T E C H N I C A L
R E P O R T

0447

A CHERNOFF-SAVAGE RESULT FOR SHAPE ON THE NON-ADMISSIBILITY OF PSEUDO-GAUSSIAN METHODS

PAINDAVEINE, D.

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

A Chernoff-Savage Result for Shape
On the Non-admissibility of Pseudo-Gaussian Methods

Davy Paindaveine

Département de Mathématique, I.S.R.O., and E.C.A.R.E.S, Université Libre de Bruxelles, Campus de la Plaine CP 210, 1050 Bruxelles, Belgium. E-mail: dpaindav@ulb.ac.be

Chernoff and Savage (1958) established that, in the context of univariate location models, Gaussian-score rank-based procedures uniformly dominate—in terms of Pitman asymptotic relative efficiencies—their pseudo-Gaussian parametric counterparts. This result, which had quite an impact on the success and subsequent development of rank-based inference, has been extended to many *location* problems, including problems involving multivariate and/or dependent observations. In this paper, we show that this uniform dominance also holds in problems for which the parameter of interest is the *shape* of an elliptical distribution. The Pitman non-admissibility of the Gaussian pseudo-maximum likelihood estimator for shape and that of the pseudo-Gaussian likehood ratio test of sphericity follow.

Keywords: Chernoff-Savage result; elliptical density; Pitman non-admissibility; semiparametric efficiency; shape matrix; test for sphericity
Running title: A Chernoff-Savage result for shape

1 Introduction

Let $\mathcal{P}^n(\boldsymbol{\Theta},\mathcal{F}) := \{\mathrm{P}^n_{\boldsymbol{\vartheta},f}, \boldsymbol{\vartheta} \in \boldsymbol{\Theta}, f \in \mathcal{F}\}$ be a sequence of semiparametric models, where $\boldsymbol{\vartheta}$ is some finite-dimensional parameter and f is some unspecified functional nuisance. The asymptotic efficiencies of competing inference procedures—on $\boldsymbol{\vartheta}$, a subset of $\boldsymbol{\vartheta}$, or more generally a function $\psi(\boldsymbol{\vartheta})$ of the parameter $\boldsymbol{\vartheta}$—that remain valid under a broad range of distributions in $\mathcal{P}^n(\boldsymbol{\Theta},\mathcal{F})$, are usually compared in terms of *Pitman* asymptotic relative efficiencies (AREs). Roughly speaking, the Pitman asymptotic relative efficiency $\mathrm{ARE}_{\boldsymbol{\vartheta},f}[T_2/T_1]$ of a procedure T_2 with respect to a procedure T_1 at $\mathrm{P}^n_{\boldsymbol{\vartheta},f}$ is defined as the limit of the ratio n_1/n_2 of observations required for T_1 and T_2 to achieve the same asymptotic performance at $\mathrm{P}^n_{\boldsymbol{\vartheta},f}$. In the particular case for which T_1 and T_2 are estimators of (the univariate quantity) $\psi(\boldsymbol{\vartheta})$ such that $\sqrt{n}(T_i-\psi(\boldsymbol{\vartheta}))$ is asymptotically normal, under $\mathrm{P}^n_{\boldsymbol{\vartheta},f}$, with mean zero and variance $v_i(\boldsymbol{\vartheta},f)$, $i=1,2$, the ARE of T_2 with respect to T_1, under $\mathrm{P}^n_{\boldsymbol{\vartheta},f}$, is given by

$$\mathrm{ARE}_{\boldsymbol{\vartheta},f}\Big[T_2\,/\,T_1\Big] = v_1(\boldsymbol{\vartheta},f)/v_2(\boldsymbol{\vartheta},f); \qquad (1.1)$$

see, e.g., Lehmann (1999). For a precise definition of the concept of Pitman ARE in the case of testing procedures, see, e.g., Lehmann (1986), Pratt and Gibbons (1981), or Nikitin (1995).

As the ARE value in (1.1) in general depends on f, no total ordering can be based on this concept of ARE. However, uniform domination may happen, in which case we adopt the following definition. Assume that the procedure T_1 is valid (by which we mean, for point estimation, that T_1 is consistent for $\psi(\boldsymbol{\vartheta})$ at the optimal rate, and, for hypothesis testing, that it has asymptotically the right nominal level α) for all $f \in \mathcal{F}_1 \subset \mathcal{F}$. We say that T_1 is *Pitman non-admissible* iff there exists some procedure T_2, valid over $\mathcal{F}_2 \supset \mathcal{F}_1$, such that

$$\mathrm{ARE}_{\boldsymbol{\vartheta},f}\Big[T_2 \,/\, T_1\Big] \geq 1 \quad \text{for all } f \in \mathcal{F}_1, \tag{1.2}$$

where the inequality is strict for at least one $f \in \mathcal{F}_1$. If (1.2) holds, we say in the sequel, for the sake of simplicity, that "T_2 beats T_1", instead of "T_2 uniformly dominates T_1 in the Pitman sense." Similarly, "T_2 strictly beats T_1 but at $\mathcal{F}_0$" means that (1.2) holds and that the equality is achieved iff $f \in \mathcal{F}_0$. Clearly, as far as semiparametric validity and asymptotic efficiency are concerned, Pitman non-admissible procedures should be avoided.

Now, assume that the parametric normal model $\mathcal{P}^n(\boldsymbol{\Theta}, \{\phi\})$, say, is contained in $\mathcal{P}^n(\boldsymbol{\Theta}, \mathcal{F})$. Then a classical approach to build inference procedures on $\psi(\boldsymbol{\vartheta})$ is to restrict to the Gaussian model $\mathcal{P}^n(\boldsymbol{\Theta}, \{\phi\})$ and invoke some method—such as, e.g., the likelihood ratio test—among the large panel of methods available for developing parametric statistical procedures that are asymptotically optimal—in some sense—within $\mathcal{P}^n(\boldsymbol{\Theta}, \{\phi\})$. Although they are of a parametric nature, the resulting procedures remain often valid outside the Gaussian model, under $\mathcal{P}^n(\boldsymbol{\Theta}, \mathcal{F}_1)$, for some $\mathcal{F}_1 \subset \mathcal{F}$, say. One then usually speaks of *pseudo-Gaussian procedures*. However, the latter in general achieves asymptotic optimality under normal distributions only.

Another—more semiparametric—approach to obtain procedures that remain valid under a broad range of distributions in $\mathcal{P}^n(\boldsymbol{\Theta}, \mathcal{F})$, consists in relying on some statistical principle, such as the invariance principle. When invariance is to be achieved with respect to a group of order-preserving transformations, this leads, typically, to the class of *rank-based procedures*. The resulting semiparametric procedures usually enjoy many desirable properties, such as broader validity (under $\mathcal{P}^n(\boldsymbol{\Theta}, \mathcal{F}_2) \supset \mathcal{P}^n(\boldsymbol{\Theta}, \mathcal{F}_1)$, say), robustness, distribution-freeness (for hypothesis testing), etc. However, it is often believed that the price to pay for these nice properties is a substantial efficiency loss when compared to the performance of pseudo-Gaussian procedures, at least at—or, in a vicinity of—the normal submodel.

Intuition in this case is misleading, as shown by the celebrated result of Chernoff and Savage (1958), which states that there is no efficiency loss at all, provided that *Gaussian scores* are used. More precisely, they show, in the context of the two-sample location problem, that the Gaussian-score rank test strictly beats the pseudo-Gaussian test—namely, the two-sample t-test—but at Gaussian distributions. The Pitman non-admissibility of the two-sample t-test follows. This celebrated result and its extensions (see below), which clearly indicate that efficiency is another advantage of rank-based methods over pseudo-Gaussian ones, had quite an impact on the success and subsequent development of rank-based inference.

This Chernoff-Savage result has been extended to many problems, including problems involving *serially dependent* and/or *multivariate* observations. Hallin (1994) show that the Gaussian-score version of the *serial* rank tests proposed by Hallin and Puri (1994) also strictly beat the corresponding pseudo-Gaussian tests, but at Gaussian innovations (those serial rank tests allow for testing for randomness against serial dependence, for testing the adequacy of an ARMA model, or for testing linear restrictions on the parameter of an ARMA model). Extensions to (possibly serial) problems involving *multivariate* observations were recently obtained by Hallin and Paindaveine (2002a, b, and 2005a), who show that the Chernoff-Savage result holds in a broad class of multivariate problems (culminating in the problem of testing linear restrictions on the parameter of the multivariate general linear model with vector ARMA errors); the Pitman non-admissibility of the corresponding everyday practice pseudo-Gaussian tests (one-sample and two-sample Hotelling tests, multivariate F-tests, multivariate Portmanteau and Durbin-Watson tests, etc.) follows.

In the review of Chernoff-Savage results above, we have focused on hypothesis testing. However rank-based methods also allow for dealing with point estimation and it can be shown that the AREs of the resulting *R-estimators*, with respect to their pseudo-Gaussian competitors, do coincide with the AREs obtained in the corresponding testing problems. Consequently, the generalized Chernoff-Savage results above also cover the estimation problem in each case, which, e.g., establishes the Pitman non-admissibility of multivariate least-squares and Yule-Walker estimators in the multivariate general linear model and in vector autoregressive models, respectively.

So far, however, Chernoff-Savage results were only established for *location* parameters (autoregressive parameters, even though they are associated with serial models, should be considered as location parameters, in the same fashion as standard regression parameters). This paper shows that the uniform Pitman dominance of Gaussian-score rank-based procedures over their pseudo-Gaussian competitors extends to the case where the parameter of interest is the *shape* of an elliptical population. We thereby establish the Pitman non-admissibility, for any space dimension $k \geq 2$, of the Gaussian pseudo-maximum likelihood estimator for the shape of a k-variate elliptical distribution, as well as that of the pseudo-Gaussian likelihood ratio tests for a specified shape (which includes the classical likelihood ratio test of sphericity as a special case). The proofs of these shape Pitman non-admissibility results however are by no means trivial, since, unlike Chernoff-Savage results for location parameters, Chernoff-Savage results for shape do not follow from standard variational arguments. We therefore propose a proof partially inspired by the "direct" method introduced by Gastwirth and Wolff (1968).

The paper is organized as follows. In Section 2, we describe the problem of estimating the shape of an elliptical distribution and that of testing for a specified shape. We recall the pseudo-Gaussian estimators and tests; we define the corresponding Gaussian-score rank-based procedures, and provide their Pitman AREs with respect to the pseudo-Gaussian estimators and tests. We state our Chernoff-Savage result for shape and its consequences in terms of Pitman admissibility. The proofs are given in Section 3, where we also explain why standard variational methods are inappropriate for the problem under consideration. Finally, Section 4 states some final comments.

2 Shape problems

2.1 Elliptical densities and shape

Let $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be a sample of independent and identically distributed k-variate observations with common elliptical density

$$\mathbf{x} \mapsto c_{k,f} \left(\det \mathbf{V}\right)^{-1/2} f\Big(\sqrt{(\mathbf{x}-\boldsymbol{\theta})' \mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\theta})}\,\Big), \tag{2.3}$$

where the *centre of symmetry* $\boldsymbol{\theta}$ is a k-vector, the *shape parameter* $\mathbf{V}$ is a symmetric positive definite real $k \times k$ matrix with $(\mathbf{V})_{11} = 1$, the *radial density* $f : \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+$ satisfies $\mu_{k-1,f} := \int_0^\infty r^{k-1} f(r)\, dr < \infty$, and $c_{k,f}$ is a normalization factor. We denote the corresponding hypothesis by $\mathrm{P}^n_{\boldsymbol{\theta},\mathbf{V},f}$. Under $\mathrm{P}^n_{\boldsymbol{\theta},\mathbf{V},f}$, the distances $d_i(\boldsymbol{\theta},\mathbf{V}) := \|\mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})\|$ (throughout, $\mathbf{V}^{1/2}$ denotes the symmetric root of $\mathbf{V}$) are i.i.d., with density and distribution function

$$r \mapsto \tilde{f}_k(r) := (\mu_{k-1,f})^{-1} r^{k-1} f(r)\, I_{[r>0]} \quad \text{and} \quad r \mapsto \tilde{F}_k(r) := \int_0^r \tilde{f}_k(s)\, ds,$$

respectively, and the *multivariate signs* $\mathbf{U}_i(\boldsymbol{\theta},\mathbf{V}) := \mathbf{V}^{-1/2}(\mathbf{X}_i - \boldsymbol{\theta})/d_i(\boldsymbol{\theta},\mathbf{V})$ are i.i.d. and uniformly distributed over the unit sphere. In the sequel, we write $d_i(\mathbf{V})$ and $\mathbf{U}_i(\mathbf{V})$ for $d_i(\hat{\boldsymbol{\theta}},\mathbf{V})$

and $\mathbf{U}_i(\hat{\boldsymbol{\theta}}, \mathbf{V})$, respectively, where $\hat{\boldsymbol{\theta}}$ stands for an asymptotically discrete root-n consistent estimator for $\boldsymbol{\theta}$. Finally, we denote by $R_i(\mathbf{V})$ the rank of $d_i(\mathbf{V})$ among $d_1(\mathbf{V}), \ldots, d_n(\mathbf{V})$.

Special cases are the k-variate multinormal distribution, with radial density $f(r) = \phi(r) := \exp(-r^2/2)$, the k-variate Student distributions, with radial densities (for ν degrees of freedom) $f(r) = f^t_\nu(r) := (1+r^2/\nu)^{-(k+\nu)/2}$, and the k-variate power-exponential distributions, with radial densities of the form $f(r) = f^e_\eta(r) := \exp(-r^{2\eta})$, $\eta > 0$. Note that, under the k-variate Gaussian distribution $\mathrm{P}^n_{\boldsymbol{\theta},\mathbf{V},\phi}$, the distances $d_i(\boldsymbol{\theta}, \mathbf{V})$ have common density and distribution function

$$r \mapsto \tilde{\phi}_k(r) := \Big(2^{(k-2)/2}\Gamma(k/2)\Big)^{-1} r^{k-1}\, \phi(r)\, I_{[r>0]} \quad \text{and} \quad r \mapsto \tilde{\Phi}_k(r) = \Psi_k(r^2),$$

respectively, where Γ stands for the Euler gamma function and Ψ_k denotes the distribution function of the χ^2_k distribution.

The parameter of interest in the sequel is throughout the shape parameter $\mathbf{V}$, which determines the shape and orientation of the equidensity contours of (2.3). In Sections 2.2 and 2.3 below, we recall the pseudo-Gaussian procedures and quickly define the Gaussian-score rank-based ones, in the problem of estimating the shape and that of testing the adequacy of a fixed shape, respectively.

2.2 Estimation of shape

Consider the problem of estimating the shape $\mathbf{V}$ under unspecified values of $\boldsymbol{\theta}$ and f. The pseudo-Gaussian estimator is the Gaussian ML estimator $\widehat{\mathbf{V}}_{\mathcal{N}} := \mathbf{S}/(\mathbf{S})_{11}$, with $\mathbf{S} := (n-1)^{-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$. This estimator is asymptotically optimal in the multinormal case, and remains root-n consistent and asymptotically normal under any elliptical distribution with finite fourth-order moments. However, it is Pitman non-admissible, since it is, as we will show, uniformly dominated by the rank-based estimator of shape $\widehat{\mathbf{V}}_{\mathrm{vdW}}$ we now proceed to define.

Define, for all $\alpha > 0$ and for some root-n consistent preliminary estimator $\widehat{\mathbf{V}}_0$, the shape matrix

$$\widehat{\mathbf{V}}_\alpha := \frac{(1-\alpha)\widehat{\mathbf{V}}_0 + \alpha\mathbf{W}}{\big((1-\alpha)\widehat{\mathbf{V}}_0 + \alpha\mathbf{W}\big)_{11}},$$

where $\mathbf{W} := n^{-1} \sum_{i=1}^n \Psi_k^{-1}(R_i(\widehat{\mathbf{V}}_0)/(n+1))\, \widehat{\mathbf{V}}_0^{1/2} \mathbf{U}_i(\widehat{\mathbf{V}}_0)\mathbf{U}'_i(\widehat{\mathbf{V}}_0)\widehat{\mathbf{V}}_0^{1/2}$. Then the *van der Waerden*—that is, Gaussian-score—R-estimator for shape proposed by Hallin *et al.* (2004, 2005) is $\widehat{\mathbf{V}}_{\mathrm{vdW}} := \widehat{\mathbf{V}}_{\alpha^*}$, where

$$\alpha^* := \arg\min_{\alpha>0} \left\{ \frac{1}{2n} \sum_{i,j=1}^n \Psi_k^{-1}\left(\frac{R_i(\widehat{\mathbf{V}}_\alpha)}{n+1}\right) \Psi_k^{-1}\left(\frac{R_j(\widehat{\mathbf{V}}_\alpha)}{n+1}\right) \left(\Big(\mathbf{U}'_i(\widehat{\mathbf{V}}_\alpha)\mathbf{U}_j(\widehat{\mathbf{V}}_\alpha)\Big)^2 - \frac{1}{k} \right) \right\}$$

(this actually consists in choosing, on the curve $\{\widehat{\mathbf{V}}_\alpha, \alpha > 0\}$, the shape estimate $\widehat{\mathbf{V}}_{\alpha^*}$ that is closest to the true shape of the underlying distribution; see Section 2.3). Provided it is based on an estimator $\widehat{\mathbf{V}}_0$ that is root-n consistent under any elliptical densities without any moment assumption (such as, e.g., Tyler (1987)'s estimator of shape), $\widehat{\mathbf{V}}_{\mathrm{vdW}}$ is root-n consistent and asymptotically normal under any elliptical distribution, satisfying some extremely mild regularity conditions (which do not involve any moment condition). It is therefore valid under broader conditions than $\widehat{\mathbf{V}}_{\mathcal{N}}$. Moreover, as the latter, $\widehat{\mathbf{V}}_{\mathrm{vdW}}$ is asymptotically optimal in the multinormal case.

Now, under $\mathrm{P}^n_{\boldsymbol{\theta},\mathbf{V},f}$, where f is such that the density (2.3) has finite fourth-order moments, $\sqrt{n}\,\mathrm{vec}(\widehat{\mathbf{V}}_{\mathrm{vdW}} - \mathbf{V})$ and $\sqrt{n}\,\mathrm{vec}(\widehat{\mathbf{V}}_{\mathcal{N}} - \mathbf{V})$ are asymptotically multinormal with *proportional* asymptotic covariance matrices ($v_2(f)\mathbf{M}(\boldsymbol{\theta}, \mathbf{V})$ and $v_1(f)\mathbf{M}(\boldsymbol{\theta}, \mathbf{V})$, respectively, say). Although the definition of Pitman ARE is somewhat more intricate in the multivariate case, it is clear, in this particular case, that the Pitman ARE may still be defined as in (1.1), that is, as the corresponding proportionality factor. Thus—see Hallin *et al.* (2004, 2005) for the scalars $v_1(f)$ and $v_2(f)$—the ARE of $\widehat{\mathbf{V}}_{\rm vdW}$ with respect to $\widehat{\mathbf{V}}_{\mathcal{N}}$ under $\mathrm{P}^n_{\boldsymbol{\theta},\mathbf{V},f}$ is given by

$$\mathrm{ARE}_{k,f} = \frac{1}{k(k+2)^3} \frac{E_k(f)}{(D_k(f))^2} \Big[J_k(\phi, f) \Big]^2, \tag{2.4}$$

where, denoting by $\varphi_f = -f'/f$ the optimal location score function, we let

$$D_k(f) := \int_0^1 \Big(\tilde{F}_k^{-1}(u) \Big)^2 du, \qquad E_k(f) := \int_0^1 \Big(\tilde{F}_k^{-1}(u) \Big)^4 du,$$

and

$$J_k(\phi, f) := \int_0^1 \Big(\tilde{\Phi}_k^{-1}(u) \Big)^2 \tilde{F}_k^{-1}(u)\, \varphi_f(\tilde{F}_k^{-1}(u))\, du;$$

note that these AREs depend on the radial density f only through its density type $\{f_a,\, a > 0\}$, where $f_a(r) := f(ar)$ for all $r > 0$.

Some numerical values of these AREs are provided in Table 1. All values in Table 1 are larger or equal than one and are equal to one in the multinormal case only (where both estimators are known to compete equally, since they are asymptotically optimal). As shown by Theorem 1 below, which is the main result of this paper, this uniform dominance holds under any elliptical distribution for which the Gaussian ML estimator for shape is root-n consistent; the latter is therefore Pitman non-admissible.

Theorem 1 *For all radial density f and all integer $k \geq 2$, we have $\mathrm{ARE}_{k,f} \geq 1$, where equality holds iff f is Gaussian (that is, iff $f = \phi_a$ for some $a > 0$). Consequently, for all integer $k \geq 2$, the pseudo-Gaussian maximum likelihood estimator $\widehat{\mathbf{V}}_{\mathcal{N}}$ for shape is Pitman non-admissible.*

	underlying density						
k	t_5	t_8	t_{12}	$\mathcal{N}$	e_2	e_3	e_5
2	2.204	1.215	1.078	1.000	1.129	1.308	1.637
3	2.270	1.233	1.086	1.000	1.108	1.259	1.536
4	2.326	1.249	1.093	1.000	1.093	1.223	1.462
6	2.413	1.275	1.106	1.000	1.072	1.174	1.363
10	2.531	1.312	1.126	1.000	1.050	1.121	1.254
∞	3.000	1.500	1.250	1.000	1.000	1.000	1.000

Table 1: AREs of the rank-based estimators $\widehat{\mathbf{V}}_{\rm vdW}$ with respect to the pseudo-Gaussian estimators $\widehat{\mathbf{V}}_{\mathcal{N}}$, under k-dimensional Student (with 5, 8, and 12 degrees of freedom), normal, and power-exponential densities (with parameter $\eta = 2, 3, 5$), for $k = 2$, 3, 4, 6, 10, and $k \to \infty$.

2.3 Testing for specified shape

The other problem we consider is that of testing that the shape $\mathbf{V}$ is equal to some given value $\mathbf{V}_0$ (admissible for a shape parameter). The special case $\mathbf{V}_0 = \mathbf{I}_k$, where $\mathbf{I}_k$ stands for the k-dimensional identity matrix, yields the problem of testing for sphericity. Hallin and

Paindaveine (2004) propose a class of rank-based tests for this problem. The van der Waerden version of their tests, ϕ_{vdW} say, rejects the null (at asymptotic level α) whenever

$$\frac{1}{2n}\sum_{i,j=1}^{n}\Psi_k^{-1}\left(\frac{R_i(\mathbf{V}_0)}{n+1}\right)\Psi_k^{-1}\left(\frac{R_j(\mathbf{V}_0)}{n+1}\right)\left(\left(\mathbf{U}_i'(\mathbf{V}_0)\mathbf{U}_j(\mathbf{V}_0)\right)^2-\frac{1}{k}\right)>\chi^2_{(k-1)(k+2)/2;1-\alpha},$$

where $\chi^2_{(k-1)(k+2)/2;1-\alpha}$ denotes the α upper-quantile of a chi-square variable with $(k-1)(k+2)/2$ degrees of freedom. In this case, the pseudo-Gaussian procedure is Muirhead and Waternaux (1980)'s version of Mauchly (1940)'s Gaussian likelihood ratio test—which, for $\mathbf{V}_0=\mathbf{I}_k$, is probably the most widely used test of sphericity. This test, $\phi_{\mathcal{N}}$ say, which requires finite fourth-order moments, rejects the null (still at asymptotic level α) whenever

$$\frac{-nk}{1+\hat{\kappa}_k}\log\left(\frac{(\det\mathbf{V}_0^{-1}\widehat{\mathbf{V}}_{\mathcal{N}})^{1/k}}{k^{-1}(\mathrm{tr}\,\mathbf{V}_0^{-1}\widehat{\mathbf{V}}_{\mathcal{N}})}\right)>\chi^2_{(k-1)(k+2)/2;1-\alpha},$$

where $\hat{\kappa}_k:=[k(n^{-1}\sum_{i=1}^n d_i^4(\mathbf{V}_0))]/[(k+2)(n^{-1}\sum_{i=1}^n d_i^2(\mathbf{V}_0))^2]-1$ is a consistent estimator of the population kurtosis parameter $\kappa_k(f):=(kE_k(f))/((k+2)D_k^2(f))-1$.

The AREs of ϕ_{vdW} with respect to $\phi_{\mathcal{N}}$ coincide with those of $\widehat{\mathbf{V}}_{\mathrm{vdW}}$ with respect to $\widehat{\mathbf{V}}_{\mathcal{N}}$; see Hallin and Paindaveine (2004). Consequently, the values provided in Table 1 do also apply in this case, and most importantly, so does Theorem 1, which proves the following corollary.

Corollary 1 *For all integrer $k\geq 2$, the pseudo-Gaussian likelihood ratio test $\phi_{\mathcal{N}}$ for specified shape is uniformly dominated by ϕ_{vdW} and therefore is Pitman non-admissible.*

3 Proof of Theorem 1

In this section, we first provide a convenient reparametrization of the variational problem under consideration. We then briefly explain why standard variational techniques are inappropriate for the problem under study, and eventually give a proof of Theorem 1 that is essentially based on the arithmetic-geometric mean inequality, Jensen's inequality, and Cauchy-Schwarz inequality.

3.1 A convenient reparametrization

Rewrite the functional $f\mapsto J_k(\phi,f)$ as

$$\begin{aligned}J_k(\phi,f) &= \int_0^\infty\left(\tilde{\Phi}_k^{-1}(\tilde{F}_k(r))\right)^2 r\,\varphi_f(r)\,\tilde{f}_k(r)\,dr\\ &= \frac{1}{\mu_{k-1,f}}\int_0^\infty\left(\tilde{\Phi}_k^{-1}(\tilde{F}_k(r))\right)^2(-f'(r))\,r^k\,dr\\ &= \int_0^\infty\left\{\frac{2r\,\tilde{\Phi}_k^{-1}(\tilde{F}_k(r))}{\tilde{\phi}_k(\tilde{\Phi}_k^{-1}(\tilde{F}_k(r)))}\,\tilde{f}_k(r)+k\left(\tilde{\Phi}_k^{-1}(\tilde{F}_k(r))\right)^2\right\}\tilde{f}_k(r)\,dr.\end{aligned}$$

For any strictly positive (over $\mathbb{R}_0^+$) density f (of class C^1), the function $R:z\mapsto\tilde{F}_k^{-1}\circ\tilde{\Phi}_k(z)$ and its inverse $R^{-1}:r\mapsto\tilde{\Phi}_k^{-1}\circ\tilde{F}_k(r)$ are monotone increasing transformations (of class C^2), mapping $\mathbb{R}_0^+$ onto itself, and satisfying $\lim_{z\to 0}R(z)=\lim_{r\to 0}R^{-1}(r)=0$ and $\lim_{z\to\infty}R(z)=$

$\lim_{r\to\infty} R^{-1}(r) = \infty$. Similarly, any monotone increasing transformation R (of class C^2) of $\mathbb{R}_0^+$ such that

$$\lim_{z\to 0} R(z) = 0 \qquad \text{and} \qquad \lim_{z\to\infty} R(z) = \infty, \tag{3.5}$$

characterizes a nonvanishing density f (of class C^1) over $\mathbb{R}_0^+$ via the relation $R = \tilde{F}_k^{-1} \circ \tilde{\Phi}_k$. The functional above thus becomes

$$J_k(\phi, R) = \int_0^\infty \left(\frac{2zR(z)}{\tilde{\phi}_k(z)} \frac{\tilde{\phi}_k(z)}{R'(z)} + kz^2 \right) \tilde{\phi}_k(z)\, dz = 2 \left(\int_0^\infty \frac{zR(z)}{R'(z)} \tilde{\phi}_k(z)\, dz \right) + k^2,$$

since $\tilde{f}_k(r) = \frac{d}{dr}\tilde{F}_k(r) = \tilde{\phi}_k(z)/(\frac{d}{dz}R)$ and $\tilde{f}_k(r)\, dr = d\tilde{F}_k(r) = \tilde{\phi}_k(z)\, dz$. In this new parametrization, the ARE functional takes the form

$$\mathrm{ARE}_{k,R} = \frac{1}{k(k+2)^3} \frac{D_k^{0,4}}{\left(D_k^{0,2}\right)^2} \left[J_k(\phi, R) \right]^2, \tag{3.6}$$

where we let

$$D_k^{a,b} = D_k^{a,b}(R) := \int_0^\infty z^a (R(z))^b\, \tilde{\phi}_k(z)\, dz.$$

The ARE functional (3.6) is to be minimized over the collection $\mathcal{R}_k$ of monotone increasing functions $R : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ of class C^2 such that (3.5) holds and $D_k^{0,4}(R) < \infty$ (the latter condition is the analog on R of the fourth-order moment condition $E_k(f) < \infty$).

Note that a density type $\{f_a,\, a > 0\}$ corresponds to a class of functions $\{R_a,\, a > 0\}$, where $R_a(z) := aR(z)$ for all $z > 0$. Also the radial density ϕ is associated with the function $R(z) = z$, $z > 0$; consequently, Gaussian distributions correspond to the class of functions $R_a(z) = az$, $a, z > 0$.

3.2 Inappropriateness of standard variational arguments

Since the AREs in (3.6) depend on R through its "R-type" $\{R_a,\, a > 0\}$ only, the variational problem under consideration consists in minimizing the functional $R \mapsto D_k^{0,4}(R)\, [J_k(\phi, R)]^2$ over the class of functions $R \in \mathcal{R}_k$ satisfying $D_k^{0,2}(R) = k$. Equivalently, letting $T(z) = (R(z))^2$ for all $z > 0$, it consists in minimizing the functional

$$T \mapsto H_k(T) := D_k^{0,2}(T)\, [\tilde{J}_k(\phi, T)]^2, \tag{3.7}$$

where

$$\tilde{J}_k(\phi, T) = 4 \left(\int_0^\infty \frac{zT(z)}{T'(z)} \tilde{\phi}_k(z)\, dz \right) + k^2,$$

over the class $\mathcal{T}_k := \{T = R^2 \mid R \in \mathcal{R}_k \text{ with } D_k^{0,2}(R) = k\}$. This new parametrization makes the problem more linear since the functional H_k is now defined over the convex subset $\mathcal{T}_k$ included in a *vectorial space*. Theorem 1 states that $H_k(T) \geq k^3(k+2)^3$ for all $T \in \mathcal{T}_k$ and that the equality only holds at $z \mapsto T_0(z) := z^2$, for all $z > 0$.

Unfortunately, the classical Euler-Lagrange first order theory does not allow to deal with the isoperimetric variational problem (3.7), as the functional H_k is a product of integrals (and not a single integral). However, ad hoc investigation of the first order variation can be achieved, and standard calculations show that the latter satisfies

$$H_k'(0) := \frac{d}{dw}(H_k((1-w)T_0 + wT))|_{w=0} = 0,$$

for all $T \in \mathcal{T}_k$, so that the function T_0—corresponding to the standard Gaussian distribution—is a critical point of the shape ARE functional. Nevertheless, unlike the ARE functional associated with location problems (see Chernoff-Savage 1958, Hallin and Paindaveine 2002a, b), this is not sufficient to conclude that T_0 is a global (not even a local) minimum, since the functional $T \mapsto H_k(T)$ is *not convex*.

To investigate further the local behavior of H_k at T_0, one can of course study the second variation

$$H_k''(0) := \frac{d^2}{dw^2}(H_k((1-w)T_0 + wT))|_{w=0},$$

which, after tedious calculations, reduces, for $T \in \mathcal{T}_k$, to

$$2k^2(k+2)^2\Big\{2\int z^2\,T(z)\,\tilde{\phi}_k(z)\,dz + \int \Big(T'(z)\Big)^2\,\tilde{\phi}_k(z)\,dz$$
$$-\frac{3}{k(k+2)}\Big(\int z^2\,T(z)\,\tilde{\phi}_k(z)\,dz\Big)^2 + (k-2)\int\Big(T(z)\Big)^2\,\frac{\tilde{\phi}_k(z)}{z^2}\,dz\Big\}.$$

Although it can be easily checked that $H_k''(0) > 0$ for all functions T of the form $z \mapsto z^a$, $a \in (0,\infty)/\{2\}$, to establish the corresponding result for an arbitrary element of $\mathcal{T}_k/\{T_0\}$ seems to be extremely difficult.

Even worse: even if it can be shown that $H_k''(0) > 0$ for all $T \in \mathcal{T}_k/\{T_0\}$, this would only prove that T_0 is a (strict) *local* minimum. According to Ewing (1977, Theorem 1.4), if $H_k'(0) = 0$ and $H_k''(0) > 0$ for all $T \in \mathcal{T}_k/\{T_0\}$, a necessary and sufficient condition for T_0 to be a *global* minimum is given by the so-called *semilocal convexity of the functional* $T \mapsto H_k(T)$ *at* T_0 (where the latter means that, for all $T \in \mathcal{T}_k/\{T_0\}$, there exists a positive number $\varepsilon(T)$ such that

$$H_k((1-w)T_0 + wT) \leq (1-w)H_k(T_0) + wH_k(T),$$

for all $w \in (0, \varepsilon(T))$). Just as the positiveness of the second variation, this weak convexity property seems hard to establish directly. Along with the fact that H_k, as a product of integrals, is incompatible with standard isoperimetric Euler-Lagrange methodology, this shows that the classical methods of the calculus of variations are inappropriate for the problem under study.

The next section therefore provides a proof which does not rely on variational methods, but is partly inspired by the "direct" method introduced by Gastwirth and Wolff (1968)—who provided a simple proof for the original non-admissibility result of Chernoff-Savage (1958). See also Paindaveine (2004) for a proof *a la* Gastwirth and Wolff (1968) of multivariate Chernoff-Savage results for *location* parameters.

3.3 A direct proof of Theorem 1

To prove Theorem 1, we come back to the R-parametrization in (3.6).

Proof of Theorem 1. Using the arithmetic-geometric mean inequality, we obtain

$$J_k(\phi, R) \geq (k+2)\left\{\Big(\int_0^\infty \frac{zR(z)}{R'(z)}\tilde{\phi}_k(z)\,dz\Big)^2 k^k\right\}^{\frac{1}{k+2}}. \tag{3.8}$$

Now, applying Jensen's inequality (with respect to the measure $(R(z))^2\,\tilde{\phi}_k(z)\,dz$ and with convex function $x \mapsto 1/x$) yields

$$\int_0^\infty \frac{zR(z)}{R'(z)}\tilde{\phi}_k(z)\,dz \geq \Big(D_k^{0,2}\Big)^2\Big(\int_0^\infty z^{-1}\,(R(z))^3\,R'(z)\,\tilde{\phi}_k(z)\,dz\Big)^{-1}. \tag{3.9}$$

Integrating by parts and using that $-\tilde{\phi}_k'(z)/\tilde{\phi}_k(z) = z - (k-1)/z$ show

$$\begin{aligned}\int_0^\infty z^{-1}\,(R(z))^3\,R'(z)\,\tilde{\phi}_k(z)\,dz &= -\frac{1}{4}\int_0^\infty (R(z))^4\left(z^{-1}\,\tilde{\phi}_k(z)\right)'dz \\ &= \frac{1}{4}\int_0^\infty \Big(1-(k-2)z^{-2}\Big)(R(z))^4\,\tilde{\phi}_k(z)\,dz \\ &= \frac{1}{4}\left(D_k^{0,4}-(k-2)D_k^{-2,4}\right).\end{aligned}$$

Substituting successively in (3.9) and (3.8), we obtain

$$J_k(\phi,R) \geq (k+2)\,k^{\frac{k}{k+2}}\left\{4\left(D_k^{0,2}\right)^2\left(D_k^{0,4}-(k-2)D_k^{-2,4}\right)^{-1}\right\}^{\frac{2}{k+2}},$$

which yields (see (3.6))

$$\mathrm{ARE}_{k,R} \geq \frac{1}{k+2}\,k^{\frac{k-2}{k+2}}\,\frac{D_k^{0,4}}{\left(D_k^{0,2}\right)^2}\left(\frac{4\left(D_k^{0,2}\right)^2}{D_k^{0,4}-(k-2)D_k^{-2,4}}\right)^{\frac{4}{k+2}}.$$

Note that this already establishes the result for $k=2$. Now, since $(D_k^{0,2})^2 \leq k\,D_k^{-2,4}$ by Cauchy-Schwarz inequality, we obtain

$$\mathrm{ARE}_{k,R} \geq \left\{\frac{4}{k}\left(\frac{k}{k+2}\,\frac{D_k^{0,4}}{\left(D_k^{0,2}\right)^2}\right)^{\frac{k+2}{4}}\frac{\left(D_k^{0,2}\right)^2}{D_k^{0,4}-\frac{k-2}{k}\left(D_k^{0,2}\right)^2}\right\}^{\frac{4}{k+2}} = \left\{\frac{(1+\kappa_k)^{\frac{k+2}{4}}}{1+(\frac{k+2}{4})\kappa_k}\right\}^{\frac{4}{k+2}}, \tag{3.10}$$

where $\kappa_k = \kappa_k(R) := kD_k^{0,4}/((k+2)(D_k^{0,2})^2)-1$ is the kurtosis parameter of the distribution associated with R; note that Cauchy-Schwarz inequality yields $\kappa_k > -2/(k+2)$. Consequently, since the function $x \mapsto g_k(x) := (1+x)^{(k+2)/4} - (1+(\frac{k+2}{4})\,x)$ has a (unique, for $k>2$) global minimum at $x=0$, with corresponding value $g_k(0)=0$, we eventually obtain that $\mathrm{ARE}_{k,R} \geq 1$ for all $R \in \mathcal{R}_k$.

It remains to prove that the equality holds at Gaussian radial densities only. Now, to have the equality in Theorem 1, Jensen's inequality in (3.9), in particular, needs to be degenerate; that is, we need to have

$$\frac{z}{R(z)R'(z)} = C, \quad \forall z > 0,$$

for some real constant C. Since R is monotone increasing and $R(0)=0$, this implies that $R(z) = az$ for some $a>0$, which means that the corresponding radial density f needs to be Gaussian (see the discussion at the end of Section 3.1). As it is trivially checked that $\mathrm{ARE}_{k,R} = 1$ for $R(z)=az$, $a,z>0$, Theorem 1 is proved.

□

4 Final comments

Note that, for $k \geq 3$, Inequality (3.10) provides a lower bound for $\mathrm{ARE}_{k,f}$ as a function of the kurtosis $\kappa_k(f)$ of the underlying elliptic distribution. Taking the limit as $k \to \infty$ shows that, with $\kappa(f) := \lim_{k\to\infty} \kappa_k(f)$, which is nonnegative (since $\kappa_k(f) > -2/(k+2)$ for all f),

$$\lim_{k\to\infty} \mathrm{ARE}_{k,f} \geq 1 + \kappa(f), \tag{4.11}$$

which is the limiting value (still as $k \to \infty$) of the ARE, under radial density f, of Tyler (1987)'s *sign* estimator of shape (resp., Ghosh and Sengupta (2001)'s *sign* test for sphericity) with respect to the pseudo-Gaussian estimator $\widehat{\mathbf{V}}_{\mathcal{N}}$ (resp., pseudo-Gaussian test of sphericity $\phi_{\mathcal{N}}$); by "sign" procedures, we mean procedures that use the observations $\mathbf{X}_i$ only through their directions $\mathbf{U}_i$ from the (estimated) centre of the distribution. Actually, it can be shown that the rank-based estimator $\widehat{\mathbf{V}}_{\mathrm{vdW}}$ and test ϕ_{vdW} defined above converge a.e., for fixed n, as $k \to \infty$, to Tyler (1987)'s estimator and Ghosh and Sengupta (2001)'s sign test, respectively. This justifies the fact that, actually, the equality holds at all f in (4.11) (in particular, it can be easily checked that the equality in (4.11) occurs in each cell of the last row of Table 1). As pointed out in Hallin, Oja, and Paindaveine (2004), this is associated with the fact that, as the dimension k of the observation space goes to infinity, the information contained in the radii d_i becomes negligible when compared with that contained in the directions $\mathbf{U}_i$.

This paper shows that Gaussian-score rank-based procedures for *shape* strictly beat their pseudo-Gaussian competitors, but at Gaussian distributions (where they compete equally). As mentioned in the introduction, this Chernoff-Savage result also holds in purely *location* problems (one-sample, two-sample, MANOVA, regression problems, etc.), as well as in *serial* models (mainly VARMA models). Table 2 provides, for the same dimensions and underlying distributions as in Table 1, the ARE figures associated with the three kinds of problems, namely, shape, location, and serial problems. A quick inspection of Table 2 reveals that the shape AREs seem to be uniformly larger than the location AREs, which themselves appear to be uniformly larger than the serial ones. While it holds true that the serial AREs are uniformly smaller than the location ones (with equality under Gaussian distributions only), there exist distributions for which the corresponding ARE values are larger in location (and even serial) cases than for shape; an example, in the bivariate case, is given by the radial density f associated with the R-function (in the sense of Section 3.1)

$$z \mapsto R(z) := \begin{cases} z^2 & \text{if} \quad 0 < z < 1 \\ 2z - 1 & \text{if} \quad z \geq 1, \end{cases} \tag{4.12}$$

for which the shape, location, and serial AREs are given by 1.067, 2.084, and 2.016, respectively. Note that, strictly speaking, this function R is not of class C^2; however it can be arbitrarily well approximated (uniformly) by a function of class C^2.

Eventually, since the Fisher information for *shape* does coincide with that for *scale* (see Hallin and Paindaveine 2004), one could wonder whether the Chernoff-Savage phenomenon extends to problems where the scale is (a part of) the parameter of interest. This includes, e.g., the problem of testing that the scales of two univariate distributions do coincide or, in the multivariate setup, that of testing the equality of the covariance matrices associated with two—or several—elliptic populations (these problems are mainly motivated by their links with (M)ANOVA problems; the corresponding null hypotheses are indeed the standard assumptions for many (M)ANOVA procedures). It can be shown (see Hallin and Paindaveine 2005b for details) that, for these

		underlying density						
k		t_5	t_8	t_{12}	$\mathcal{N}$	e_2	e_3	e_5
2	shp	2.204	1.215	1.078	1.000	1.129	1.308	1.637
	loc	1.171	1.059	1.025	1.000	1.097	1.218	1.414
	ser	1.125	1.047	1.021	1.000	1.086	1.196	1.375
3	shp	2.270	1.233	1.086	1.000	1.108	1.259	1.536
	loc	1.194	1.069	1.030	1.000	1.077	1.176	1.339
	ser	1.140	1.054	1.024	1.000	1.069	1.158	1.307
4	shp	2.326	1.249	1.093	1.000	1.093	1.223	1.462
	loc	1.212	1.077	1.034	1.000	1.064	1.148	1.287
	ser	1.153	1.061	1.028	1.000	1.057	1.132	1.260
6	shp	2.413	1.275	1.106	1.000	1.072	1.174	1.363
	loc	1.242	1.092	1.042	1.000	1.048	1.111	1.219
	ser	1.172	1.071	1.034	1.000	1.042	1.100	1.199
10	shp	2.531	1.312	1.126	1.000	1.050	1.121	1.254
	loc	1.283	1.112	1.053	1.000	1.032	1.074	1.149
	ser	1.197	1.086	1.042	1.000	1.028	1.067	1.135
∞	shp	3.000	1.500	1.250	1.000	1.000	1.000	1.000
	loc	1.509	1.253	1.151	1.000	1.000	1.000	1.000
	ser	1.281	1.153	1.095	1.000	1.000	1.000	1.000

Table 2: AREs of Gaussian-score rank-based estimators for shape (sph), location (loc), and autoregressive (ser) parameters, with respect to their pseudo-Gaussian comptetitors, under k-dimensional Student (with 5, 8, and 12 degrees of freedom), normal, and power-exponential densities (with parameter $\eta = 2, 3, 5$), for $k = 2$, 3, 4, 6, 10, and $k \to \infty$.

problems, the Gaussian-score rank-based tests *do not* uniformly dominate, in the Pitman sense, the corresponding pseudo-Gaussian ones. For instance, when testing the equality of the scales of two univariate populations, the ARE of the Gaussian-score rank test with respect to the pseudo-Gaussian test, under the symmetric univariate density f associated with the function R in (4.12), is 0.947. *Location* and *scale* thus play distinct roles with respect to the Chernoff-Savage phenomenon. This leads us to conjecture that the latter is some kind of miracle that is specific to location parameters, such as location centres, regression or autoregression parameters, moving-average coefficients, and, in some sense... *Shape*, which, roughly speaking, in the orthogonal decomposition (see Hallin and Paindaveine 2004 for details) of a covariance matrix $\mathbf{\Sigma}$ into *scale* σ *and shape* $\mathbf{V}$, can be considered as the "location component" of $\mathbf{\Sigma}$.

Acknowledgements

This research was supported by a P.A.I. contract of the Belgian Federal Government and an Action de Recherche Concertée of the Communauté française de Belgique.

References

[1] Chernoff, H. and Savage, I.R. (1958), Asymptotic normality and efficiency of certain non-parametric tests, *Ann. Math Statist.*, **29**, 972-994.

[2] Ewing, G.M. (1977). Sufficient conditions for global minima of suitably convex functionals from variational and control theory, *SIAM review*, **19**, 202-220.

[3] Gastwirth, J.L. and Wolff, S.S. (1968), An elementary method for obtaining lower bounds on the asymptotic power of rank tests, *Ann. Math Statist.*, **39**, 2128-2130.

[4] Ghosh, S.K. and Sengupta, D. (2001). Testing for proportionality of multivariate dispersion structures using interdirections, *J. Nonparam. Statist.*, **13**, 331-349.

[5] Hallin, M. (1994), On the Pitman-nonadmissibility of correlogram-based methods, *J. Time Ser. Anal.*, **15**, 607-612.

[6] Hallin, M., Oja, H. and Paindaveine, D. (2004). Optimal R-estimation of shape, submitted.

[7] Hallin, M., Oja, H. and Paindaveine, D. (2005). Affine-equivariant R-estimation of shape. Manuscript in preparation.

[8] Hallin, M. and Paindaveine, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks, *Ann. Statist.*, **30**, 1103-1133.

[9] Hallin, M. and Paindaveine, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence, *Bernoulli*, **8**, 787-816.

[10] Hallin, M. and Paindaveine, D. (2004). Optimal rank-based tests for sphericity. Tentatively accepted in *Ann. Statist.*

[11] Hallin, M. and Paindaveine, D. (2005a). Affine invariant aligned rank tests for the multivariate general linear model with ARMA errors, *J. Multivariate Anal.*, **93**, 122-163.

[12] Hallin, M. and Paindaveine, D. (2005b). Optimal rank-based tests for the equality of covariance matrices. Manuscript in preparation.

[13] Hallin, M. and Puri, M.L. (1994). Aligned rank tests for linear models with autocorrelated error terms, *J. Multivariate Anal.*, **50**, 175-237.

[14] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, J. Wiley, New York.

[15] Lehmann, E.L. (1999). *Elements of Large Sample Theory*, Springer Verlag, New York.

[16] Mauchly, J.W. (1940). Test for sphericity of a normal n-variate distribution, *Ann. Math Statist.*, **11**, 204-209.

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

[18] Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York.

[19] Paindaveine, D. (2004). A unified and elementary proof of serial and nonserial, univariate and multivariate, Chernoff-Savage results. *Statist. Methodol.*, **1**, 81-91.

[20] Pratt, J.W. and Gibbons, J.D. (1981). *Concepts of Nonparametric Theory*, Springer Verlag, New York.

[21] Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter, *Ann. Statist.*, **15**, 234-251.