insightful comments and suggestions that helped to improve the paper.

This research was supported by the Academy of Finland.

References

Bera, A., John, S., 1983. Tests for multivariate normality with Pearson alternatives. Communications in Statistics - Theory and Methods 12, 103–117.

Cardoso, J.F., 1989. Source separation using higher moments, in: Proceedings of IEEE international conference on acoustics, speech and signal processing, pp. 2109–2112.

Chakraborty, B., Chaudhuri, P., 1999. On affine invariant sign and rank tests in one sample and two sample multivariate problems, in: Ghosh, S. (Ed.), Multivariate, Design and Sample Survey, Marcel-Dekker, New York. pp. 499–522.

Kankainen, A., Taskinen, S., Oja, H., 2007. Tests of multinormality based on location vectors and scatter matrices. Statistical Methods & Applications 16, 357–379.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

Nordhausen, K., Oja, H., Ollila, E., 2010. Multivariate models and the first four moments, in: Hunter, D., Rosenberger, J., Richards, D. (Eds.), Festschrift for Thomas P. Hettmansperger. volume in press.

Nordhausen, K., Oja, H., Paindaveine, D., 2009. Signed-rank tests for location in the symmetric independent component model. Journal of Multivariate Analysis 100, 821–834.

Oja, H., Paindaveine, D., Taskinen, S., 2010. Parametric and nonparametric tests for multivariate independence in the independent component model. Submitted .

Oja, H., Sirkiä, S., Eriksson, J., 2006. Scatter matrices and independent component analysis. Austrian Journal of Statistics 35, 175–189.

Serfling, R.J., 2010. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. Journal of Nonparametric Statistics in press.

Tyler, D.E., Critchley, F., Dümbgen, L., Oja, H., 2009. Invariant co-ordinate selection. Journal of the Royal Statistical Society, Series B 71, 549–592.