Generalized Skew-Elliptical Distributions and their Quadratic Forms

Marc G. Genton* and Nicola M. R. Loperfido†

October 9, 2003

Abstract

This paper introduces generalized skew-elliptical distributions (GSE), which include the multivariate skew-normal, skew-t, skew-Cauchy, and skew-elliptical distributions as special cases. GSE are weighted elliptical distributions but the distribution of any even function in GSE random vectors does not depend on the weight function. In particular, this holds for quadratic forms in GSE random vectors. This property is beneficial for inference from non-random samples. We illustrate the latter point on a data set of Australian athletes.

Key words: elliptical distribution, invariance, kurtosis, selection model, skewness, weighted distribution

1 Introduction

Probability distributions that are more flexible than the normal are often needed in statistical modeling (Hill and Dixon, 1982). Skewness in datasets, for example, can be modeled through the multivariate skew-normal distribution introduced by Azzalini and Dalla Valle (1996), which appears to attain a reasonable compromise between mathematical tractability and shape flexibility. Its probability density function (pdf) is

$$2\phi_p\left(z;\xi,\Omega\right)\cdot\Phi\left(\alpha^T\left(z-\xi\right)\right),\qquad z\in\mathbb{R}^p,\tag{1}$$

*Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA. E-mail: genton@stat.ncsu.edu

†Instituto di scienze Economiche, Facoltá di Economia, Università degli Studi di Urbino, Via Saffi 42, 61029 (PU), Italy. E-mail: nicola@econ.uniurb.it

where ϕ_p denotes the pdf of a p-dimensional normal distribution centered at $\xi \in \mathbb{R}^p$ with scale matrix $\Omega \in \mathbb{R}^{p \times p}$ and Φ denotes the cumulative distribution function (cdf) of a standard normal distribution. When (1) is the pdf of a random vector Z we write $Z \sim SN_p(\xi, \Omega, \alpha)$. The vector $\alpha \in \mathbb{R}^p$ controls the shape and the special case $\alpha = 0$ corresponds to the multivariate normal distribution. Despite the presence of an additional parameter, skew-normal distributions resemble the normal ones in several ways, for instance they are unimodal and $(Z-\xi)^T \Omega^{-1} (Z-\xi) \sim \chi_p^2$.

The kurtosis of skew-normal distributions is however bounded. Data from heavy tailed distributions can be better modeled through continuous elliptical pdfs

$$|\Omega|^{-1/2} \cdot g\left(\Omega^{-1/2}(z-\xi)\right), \qquad z \in \mathbb{R}^p, \tag{2}$$

where $\xi \in \mathbb{R}^p$ is the location vector parameter, $\Omega \in \mathbb{R}^{p \times p}$ is the scale matrix parameter, and g is the pdf of a spherical distribution, that is $g(a)$ depends on a only through $a^T a$. It follows that elliptical densities are symmetric around their location parameters.

In order to model both skewness and kurtosis, Branco and Dey (2001) introduced skew-elliptical distributions by means of the pdf

$$\frac{2}{\sqrt{|\Omega|}} g\left(\Omega^{-1/2}(z-\xi)\right) \cdot G\left(\Omega^{-1/2}(z-\xi)\right), \qquad z \in \mathbb{R}^p, \tag{3}$$

where g is the same as in (2) and G is the cdf of a univariate marginal pdf of g. Skew-elliptical distributions include skew-normal ones as well as elliptical ones.

The purpose of the present paper is to analyze invariance properties of generalized skew-elliptical distributions (GSE), a class of distributions which includes the skew-elliptical ones. More precisely, Section 2 introduces GSE and examines their relationships with other skew distributions in the literature. Section 3 presents a distributional invariance property of even functions in GSE random vectors. Section 4 discusses likelihood-based inference for generalized skew-normal (GSN) distributions. Section 5 applies the results of the previous section to inference from non-random samples, that is a set of non i.i.d. random variables. Section 6 concludes and presents some open problems and conjectures.

2 Generalized skew-elliptical distributions

Azzalini and Capitanio (1999) defined skew-elliptical distributions by means of pdfs of the form

$$\frac{2}{\sqrt{|\Omega|}} \cdot g\left(\Omega^{-1/2}(z-\xi)\right) \cdot H\left(\alpha^T (z-\xi)\right), \qquad z \in \mathbb{R}^p, \tag{4}$$

where g is the pdf of a spherical distribution and H is the cdf of a distribution symmetric around 0. The above class of densities includes the skew-elliptical ones (Branco and Dey, 2001). Note also that (4) is twice an elliptical pdf at $z \in \mathbb{R}^p$ multiplied by a cdf evaluated at a linear function of $z - \xi$. The pdf of a GSE distribution is twice an elliptical pdf at $z \in \mathbb{R}^p$ multiplied by a function of $z - \xi$. The latter function is not necessarily a cdf and does not necessarily depend on z only through $\alpha^T (z - \xi)$. More formally, we give the following definition.

Definition 1 *A p-dimensional random vector Z has a generalized skew-elliptical (GSE) distribution with location vector parameter $\xi \in \mathbb{R}^p$, positive definite scale matrix parameter $\Omega \in \mathbb{R}^{p \times p}$, and skewing function π, if its pdf is*

$$\frac{2}{\sqrt{|\Omega|}} g\left(\Omega^{-1/2}(z - \xi)\right) \cdot \pi\left(\Omega^{-1/2}(z - \xi)\right), \qquad z \in \mathbb{R}^p, \tag{5}$$

where g is the pdf of a spherical distribution, and π satisfies $0 \leq \pi(z) \leq 1$ and $\pi(-z) = 1 - \pi(z)$, $\forall z \in \mathbb{R}^p$. We write $Z \sim GSE_p(\xi, \Omega, g, \pi)$.

The location vector ξ and the scale matrix Ω are not, in general, the expected value and the covariance matrix of Z, since GSE distributions may not be symmetric with respect to ξ. Moreover, they may not have finite second moments. Note also that we could write $\pi\left(\Omega^{-1/2}(z - \xi)\right)$ as $\tilde{\pi}(z - \xi)$ where $\tilde{\pi}$ satisfies exactly the same constraints as π, i.e. $0 \leq \tilde{\pi}(z) \leq 1$ and $\tilde{\pi}(-z) = 1 - \tilde{\pi}(z)$, $\forall z \in \mathbb{R}^p$. Hence, the form (5) and the form

$$\frac{2}{\sqrt{|\Omega|}} g\left(\Omega^{-1/2}(z - \xi)\right) \cdot \tilde{\pi}(z - \xi), \qquad z \in \mathbb{R}^p, \tag{6}$$

can be used interchangeably, see Section 4.

The skewing function π is flexible enough for the GSE class to include many well known skew distributions, some of which appear in Table 1. In order to keep the presentation simple, location vectors and scale matrices are set equal to 0 and I_p respectively. In Table 1, ϕ_p, τ_{p,ν_1,ν_2}, γ_p, and ϵ_p denote the p-dimensional pdf of a normal, generalized Student t, Cauchy, and elliptical distribution respectively. Similarly, Φ, $T_{\nu_1^*,\nu_2^*}$, C, and El denote the univariate cdf of the standard normal, the generalized Student t with $\nu_1^* = \nu_1 + z^T z$ and $\nu_2^* = \nu_2 + p$, the Cauchy, and an elliptical distribution. All the distributions in Table 1 are defined on $\mathbb{R}^p$.

The skewing functions in Table 1 are monotone functions of their arguments, but this does not need to hold for every GSE distribution. For instance, consider the skewing function $\pi(z) = \Phi(z^3 - z)$ defined on $\mathbb{R}$, where Φ is the cdf of a univariate standard normal distribution. It readily follows that $\pi(z)$ is increasing in z only when $|z| > 1/\sqrt{3}$.

Table 1: Some well known skew distributions belonging to the GSE family.

	$g(z)$	$\pi(z)$
skew-normal	$\phi_p(z)$	$\Phi\left(\alpha^T z\right)$
skew-t	$\tau_{p,\nu_1,\nu_2}(z)$	$T_{\nu_1^*,\nu_2^*}\left(\alpha^T z\right)$
skew-Cauchy	$\gamma_p(z)$	$C\left(\alpha^T z\right)$
skew-elliptical	$\epsilon_p(z)$	$El\left(\alpha^T z\right)$

The skewing function π can also be a constant. Then it follows that $\pi(z) = 1/2$, $\forall z \in \mathbb{R}^p$, and the corresponding distribution is elliptical. Therefore, the class of GSE distributions includes elliptical distributions as special cases. The skew-normal, skew-Cauchy, skew-t, and skew-elliptical distributions in Table 1 assume some relationship between g and π. The GSE class, however, includes all possible combinations of elliptical distributions and skewing functions. Flexibility of the GSE class in modeling skewness can be appreciated by considering the univariate skew-t distribution. Its third cumulant is unbounded, as can easily be seen by considering the pdf of the t distribution truncated at the origin, a limit of the skew-t, with ν degrees of freedom (personal communication from Donald Rubin):

$$\frac{\Gamma\left[(\nu+1)/2\right]}{\Gamma(\nu/2)} \frac{2}{\sqrt{\nu\pi}} \left(1+\frac{z^2}{\nu}\right)^{\frac{-\nu-1}{2}}, \qquad z \geq 0.$$

This property represents an advantage of the GSE family over the skew-normal distribution when modeling highly skewed data. Indeed, the skewness of the skew-normal distribution is bounded, see Azzalini and Dalla Valle (1996).

GSE distributions arise in inference from non-random samples, that is sets of observations which are not i.i.d. according to the population's density. This happens when the value of an observation influences its probability of being included in the sample (Copas and Li, 1997). Under these circumstances, an observation's distribution can be better described by a selection model (Bayarri and De Groot, 1992):

$$\frac{f(z;\theta) \cdot w(z)}{E\left[w(Z)\right]}, \qquad z \in \mathbb{R}^p, \tag{7}$$

where f is the pdf of the sampled population, θ is the parameter of interest, w is a nonnegative weight function and $E\left[w(Z)\right]$ is the expected value of the random variable $w(Z)$. The representation of a $GSE_p(0, \Omega, g, \pi)$ distribution as a selection model is straightforward:

$$f(z;\theta) = \frac{g\left(\Omega^{-1/2} z\right)}{|\Omega|^{1/2}}, \qquad w(z) = \pi\left(\Omega^{-1/2} z\right), \qquad E[w(Z)] = 1/2. \tag{8}$$

It follows that the multivariate skew-normal distribution is a selection model as well. Indeed, the p-dimensional skew-normal distribution can be generated through a $(p+1)$-dimensional normal distribution, by conditioning the first p variates on the event that the last variate is larger than its expected value. From the inferential point of view, this means that skew-normal samples arise when multivariate normal observations are included in the sample only if a given component of the observation itself is larger than it is expected to be. Similar comments hold for the multivariate skew-elliptical distribution (Azzalini and Capitanio, 1999; Branco and Dey, 2001; Sahu *et al.*, 2003), the multivariate skew-t distribution (Branco and Dey, 2001; Sahu *et al.*, 2003), and the multivariate skew-Cauchy (Arnold and Beaver, 2000).

GSE distributions also arise in prospective studies (Weinberg and Sandler, 1991; Weinberg and Wacholder, 1993; Wacholder and Weinberg, 1994; Zhang, 2000). Consider a random sample $Z_1, \ldots, Z_n$ from a p-dimensional elliptical distribution with pdf g. Let $d_i \in \{0, 1\}$, $i = 1, \ldots, n$, be the observed value of a dichotomous random variable D_i associated with the i-th observation, and $P(D_i = 0 | Z_i = z_i) = \pi(z_i)$. Prospective studies focus on the conditional distribution of Z_i given $D_i = d_i$. From Bayes' theorem we get:

$$f(z_i | d_i = 0) = \frac{g(z_i) \cdot \pi(z_i)}{E[\pi(Z)]}, \quad f(z_i | d_i = 1) = \frac{g(z_i) \cdot (1 - \pi(z_i))}{1 - E[\pi(Z)]}. \tag{9}$$

It easily follows that $\pi(-z_i) = 1 - \pi(z_i)$ implies that $f(z_i | d_i)$ is GSE. Relevant examples include the logistic regression model:

$$P(D_i = 0 | Z_i = z_i) = \frac{\exp\left(\beta^T z_i\right)}{1 + \exp\left(\beta^T z_i\right)}, \tag{10}$$

and the probit regression model:

$$P(D_i = 0 | Z_i = z_i) = \Phi\left(\beta^T z_i\right), \tag{11}$$

when the intercept is assumed to be equal to zero.

3 Invariance properties

Azzalini (1985) shows that if $Z \sim SN_1(0, 1, \alpha)$, then $Z^2 \sim \chi^2_1$ for any value of the shape parameter α. This result has been generalized to the multivariate skew-normal distribution by

Azzalini and Dalla Valle (1996): if $Z \sim SN_p(0, \Omega, \alpha)$, then $Z^T \Omega^{-1} Z \sim \chi^2_p$. A similar result holds for skew-elliptical random vectors (Branco and Dey, 2001). A further generalization is given by Azzalini and Capitanio (1999): if $Z \sim SN_p(0, \Omega, \alpha)$, then $Z^T A Z \sim \chi^2_p$ when $A\Omega A = A$. Genton *et al.* (2001) compute moments of quadratic forms in skew-normal random vectors. The joint distribution of several quadratic forms is examined in Azzalini and Capitanio (1999), as well as in Genton *et al.* (2001). These results are related to the following distributional invariance property (Loperfido, 2001): if $Z \sim SN_p(0, \Omega, \alpha)$, then the joint distribution of the products $Z_i Z_j$ $(i, j = 1, \ldots, p)$ does not depend on the shape parameter α . In this section, we show that a similar property holds for any even function $\tau(Z)$, i.e. a function such that $\tau(-z) = \tau(z)$ $\forall z \in \mathbb{R}^p$, of a GSE random vector Z centered at 0.

Proposition 1 *If* $Z \sim GSE_p(0, \Omega, g, \pi)$*, then the distribution of* $\tau(Z)$*, where* τ *is an even function, does not depend on the skewing function* π*.*

Proof: We derive the proof for a real-valued function τ since the vector- or matrix-valued case is similar and straightforward. Without loss of generality, we can assume that $\Omega = I_p$. It suffices to prove that the characteristic function $c(q)$ of $\tau(Z)$,

$$c(q) = \int_{\mathbb{R}^p} \exp\left(i\tau(z)q\right) 2g(z)\,\pi(z)\,dz, \qquad q \in \mathbb{R}, \tag{12}$$

does not depend on π. Let A^+ (A^-) be the set of vectors in $\mathbb{R}^p$ whose first component is not negative (is negative):

$$A^+ = \{(z_1, z_2, \ldots, z_p) : z_1 \geq 0\}, \qquad A^- = \{(z_1, z_2, \ldots, z_p) : z_1 < 0\}. \tag{13}$$

It follows that $A^+ \cup A^- = \mathbb{R}^p$ and $A^+ \cap A^- = \emptyset$. Hence:

$$c(q) = \int_{A^+} \exp\left[i\tau(z)q\right] 2g(z)\,\pi(z)\,dz + \int_{A^-} \exp\left[i\tau(z)q\right] 2g(z)\,\pi(z)\,dz.$$

Consider now the transformation $w = -z$ and notice that $z \in A^+ \Longleftrightarrow -z \in \bar{A}^-$, the closure of A^-. Because $\bar{A}^-$ and A^- have the same measure, we have:

$$\int_{A^-} \exp\left[i\tau(z)q\right] 2g(z)\,\pi(z)\,dz = \int_{A^+} \exp\left[i\tau(-w)q\right] 2g(-w)\,\pi(-w)\,dw. \tag{14}$$

By assumption, $\tau(-z) = \tau(z)$, $g(-z) = g(z)$, and $\pi(-z) = 1 - \pi(z)$. Thus:

$$c(q) = 2\int_{A^+} \exp\left[i\tau(z)q\right] 2g(z)\,dz, \tag{15}$$

which does not depend on π. ■

Corollary 1 *If $Z \sim GSE_p(0, \Omega, g, \pi)$, then the distribution of ZZ^T does not depend on the skewing function π.*

Proof. This is a direct consequence of ZZ^T being an even function and Proposition 1. ■

Corollary 2 *Let $A_1, \ldots, A_m$ be $p \times p$ real matrices and let $Z \sim GSE_p(0, \Omega, g, \pi)$. Then the joint distribution of the quadratic forms $(Z^T A_1 Z, \ldots, Z^T A_m Z)$ does not depend on the skewing function π.*

Proof. This is a direct consequence of quadratic forms being even functions and Proposition 1. ■

The following proposition presents an invariance property that will be useful for inference in the GSE family, see Section 4.

Proposition 2 *If $Z_i \sim GSE_p(0, \Omega, g, \pi_i)$, $i = 1, \ldots, n$, are independent random vectors and $S = \sum_{i=1}^n Z_i {Z_i}^T / n$, then the joint distribution of ${Z_1}^T S^{-1} Z_1, \ldots, {Z_n}^T S^{-1} Z_n$ does not depend on Ω , $\pi_1, \ldots, \pi_n$.*

Proof. Without loss of generality, we can assume that $\Omega = I_p$. Then it suffices to show that the joint distribution of the vector of quadratic forms $(Z_1^T S^{-1} Z_1, \ldots, Z_n^T S^{-1} Z_n)$ does not depend on the skewing functions $\pi_1, \ldots, \pi_n$. The following set of determinants is a function of $(Z_1 Z_1^T, \ldots, Z_n Z_n^T)$:

$$\frac{|Z_1 Z_1^T + S|}{|S|} - 1, \ldots, \frac{|Z_n Z_n^T + S|}{|S|} - 1. \tag{16}$$

From elementary properties of determinants, we know that $|A + yy^T| = |A|(1 + y^T A^{-1} y)$. Then:

$$\left(\frac{|Z_1 Z_1^T + S|}{|S|} - 1, \ldots, \frac{|Z_n Z_n^T + S|}{|S|} - 1 \right) = \left({Z_1}^T S^{-1} Z_1, \ldots, {Z_n}^T S^{-1} Z_n \right). \tag{17}$$

By Corollary 1, the joint distribution of $(Z_1 Z_1^T, \ldots, Z_n Z_n^T)$ does not depend on the skewing functions $\pi_1, \ldots, \pi_n$, which is therefore also the case for $\left({Z_1}^T S^{-1} Z_1, \ldots, {Z_n}^T S^{-1} Z_n \right)$. ■

4 Generalized skew-normal distributions

Invariance properties of GSE distributions allow one to ignore the sampling bias when the distribution of the data are sample selection models. They lead to several extensions of well-known inferential results from normal distribution theory.

Consider a GSE pdf (6) which factors into a $p-$dimensional normal pdf ϕ_p and a skewing function π, that is:

$$2\phi_p\left(z;\xi,\Omega\right)\cdot\pi\left(z-\xi\right),\qquad z\in\mathbb{R}^p. \tag{18}$$

A random vector Z with pdf (18) has a generalized skew-normal distribution with location vector parameter ξ, scale matrix parameter Ω, and skewing function π. We denote it by $Z\sim GSN_p\left(\xi,\Omega,\pi\right)$. Note that the dependence of π on Ω has been removed, and thus inference is simplified as shown below. The multivariate normal distribution and the skew-normal distribution (1) are special cases of GSN distributions.

The skewing function π can be interpreted as a parameter, since different skewing functions lead to different GSN distributions. It follows that sufficient and ancillary statistics are defined for GSN distributions too. For example, consider a random variable Z whose pdf $f\left(z;\theta,\gamma\right)$ depends on the parameters θ and γ. A statistic t is said to be partially sufficient for θ if its distribution depends on θ only and it is sufficient for θ for any given value of γ (Basu and Pereira, 1983; Reid, 1995). Loperfido (2001) shows that the sum of squares and products (SSP) matrix is partially sufficient for the scale parameter Ω when the rows of the data matrix are independent skew-normal $SN_p\left(0,\Omega,\alpha\right)$ random vectors. Similar results hold for GSN random vectors.

Under the same assumptions, the SSP matrix divided by the number of observations is the maximum likelihood (ML) estimator and the uniform minimum variance unbiased (UMVU) estimator of the scale matrix. Moreover, its distribution is Wishart. The above statements can be formalized as follows:

Proposition 3 *Let $\widehat{\Omega}=Z^TZ/n$, where Z is an $n\times p$ matrix whose rows $Z_1,\ldots,Z_n$ are independent and $Z_i\sim GSN_p\left(0,\Omega,\pi_i\right)$, $i=1,\ldots,n$. Then*

(1) $\widehat{\Omega}$ is the ML estimator of Ω.

(2) The distribution of $\widehat{\Omega}$ is Wishart: $\widehat{\Omega}\sim W\left(\Omega/n,n\right)$.

(3) $\widehat{\Omega}$ is partially sufficient for Ω.

(4) $\widehat{\Omega}$ is the UMVU estimator of Ω.

Proof:

(1) Because the random vectors $Z_1, \dots, Z_n$ are independent and $Z_i \sim GSN_p(0, \Omega, \pi_i)$, $i = 1, \dots, n$, the likelihood function l is the product of a function of Ω only and a function of $\pi_1, \dots, \pi_n$ only. Therefore, the supremum of l is:

$$\sup_{\Omega, \pi_1, \dots, \pi_n} l(\Omega, \pi_1, \dots, \pi_n) = \sup_{\Omega} |\Omega|^{-n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1} Z^T Z\right)\right] \cdot \sup_{\pi_1, \dots, \pi_n} \prod_{i=1}^{n} \pi_i(Z_i).$$

The above equation implies that the functional form of $\widehat{\Omega}$ does not depend on the choice of $\pi_1, \dots, \pi_n$. Hence standard maximization procedures (e.g. Mardia *et al.*, 1979, p. 104) lead to $\widehat{\Omega} = Z^T Z/n$ and the first part of the proof is complete.

(2) If the n rows of a matrix Y are i.i.d. $N_p(0, \Omega)$ then $Y^T Y \sim W(\Omega, n)$, see e.g. Mardia *et al.* (1979, p. 66). Equivalently, if the n rows of the matrix $Y/\sqrt{n}$ are i.i.d. $N_p(0, \Omega/n)$ then $Y^T Y/n \sim W(\Omega/n, n)$. Since $Z^T Z/n$ is an even function in the rows Z_i of Z, $Z^T Z/n \sim W(\Omega/n, n)$, see Proposition 1 in Section 3. We already proved that $\widehat{\Omega} = Z^T Z/n$ and this completes the second part of the proof.

(3) We already proved that the distribution of $\widehat{\Omega}$ depends on Ω only. Then it suffices to prove that the conditional distribution of the sample given $Z^T Z$ does not depend on Ω for any given choice of $\pi_1, \dots, \pi_n$. Since $Z^T Z \sim W(\Omega, n)$, the pdf of $Z^T Z$ is proportional to (Mardia *et al.*, 1979, p. 85):

$$|\Omega|^{-n/2} \sqrt{|Z^T Z|^{n-p-1}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1} Z^T Z\right)\right].$$

The conditional pdf of Z given $Z^T Z = D$ is proportional to:

$$\frac{|\Omega|^{-n/2} \cdot \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1} D\right)\right] \cdot \prod_{i=1}^{n} \pi_i(Z_i)}{|\Omega|^{-n/2} \sqrt{|D|^{n-p-1}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Omega^{-1} D\right)\right]} = \sqrt{|D|^{1+p-n}} \prod_{i=1}^{n} \pi_i(Z_i).$$

The last relationship does not depend on Ω for any given choice of $\pi_1, \dots, \pi_n$. Hence the third part of the proof is complete.

(4) The Wishart distribution belongs to the exponential family and hence is complete (e.g. Mardia *et al.*, 1979, p. 46). The UMVU property of $\widehat{\Omega}$ easily follows from sufficiency of $Z^T Z$, unbiasedness of $\widehat{\Omega}$ with respect to Ω, and completeness of the Wishart distribution. ■

Invariance properties of GSN distributions can be applied to hypothesis testing as well. Many likelihood-based tests for scale matrices with normal data $Z_i \sim N_p(0, \Omega)$ maintain their properties when the data are GSN, i.e. $Z_i \sim GSN_p(0, \Omega, \pi_i)$. The following proposition focuses on likelihood ratio tests (LRT) for the equality of two scale matrices. It shows that the functional form of the LRT statistic, as well as the power function of the LRT test, do not depend on the skewing functions $\pi_1, \dots, \pi_n$. Hence the skewing functions do not need to be specified, allowing a high degree of robustness with respect to departures from normality and maintaining many optimality properties of the LRT. More formally:

Proposition 4 *Let $Z_1, \ldots, Z_n$ and $Y_1, \ldots, Y_m$ be independent p-dimensional random vectors such that $Z_i \sim GSN_p(0, \Omega_Z, \pi_i)$, $i = 1, \ldots, n$, and $Y_j \sim GSN_p(0, \Omega_Y, \omega_j)$, $j = 1, \ldots, m$. Moreover, let Λ be the likelihood ratio test statistic for the null hypothesis $H_0 : \Omega_Z = \Omega_Y$ against the alternative $H_1 : \Omega_Z \neq \Omega_Y$. Then the followings hold:*

(1) Λ is invariant with respect to the skewing functions $\pi_1, \ldots, \pi_n$ and $\omega_1, \ldots, \omega_m$.

(2) The power function of the LRT does not depend on the skewing functions $\pi_1, \ldots, \pi_n$ and $\omega_1, \ldots, \omega_m$.

(3) When the null hypothesis is true and the sample size is large, the distribution of $-2\log\Lambda$ is chi-square: $-2\log\Lambda \overset{a}{\sim} \chi^2_{p(p+1)/2}$ under H_0.

Proof: Denote by A, B, and C the following matrices:

$$A = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T, \qquad B = \frac{1}{m}\sum_{j=1}^{m} Y_j Y_j^T, \qquad C = \frac{nA + mB}{n+m},$$

and denote by W the vector of weight functions $\pi_1, \ldots, \pi_n, \omega_1, \ldots, \omega_m$.

(1) From the previous proposition's proof we know that:

$$\sup_{H_0} l(\Omega_Z, \Omega_Y, W) = \sqrt{|C|^{-n-m}} \cdot \sup_W \prod_{i=1}^{n} \pi_i(Z_i) \prod_{j=1}^{m} \omega_j(Y_j).$$

Following a similar argument, we can write:

$$\sup_{\Omega_Z, \Omega_Y, W} l(\Omega_Z, \Omega_Y, W) = \sqrt{|A|^{-n}} \cdot \sqrt{|B|^{-m}} \cdot \sup_W \prod_{i=1}^{n} \pi_i(Z_i) \prod_{j=1}^{m} \omega_j(Y_j)$$

By definition, the LRT statistic is the ratio of the above suprema:

$$\Lambda = \sqrt{\frac{|C|^{n+m}}{|A|^n \cdot |B|^m}}$$

Its functional form does not depend on W and hence it is invariant with respect to it. This completes the first part of the proof.

(2) The statistic Λ depends on $Z^T Z$ and $Y^T Y$ only, which are even functions of the data. Hence Λ is an even function of the data. By Proposition 1 in Section 3, this implies that the distribution of Λ, and hence the power function of the test, depends on Ω_Z and Ω_Y only, and not on the skewing functions $\pi_1, \ldots, \pi_n, \omega_1, \ldots, \omega_m$.

(3) Proof of the asymptotic properties of $-2\log\Lambda$ easily follows from ordinary properties of likelihood ratio tests and from the invariance property of Λ. The proof is then complete. ■

It is well known that the joint distribution of the random variables:

$$Z_1^T\left(Z^TZ\right)^{-1}Z_1,\ldots,Z_n^T\left(Z^TZ\right)^{-1}Z_n$$

does not depend on Ω, when the rows $Z_1,\ldots,Z_n$ of the data matrix Z are independent and $Z_i\sim N_p\left(0,\Omega\right).$ The same result holds if $Z_i\sim GSN_p\left(0,\Omega,\pi_i\right)$: they are ancillary with respect to the scale matrix Ω and skewing functions $\pi_1,\ldots,\pi_n$. Therefore they can be used to check generalized skew-normality.

Proposition 5 *If $Z_i\sim GSN_p\left(0,\Omega,\pi_i\right)$, $i=1,\ldots,n$ are independent random vectors and $S=\sum_{i=1}^n Z_i{Z_i}^T/n$, then the joint distributions of $Z_1^TS^{-1}Z_1,\ldots,{Z_n}^TS^{-1}Z_n$ does not depend on Ω, $\pi_1,\ldots,\ \pi_n$.*

Proof: This is a special case of Proposition 2. ■

Notice that all the above propositions hold even when the observations are not identically distributed, since the skewing functions $\pi_1,\ldots,\pi_n$ might not be equal. Moreover, the assumption of the location parameter ξ being zero is not restrictive when ξ is known or the sample is large and a consistent estimator of ξ is available.

5 Inference from non-random samples

This section applies invariance properties of GSN distributions to inference on heights and weights of Australian adults, using a non-random sample drawn from the same population. From official statistics (Australian Bureau of Statistics, 1995) we know that the average height (in centimeters) and weight (in kilograms) of adult Australian males (females) is 174.8 and 82.0 (161.4 and 67.0) respectively. We make the standard assumption that the joint distributions of heights and weights for adult Australian males (H_M and W_M) and females (H_F and W_F) are bivariate normal:

$$\begin{pmatrix} H_M \\ W_M \end{pmatrix} \sim N_2\left[\begin{pmatrix} 174.8 \\ 82.0 \end{pmatrix},\Omega_M\right]$$
$$\begin{pmatrix} H_F \\ W_F \end{pmatrix} \sim N_2\left[\begin{pmatrix} 161.4 \\ 67.0 \end{pmatrix},\Omega_F\right]$$

Under the above model, inference on heights and weights of Australian adults reduces to inference on Ω_M and Ω_F. We shall focus on point estimation of Ω_M and Ω_F and hypothesis testing for their equality. Inference will be based on the heights and weights collected by the Australian Institute of Sport (AIS) from 202 athletes of both sexes (102 males and 100 females), competing in different events. Cook and Weisberg (1994) analyzed this data within the framework of a regression model, under the normality assumption. Azzalini and Dalla Valle (1996) showed that the bivariate skew-normal distribution gave a better fit, the data being slightly skewed. Arnold and Beaver (2000) modeled skewness through the skew-Cauchy distribution, which also accounts for heavier tails. AIS data, however, are not a random sample from the adult Australian population. Individuals are included in the sample only if they are gifted athletes. Moreover, the sample includes athletes of both sexes, competing in different disciplines (i.e. basketball, gymnastic, and rowing). Indeed, the AIS data are significantly skewed: p-values associated with Mardia's $b_{1,2}$ measure of multivariate skewness (Mardia, 1970) for male athletes and female athletes are 0.003 and 0.046 respectively.

We now motivate the GSN model for the AIS data. Without loss of generality we can represent the joint distribution of height H_S and weight W_S of an adult Australian individual of sex S ($S = M$ for males and $S = F$ for females) as follows:

$$\begin{aligned} H_S &= \mu_{HS} + \lambda_{HS} \cdot U_S + \gamma_{HS} \cdot \xi_{HS}, \\ W_S &= \mu_{WS} + \lambda_{WS} \cdot U_S + \gamma_{WS} \cdot \xi_{WS}, \end{aligned}$$

where U_S, ξ_{HS}, and ξ_{WS} are independent standard normal variables and:

$$\Omega_S = \begin{pmatrix} \lambda_{HS}^2 + \gamma_{HS}^2 & \lambda_{HS} \cdot \lambda_{WS} \\ \lambda_{HS} \cdot \lambda_{WS} & \lambda_{WS}^2 + \gamma_{WS}^2 \end{pmatrix},$$

$$\mu_{HS} = \begin{cases} 174.8 & S = M, \\ 161.4 & S = F, \end{cases} \qquad \mu_{WS} = \begin{cases} 82.0 & S = M, \\ 67.0 & S = F. \end{cases}$$

It easily follows that the above model is a single factor model, where U_S is the common factor and ξ_{HS}, ξ_{WS} are the specific factors for height and weight respectively. The former can be interpreted as a proxy for "physical fitness", which is clearly above average for all individuals in the AIS data set. More formally:

$$\left.\begin{pmatrix} H_S \\ W_S \end{pmatrix}\right| \text{individual in the data set} \quad \Leftrightarrow \quad \left.\begin{pmatrix} H_S \\ W_S \end{pmatrix}\right| U_S > 0$$

Azzalini and Dalla Valle (1996) show that the distribution of $H_S, W_S|U_S > 0$ (and hence the joint distribution of heights and weights for the athletes in the sample) is bivariate skew-normal, under the above assumptions. More precisely:

$$\left.\begin{pmatrix} H_S \\ W_S \end{pmatrix}\right| \text{individual in the data set} \sim SN_2\left[\begin{pmatrix} \mu_{HS} \\ \mu_{WS} \end{pmatrix}, \Omega_S, \alpha_S\right]$$

where α_S is a function of λ_{HS}, γ_{HS}, λ_{WS}, and γ_{WS}. In order to achieve some robustness and to account for individual differences (i.e. for different sport events) we propose the more general GSN bivariate model:

$$\begin{aligned} Z_i &\sim GSN_2\left(0, \Omega_M, \pi_i\right) & i &= 1, \ldots, 102, \\ Y_j &\sim GSN_2\left(0, \Omega_F, \omega_j\right) & j &= 1, \ldots, 100. \end{aligned}$$

Here $Z_1, \ldots, Z_{102}$ and $Y_1, \ldots, Y_{100}$ are the vectors of differences between the observed heights, weights and the corresponding populations' averages:

$$\begin{aligned} Z_i &= \begin{pmatrix} \text{Height of the } i\text{-th male athlete} - 174.8 \\ \text{Weight of the } i\text{-th male athlete} - 82.0 \end{pmatrix}, \\ Y_j &= \begin{pmatrix} \text{Height of the } j\text{-th female athlete} - 161.4 \\ \text{Weight of the } j\text{-th female athlete} - 67.0 \end{pmatrix}. \end{aligned}$$

Despite the above arguments, the GSN distribution might not be an appropriate model for the AIS data. The sampled population might not be normal, nor the sampling bias could be modeled through a skewing function $\pi(z) = 1 - \pi(-z)$. An example of the latter case is:

$$f(z) = \frac{\phi_1(z)}{\Phi(\delta_0)} \cdot \Phi\left(\frac{\delta z + \delta_0}{\sqrt{1-\delta^2}}\right), \qquad z \in \mathbb{R}; \delta_0 \in \mathbb{R}, \quad -1 < \delta < 1. \tag{19}$$

The above distribution is a selection model with respect to the normal pdf, but it is not GSN since in general:

$$\Phi\left(\frac{-\delta z + \delta_0}{\sqrt{1-\delta^2}}\right) \neq 1 - \Phi\left(\frac{\delta z + \delta_0}{\sqrt{1-\delta^2}}\right).$$

Simple counterexamples show that invariance properties introduced in Section 3 do not hold for (19) unless $\delta_0 = 0$.

Invariance properties of GSE distributions can be used to check the adequacy of the GSN model. Azzalini and Capitanio (1999) analyzed body fat and body mass indices of the athletes in the AIS data set, under the assumption that they came from a skew-normal distribution. They computed the quadratic forms of the data and used Healy's plot (Healy, 1968) to obtain a graphical display of fit. We also use quadratic forms to check the adequacy of the GSN model for the heights and weights in the AIS data set. We consider the Mahalanobis distances of the observations from their official average values, with respect to the ML estimates $\widehat{\Omega}_M$ and $\widehat{\Omega}_F$, that is:

$$Z_1^T\widehat{\Omega}_M^{-1}Z_1,\ldots,Z_{102}^T\widehat{\Omega}_M^{-1}Z_{102} \text{ and } Y_1^T\widehat{\Omega}_F^{-1}Y_1,\ldots,Y_{100}^T\widehat{\Omega}_F^{-1}Y_{100}.$$

By Proposition 5, the joint distribution of these statistics does not depend neither on Ω_M,Ω_F nor on $\pi_1,\ldots,\pi_{102}$ and $\omega_1,\ldots,\omega_{100}$. The same holds for any function of the above statistics, including Mardia's measure of multivariate kurtosis (Mardia, 1970):

$$b_{2,p}(M) = \frac{1}{102}\sum_{i=1}^{102} Z_i^T\widehat{\Omega}_M^{-1}Z_i, \tag{20}$$

$$b_{2,p}(F) = \frac{1}{100}\sum_{j=1}^{100} Y_j^T\widehat{\Omega}_F^{-1}Y_j. \tag{21}$$

Mardia (1970) obtains the asymptotic distribution of $b_{2,p}$ under normality. Proposition 5 implies that the same result holds under generalized skew-normality. Hence, we can use $b_{2,p}(M)$ and $b_{2,p}(F)$ to check the adequacy of the GSN model. The p-values of the $b_{2,2}$ statistics for male and female athletes are 0.098 and 0.016 respectively. Hence, at the 0.05 level, we do not reject the GSN hypothesis for male athletes, but we do reject it for female athletes. If we do not reject the GSN model for both male and female athletes, we can test the hypothesis $H_0 : \Omega_M = \Omega_F$ through the likelihood ratio test. Since the sample size is large, the sampling distribution of $-2\log\Lambda$ is approximately chi-square with 3 degrees of freedom under the null hypothesis (Proposition 4 in the previous section). The observed value of $-2\log\Lambda$ is 11.166, and the corresponding p-value is 0.011. We can therefore reject the hypothesis $\Omega_M = \Omega_F$ at the 0.05 level.

6 Conclusions

In this article, we have introduced a new class of skewed distributions: generalized skew-elliptical (GSE) distributions. These are weighted elliptical distributions that include the multivariate skew-normal, skew-t, skew-Cauchy, and skew-elliptical distributions as special cases. We have shown that the distribution of any even function in GSE random vectors does not depend on the skewing function, which holds in particular for quadratic forms. This property is beneficial for inference from non-random samples. We have developed inference for the special case of generalized skew-normal (GSN) distributions and illustrated our results on the Australian athletes data set.

At present time there does not exist a general method for computing ordinary measures of skewness and kurtosis (i.e. cumulants) for GSE distributions, due to the great flexibility of the skewing function and the extent of the elliptical family. The problem is strictly related to the characteristic function of a GSE pdf, which we conjecture to have the following form:

$$c(t) = \int_{\mathbb{R}^p} \exp\left(it^T z\right) 2g\left(z\right)\pi\left(z\right)dz = 2\Psi(t^T t)k(t), \qquad t \in \mathbb{R}^p, \tag{22}$$

where Ψ is the characteristic function corresponding to the elliptical pdf g and the function k is a function such that $k\left(-t\right) = 1 - k\left(t\right)$ and $0 \leq k\left(t\right) \leq 1$. The conjecture is true for the skew-normal distribution (Azzalini and Dalla Valle, 1996) where $k(t) = \Phi(\delta^T t)$, and it is worth asking whether it holds for a more general subclass of GSE distributions.

7 Acknowledgments

The authors would like to thank the Editor, the Associate Editor, and three anonymous referees for comments that improved this manuscript. The authors are also grateful to Adelchi Azzalini and Donald Rubin for useful remarks. This research was partially supported by MURST (Grant PRIN 2000), Italy.

8 References

1. Arnold, B. C. & Beaver, R. J. (2000). The skew-Cauchy distribution. *Statistics and Probability Letters* **49**, 285-290.

2. Australian Bureau of Statistics (1995). *How Australians measure up.* ABS Publications, Canberra.

3. Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171-178.

4. Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

5. Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Statist. Soc. B* **61**, 579-602.

6. Basu, D. & Pereira, C. A. B. (1983). Conditional independence in statistics. *Sankhya A*, **45**, 324-37.

7. Bayarri, M. J. & De Groot, M. (1992). A BAD view of weighted distributions and selection models. In *Bayesian Statistics 4* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith).

8. Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Analysis* **79** , 99-113.

9. Cook, R. D. & Weisberg, S. (1994). *An introduction to regression graphics.* Wiley, New York.

10. Copas, J. B. & Li, H. G. (1997). Inference from non-random samples (with discussion). *J. R. Statist. Soc. B*, **59**, 55-95.

11. Genton, M. G., He, L. & Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statistics and Probability Letters* **51**, 319-325.

12. Healy, M. J. R. (1968). Multivariate normal plotting. *Appl. Statist.* **17**, 157-161.

13. Hill, M. A. & Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrics* **38**, 377-396.

14. Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statistics and Probability Letters* **54**, 381-387.

15. Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519-530.

16. Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate analysis.* Academic Press, London.

17. Reid, N. (1995). The roles of conditioning in inference. *Statistical Science*, **10**, 138-157.

18. Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics.* In press.

19. Wacholder, S. & Weinberg, C. R. (1994). Flexible maximum likelihood methods for assessing joint effects in case-control studies with complex sampling. *Biometrics* **50**, 350-357.

20. Weinberg, C. R. & Sandler, D. P. (1991). Randomized recruitment in case-control studies. *Amer. J. Epidemiology* **134**, 421-433.

21. Weinberg, C. R. & Wacholder, S. (1993). Prospective analysis of case-control data under general multiplicative-intercept risk models. *Biometrika* **80**, 461-465.

22. Zhang, B. (2000). A goodness of fit test for multiplicative-intercept risk models based on case-control data. *Statistica Sinica* **10**, 839-865.