Elliptical symmetry

Abstract: This article first reviews the definition of elliptically symmetric distributions and discusses identifiability issues. It then presents results related to the corresponding characteristic functions, moments, marginal and conditional distributions, and considers the absolutely continuous case. Some well known instances of elliptical distributions are provided. Finally, inference in elliptical families is briefly discussed.

Keywords: Elliptical distributions; Mahalanobis distances; Multinormal distributions; Pseudo-Gaussian tests; Robustness; Shape matrices; Scatter matrices

Definition

Until the 1970s, most procedures in multivariate analysis were developed under multinormality assumptions, mainly for mathematical convenience. In most applications, however, multinormality is only a poor approximation of reality. In particular, multinormal distributions do not allow for heavy tails, that are so common, e.g., in financial data. The class of elliptically symmetric distributions extends the class of multinormal distributions by allowing for both lighter-than-normal and heavier-than-normal tails, while maintaining the elliptical geometry of the underlying multinormal equidensity contours. Roughly, a random vector $\mathbf{X}$ with elliptical density is obtained as the linear transformation of a *spherically* distributed one $\mathbf{Z}$—namely, a random vector with spherical equidensity contours, the distribution of which is invariant under rotations centered at the origin; such vectors always can be represented under the form $\mathbf{Z} = R\mathbf{U}$, where R is a positive real variable, and $\mathbf{U}$, independent of R, is uniform over the unit sphere.

More formally, elliptically distributed random vectors, as introduced by [1] and further studied by [2, 3], do not necessarily possess a density, and are defined as follows.

Definition 1. The random d-vector $\mathbf{X}$ is elliptically distributed if and only if there exists a d-vector $\boldsymbol{\mu}$, a $d \times r$ matrix $\mathbf{A}$ with maximal rank r, and a nonnegative random variable R, such that (denoting by $\stackrel{\mathcal{D}}{=}$ equality in distribution)

$$\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{U},$$

where the random r-vector $\mathbf{U}$ is independent of R and is uniformly distributed on the unit sphere $\mathcal{S}^{r-1} \subset \mathbb{R}^r$. One may then write $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$, where $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$ and F_R is the cumulative distribution function of the *generating variate* R. The distribution $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$ is said to be *elliptically symmetric*—or simply *elliptical*.

Several comments are in order. First, if $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$, then $\mathbf{X} - \boldsymbol{\mu}$ takes its values in the column space $\mathcal{M}(\mathbf{A})$ of $\mathbf{A}$, an r-dimensional subspace of $\mathbb{R}^d$. The support of the distribution of $\mathbf{X} - \boldsymbol{\mu}$ will be $\mathcal{M}(\mathbf{A})$ itself if and only if the support of the distribution of R is the whole positive real half line $\mathbb{R}^+$ (the latter support, however, may be a bounded interval or an at most countable subset of $\mathbb{R}^+$). Second, strictly speaking, $\mathbf{A}$ (equivalently, $\boldsymbol{\Sigma}$) is only defined up to a positive scalar factor, as replacing the couple $(R, \mathbf{A})$ with $(\lambda^{-1}R, \lambda\mathbf{A})$, for any $\lambda > 0$, does not affect the distribution of $\mathbf{X}$. This may be solved by imposing that the covariance matrix of $\mathbf{X}$ is equal to $\boldsymbol{\Sigma}$ or by imposing that R has median one (the latter solution has the advantage of avoiding finite moment assumptions; yet, under finite second-order moments, the former may be more convenient).

2 Elliptical symmetry

The *uniqueness theorem* (see, e.g., pages 346 and 351 in [4]) implies that elliptically distributed random vectors alternatively may be defined in terms of characteristic functions.

Definition 2. The random d-vector $\mathbf{X}$ is elliptically distributed if and only if there exists a d-vector $\boldsymbol{\mu}$, a $d \times d$ symmetric and positive semi-definite matrix $\boldsymbol{\Sigma}$, and a function $\varphi : \mathbb{R}^+ \to \mathbb{R}$, such that the characteristic function of $\mathbf{X}$ is of the form

$$\mathbf{t} \mapsto \phi_{\mathbf{X}}(\mathbf{t}) = \exp(i\mathbf{t}'\boldsymbol{\mu})\varphi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}), \qquad \mathbf{t} \in \mathbb{R}^d.$$

One may then write $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$. The function φ is called *the characteristic generator* of $\mathbf{X}$.

This latter definition is the one adopted in [1, 2], where Definition 1 then has the status of a *stochastic representation result*. The uniqueness theorem further implies that there is a one-to-one relation between F_R and φ; see [3] for an explicit formula. Note that Definition 2 justifies that the notation $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$ may be used instead of $\mathcal{E}_d(\boldsymbol{\mu}, \mathbf{A}, F_R)$, since it shows that this distribution does depend on $\mathbf{A}$ only through $\boldsymbol{\Sigma}$.

Moments

If $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R) = \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$, then it is clear that $\mathbf{X}$ admits finite moments of order $s > 0$ if and only if $\mathrm{E}[R^s] < \infty$. Assuming that the corresponding moments are finite, one has

$$\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu} \quad \text{and} \quad \mathrm{Var}[\mathbf{X}] = \frac{1}{r}\mathrm{E}[R^2]\boldsymbol{\Sigma} = -2\varphi'(0)\boldsymbol{\Sigma} \tag{1}$$

(under finite second-order moments, the normalization $\mathrm{Var}[\mathbf{X}] = \boldsymbol{\Sigma}$ is thus strictly equivalent to $\mathrm{E}[R^2] = r$).

To consider higher-order moments, let $\sigma_{i_1 i_2 \ldots i_\ell} := \mathrm{E}[(\mathbf{X} - \boldsymbol{\mu})_{i_1}(\mathbf{X} - \boldsymbol{\mu})_{i_2} \ldots (\mathbf{X} - \boldsymbol{\mu})_{i_\ell}]$. By symmetry, all third-order moments σ_{ijk} are zero, and this extends to all odd moments (still provided that the corresponding moments do exist). As for fourth-order moments, they *all* (in the full-rank case $r = d$) satisfy the identity

$$\sigma_{ijk\ell} = (\kappa + 1)(\sigma_{ij}\sigma_{k\ell} + \sigma_{ik}\sigma_{j\ell} + \sigma_{i\ell}\sigma_{jk}),$$

where

$$\kappa := \frac{1}{3}\Big(\frac{\sigma_{iiii}}{\sigma_{ii}^2} - 3\Big) \quad \Big(= \frac{d}{d+2}\frac{\mathrm{E}[R^4]}{(\mathrm{E}[R^2])^2} - 1\Big) \tag{2}$$

is the *kurtosis parameter* of the elliptical distribution at hand; see [5, 6]. It is remarkable that κ indeed does not depend on i, nor on the scatter matrix $\boldsymbol{\Sigma}$. Note that κ does not reduce to the usual kurtosis coefficient used in the univariate case. It is shown in [7] that $\kappa > -2/(d+2)$.

The absolutely continuous case

As mentioned above, $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$ almost surely takes its values in the column space of the $d \times r$ matrix $\mathbf{A}$, translated at $\boldsymbol{\mu}$. Therefore a necessary condition for $\mathbf{X}$ to be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ is $r = d$—in which case both $\mathbf{A}$ and $\boldsymbol{\Sigma}$ are invertible matrices. Under the condition that $r = d$, it can be shown that absolute continuity of $\mathbf{X}$ holds if and only if R is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_0^+$.

In the absolutely continuous case, the density of $\mathbf{X}$ is of the form

$$\mathbf{x} \mapsto f_{\mathbf{X}}(\mathbf{x}) = c_{d,f}\,(\det \boldsymbol{\Sigma})^{-1/2} f\Big(\sqrt{(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})}\Big),$$

where $f : \mathbb{R}_0^+ \to \mathbb{R}^+$ is the so-called *radial density* and $c_{d,f} > 0$ is a normalization factor (ensuring that $f_{\mathbf{X}}$ integrates to one). Putting $\mu_{\ell,f} := \int_0^\infty r^\ell f(r)\,dr$, the density of R is then $r \mapsto f_R(r) := \mu_{d-1,f}^{-1} r^{d-1} f(r) I_{[r>0]}$, where I_B stands for the indicator function of B. Integrability of $f_{\mathbf{X}}$ is equivalent to $\mu_{d-1,f} < \infty$, while $\mathbf{X}$ admits finite moments of order $s > 0$ if and only if $\mu_{d+s-1,f} < \infty$. Under finite second-order moments, $\boldsymbol{\Sigma}$ and f can be properly identified by imposing that $\mathrm{Var}[\mathbf{X}] = \boldsymbol{\Sigma}$, which, in view of (1) and the density of R just given, is equivalent to $\mu_{d+1,f} = d\mu_{d-1,f}$. As already mentioned, rather requiring that R has median one, which rewrites

$$\frac{1}{\mu_{d-1,f}} \int_0^1 r^{d-1} f(r)\, dr = \frac{1}{2}, \tag{3}$$

would achieve identifiability without imposing any moment condition. Unless otherwise stated, radial densities below are standardized through (3).

The quantity $d_{\boldsymbol{\Sigma}}(\mathbf{x}, \boldsymbol{\mu}) := \sqrt{(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$ is usually called the *Mahalanobis distance* between $\mathbf{x}$ and $\boldsymbol{\mu}$ in the metric associated with $\boldsymbol{\Sigma}$. If $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$, then $d_{\boldsymbol{\Sigma}}(\mathbf{X}, \boldsymbol{\mu})$ is equal to R in distribution, hence admits the density f_R. The level sets of the elliptical density $f_{\mathbf{X}}$ are those of the mapping $\mathbf{x} \mapsto d_{\boldsymbol{\Sigma}}(\mathbf{x}, \boldsymbol{\mu})$, hence are a family of hyper-ellipsoids (that are centered at $\boldsymbol{\mu}$ and whose shape and orientation are determined by $\boldsymbol{\Sigma}$). This is of course what justifies the terminology *elliptical distributions*.

Well-known instances of elliptical densities are the d-variate multinormal distributions, with radial density

$$f(r) = \exp(-a_d r^2/2),$$

and the d-variate Student distributions, with radial densities

$$f_\nu(r) = (1 + a_{d,\nu} r^2/\nu)^{-(d+\nu)/2}, \qquad \nu > 0.$$

The latter present heavier-than-normal tails : the smaller the *degree of freedom* ν, the heavier the tails, and the multinormal case is obtained as $\nu \to \infty$. A parametric family of elliptical densities presenting both heavier-than-normal and lighter-than-normal tails is the class of d-variate power-exponential distributions, with radial densities of the form

$$f_\xi(r) = \exp(-b_{d,\xi} r^{2\xi}), \qquad \xi > 0.$$

Finally, the d-variate "ε-contaminated" multinormal distributions, that play an important role in robust statistics, are obtained with

$$f_\varepsilon(r) = (1-\varepsilon)\exp(-c_{d,\varepsilon} r^2/2) + \varepsilon\sigma^{-d}\exp(-c_{d,\varepsilon} r^2/2\sigma^2), \qquad \varepsilon \in (0,1),\ \sigma > 0;$$

see, e.g., [8]. Above, the positive constants a_d, $a_{d,\nu}$, $b_{d,\xi}$, and $c_{d,\varepsilon}$ are such that (3) is satisfied.

Marginal and conditional distributions

Consider a random d-vector $\mathbf{X}$ with distribution $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R) = \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$. Then, $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{U}$, where R, $\mathbf{A}$, and $\mathbf{U}$ are as in Definition 1. It directly follows that, for any $k \times d$ matrix $\mathbf{L}$ and any k-vector $\mathbf{b}$,

$$\mathbf{L}\mathbf{X} + \mathbf{b} \stackrel{\mathcal{D}}{=} (\mathbf{L}\boldsymbol{\mu} + \mathbf{b}) + R(\mathbf{L}\mathbf{A})\mathbf{U} \sim \mathcal{E}_k(\mathbf{L}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}', F_R) = \mathcal{E}_k(\mathbf{L}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}', \varphi).$$

The class of elliptical distributions is therefore closed under affine transformations. Note that the induced transformations of the parameters $\boldsymbol{\mu} \mapsto \mathbf{L}\boldsymbol{\mu} + \mathbf{b}$ and $\boldsymbol{\Sigma} \mapsto \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}'$ are the natural transformations for multivariate location and scatter parameters, respectively, and that the generating variate and characteristic generator are invariant under affine transformations. Of course, it directly follows that (possibly multivariate) marginals of an elliptically distributed random vector also are elliptically distributed.

Theorem 1. Let $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')' \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$, where $\mathbf{X}_1$ is a random d_1-vector and $\mathbf{X}_2$ is a random d_2-vector ($d = d_1 + d_2$). Partition the arrays $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly, with blocks $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$, and $\boldsymbol{\Sigma}_{22}$, of size $d_1 \times 1$, $d_2 \times 1$, $d_1 \times d_1$, $d_1 \times d_2$, and $d_2 \times d_2$, respectively. Then $\mathbf{X}_1 \sim \mathcal{E}_{d_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, F_R)$.

Conditional distributions are more difficult to describe, and, for the sake of simplicity, only the full-rank case $r = d$ is treated here; see [2] for more general statements. As shown by the following result, conditional distributions are still elliptical, with a conditional mean function that does not depend on the parent (joint) elliptical distribution, hence simply coincides with the linear conditional mean function that is well known in the multinormal case.

Theorem 2. Consider the same setup as in Theorem 1 and further assume that $r = \mathrm{rank}(\boldsymbol{\Sigma}) = d$. Then $\mathbf{X}_1 | [\mathbf{X}_2 = \mathbf{x}_2] \sim \mathcal{E}_{d_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}, F_{R_{1|2}})$, where

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \qquad \boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21},$$

and $F_{R_{1|2}}$ is the cdf of $(R^2 - d^2_{\boldsymbol{\Sigma}_{22}}(\mathbf{x}_2, \boldsymbol{\mu}_2))^{1/2} \,|\, [\mathbf{X}_2 = \mathbf{x}_2]$ (here, R is the generating variate of $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$).

The conditional scatter matrix $\boldsymbol{\Sigma}_{1|2}$ does not depend on the fixed value $\mathbf{x}_2$ of $\mathbf{X}_2$. However, whenever it exists, the conditional covariance matrix, that—in view of (1)—is given by

$$\mathrm{Var}[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] = \frac{1}{d_1}\big(\mathrm{E}[R^2 | \mathbf{X}_2 = \mathbf{x}_2] - d^2_{\boldsymbol{\Sigma}_{22}}(\mathbf{x}_2, \boldsymbol{\mu}_2)\big)\boldsymbol{\Sigma}_{1|2},$$

in general is a non-trivial function of $\mathbf{x}_2$. Actually, the only case where this heteroskedasticity is never present is the multinormal case; see Theorem 7 in [1].

Inference in elliptical families

Elliptical families have been introduced in multivariate analysis as a reaction against pervasive Gaussian assumptions. Most classical (Gaussian) procedures in that field—one-sample location, multivariate regression, MANOVA/MANOCOVA, principal components, discriminant analysis, canonical correlations, etc.—can be turned into elliptical *pseudo-Gaussian* procedures, that is, into procedures that apply under a broad class of elliptical distributions while remaining asymptotically equivalent to the corresponding original Gaussian procedures under multinormality.

For location problems (one-sample location, multivariate regression, MANOVA/MANOCOVA, etc.), there is actually no need to define such pseudo-Gaussian procedures. In a one-sample context, traditional Gaussian methods, based on the asymptotic normality of the sample mean $\bar{\mathbf{X}} := \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$, indeed typically remain valid under any elliptical distribution with finite second-order moments—and actually under any distribution meeting the same moment restrictions. However, the situation is much different for inference problems related to scatter matrices, such as principal components, discriminant analysis, canonical correlations, testing for sphericity, testing for the equality of the scatter matrices associated with m independent populations, etc. For such problems, Gaussian procedures are notoriously sensitive to violations of Gaussian assumptions, and extending the normal-theory standard procedures—Gaussian MLEs and likelihood ratio (or Wald) tests—to the elliptical case may be much more difficult; see [9] for a classical discussion of this fact, or [10] for a more recent overview.

This is illustrated in the generic problem where, on the basis of n mutually independent observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with common distribution $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, F_R)$, one wants to test the null hypothesis $\mathcal{H}_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, for some fixed symmetric positive definite matrix $\boldsymbol{\Sigma}_0$ (for the sake of simplicity, scatter matrices $\boldsymbol{\Sigma}$ are assumed invertible here, and identifiability issues are solved by imposing that $\mathrm{Var}[\mathbf{X}_1] = \boldsymbol{\Sigma}$). A Wald test can be based on the asymptotic distribution of the regular sample covariance matrix $\mathbf{S} :=$

$\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i-\bar{\mathbf{X}})(\mathbf{X}_i-\bar{\mathbf{X}})'$, which, under finite fourth-order moments, reads

$$\sqrt{n}\,\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}\big(\mathbf{0},\boldsymbol{\Gamma}_{\mathrm{scatter}}(\boldsymbol{\Sigma})\big), \quad \text{with } \boldsymbol{\Gamma}_{\mathrm{scatter}}(\boldsymbol{\Sigma}) := (1+\kappa)(\mathbf{I}_d+\mathbf{K}_d)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \kappa\,(\mathrm{vec}\,\boldsymbol{\Sigma})(\mathrm{vec}\,\boldsymbol{\Sigma})',$$

where vec and $\otimes$ stand for the usual vectorization operator and Kronecker product, respectively, κ is the kurtosis parameter of the underlying elliptical distribution (see (2)), $\mathbf{I}_d$ is the d-dimensional identity matrix, and $\mathbf{K}_d := \sum_{i,j=1}^{d}(\mathbf{e}_i\mathbf{e}_j')\otimes(\mathbf{e}_j\mathbf{e}_i')$ is the so-called $d^2\times d^2$ *commutation* matrix (here, $\mathbf{e}_\ell$ is the ℓth vector in the canonical basis of $\mathbb{R}^d$); see, e.g., [11, 12]. The resulting Wald test for $\mathcal{H}_0 : \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$ typically rejects the null for large values of $Q_{\mathrm{scatter}} := n\,(\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma}_0))'\,(\boldsymbol{\Gamma}_{\mathrm{scatter}}(\boldsymbol{\Sigma}))^{-1}\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma}_0)$, a statistic whose asymptotic null distribution unfortunately is $(1+\kappa+d\kappa/2)\chi^2_1+(1+\kappa)\chi^2_{d(d+1)/2-1}$, and there is no easy way to modify this test statistic to let it agree—under any elliptical distribution with finite fourth-order moments—with the asymptotic null $\chi^2_{d(d+1)/2}$ distribution obtained in the multinormal case; see, e.g., [9]. This makes it difficult to extend the Gaussian Wald test—equivalently, the Gaussian LRT test, since both tests are actually asymptotically equivalent under the null—to the elliptical case. While the testing problem considered here may seem of academic interest only, the same difficulty holds for the problem of testing equality of several covariance matrices, which is of very high practical relevance. A classical way out of this difficulty consists in bootstrapping the Wald or LRT tests, but a preferable solution is the one developed in [13] (see also [14]) that modifies the test statistic Q_{scatter} in such a way that its asymptotic null distribution agrees, irrespective of the underlying elliptical distribution with finite fourth-order moments, with the one obtained in the multinormal case.

Quite fortunately, not all problems related to scatter or covariance matrices are as "hard" as the ones above. It actually turns out that the numerous problems in multivariate analysis where the parameter of interest is not the scatter matrix $\boldsymbol{\Sigma}$ but rather a properly normalized version of the scatter matrix are "easier"*. In principal components, for instance, it is sufficient to know $\boldsymbol{\Sigma}$ up to a positive scalar factor to compute the principal directions or to decide how many principal directions are needed to explain some fixed proportion of the total variance. The parameter of interest is then the so-called *shape matrix* $\mathbf{V}$, that is defined as the only matrix proportional to $\boldsymbol{\Sigma}$ having upper-left entry equal to one, or alternatively having trace equal to d, or alternatively having determinant equal to one†. If one adopts the normalization $\mathbf{V} := \boldsymbol{\Sigma}/(\det\boldsymbol{\Sigma})^{1/d}$, then the MLE for $\mathbf{V}$, namely $\hat{\mathbf{V}} := \mathbf{S}/(\det\mathbf{S})^{1/d}$, satisfies (still under any elliptical distribution with finite fourth-order moments)

$$\sqrt{n}\,\mathrm{vec}\big(\hat{\mathbf{V}}-\mathbf{V}\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(\mathbf{0},\boldsymbol{\Gamma}_{\mathrm{shape}}(\mathbf{V})\big), \quad \text{with } \boldsymbol{\Gamma}_{\mathrm{shape}}(\mathbf{V}) := (1+\kappa)\Big((\mathbf{I}_d+\mathbf{K}_d)(\mathbf{V}\otimes\mathbf{V}) + \frac{2}{d}\,(\mathrm{vec}\,\mathbf{V})(\mathrm{vec}\,\mathbf{V})'\Big),$$

and testing for $\mathcal{H}_0 : \mathbf{V}=\mathbf{V}_0$ can simply be based on the asymptotic null $(1+\kappa)\chi^2_{d(d+1)/2-1}$ distribution of $Q_{\mathrm{shape}} := n\,(\mathrm{vec}(\hat{\mathbf{V}}-\mathbf{V}_0))'\,(\boldsymbol{\Gamma}_{\mathrm{shape}}(\mathbf{V}))^{-1}\mathrm{vec}(\hat{\mathbf{V}}-\mathbf{V}_0)$ (in practice, the test is based on the asymptotic null $\chi^2_{d(d+1)/2-1}$ distribution of $Q_{\mathrm{shape}}/(1+\hat{\kappa})$, where $\hat{\kappa}$ is an arbitrary consistent estimate of κ). This allows to test sphericity of the underlying distribution (by choosing $\mathbf{V}_0=\mathbf{I}_d$), but also similarly applies to principal component analysis, canonical correlations, and to the problem of testing for multivariate non-correlation, among others; see, e.g., [9, 12, 16] and [17].

Being based on sample means and sample covariance matrices, the pseudo-Gaussian tests mentioned above unfortunately remain poorly robust. They still are very sensitive to the presence of outliers and they do require finite moments of order two (resp., four) for location problems (resp., for problems on scatter or shape matrices). Robust procedures resisting arbitrarily heavy tails, however, may be based on multivariate *signed*

---

* The terminology "easy" and "hard" goes back to [9].

†Among these three possible definitions of shape, the determinant-based one was shown canonical in [15], in the sense that it is the only one for which the shape parameter $\mathbf{V}$ and the scale parameter σ^2 arising from the decomposition of scatter into $\boldsymbol{\Sigma}=\sigma^2\mathbf{V}$ are orthogonal in the sense of Fisher information.

ranks, that allow to build (asymptotically) distribution-free and invariant procedures not only for all inference problems mentioned above, but also for some time series problems; see [18–22] and [23].

In all inference problems considered above, the parameter of interest is either the location centre $\boldsymbol{\mu}$, the scatter matrix $\mathbf{\Sigma}$, or the shape matrix $\mathbf{V}$, and the radial density f therefore always plays the role of an unspecified nuisance. Sometimes it is also of interest to estimate the radial density. This may be done in a parametric way (by assuming, e.g., that the underlying elliptical distribution is a Student distribution and by estimating the corresponding degree of freedom via maximum likelihood) or in a nonparametric way (by adopting, e.g., a kernel-based estimator of the radial density). Unlike in traditional density estimation, the latter approach is not affected by the curse of dimensionality, since, irrespective of the dimension d, only a univariate density (namely, that of the generating variate R) has to be estimated; see, e.g., [24].

References

[1] Kelker D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization, *Sankhya A* 32, 419–430.

[2] Cambanis S., Huang S. & Simons G. (1981). On the theory of elliptically contoured distributions, *Journal of Multivariate Analysis* 11, 368–385.

[3] Fang K.T., Kotz S. & Ng K.W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman & Hall.

[4] Shorack G.R. (2000). *Probability for Statisticians*, Springer, New York.

[5] Anderson T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edition, Wiley-Interscience, New York.

[6] Bentler P.M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures, *Psychometrika* 48, 493–517.

[7] Bentler P.M. & Berkane M. (1986). Greatest lower bound to the elliptical theory kurtosis parameter, *Biometrika* 73, 240–241.

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

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

[10] Yanagihara H., Tonda T. & Matsumoto C. (2005). The effects of nonnormality on asymptotic distributions of some likelihood ratio criteria for testing covariance structures under normal assumption, *Journal of Multivariate Analysis* 96, 237–264.

[11] Tyler D.E. (1982). Radial estimates and the test for sphericity, *Biometrika* 69, 429–436.

[12] Tyler D.E. (1983). Robustness and efficiency properties of scatter matrices, *Biometrika* 70, 411–420.

[13] Hallin M. & Paindaveine D. (2009). Optimal tests for homogeneity of covariance, scale, and shape, *Journal of Multivariate Analysis* 100, 422–444.

[14] Hallin M. (2010). On the non Gaussian asymptotics of the likelihood ratio test statistic for homogeneity of covariance. In D. Hunter, D. Richards, and J.L. Rosenberger, Eds, *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger*, World Scientific, Singapore, 162–173.

[15] Paindaveine D. (2008). A canonical definition of shape, *Statistics and Probability Letters* 78, 2240–2247.

[16] Davis A.W. (1977). Asymptotic theory for principal component analysis: the non-normal case, *Austrian Journal of Statistics* 19, 206–212.

[17] Tyler D.E. (1981). Asymptotic inference for eigenvectors, *Annals of Statistics* 9, 725–736.

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

[19] Hallin M. & Paindaveine D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model, *Annals of Statistics* 32, 2642–2678.

[20] Hallin M. & Paindaveine D. (2005). Affine invariant aligned rank tests for the multivariate general linear model with ARMA errors, *Journal of Multivariate Analysis* 93, 122–163.

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

[22] Hallin M., Oja H., & Paindaveine D. (2006). Semiparametrically efficient rank-based inference for shape II: Optimal R-estimation of shape, *Annals of Statistics* 34, 2757–2789.

[23] Hallin M. & Paindaveine D. (2008). Optimal rank-based tests for homogeneity of scatter, *Annals of Statistics* 36, 1261–1298.

[24] Liebscher E. (2005). A semiparametric density estimator based on elliptical distributions, *Journal of Multivariate Analysis* 92, 205–225.

Davy Paindaveine

Université Libre de Bruxelles, E.C.A.R.E.S. and Département de Mathématique,
Av. F.D. Roosevelt 50, CP114/04, B-1050 Brussels, Belgium