METRON - *International Journal of Statistics*
2005, vol. LXIII, n. 2, pp. 225-261

SAMUEL KOTZ – DONATELLA VICARI

Survey of developments in the theory of continuous skewed distributions

Summary - In this paper we trace developments in the theory of skewed continuous distributions (univariate and multivariate) which commenced in the late 19th century and —after some dormant period during the most of the 20th century— were invigorated in the middle 80's of the 20th century and has become in the last 20 years an area of rapid advances.

Key Words - Gaussian distribution; Pearson curves; Thiele-Gram-Charlier expansion; Johnson transformation; Hidden truncation; Elliptical distributions.

1. From the 18th until the middle of the 20th century

Our purpose is to highlight the milestones in the development of somewhat non-homogeneous area in statistical distributions which can be traced to the 19th century, but has gained speed and attention only in the last 20 years.

We shall take this opportunity to present a number of remarks on the early history of skewed (continuous) distributions which seems to be presented in the literature in a somewhat fragmented manner.

We start with a swift historical overview of the subject matter. Unfortunately it is quite common in the statistical literature that many developments are carried by researchers in different countries more or less simultaneously without any coordination. This is especially valid for the initial studies of general continuous distributions at the very end of the 19th century. This may explain in part a somewhat disconnected presentation of this first part which is, in our intention, just an historical background of the following developments.

It is well known that so-called normal or Gaussian or Laplace-Gaussian distribution, law of error (or the "probability curve"—the term coined by Chauvenet (1863) and popularized by F. Y. Edgeworth (1845–1926) in his 1883 paper) has, for over 250 years, dominated developments in probability theory and as well as much of practical statistical work— often uncritically.

Received December 2004 and revised June 2005.

By now the bookshelves in university libraries cry for mercy under the weight of statistical textbooks based on the normal distribution assumptions. As J. Aldrich (2003) points out: The word "normal" had been used by F. Galton (1822–1911) but K. Pearson (1857–1936) made it a standard term. The books K. Pearson used as a student in 1874–1877 referred to the "probability curve" or "law of error." The later term was used by F. Bessel (1784–1865) in his 1815 paper on the positions of stars. "Normal" may be a sound abbreviation for *according to the law* (*of error*) and K. Pearson (1893) may have begun by thinking that "as a rule" the normal held but he kept the name even after he had concluded that it did not. By 1895 he was saying, "*to deal effectively with statistics* [data] *we require generalised probability curves which include the factors of skewness and range.*"

The earliest published derivation of the normal distribution (as an approximation to a binomial) is due to A. de Moivre (1667–1754) in his pamphlet of November in 1733 in Latin, which was translated by him into English five years later. The original seven-page Latin treatise was discovered by Karl Pearson in the library of University College, London, in 1924. P. S. Laplace in 1774 obtained the normal distribution as an approximation of hypergeometric distribution and four years later suggested tabulation of the normal probability integral $\Phi(x)$.

C. F. Gauss (1777–1855) in 1809 and 1816 established estimation techniques based on the normal distribution that became standard methods used during the 19^{th} and early 20^{th} centuries. More recently, starting from the middle of the 20^{th} century, a great deal of research in statistics has been devoted to testing for and applying transformations to produce approximate normality in data sets.

However, in the late 19^{th} century the increasing collection, tabulation, and publication of data by government and private institutions and agencies in demography, social sciences and medicine, biology, economics, and insurance revealed that the normal distribution is not sufficient for describing phenomena (homogeneous with respect to all but random factors) in real world situations.

A. Quetelet (1796–1874) was one of the first to encounter such data (Quetelet, 1846). In 1879, the lognormal distribution was introduced by McAlister at F. Galton's instigation. J. Venn in 1887 criticized the unquested application of the normal distribution and presented examples of meteorological measurements (pressure and temperature) that seem to show marked departures from normality. This criticism was deflected by F. Y. Edgeworth in a letter to *Nature*. However, J. Venn's criticism touched a raw nerve, and in his first statistical publication in 1893 (also a letter to *Nature*), Karl Pearson mentions Venn's objections and announces that he obtained a generalization of "probability curve" (*i.e.*, the normal distribution) using a method of "higher moments." The full account appeared in 1894 and 1895 in two voluminous memoirs in the *Philosophical Transactions of the Royal Society* where Pearson develops:

(a) a mixture of two normal curves;
(b) derived a "generalized form of the normal curve of an asymmetrical character", currently referred to as Pearson's type III curve (closely related to the Gamma distribution).

This curve was earlier derived and presented by the American scholar E. De Forest in 1882–1883 (and K. Pearson acknowledged this in his letter in *Nature* in August 1895). In his monumental work, K. Pearson generalized the geometrical relation between binomial polygon and the normal curve of frequency, *i.e.*, generalizing the differential equation

$$\frac{1}{y}\,\frac{dy}{dx} = -\,\frac{x}{c_1}$$

to

$$\frac{1}{y}\,\frac{dy}{dx} = \frac{-x}{c_1 + c_2 x}$$

and later on provided analogous analysis of the hypergeometric distribution leading to the famous differential equation

$$\frac{1}{y}\,\frac{dy}{dx} = \frac{-x}{c_1 + c_2 x + c_3 x^2}$$

for the *Pearson curves*. The 1894 and 1895 memoirs which contained 30 pages of examples received worldwide attention and these curves are still being extensively used after 110 years since their "inauguration". Johnson *et al.* (1994) provide a detailed account of the Pearson's curves. The type of a curve depends on the nature of the roots of the quadratic in the denominator which can be expressed in terms of moments. Pearson (1895) started with five types of curves (treated a frequency curve) as an object with mechanical properties. The number of types grew to seven, and finally to twelve.

Independently and somewhat earlier during the years 1873–1897, Danish astronomer, actuary, mathematician and statistician (a person of remarkable erudition and profound mental abilities), T. N. Thiele (1838–1910) who was unintentionally neglected by the leaders of the British statistical establishment at the turn of the 19^{th} century, introduced three new frequency functions that were generalizations of the normal and binomial distributions. He needed a skew distribution for the demographic and actuarial data from a life insurance company where he worked as an actuary. This led him to the form

$$(a_0 + a_1 x + a_2 x^2 + \cdots)\, b^x e^{-(x-m)^2/2k}$$

(but he used this model only for $b = 1$).

This is a linear combination of a normal distribution and its successive derivatives known as Gram-Charlier series. Thiele's approach of generating skew distributions is radically different from Pearson's method of inventing a four-parameter system of continuous distributions, among which the normal distribution is a special (two-parameter) case. Thiele discusses a partial sign of the Gram-Charlier series (the *A series* in A. Hald's (1981) terminology) with orthogonal terms and a series defined as a linear combination of symmetric binomial distributions and its successive differences (the *B series*). Thiele's idea of writing a *frequency* function as a linear combination of known functions allowed to bring the analysis of frequency functions within the realm of the method of least squares. Gram (1850–1916) also added the requirement of orthogonality. This resulted in the emergence of the so-called Thiele-Gram method of series expansion and estimation of frequency functions.

Thiele's C series (1897 and 1903) proposes to use polynomial interpolation on the logarithm of a given density $g(x)$, leading to

$$\ln g(x) = \sum_{r=0}^{2m} c_r x^r, \quad c_{2m} < 0, \;\; m = 1, 2, \dots, \quad -\infty < x < \infty.$$

For $m = 1$, we arrive at the normal distribution.

On the European continent, Thiele's ideas were "in the air" as it is witnessed by independent, related investigations of Gram-Charlier series by Bruns (1897, 1906), Hausdorff (1901) and Charlier (1906). Bruns tables of the first six derivatives of the normal density were published in Czuber's (1903) classical text on probability theory.

In his book, posthumously published in 1897, *Kollectivemasslehre*, the founder of psychophysics, G. T. Fechner (1801–1887) proposed a two-sided Gaussian law which may have different scales for positive and negative deviations, being a composition of two normal distributions with different standard deviations and common mode.

In 1908 at the 4th International Mathematical Congress in Rome, an Italian mathematician F. de Helguero presented a paper on the analytical representation of the "abnormal" distribution curves which seems to be one of the earliest (if not the earliest) predecessors in the spirit, along the path of the univariate skew-normal distribution that was explicitly introduced only in the second half of the 20th century (see Section 2). De Helguero (1909) noted that the theoretical curves studied by Pearson and Edgeworth are lacking in considering that the causes of the variation are interdependent, but they assume nothing about the law of dependence. He suggested that could be very helpful to consider distributions including a variability due to external causes. De Helguero considered two forms of departure from normality: the first one, deals with the mixture of two normal populations and was motivated by a paper by A. Giard (1894), who firstly formulated a similar hypothesis in biological context. The second formulation refers to a selection mechanism of a normal population. Starting from the normal density, both a constant measuring the effect of the perturbation cause and the probability of an individual being hit by such a perturbation are included. That it could be the first occurrence in the literature of distribution generated by selective sampling which leads to a unimodal and asymmetric curve. Some 100 years later, G. Barnard (1915–2002) defined the Fechner family of distributions which was investigated by K. J. Thomas (1993). A Dutch scientist, J. C. Kapteyn, in 1903 published an influential book, *Skew Frequencies in Biology and Statistics*, in which he considered the genesis of the lognormal distribution along the lines of McAlister (1879). This distribution is also mentioned in G. T. Fechner's (1897) book on *Kollectivemasslehre*.

The early history of bivariate (and multivariate) skewed continuous distributions is associated with Karl Pearson's *Notes on Skew Frequency Surfaces* (1923). After an unsuccessful attempt to replace a pair of correlated variables by a pair of independent ones, in his 1905 work mentioned above, K. Pearson considered methods of construction of joint distributions starting from specific requirements on the regression and scedastic functions. Attempts by the British researchers Filon, Rhodes, and Isserlis, who were associated at one time or another with K. Pearson's laboratory, to generalize Pearson's curves to skewed frequency surfaces via double hypergeometric series or by rotation of axis were not fully successful, although Filon-Isserlis and Rhodes surfaces provided impetus for a more general derivation by S. Narumi (a Japanese mathematician) which was hailed by K. Pearson and published in *Biometrika* 1923. Rhodes' work has appeared in *Biometrika* (also in 1923) but was criticized in the *Current Notes of the J.R.S.S.* (1923), Vol. 86, pp. 460–461 [May 1923]. Filon's and Isserlis' results were somewhat neglected and only briefly and critically discussed by K. Pearson in his *Notes on Skew Frequency Surfaces* (1923). Narumi (1923) imposes somewhat stronger requirements on the model namely that the shape of each conditional (array) distribution of one variable, given the others, should be the same for all values of the conditioning variables and places restrictions on the *median* regression function. This allows him to construct some specific distributions of the form

$$Z = \frac{z_0 \left(1 + \frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)^{-\lambda-2}}{2\lambda(1-r^2)},$$

while Rhodes surfaces (which are strictly speaking not a generalization of Pearson's skew frequency curves) are of the type

$$e^{\ell x + my}(1 + ax + by)^p (1 + cs + dy)^q .$$

Attempts to generalize the univariate Pearson's differential equation

$$\frac{1}{p}\ \frac{dp}{dx} = -\ \frac{a+x}{c_0 + c_1 x + c_2 x^2}$$

to bivariate and multivariate cases were carried out by very diligent Dutch researcher, M. J. van Uven for over the 20-year period (from 1925 to 1948) and by H. S. Steyn (1960), and more recently by French scholar R. Risser (1945–1950) and S. N. Sagrista (1952).

Bivariate Pearson surfaces are presented in Elderton and Johnson (1969). They all have *linear* regression of each variable on the other.

1.1. Transformed skewed distributions

It is handy to consider constructing skewed univariate distributions by assuming that a *transformation* of the random variable at hand has a standard normal distribution.

This is known as the *method of translation*. Over 100 years ago, F. Y. Edgeworth (1896) investigated transformations represented by polynomials while Wicksell (1917, 1923) and Rietz (1922) considered more general transformations including logarithmic. However, the curves based on these transformations cover a rather limited variety of shapes as compared to the Pearson curves.

In 1949, N. L. Johnson (1917-2004) applied the method of translation to generate three families of frequency distributions which assume a wide variety of shapes and cover the feasible (β_1, β_2) plane, *i.e.*, all ordered pairs (β_1, β_2) which are below the limiting line $\beta_2 - \beta_1 - 1$.

The transformations are of the form:

$$z = \nu + \delta f(y; \gamma, \sigma)\,, \tag{1}$$

where z is the standard normal variable and $f(y; \gamma, \sigma)$ is a simple monotone function. The four parameters are ν, δ, γ, and σ.

The specific families of Johnson distributions are:

1. The S_L generated by $f(y; \gamma, \sigma) = \ln[(y-\gamma)/\sigma]$.
2. The S_B generated by $f(y; \gamma, \sigma) = \ln[(y-\gamma)/(\sigma + \gamma - y)]$.
3. The S_U family generated by $f(y; \gamma, \sigma) = \sinh^{-1}[(y-\gamma)/\sigma]$.

More recently Rieck and Nedelman (1991) proposed two additional distributions: non-central sinh-normal distributions $(SN(\alpha, \gamma, \sigma, \nu))$ defined by the transformation

$$z = \nu + \left(\frac{2}{\alpha}\right) \sinh\left[\frac{y-\gamma}{\sigma}\right]\,, \tag{2}$$

where z is a standard normal variable, and the central sinh-normal distribution $(SN(\alpha, \gamma, \sigma))$ defined by setting $\nu = 0$ in (2).

The non-central distribution admits the c.d.f. of the form

$$F(y) = \Phi\left\{\nu - \left(\frac{2}{\alpha}\right)\sinh\left[\frac{y-\gamma}{\sigma}\right]\right\}; \tag{3}$$

here γ and σ are location and scale parameters respectively, and α and ν affect kurtosis and skewness. The density is

$$f(y) = \left[\frac{2}{\alpha\sigma\sqrt{2\pi}}\right]\cosh\left[\frac{y-\gamma}{\sigma}\right]\exp\left\{-.5\left[\nu + \frac{2}{\alpha}\sinh\left[\frac{y-\gamma}{\sigma}\right]\right]^2\right\}. \tag{4}$$

The parameters α and σ are positive while ν and γ are unrestricted.

The central distribution ($\nu = 0$) is symmetric but can be bimodal. The majority of (β_1, β_2) points of $SN(\alpha, \gamma, \sigma, \nu)$ distribution for α in the range of 0.2 to 10 and ν from 0 to 10 fall in the same region as the Johnson's S_B family, although there are some points in the S_L and S_U regions as well. The logarithm of a well-known Birnbaum-Saunders fatigue-life distribution is in the family of $SN(\alpha, \gamma, \sigma)$.

As in the case of univariate distributions, bivariate translation systems are generated by supposing that certain functions of variables are normally distributed.

The pioneering work along these lines has been carried out by N. L. Johnson (1949) and the bivariate joint-skewed distributions found recent applications in describing the structure of tree heights and diameters, Schreuder and Hafley (1977), Knoebel and Burkhart (1991), among other applications. Some details are given in Kotz *et al.* (2000).

2. Univariate skew-normal (and skew non-normal) distributions. The modern era

The idea of modelling skewness by constructing a mathematically tractable family of distributions starting from the symmetric normal distribution *by modifying it* in the context of Bayesian analysis can be perhaps traced to Birnbaum (1950) and, independently, much later to O'Hagan and Leonard (1976). Birnbaum (1950) suggested to apply what is now known as "conditioning method" (discussed below). Two short notes by M. A. Weinstein (1964) and by M. Lipow, N. Mantel, and J. W. Wilkinson, as reported by the editor L. S. Nelson (1964), deal with analogous problem catched in slightly different language. The paper by C. Roberts (1966) where a correlation model "useful in the study of twins" was developed by taking maxima of normal variables resulted in an equivalent representation. Aigner *et al.* (1977) tackled the same problem using the “transformation method” (discussed below) involving two normal variables with applications in econometrics.

Much later Mukhopadhyah and Vidakovic (1995) provided further examination of this problem with application to Bayesian analysis for constructing skewed prior classes. These authors also suggested generalizations to be mentioned below.

However, the initiator and the driving force behind the development of the modern theory and applications of skewed-normal distributions undoubtedly was A. Azzalini (and his students and associates). It starts with the ground-breaking paper in the 1985 issue of the *Scandinavian Journal of Statistics*, 12, 171–178, which is probably the most quoted paper in the field of skewed distributions. (Some ten years later, Azzalini and Dalla Valle (1996), extended the original result to the multivariate cases which also generated widespread attention).

An important contribution by B. C. Arnold *et al.* (1993) provided applications and further amplifications and interpretations (followed by a number of papers by Arnold and Beaver (2000, 2002) exploring the multivariate case). More recently, M. G. Genton and his coworkers initiated further investigations in the multivariate case.

The basic definition proposed by Azzalini (1985) states that a random variable Z (sometimes denoted Z_λ) has skew-normal (SN) distribution with asymmetric parameter λ (*i.e.*, $Z_\lambda \sim SN(\lambda)$) if the density of Z_λ is given by

$$f_{Z_\lambda}(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R},\ \lambda \in \mathbb{R}, \tag{5}$$

where ϕ and Φ are the $N(0, 1)$ p.d.f. and c.d.f., respectively. Azzalini (1985) and Azzalini and Dalla Valle (1996) provide expressions for the mean and variance of a standardized skew-normal random variable Z_λ with density (5):

$$E(Z_\lambda) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{\lambda}{(1+\lambda^2)^{\frac{1}{2}}},$$

$$\mathrm{Var}(Z_\lambda) = 1 - \frac{2}{\pi}\ \frac{\lambda^2}{(1+\lambda)}.$$

Henze (1986) cites expressions for all odd moments in a closed form. The even moments (as it is evident from $Z_\lambda^2 \sim \chi_1^2$ and density (5)) are the same as those of $N(0, 1)$ variable. Henze (1986) and simultaneously Azzalini (1986), building on Azzalini (1985), provide the stochastic representation of Z_λ:

$$Z_\lambda \stackrel{d}{=} \frac{\lambda}{\sqrt{1+\lambda^2}}\,|U| + \frac{1}{\sqrt{1+\lambda^2}}\,V, \tag{6}$$

where U and V are independent $N(0, 1)$ random variables. The notation $X \stackrel{d}{=} Y$ means that X and Y have the same distributions. It is easy to verify that the r.h.s. of (5) is a proper density and the case $\lambda = 0$ results in $N(0, 1)$ density. It was stated by Azzalini (1985), and later emphasized by Mukhopadhyah and Vidakovic (1995), among others, that for any $\lambda \in \mathbb{R}$, the function

$$2f(u)G(\lambda u)\,, \tag{7}$$

is a p.d.f. where f is the density of a variable symmetric around 0, and G is the c.d.f. of another independent random variable. By combining different symmetric distributions (t, logistic, uniform, double exponential (Laplace), etc.) numerous families of skewed distributions may be generated (Nadarajah and Kotz (2003, 2004)).

Specifically following (5), if the densities ψ_0 and ψ_1 are, for example, standard Cauchy densities, the skew-Cauchy density (of "the Azzalini" type) is given by

$$f(z, \lambda) = \frac{2}{\pi(1+z^2)} \left[\frac{1}{2} + \frac{1}{\pi}\tan^{-1}(\lambda z)\right], \quad z \in \mathbb{R}\,. \tag{8}$$

Alternatively, defining

$$Z_\lambda = \delta|Y_0| + \sqrt{1-\delta^2}\,Y_1, \qquad \left(\lambda = \frac{\delta}{1-\delta^2}\right), \tag{9}$$

see (6), where Y_0 and Y_1 has symmetric density ψ we obtain

$$f_{Z_\lambda}(z) = \frac{2}{\delta\sqrt{1-\delta^2}} \int_0^\infty \psi\left(\frac{z-u}{\sqrt{1-\delta^2}}\right)\psi\left(\frac{u}{\delta}\right) du\,, \tag{10}$$

and substituting for ψ_0 and ψ_1 Cauchy (0, 1), we have for $\delta > 0$,

$$f_Z(z) = \frac{2}{\delta\sqrt{1-\delta^2}} \int_0^\infty \frac{1}{\pi\left(1+\frac{(z-u)^2}{1-\delta^2}\right)} \frac{1}{\pi\left(1+\frac{u^2}{\delta^2}\right)}\, du\,. \tag{11}$$

After partial fraction expansion of the integrand, we obtain the closed form:

$$\begin{aligned} f_Z(z) = &\frac{1}{\pi(\delta+\sqrt{1-\delta^2})} \frac{1}{\left(1+\frac{z}{(\delta+\sqrt{1-\delta^2})}\right)} \\ &\times \left[1 + \frac{2\delta}{\pi(\delta+\sqrt{1-\delta^2})}\tan^{-1}\left(\frac{z}{\sqrt{1-\delta^2}}\right)\right] \\ &- \frac{2\delta\sqrt{1-\delta^2}z}{\pi^2(z^2+(\delta+\sqrt{1-\delta^2})^2)(z^2+(\delta-\sqrt{1-\delta^2})^2)} \\ &\times \log\left(\frac{\delta^2}{z^2+1-\delta^2}\right), \quad z \in \mathbb{R}\,. \end{aligned} \tag{12}$$

Evidently, the “Azzalini”-type density (8) is simpler. See Arnold and Beaver (2002).

Analogously, assuming that ψ_0 and ψ_1 are standard Laplace densities: $\psi_0(u) = \psi_1(u) = e^{-|u|}$, $u \in \mathbb{R}$, we can arrive at two “skewed-Laplace” models. We note here that two different generating mechanisms resulted in two different forms of skewed distributions. We shall dwell on various generating approaches involving normal distributions in the next section.

Another property (known as the square root property, see, *e.g.*, Arellano-Valle *et al.*, 2002) is that

$$Z_\lambda \sim SN(\lambda) \text{ implies that } Z_\lambda^2 \sim \chi_1^2,$$

as it is the case for normal (0, 1) random variable. This property is extended by Arellano-Valle *et al.*, 2002) to the multivariate case. (See the next section).

Wang *et al.* (2004b) provides, inter alia, a chi-square characterization of Z_λ which improves on the square root property: let $g(x)$ be the probability density function of a random variable X. Then $X^2 \sim \chi_1^2$ if and only if there exists a skewing function $\pi(x)$ such that $g(x) = 2\phi(x)\pi(x)$, where $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}$.

This allows us to interpret *skewness* in terms of conditional distribution of $aX_0 + bX$ given $x_0 > 0$ where $(x_0, X)^T$ is the so-called $\mathcal{C}$-random vector. Details are presented in Arellano-Valle *et al.* (2002).

A characterization of the $SN(\lambda)$ density f given by

$$f(z, \lambda) = 2\Phi(\lambda z)\phi(z)$$

in the spirit of Arellano-Valle *et al.* (2002) and Wang *et al.* (2004) was recently developed by Gupta and Chen (2004).

Let X and Y be i.i.d. skew normal, and let F be the distribution function of two i.i.d. random variables X_1 and X_2 having all moments finite. Then F is skew normal iff X_1^2 and X_2^2 are distributed as X^2 (*i.e.* chi squared) and $(X_1 + X_2)^2$ as $(X + Y)^2$. The proof, by induction, is rather complicated. It is based on a result saying basically that distributions having all moments are determined by the distributions of X_1^2, X_2^2 and $(X_1 + X_2)^2$.

Ma and Genton (2004) propose a flexible class of skew-symmetric distributions with the probability density function of the form of a product of a symmetric density (non necessarily normal or Laplace) and a skewing function. They illustrate their approach by examples for the fiberglass data and the well-known Swiss bills data.

Ali and Woo (2005) discuss skew reflected Gamma distribution, skew-symmetric double Weibull distribution and skew-symmetric beta prime distribution and derive their moments. For example, in the Gamma case, the reflected variable has the p.d.f.

$$f(x) = \frac{1}{2\Gamma(\alpha)}|x|^{\alpha-1}e^{-|x|}, \quad x \in \mathbb{R}, \quad \alpha > 0$$

(note that $f(x) = f(-x)$ for all $x \in \mathbb{R}$) and the c.d.f.

$$F(x) = \frac{1}{2} + \frac{\operatorname{sgn}(x)}{2\Gamma(\alpha)}\gamma(\alpha, |x|), \quad x \in \mathbb{R}$$

where $\gamma(\alpha, x)$ is the incomplete Gamma function

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1}e^{-t}dt\,.$$

The skew reflected Gamma distribution has the density $2f(x)F(cx)$ for any real c.

2.1. Estimation of univariate skew normal distributions

It was already mentioned that the standard skew-normal distribution with parameter λ, $SN(\lambda)$, represented by r.v. X, can be generalized to

$$Y = \xi + \eta X\,.$$

This is called *direct* parametrization and distribution of Y is denoted by $SN_D(\xi, \eta, \lambda)$. Consider a *standardized* sample $Y_S = (y_{0_1}, \dots, y_{S_n})$ where $Y_{s_i} = (y_i - \bar{y})/s$ and s^2 is the sample variance) of n observations from a $SN_D(\xi, \eta, \lambda)$ distribution.

The method of moments yields the following estimators of the parameters:

$$\tilde{\xi}_s = -cm_3^{\frac{1}{3}}/s, \quad \tilde{\eta}_s = (1 + \tilde{\xi}_s^2)^{\frac{1}{2}} \text{ and } \tilde{\delta} = -\frac{\tilde{\xi}_s}{b\tilde{\eta}_s}\,, \tag{13}$$

where $b = \sqrt{\frac{2}{\pi}}$, m_3 is the third central moment,

$$c = \left(\frac{2}{4-\pi}\right)^{\frac{1}{3}}, \text{ and } \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}\,. \tag{14}$$

(Note that $\tilde{\lambda} = \tilde{\delta}/(1 - \tilde{\delta}^2)^{\frac{1}{2}}$, provided $|\tilde{\delta}| < 1$).

The method of moments estimators of the location and scale parameters for a normal distribution are $\bar{y}$ and s. Thus, if we use (13) to estimate the parameters of an *assumed skew-normal distribution* when data actually comes from a normal distribution, we shall overestimate η and, depending on the sign of $\tilde{\xi}_s$, underestimate (or overestimate) ξ. Azzalini (1985) introduces the "centered" parametrization (μ, σ, γ_1) and defines skew-normal variable Y with $E(Y) = \mu$ and $\text{Var}(Y) = \sigma^2$ by

$$Y = \mu + \sigma \frac{X - E(X)}{(\text{Var}(X))^{\frac{1}{2}}}, \ -\infty < \mu < \infty, \tag{15}$$

where $X \sim SN(\lambda)$.

In this reparameterization the methods of moments estimates are given by $\tilde{\mu} = \bar{y}$, $\tilde{\sigma} = s$ and $\tilde{\gamma} = \frac{m_3}{s^2}$. Since the coefficient of skewness $\gamma_1 \in (-0.99527, +0.99527)$, the inadmissible values for $\tilde{\gamma}_1$ satisfy $|\tilde{\gamma}_1| > 0.99527$. Such values occur often for highly skewed populations for large sample sizes. They are an indication that *negative* folded normal distribution (rather than a skewed-normal) is the underlying generating mechanism (see Pewsey, 2000, for details).

The problem when estimating parameters of the SN model using maximum likelihood approach is that the information matrix becomes singular when the data are generated by a normal distribution (Azzalini, 1985, Pewsey, 2000). The singularity of the information matrix can be removed by the reparameterization of the model in terms of the centered parameters (μ, σ, γ). These reparameterizations made the shape of the likelihood function closer to quadratic and provides less correlated estimators.

However, this estimation procedure may lead to $\hat{\lambda} = \pm\infty$ even if the generating model has finite λ. The frequency of these *boundary estimates* decreases as n increases but is substantial even for large samples, especially if $|\lambda|$ is large. (This is due to the fact that shape of the density is almost unchanged when $|\lambda|$ is greater than 20). In the general three parameter case the Fisher information is singular as $\lambda \to 0$ (Azzalini and Capitanio, 1999, Liseo and Loperfido, 2004).

Azzalini and Capitanio (1999) propose to stop the maximization procedure when the log-likelihood value is "not substantially" lower than the maximum. (Since the estimates also depend on the significance level this proposal may be somewhat arbitrary.) A two-step procedure to obtain a finite estimate was proposed by Sartori (2003) using Firth's (1993) bias reduction method. Liseo and Loperfido (2003, 2004) develop a procedure for Bayesian analysis of the skew-normal distribution. In the standard case $SN(\lambda)$, they propose to calculate the Jeffrey's prior for λ and show that it is a proper density with tails of order $O(n^{-\frac{3}{2}})$ (as a rule non-informative priors for real parameters are improper). This allows us to use a Metropolis-Hastings type algorithm to obtain a sample from the posterior distribution of λ. The

Bayes estimate (the posterior mean of λ) seems to be superior over that of the MLE.

In the general three-parameter case, a reference prior (Berger and Bernardo, 1992) is derived. It involves the product of non-informative prior $\frac{1}{\sigma}$ for the location-scale parameter times a marginally proper density of λ. After the integrated likelihood for λ is calculated, the location and scale parameters are eliminated using a normal-Gamma type prior (of which $\frac{1}{\sigma}$ is a special case). A comparison with Azzalini and Capitanio (1999) results is then carried out.

A. C. Monti (2003) proposes a variant of the *minimum chi-square estimation* method that is asymptotically equivalent to maximum likelihood method to estimate SN parameters. Here a random sample of size n is allocated to k classes, and we are searching for the value of the parameter corresponding to the distribution that is the closest to the empirical distribution in accordance with the χ^2 criterion. This is equivalent (asymptotically in probability) to maximizing the likelihood function of a *multinomial* distribution defined on these classes. See, *e.g.*, Neyman (1949), Serfling (1980), Harris and Kanji (1983). These classes considered are:

$$C_1 = (-\infty, u_1), C_2 = (u_1, u_2), \dots, C_{n-1} = (u_{n-2}, u_{n-1}), C_n = (u_{n-1}, +\infty),$$

where $u_i = (x_{(i)} + x_{(i+1)/2})$ for $i = 1, 2, \dots, n-1$.

It turns out that means, standard deviations, and root mean square *errors* of estimators of λ are quite inflated by large estimates, while medians and median absolute deviations are unaffected by them. As it was mentioned, simulation results indicate that the original parametrization can produce unreasonably large estimates of λ. The method proposed by Monti (2003), seems to reduce substantially the presence of boundary values of estimators of γ.

Arnold and his coworkers (1993, 2000a) emphasize the concept of hidden truncation: the motivation is illustrated by the distribution of waist sizes for uniforms of elite troops who are selected only if they meet a specific minimal height requirement. The joint distribution of height and waist measurements could well be bivariate normal but imposing the height restriction will result in a positively skewed distribution for the waist sizes of the selected individuals. In some situations this kind of truncation may occur even without us having knowledge of its occurrence. An alternative expression for this procedure is "selective reporting." The latter concept is somewhat broader than hidden truncation (when, for example, each reported observation is actually the maximum of two independent, identically distributed normal observations obtained by picking the "best" one of each pair of observations). Azzalini and Dalla Valle (1996), Arnold and Beaver (2000) describe four scenarios of generating skewed normal distributions:

(a) Let Y and W be two independent i.d. standard normal variables. Define $Z = \{Y|\lambda Y > W\}$, then Z has skew-normal density: $f(z|\lambda) = 2\varphi(z)\Phi(\lambda z)$, $z \in \mathbb{R}$.

(b) (*Conditioning method*) Let (Z, Y) have bivariate normal distribution with $N(0, 1)$ marginals and correlation δ. Consider the conditional density of Z given $Y > 0$, namely keeping only those Z's whose corresponding Y value is above average. The variable Y can be expressed as:

$$Y = \delta Z - \sqrt{1-\delta^2}\, W\,, \tag{16}$$

where W is i.d. $N(0, 1)$ variable independent of Z. (This is because the event $\{Y > 0\}$ is equivalent to $\{\delta Z - \sqrt{1-\delta^2}\, W > 0\}$ or $\{\frac{\delta}{\sqrt{1-\delta^2}} Z > W\}$. Hence conditional distribution of Z given $\{Y > 0\}$ is skew normal with parameter $\lambda = \frac{\delta}{\sqrt{1-\delta^2}}$. (As observed initially by Henze (1986) mentioned above.)

(c) (*Transformation method*) Let Y_1, Y_2 be independent standard normal variables and a constant $\delta \in (-1, 1)$ be given. Then

$$Z = \delta|Y_1| + \sqrt{1-\delta^2}\, Y_2 \tag{17}$$

is skew-normal with parameter $\lambda = \frac{\delta}{\sqrt{1-\delta^2}}$. Note that (c) seems to be conceptually different from (a) and (b).

(d) Consider (W_1, W_2) to be a bivariate normal random vector with standard normal $N(0, 1)$ marginals and correlation δ. Let $Z = \max(W_1, W_2)$, then Z is skew-normal random variable with parameter $\lambda = \sqrt{(1-\delta)/(1+\delta)}$.

The basic skew normal density can be extended to

$$f(z; \lambda_0, \lambda_1) = \frac{\varphi(z)\Phi(\lambda_0 + \lambda_1 z)}{\Phi(\lambda_0/\sqrt{1+\lambda_1^2})} \tag{18}$$

(discussed in some detail in Arnold, *et al.* (1993)), and by introducing scale and location parameter

$$f(z, \lambda_0, \lambda_1, \mu, \sigma) = \frac{\varphi\left(\frac{z-\mu}{\sigma}\right)\Phi\left(\lambda_0 + \lambda_1\left(\frac{z-\mu}{\sigma}\right)\right)}{\Phi(\lambda_0/\sqrt{1+\lambda_1^2})} \text{ with } \lambda_0, \lambda_1 \in \mathbb{R}\,. \tag{19}$$

The moment generating function corresponding to a variable given by density (18) is:

$$M(t) = e^{\frac{t^2}{2}} \frac{\Phi\left(\frac{\lambda_0+\lambda_1 t}{\sqrt{1+\lambda_1^2}}\right)}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\lambda_1^2}}\right)}\,. \tag{20}$$

See Azzalini (1985) for elegant derivations (his equation (9)). With a different notation, the moment generating function of the basic skew normal density can be found in Arnold, *et al.* (1993) and in the earlier paper by Chou and Owen (1984). The basic paper by Azzalini (1985) also deals with the case $\lambda = 0$ (where the denominator becomes $\frac{1}{2}$).

For non-normal univariate skew distributions using the hidden truncation method, we consider two independent random variables W, U assuming that W has density (distribution) function $\psi_1(\Psi_1)$ and U has $\psi_2(\Psi_2)$, respectively. The conditional density of W given $\{\lambda_0 + \lambda_1 W > U\}$ is

$$f(w; \lambda_0, \lambda_1) = \frac{\psi_1(w)\Psi_2(\lambda_0 + \lambda_1 w)}{P(\lambda_0 + \lambda_1 W > U)} . \tag{21}$$

Computation of the denominator may be troublesome in certain cases. Arellano-Valle, *et al.* (2003) and Genton and Loperfido (2005), extending the results of Azzalini and his coworkers, discuss all the skew p.d.f.'s F_* which can be written in the form $f_* = 2f(z)Q(z)$ where f is a symmetric p.d.f. about zero and $0 \leq Q(z) \leq 1$ and $Q(-z) = 1 - Q(z)$, which is called a skewing function.

These authors refer to them as generalized skew-distributions. In particular, for some special skewing function, Arellano-Valle, *et al.*, obtain the following generalized skew-normal p.d.f.: $f_*(z|\eta, \tau) = 2\varphi(z)\Phi\left(\frac{\eta z}{\sqrt{1+\tau^2 z^2}}\right)$, where $z \in \mathbb{R}$ and $\eta \in \mathbb{R}$ is a location parameter, and $\tau > 0$ is a scale one. (Compare with the density given by (18)). The corresponding variable is denoted by $Z \sim GSN(\eta, \tau)$. Some properties of this model such as $Z \overset{d}{=} \frac{X}{\sqrt{1+X^2}} U + \frac{1}{\sqrt{1+X^2}} V$, where $X \sim N(\eta, \tau^2)$ and is independent of U and V which are i.i.d. $N(0, 1)$ variables, are discussed in Arellano-Valle, *et al.* (2003a,b).

The standard skew-t distribution possesses the density

$$2t(z; \nu) \cdot T\left(\lambda z \left(\frac{1+\nu}{z^2+\nu}\right)^{\frac{1}{2}}; \nu + 1\right) , \tag{22}$$

where $t(z, \nu)$ is the density of the "ordinary" t-random variable and T is the t c.d.f. with $\nu + 1$ degrees of freedom. The random variable Z with density (22) satisfies the square root property in the sense that $\tilde{Z}^2$ is distributed as F with $(1, \nu)$ degrees of freedom. Details are given in Azzalini and Capitanio (2003). This distribution (actually its log transformation) has been recently successfully applied as a model for the yearly family income data in 13 European countries and the USA in the late 20^{th} century (Azzalini, Dal Cappello, and Kotz, 2003). Jones and Faddy (2003) also discuss a skew extension of the t-distribution (emphasizing it as a heavy-tailed alternative to the scale (univariate) skew-normal class).

The linear regression model $Y_i = \alpha + \beta x_i + \epsilon_i$ relating the variables Y_i and X (assuming that x_i are directly observed but the errors ϵ_i are i.i.d. $SN(\lambda)$ and λ has an SN prior distribution has been discussed by Arellano Valle, *et al.* (2003) and a more elaborate regression models involving SN distributions by Arellano-Valle, Bolfarine and Ozán, 2003).

An investigation of properties of scaled (univariate) skew-normal distributions were carried out by Chiogna (1998). She presents recurrence relations for incomplete moments of the skew-normal variable and provides moments of order statistics. While explicit derivation of moments of normal order statistics fails for sample sizes of greater than 6 (Ruben, 1954), derivation of a single moment of order statistics: $\alpha_{i,n}^{(r)} = E(z_{i,n}^r)$ for scalar *skew*-normal distributions seems to fail already for samples of size 3 and higher. In general, it is possible to adapt relations satisfied by normal order statistics to skew-normal order statistics by introduction of correction terms which account for the skewness.

Pewsey (2000) proposed the wrapped skew-normal distribution on the circle. Let $X \sim SN(\lambda)$. The variable $\theta = Y(\mathrm{mod}\, 2\pi)$ where $Y = \xi + \eta X$ is given by the density

$$f(\theta;\xi,\eta,\lambda) = \frac{2}{\eta}\sum_{t=-\infty}^{\infty}\varphi\left(\frac{\theta+2\pi r-\xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{\theta+2\pi r-\xi}{\eta}\right)\right\}, 0 \le \theta \le 2\pi. \quad (23)$$

An application is provided by fitting the distribution to the “headings” (the direction of a bird body during flight of migrating birds) of some 1655 birds recorded near Stuttgart, Germany, during the autumnal migratory period of 1987.

Alternative approaches to generating univariate skew distribution from symmetric ones (including univariate normal distributions) have been discussed by Mudholkar and Huston (2000), among others.

Fernández and Steel (1998) propose to convert a symmetric distribution into a skewed one by introducing inverse scale factors into the negative and positive parts of the real line.

Suppose that f is a symmetric p.d.f. defined on $\mathbb{R}$. Then, for $c > 0$ and $\kappa > 0$, we define a new density function as follows:

$$g(x) = \begin{cases} cf(x\kappa), & x \ge 0, \\ cf\left(\dfrac{x}{\kappa}\right), & x < 0. \end{cases} \quad (24)$$

This distribution is skewed. To find the normalizing constant c, note that

$$\int_{-\infty}^{0} cf(x\kappa)dx + \int_{0}^{\infty} cf\left(\frac{x}{\kappa}\right)dx = 1, \quad (25)$$

and the normalizing constant c is:

$$c = \frac{2\kappa}{1+\kappa^2}. \quad (26)$$

Thus, the density $g(x)$ becomes

$$g(x) = \frac{2\kappa}{1+\kappa^2} \begin{cases} f(x\kappa), & \text{for } x \geq 0\,, \\ f\left(\dfrac{x}{\kappa}\right), & \text{for } x < 0\,. \end{cases} \tag{27}$$

Note that a more general model can be obtained replacing κ with κ_1 and $\frac{1}{\kappa}$ with κ_2. When f is the normal density, we obtain a class of skew distribution different from the one given by density (5) (see also Mudholkar and Huston, 2000 and references therein). When f is the Laplace density, we obtain a class of skew Laplace distributions (see Kotz *et al.*, 2001).

A comparison between $h(x) = 2f(x)F_x(\lambda x)$ and $g(x)$ apparently has not been carried out. Mudholkar and Huston (2000) introduce and investigate estimation in the so-called *epsilon-SN* family:

$$f(z|\epsilon) = \phi\left(\frac{2}{1+\epsilon}\right) I_{\{z<0\}} + \phi\left(\frac{z}{1-\epsilon}\right) I_{\{z\geq 0\}}, \quad z \in \mathbb{R},\ |\epsilon| < 1\,. \tag{28}$$

This family in the literature goes under different names and Mudholkar and Huston (2000) include a detailed list of early occurrences of the same distribution going back to Fechner (1897). Polynomially skewed-normal models suggested by Arnold, Castillo, and Sarabia (2002) are of the form

$$f(x) \propto f_0(x) \cdot G_0\left(\sum_{i=1}^{k} \lambda_i x^i\right)\,, \tag{29}$$

where f_0 is the basic density, G_0 is usually an unrelated c.d.f. and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$ is a k-dimensional skewness parameter.

Choosing f_0 and G_0 to be standard normal density ϕ and c.d.f. Φ, respectively, Arnold, Castillo, and Sarabia (2002) investigate the polynomial skewed-normal density of the form:

$$f(x) = \phi(x)\Phi(x(x-1)(x+1))\,. \tag{30}$$

(See Fig. 1).

As we shall observe in the following section, it is worthwhile to note that although the initial purpose of the research reviewed in this Survey was the idea of adding appropriately skewness to a continuous symmetric distribution, the more recent results are directed to wider aspects of multivariate distributions which are not totally dominated by skewness considerations.

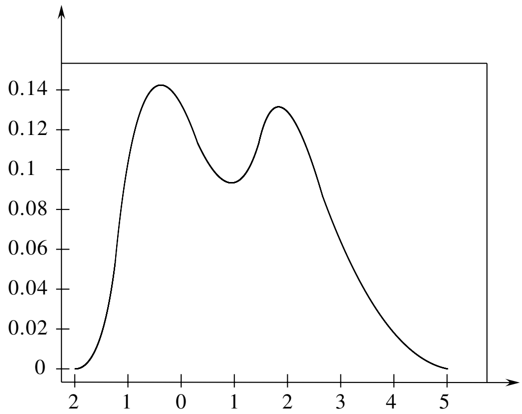

Figure 1. Graph of the function (30).

3. Multivariate skew-normal and related distributions

3.1. Skew-normal case

As we have seen —when discussing univariate skew-normal distributions $SN(\lambda)$— if $Z \sim SN(\lambda)$,

$$Z = \frac{\lambda}{\sqrt{1+\lambda^2}}\ |X| + \frac{1}{\sqrt{1+\lambda^2}}\ Y\,. \tag{31}$$

The representation (31) was used by Azzalini and Dalla Valle (1996) to extend the density of Z

$$f(z|\lambda) = 2\phi(z)\Phi(\lambda z), \quad z \in \mathbb{R},\ \lambda \in \mathbb{R}$$

(ϕ and Φ are the $N(0,1)$ p.d.f. and c.d.f., respectively) to the multivariate case given by

$$f(\mathbf{z}|\boldsymbol{\lambda}) = 2\phi_d(\mathbf{z})\Phi_d(\boldsymbol{\lambda}'\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d, \quad \boldsymbol{\lambda} \in \mathbb{R}^d\,,$$

where ϕ_d and Φ_d are the p.d.f. and c.d.f. of the d-dimensional $N^d(\mathbf{0}, \mathbf{I}_d)$ distribution. This is the standard multivariate $SN^d(\lambda)$ distribution. Note that this approach —based on transformation— leads to identical result obtained by conditioning (see Section 2.1).

Extensions for multivariate *location-scale* SN-distributions denoted SN^d $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ were also considered by Azzalini and Dalla Valle (1996) and various properties were discussed by Azzalini and Capitanio (1999). A general form of $SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ density is given by:

$$f(\mathbf{y}; \boldsymbol{\mu}; \boldsymbol{\Sigma}, \boldsymbol{\alpha}) = 2\phi_d(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\alpha}'\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}))\,, \tag{32}$$

where $\phi_d(\,\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the d-dimensional normal density with vector of expected values $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\Phi(\,\cdot\,)$ the $N(0,1)$ distribution function; $\boldsymbol{\omega}$ the diagonal matrix containing the standard deviations of $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ a d-dimensional vector which regulates the "shape" of the distribution. (Note that the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in expression (32) are not the vector of expected values and the covariance matrix of the random vector $\mathbf{Y}$ —except in the case $\boldsymbol{\alpha} = \mathbf{0}$ when the distribution reduces to $N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Azzalini and Capitanio (1999) present expressions for the mean and variance of the d-dimensional r.v. $\mathbf{Y}$ with density (32) which are:

$$\boldsymbol{\mu}_{\mathbf{Y}} = E(\mathbf{Y}) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \boldsymbol{\delta}\,,$$

where

$$\boldsymbol{\delta} = \frac{1}{(1 + \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha})^{\frac{1}{2}}}\ \boldsymbol{\Omega}\boldsymbol{\alpha}\,,$$

and

$$\mathrm{Var}(\mathbf{Y}) = \boldsymbol{\Omega} - \boldsymbol{\mu}_{\mathbf{Y}}\boldsymbol{\mu}'_{\mathbf{Z}}\,.$$

Compare with the univariate case which shows that the multivariate expressions are direct extensions of the univariate ones.

A linear transformation property can be expressed as follows: If $\mathbf{Y} \sim SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\alpha})$ and $\mathbf{A}$ is a $d \times h$ matrix of constants ($h \le d$), then $\mathbf{B} = \mathbf{A}'\mathbf{Y} \sim SN^h(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \boldsymbol{\alpha}^*)$, with

$$\boldsymbol{\mu}^* = \mathbf{A}'\mu, \quad \boldsymbol{\Sigma}^* = \mathbf{A}'\boldsymbol{\Sigma}\mathbf{A}, \quad \boldsymbol{\alpha}^* = \frac{\boldsymbol{\omega}^*(\boldsymbol{\Sigma}^*)^{-1}\mathbf{B}'\boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}'(\boldsymbol{\omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\omega}^{-1} - \mathbf{B}(\boldsymbol{\Sigma}^*)^{-1}\mathbf{B}')\boldsymbol{\alpha}}}\,,$$

where $\mathbf{B} = \boldsymbol{\omega}^{-1}\boldsymbol{\Sigma}\mathbf{A}$ and $\boldsymbol{\omega}^*$ is a diagonal matrix containing the standard deviations of the diagonal of $\boldsymbol{\Sigma}^*$. In particular, if $\mathbf{A}$ is a non-singular $d \times d$ matrix, then $\boldsymbol{\alpha}^* = \boldsymbol{\omega}^*\mathbf{A}^{-1}\boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$. Furthermore, the following result is also valid: $(\mathbf{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2_d$.

For a d-dimensional random variable with the normal distribution with zero mean and correlation matrix $\boldsymbol{\Omega}$ the region of preassigned probability p and minimum geometric measure is

$$R_N = \{\mathbf{x} : \mathbf{x}'\boldsymbol{\Omega}^{-1}\mathbf{x} \le c_p\}\,, \tag{33}$$

where c_p is the p^{th} quantile of the χ^2_d distribution.

For the d-dimensional *skew*-normal distribution $SN^d(\boldsymbol{\alpha})$ given by the density

$$f(\mathbf{x}) \equiv 2\phi_d(\mathbf{x};\boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'x)\,, \tag{34}$$

where $\phi_d(\mathbf{x};\boldsymbol{\Omega})$ denotes the $N^d(0,\boldsymbol{\Omega})$ density at $\mathbf{x}$, Φ is the $N(0,1)$ c.d.f., and $\boldsymbol{\alpha}$ is a vector of shape parameters (cfr. (32)). The region (33) is *still* a region of probability p but does not have minimum volume since it does not necessarily correspond to the set of points with maximal values of the density function.

Azzalini (2001) suggests to consider the region

$$\{\mathbf{x} :\ 2\log f(\mathbf{x}) \geq c_p - d\log(2\pi) - \log(|\boldsymbol{\Omega}|)\}\,,$$

as an appropriate region corresponding to (33) in the $SN^d(\boldsymbol{\alpha})$ case and examines its empirical performance by means of simulation experiments.

Denote $\alpha^* = \sqrt{\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}}$ (a summary measure of skewness). Then, quite an accurate approximation to an optimal region is:

$$R_{SN} \approx \{x : 2\log f(x) \geq -c_p - d\log(2\pi) - |\boldsymbol{\Omega}| + 2\log[1+\exp(-b/\alpha^*)]\}\,,$$

here $b = 1.854, 1.544, 1.498, 1.396$ when the dimension d varies from 1 to 4, respectively.

Dunajeva *et al.* (2003) derive an estimator of $\boldsymbol{\alpha}$ in (34) using the method of moments and calculate its bias.

Kollo and Traat (2001) and Gupta and Kollo (2000) slightly modify Azzalini's definition (34) by assuming $\boldsymbol{\Omega}$ to be a $d \times d$ positive definite matrix. They provide the moment-generating function

$$M(t) = 2\exp\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)\Phi\left(\frac{\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha}}{(1+\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha})^{\frac{1}{2}}}\right)$$

which allows for easy calculations of dispersion matrix, moments and cumulants. The authors assert that using their parametrization it is possible to describe distributions of estimators of the parameters in a simpler manner.

Mateu-Figueras *et al.* (1998) define the additive logistic skew-normal distributions on a simplex S^D, where $D = d+1$, d being the dimension, denoted $LS^d(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha})$. Logistic-normal theory is based on the transformation of compositional data from $\mathbf{S}^D$ to $\mathbf{R}^d$ and modelling the transformed data by means of multivariate normal distributions $N^d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ (Aitchison, 1986). The additive logistic *skew*-normal distributions $LS^d(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha})$ is represented by a D-part composition $\mathbf{X}$ where $\mathbf{Y} = \text{logratio}(\mathbf{X})$ has the regular $SN^d(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha})$ distribution. The definition is independent of the component used as denominator in the logratio transformation.

The modelling advantages of the LS^d class were illustrated by Mateu-Figueras *et al.* (1998). It turns out that data sets originating from convex linear combinations of additive logistic *normal* distributions are adequately fitted by a distribution from the LS^d class.

Azzalini and Capitanio (1999) describe cases in which quadratic forms in skew-normal variates have independent χ^2 distributions (analogously to the Cochran theorem for multivariate normal variables).

Genton *et al.* (2001) derive the moments of a random vector with multivariate $SN^d(\boldsymbol{\lambda})$ distribution and their quadratic forms. (See also a somewhat related paper by Loperfido, 2001). As in the case of univariate skew-normal distributions, there seem to be minor differences between the "Azzalini school" and "Arnold's school" in their interpretation of multivariate skew-normal distributions. Arnold and his coworkers (2000, 2002) describe three scenarios.

(1) Consider i.i.d. skew-normal variables

$$Y_1, Y_2, \dots, Y_d \text{ and } W\,.$$

Take $\mathbf{Z} = (\mathbf{Y} | \sum_{i=1}^{d} \lambda_i Y_i > W)$. Then $\mathbf{Z}$ (or its affine transformation $\mathbf{X} = \mu + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z}$) has d-dimensional skew-normal distribution.

(2) Consider $(d+1)$-dimensional normal vector $(X_1, Y_1, Y_2, \dots, Y_d)$ with zero means and $(d+1) \times (d+1)$ covariance matrix $\boldsymbol{\Lambda}$. (Note that Y_i, $i = 1, \dots, d$ are not necessarily independent.) The *hidden truncation model* is represented by the conditional distribution of $\mathbf{Y} = (Y_1, \dots, Y_d)$ *given* $\mathbf{X} > 0$ and provides a d-dimensional skew-normal distribution.

(3) Label $(d+1)$ i.i.d. $N(0, 1)$ random variables $Y_1, Y_2, \dots, Y_d, Y_0$.

The vector $\mathbf{Z}$ whose components are

$$Z_j = \delta_j |Y_0| + \sqrt{1 - \delta_j^2}\; Y_j, \quad \delta_j \in [-1, 1], \;\; j = 1, \dots, d\,,$$

has (a standard) d-dimensional skew-normal distribution.

Equation (32) is described in a slightly different notation ($\boldsymbol{\Sigma} \equiv \boldsymbol{\Omega}$, $\boldsymbol{\xi} \equiv \boldsymbol{\mu}$) in Azzalini and Dalla Valle (1996).

Numerous variants of multivariate skew-normal distribution are presented in the recent literature (see Arnold and Beaver, 2002 and Azzalini, 2003 for comprehensive references). Arnold and Beaver (2000a, 2002) define a multivariate skewed normal distribution via a linear transformation of the hidden truncation model (scenario 2) above and arrive at the multivariate normal-skewed distribution which is a generalization of the *univariate* skew density (18) (see Azzalini, 1985 and Arnold *et al.*, 1993). Arnold *et al.* (2002) calls (18) to be *linearly* skewed normal density.

The multivariate version of (18) is given by

$$f(\mathbf{Z}, \lambda_0, \boldsymbol{\lambda}_1) = \frac{\left[\prod_{i=1}^{d} \phi(z_i)\right] \Phi(\lambda_0 + \boldsymbol{\lambda}_i \mathbf{Z})}{\Phi\left(\frac{\lambda_0}{\sqrt{1+\boldsymbol{\lambda}_1' \boldsymbol{\lambda}_1}}\right)} . \tag{35}$$

A generalized class of multivariate skew-normal distributions discussed by Gupta *et al.* (2001, 2004) and González-Farías *et al.* (2004) is defined by the density

$$f(\mathbf{z}; \mathbf{D}, \Phi_m) = \frac{\phi_d(\mathbf{z})\Phi_d(\mathbf{Dz})}{\Phi_m(\mathbf{0}|\mathbf{I}_m + \mathbf{DD}')}, \quad \mathbf{z} \in \mathbb{R}^d , \tag{36}$$

where $\phi_d(\,\cdot\,|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_d(\,\cdot\,|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the p.d.f. and c.d.f. of the $N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, and $\mathbf{D}$ is a matrix of dimension $m \times d$. For special forms of $\mathbf{D}$, (36) is reduced to multivariate $SN(\boldsymbol{\lambda})$ family and to a product of univariate skew-normal distributions. Further far-reaching generalizations are given in the important paper by Arellano-Valle *et al.* (2002) based on the concepts of a C-random vector $(\mathbf{X}, \mathbf{Y})$, and a skewing function of the form Prob$(\mathbf{X} > \mathbf{0}|\mathbf{Z} = \mathbf{z})$ for some random vectors of dimensions $m \times 1$ and $d \times 1$, respectively.

A related generalization was recently given by Sahu *et al.* (2003): let $\boldsymbol{\epsilon}$ and $\mathbf{z}$ be d-dimensional *random* vectors, $\boldsymbol{\mu}$ be a d-dimensional vector and $\boldsymbol{\Sigma}$ be a $d \times d$ positive definite matrix. Define

$$\mathbf{Y} = \mathbf{DZ} + \boldsymbol{\epsilon} ,$$

where $\mathbf{D}$ is a $d \times d$ diagonal matrix. The skew normal distribution proposed by Sahu *et al.* (2003) is given by the density:

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}) = 2^d|\boldsymbol{\Sigma} + \mathbf{D}^2|^{-\frac{1}{2}}\phi_d\{(\boldsymbol{\Sigma} + \mathbf{D}^2)^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu})\}\text{Prob}(\mathbf{V} > \mathbf{0}) , \tag{37}$$

where

$$\mathbf{V} \sim N^d\{\mathbf{D}(\boldsymbol{\Sigma} + \mathbf{D}^2)^{-1}(\mathbf{y} - \boldsymbol{\mu}), \mathbf{I} - \mathbf{D}(\boldsymbol{\Sigma} + \mathbf{D}^2)^{-1}\mathbf{D}\} ,$$

and ϕ_d is the density of d-dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}$ and is denoted by $SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D})$. $SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D})$ coincides with the Azzalini multivariate skew-normal distribution only for $d = 1$. For $SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D})$ with $d > 1$, the skewness does *not* affect the correlation structure unlike the situation in the case of the Azzalini distribution. For $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \cdots, \sigma_d^2)$, (37) becomes a distribution with independent marginals

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}) = \prod_{i=1}^{d} \left\{ 2(\sigma_i^2 + \delta_i^2)^{-\frac{1}{2}}\phi\left(\frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}}\right) \Phi\left(\frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}}\right) \right\} . \tag{38}$$

For $d = 2$, compare this with the bivariate skew-normal distribution of Azzalini (1996):

$$f(\mathbf{y}|\lambda) = 2\phi(y_1)\phi(y_2)\Phi\left\{\frac{\lambda}{\sqrt{1-2\lambda^2}}(y_1+y_2)\right\}, \quad (39)$$

where λ is the skewness parameter. This is also a distribution with identical marginals but with *different* variances than the $SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D})$ for $d = 2$.

Gupta and Chen (2004) define a multivariate skew-normal distribution with the density function $f(\mathbf{y}, \boldsymbol{\Omega}, \mathbf{d}) = 2^d \phi_d(\mathbf{y}, \boldsymbol{\Omega}) \prod_{j=1}^{d} \Phi(\underline{\boldsymbol{\lambda}}_j', \mathbf{y})$ where $\phi_d(\cdot, \boldsymbol{\Omega})$ denotes the density function of a multivariate normal variable with $d \times d$ correlation matrix $\boldsymbol{\Omega}$, $\Phi(\cdot)$ is as usual the c.d.f. of the standard normal variable, $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ and the vectors $(\underline{\boldsymbol{\lambda}}_1, \dots, \underline{\boldsymbol{\lambda}}_d)$ are defined by $(\underline{\boldsymbol{\lambda}}_1, \dots, \underline{\boldsymbol{\lambda}}_d) = \boldsymbol{\Omega}^{-\frac{1}{2}} \mathrm{diag}(\mathbf{d})$. Here $\mathbf{d}$ denotes the d-dimensional vector of skewness parameters of the marginal distributions.

This distribution is then compared with the Azzalini's version and it is shown that it possesses a stochastic representation in terms of multivariate normal distributions.

The most general multivariate skew-normal distribution ($GMSN$) so far is apparently the one due to Gupta *et al.* (2004) given by the density

$$f_{p,q}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \Delta) = \frac{1}{\Phi_q(\mathbf{D}\boldsymbol{\mu}; \boldsymbol{\nu}, \Delta + \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}')} \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi_q(\mathbf{D}\mathbf{y}; \boldsymbol{\nu}, \Delta), \quad (40)$$

where $\phi_p(\,\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the p.d.f. and the c.d.f. of p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\nu} \in \mathbb{R}^q$, $\boldsymbol{\Sigma}(p \times p)$ and $\Delta(q \times q)$ are two covariance matrices and $\mathbf{D}(q \times p)$ is an arbitrary matrix. A random vector $\mathbf{Y}$ having this density is denoted by $\mathbf{Y} \sim GSN^{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \Delta)$.

This class is closed under marginalization and conditioning. Some mathematical and statistical properties of this class are quite appealing. Since the standard SN distribution can be very naturally associated to the conditioning mechanism on one observable variable, then it is a conceptually simple extension to consider q conditioning events, and the interpretation involves a selection process based on q linear constraints that the population units must satisfy. Calculations of the mean value and variance are rather cumbersome. For case $q = 1$ and $\nu = 0$ the density reduces to (32).

In the special bivariate case with $\mathbf{D} = \begin{pmatrix} \delta_1 & \delta_2 \\ 0 & 0 \end{pmatrix}$ and $\boldsymbol{\nu} = \mathbf{0}$, the density reduces to the one given by Azzalini and Dalla Valle (1996).

Gupta *et al.* (2004) summarizing the results for the bivariate case scattered in the technical reports by the same authors (which is not easy to follow) provide a slightly simplified density of the bivariate general skew normal (BGSN) distribution

$$f_2(x, y; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}) = \frac{\exp\left\{\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\left[\frac{1}{2} - \frac{1}{2\pi}\arccos(\rho_{\mathbf{D}\boldsymbol{\Sigma}})\right]} \times \Phi[\delta_{11}(x-\mu_1) + \delta_{12}(y-\mu_2)] \times \Phi[\delta_{21}(x-\mu_1) + \delta_{22}(y-\mu_2)], \tag{41}$$

where $\mu = (\mu_1, \mu_2)'$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$, and

$$\rho_{\mathbf{D}\boldsymbol{\Sigma}} = \frac{\delta_{21}\delta_{11}\sigma_1^2 + \delta_{22}\delta_{12}\sigma_2^2 + (\delta_{12}\delta_{21} + \delta_{22}\delta_{11})\sigma_1\sigma_2\rho}{\sqrt{(1+\delta_{11}^2\sigma_1^2 + 2\delta_{11}\delta_{12}\sigma_1\sigma_2\rho + \delta_{12}^2\sigma_2^2)(1+\delta_{21}^2\sigma_1^2 + 2\delta_{21}\delta_{22}\sigma_1\sigma_2\rho + \delta_{22}^2\sigma_2^2)}}.$$

The contours of the distribution (41) are not elliptical.

For $\boldsymbol{\Sigma} = \mathbf{I}$ and $\boldsymbol{\Delta} = \mathbf{I}$ (41) reduces to the density in González-Farías *et al.* (2004). Ma and Genton (2004) have recently proposed a substantial generalization of (32) which systematically captures skewness, heavy tails, and multimodality.

The literature dealing with multivariate skew-normal distributions has recently grown rapidly including different proposals and various generalizations have been proposed. In particular Arellano-Valle and Azzalini (2004) have suggested a Unified Skew-Normal model (SUN) using both a convolution and a conditioning mechanism for its genesis. This model includes as particular cases up to a parametrization the original SN family, the Closed Skew-Normal family (González-Farías *et al.*, 2004) and the Hierarchical Skew-Normal (Liseo and Loperfido, 2003).

3.2. Brief remarks on estimation in multivariate skew-normal case

Analogously to the univariate case, there are difficulties with maximum likelihood estimation of the $SN^d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ parameters due to behavior of the likelihood function and that of the information matrix at $\boldsymbol{\lambda} = \mathbf{0}$. (For small values of n also elsewhere.)

As we have seen, reparametrization is a powerful device at least for $d = 1$. In some cases the maximum likelihood estimate is attained on a boundary $\boldsymbol{\alpha} = \pm\infty$. For large n, however, the log-likelihood function usually behaves regularly. These aspects were investigated in Azzalini and Capitanio (1999) and in Pewsey (2000). Bayesian estimation is discussed in Liseo (1990). Possibly Monti's (2003) method described above can be extended to the multivariate case. Capitanio *et al.* (2003) investigate graphical models involving skew-normal variates (including multivariate).

4. Elliptical (elliptical-contoured) ECD symmetric distributions

Let $N^d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate random variable having mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Let the probability density be a function of positive definite quadratic form

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) ,$$

then the contours of the density are ellipses. Here $\mathbf{x}' = (x_1, x_2, \dots, x_d)$, $\boldsymbol{\mu}' = (\mu_1, \mu_2, \dots, \mu_d)$ and $\boldsymbol{\Sigma}$ is a nonsingular $d \times d$ scaling matrix (determined up to a multiplicative constant). In this case the distribution is called *elliptical-contoured symmetrical distribution*. The elliptical contours here are concentric with constant eccentricities. (Note that multivariate hyperbolic distribution has elliptic contours without the latter property and hence is not symmetrical.)

If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$, the distribution is called *spherical or radial*. As Bentler and Berkane (1985) emphasize: "It is becoming apparent that [elliptical] theory has the potential to displace multinormal theory in a variety of applications such as linear structural modelling."

One of the earliest works on *spherical* distributions is Hartman and Wintner (1940). A later Kelker's (1970) paper has been an important contribution to this subject.

An early paper on *elliptical symmetric* distributions is due to McGraw and Wagner (1968). Chmielewski (1981) provides an excellent review and bibliography until the late 70's of the 20th century. See also Cambanis *et al.* (1981) for a lucid and comprehensive discussion.

The modern era of research on elliptical distributions commences with the works of K. T. Fang and his co-authors (see Fang *et al.*, 1990 and the references cited in the bibliography). Gupta and Varga (1993) provided an important addition to the literature especially from theoretical aspects. These distributions model *kurtosis* inference by including light-tailed (multivariate uniform), heavy-tailed (multivariate t) and the multivariate normal. We emphasize that they are all *symmetric* in the multivariate setting and do not model skewness, thus can be applied in practical situations *only* where the symmetry seems to be appropriate. In the last 10 years numerous contributions to ECDs have been in various publications especially in the *Journal of Multivariate Analysis*. Estimation of variance components in (ECD) distribution was discussed in detail by Kubokawa (2000). Dominance results of estimating covariance and variance components in a decision-theoretic set-up for normal models remain valid for a class of ECD distributions.

5. Skew elliptical (SE) (non-normal) multivariate distributions

We shall briefly survey some selected results related to SE multivariate distribution which at present has attracted substantial attention. Extending multivariate skew-normal distribution and using conditioning method, Azzalini and Capitanio (1999) (and somewhat later Branco and Dey, 2001) generate a class multivariate *skew-elliptical* distributions utilizing the generating function of the form:

$$g^{(k)}(u) = \frac{2\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})} \int_0^\infty g^{(k+1)}(r^2 + u) r^{k-1} dr, \quad u \geq 0\,, \tag{42}$$

(see Fang *et al.*, 1990 for a discussion of generators for multivariate elliptical distributions). The density can be expressed as

$$f_Y(\mathbf{y}) = 2|\mathbf{\Sigma}|^{-\frac{1}{2}} \int_{-\infty}^{\lambda^T(\mathbf{y}-\boldsymbol{\mu})} g^{(k+1)}(r^2 + (\mathbf{y}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}))dr\,,$$

where $\mathbf{\Sigma}$ is the scale matrix associated with the vector $\mathbf{X} = (X_0, \dots, X_k)$ and $\mathbf{Y} = [\mathbf{X}|X_0 > 0]$ is the skew-elliptical vector under consideration. Numerous examples: skew-scale mixture of normal distribution, skew-logistic, skew-t, and skewed Pearson Type II distributions are provided in Branco and Dey (2001). (The original motivation is to apply these distributions in regression and calibration problems when the error distributions are skewed). Sahu *et al.* (2003) present an analogous but slightly more general definition of SE distributions also based on conditioning.

The early years of the 21st century produced a number of valuable results dealing with generalized skew elliptical distributions (which led to the volume edited by Genton in 2004 on this subject). Firstly, Genton and Loperfido provided a manuscript in 2001 in which they introduced a class of *generalized* skew-elliptical (GSE) distributions given by the density

$$f(\mathbf{z}|Q) = 2 f_k(\mathbf{z}) Q(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^k\,, \tag{43}$$

where $f_k(\,\cdot\,)$ is the density of k-dimensional elliptic-contour distribution and Q is a *skewing function* satisfying

$$Q(\mathbf{z}) \geq 0 \text{ and } Q(-\mathbf{z}) = 1 - Q(\mathbf{z})\,, \tag{44}$$

for all $\mathbf{z} \in \mathbb{R}^k$. This work is due to be published in 2005. Many of the $SN(\boldsymbol{\lambda})$ properties are also valid for distributions in the class given by (43). Presumably independently Azzalini and Capitanio (2003) consider a family of GSN distributions which generalizes Genton and Loperfido (2005). Their results are similar to Wang *et al.* originally presented in a technical report (2002) and more recently as a paper in *Statistica Sinica* (2004a).

The density (43) is extended to any symmetric density f_k in $\mathbb{R}^k$ (namely, assuming that $f_k(-\mathbf{z}) = f_k(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{R}^k$). For the class given by (43) the classical distribution theory of linear and quadratic forms, for the most part, remains valid. The skewing function Q is flexible enough to include skew-normal, skew-t, skew-Cauchy, and also other skew-elliptical forms. Here skewing functions are monotone function of their argument, but as shown by Genton and Loperfido (2005) this is not a necessary condition. Also the skewness of a multivariate skew-normal distribution is bounded (Azzalini and Dalla Valle, 1996) while GSE distributions may have unbounded skewness.

B. Q. Fang (2003) proposes and studies a family of the skew elliptic distributions which includes the skew normal distributions in Azzalini and Capitanio (1999) and Azzalini and Dalla Valle (1996) as well as Branco and Dey (2001) and some elliptical distributions. This family-like skew normal distributions are closed under marginalization, conditioning and linear transformations. We first define briefly an elliptic distribution.

Let a random vector $\tilde{\mathbf{v}} = (v_0, \mathbf{v}')$ in $\mathbb{R}^n$ have spherical distribution, where $\mathbf{v} \in \mathbb{R}^k$, $n = k + 1$. Then it has a stochastic representation (Fang *et al.*, 1990)

$$\begin{pmatrix} v_0 \\ \mathbf{v} \end{pmatrix} \stackrel{d}{=} R \begin{pmatrix} u_0 \\ \mathbf{u} \end{pmatrix} \tag{45}$$

where $R \stackrel{d}{=} ||(v_0, \mathbf{v}')'||$, $\quad (u_0, \mathbf{u}')' \stackrel{d}{=} (v_0, \mathbf{v}')'/||(v_0, \mathbf{v}')'||$ has uniform distribution on the sphere in $\mathbb{R}^n$, independent to each other, $||\cdot||$ denotes L_2 norm. The relationship of (45) is one to one in the sense that, if $(v_0, \mathbf{v}')'$ has two such representations, then the two R's must have the same distribution (Fang *et al.*, 1990, p. 38). If $(v_0, \mathbf{v}')'$ has p.d.f. $f(v_0^2 + \mathbf{v}'\mathbf{v})$ on $\mathbb{R}^n$, then R has p.d.f. $g(r)$ on $\mathbb{R}^+$ given by

$$g(r) = \frac{2\pi^{\frac{d}{2}}}{\Gamma(d/2)} r^{d-1} f(r^2), \quad r > 0,$$

(Fang *et al.*, 1990 p. 35). The distribution of $\boldsymbol{\mu} + \mathbf{A}'\tilde{\mathbf{v}}$ is by definition an elliptical distribution, which depends on $\mathbf{A}$ only through $\mathbf{A}'\mathbf{A}$.

Denote by F_1 the one-dimensional marginal distribution function of $(v_0, \mathbf{v})'$.

Let $\lambda \in \mathbb{R}$, $\boldsymbol{\alpha} \in \mathbb{R}^k$ be constants and $\boldsymbol{\Omega}$ be a $k \times k$ constant positive-definite matrix.

Let $c_0 = (1 + \boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha})^{1/2}$ by the property of the spherical distribution (Fang *et al.*, 1990),

$$\int_{-\infty}^{\lambda + \boldsymbol{\alpha}'\boldsymbol{\Omega}^{1/2}\mathbf{v}} f(v_0^2 + \mathbf{v}'\mathbf{v})\, dv_0\, d\mathbf{v} = F_1(\lambda/c_0).$$

The random vector $\mathbf{z} \in \mathbb{R}^k$ with the p.d.f.:

$$\int_{-\infty}^{\lambda+\alpha'} f(y_0^2 + (\mathbf{z} - \zeta)' \mathbf{\Omega}^{-1} (\mathbf{z} - \zeta)) dy_0 |\mathbf{\Omega}|^{-1/2} / F_1(\lambda / c_0)$$

is called the Fang (2003) skew elliptical distribution and is denoted by $\mathbf{z} \sim S_k(\psi, \mathbf{\Omega}, \lambda, \boldsymbol{\alpha}; f)$ where $\zeta \in \mathbb{R}^k$ is a constant. For $\lambda = 0$ such a distribution reduces to Branco and Dey (2001).

The invariance property of $SN^k(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda})$ distributions asserts that if

$$\mathbf{Z} \sim SN^k(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda}),$$

then the product $Z_i Z_j$ $(i, j = 1, \dots, n)$ does not depend on the skewness parameter $\boldsymbol{\lambda}$. This property can be extended to the GSE distribution, replacing the product by an even function $\tau(\mathbf{Z})$ and the skewing function $Q(\,\cdot\,)$ replaces the parameter $\boldsymbol{\lambda}$. This is useful for derivation of the distribution of quadratic forms which are also independent of $Q(\,\cdot\,)$.

Non-normal skewed multivariate distribution are defined by Arnold and Beaver (2002) via $k+1$ independent random variables $W_1, W_2, \dots, W_k$ and U with densities ψ_i, $i = 1, \dots, k, 0$, and c.d.f.'s Ψ_i, $i = 1, \dots, k, 0$, respectively. With $\lambda_0 \in \mathbb{R}$, $\boldsymbol{\lambda}_1 \in \mathbb{R}^k$, the skewed distribution is defined by

$$f_W(\boldsymbol{\omega}) = \frac{\prod_1^k \psi_i(\omega_i) \Psi_0(\lambda_0 + \lambda_i \boldsymbol{\omega})}{\text{Prob}(A)}, \tag{46}$$

where $\boldsymbol{\omega} \equiv (\omega_1, \omega_2, \dots, \omega_k)$ and $A = \{\lambda_0 + \boldsymbol{\lambda}_1' \mathbf{W} > U\}$.

In the case when $\lambda_0 = 0$ and ψ_i, $i = 1, \dots, k$, are symmetric (about 0) we have $\text{Prob}(A) = \frac{1}{2}$ and $f_{\mathbf{W}}(\boldsymbol{\omega})$ becomes $2 \left[\prod_1^k \phi_i(\omega_i)\right] \Phi(\boldsymbol{\lambda}_1' \boldsymbol{\omega})$. Variants of this density without assuming independence of W_i's are also available. Sahu *et al.* (2003) suggest multivariate t-skewed distribution $ST_\nu(\mu, \mathbf{\Sigma}, \mathbf{D})$ which even in the special case $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$, and $\mathbf{D} = \delta \mathbf{I}$ has a rather formidable expression for its density involving the density of k-variate t-distribution with parameters $\boldsymbol{\theta}$, $\mathbf{\Omega}^{-1}$ and degrees of freedom $\nu + k$ of the form:

$$\frac{\Gamma\left(\frac{\nu+k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{k}{2}}} |\mathbf{\Omega}|^{-\frac{1}{2}} \left\{1 + \frac{(\mathbf{x} - \boldsymbol{\theta})' \mathbf{\Omega}^{-1} (\mathbf{x} - \boldsymbol{\theta})}{\nu}\right\}^{-(\nu+k)/2}, \quad \mathbf{x} \in \mathbb{R}^k,$$

and also give the corresponding c.d.f.

Unlike in the multivariate skew-normal case, the k variate skew-t distribution with $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$ and $\mathbf{D} = \delta \mathbf{I}$ cannot be written as the product of univariate skew-densities. Here the components are uncorrelated but not independent.

Using the definitions and arguments above, Arellano-Valle *et al.* (2002) prove the following result (which substantially generalizes the original Azzalini, 1985 definition of skewed normal distribution).

Let $\mathbf{Z}$ be a $k \times 1$ random vector with density $f_{\mathbf{Z}}$ and let $\mathbf{X}_0$ be an $m \times 1$ random vector. Let $f_{\mathbf{Z}^*}$ be the conditional density of $\mathbf{Z}$ given $\mathbf{X}_0 > 0$. Then

$$f_{\mathbf{Z}^*}(\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z}) \frac{\text{Prob}(\mathbf{X}_0 > \mathbf{0} | \mathbf{Z} = \mathbf{z})}{\text{Prob}(\mathbf{X}_0 > \mathbf{0})} . \tag{47}$$

If $\text{sgn}(\mathbf{X}_0)$ is uniformly distributed, then

$$f_{\mathbf{Z}^*}(\mathbf{z}) = 2^m f_{\mathbf{Z}}(\mathbf{z}) \text{Prob}(\mathbf{X}_0 > \mathbf{0} | \mathbf{Z} = \mathbf{z}) . \tag{48}$$

When $\mathbf{Z}$ has a symmetric distribution, the term $\text{Prob}(\mathbf{X}_0 > \mathbf{0} | \mathbf{Z} = \mathbf{z})$ determines the degree of skewness of the distribution of $\mathbf{Z}^*$ (the case $m = 1$ includes the skew-normal distributions).

This is also a far-reaching extension of the "conditioning method" for a bivariate normal variable (X, Z) originally suggested by Birnbaum (1950) and later elaborated by Arnold *et al.* (1993). A later paper by Copas and Lee (1997) indicates that this method is closely related to selective sampling.

Arnold and Beaver (2002) study multivariate survival models when the given random variables $(X_1, X_2, \ldots, X_k, Y)$ are independent (but not necessarily identically distributed) with the densities $\psi_1, \psi_2, \ldots, \psi_k, \psi_0$ and survival functions $\Psi_1, \Psi_2, \ldots, \Psi_k, \Psi_0$.

Under the restriction that the densities $\psi_0, \psi_1, \ldots, \psi_k$ all have the positive half line as their support (viewed as representing survival times of k components in a system) and Y being a concomitant variable, the conditional distribution of $\mathbf{X}$ given $\boldsymbol{\lambda}'\mathbf{X} < Y$ (the hidden truncation model) is:

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\lambda}) \propto \left[\prod_{i=1}^{k} \psi_i(x_i) \right] \Psi_0(\boldsymbol{\lambda}'\mathbf{x}) , \tag{49}$$

and $X_i \sim \exp(\delta_i)$, $i = 1, 2, \ldots, k$ and $Y \sim \exp(\delta_0)$ ($\boldsymbol{\delta}$ represents the intensity vector).

Namely we have

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\lambda}) \propto \left[\prod_{i=1}^{k} \delta_i e^{-\delta_i x_i} \right] e^{-\delta_0 \boldsymbol{\lambda}'\mathbf{x}} I(\mathbf{x} > \mathbf{0}) , \tag{50}$$

where $I(x > 0)$ is an indicator function. Thus, this (hidden truncation) model has independent exponential $(\delta_i + \delta_0 \lambda_i)$, $i = 1, \ldots, k$, marginals which is equivalent to a scale change on $\mathbf{X}$. The two properties: (a) independent marginals, and (b) rescaling of **X**, are also valid when X_i's have gamma distributions and Y has the exponential distribution. Also, if the joint density of the X_i's is not an exponential family, then the hidden truncation model will be of the same form. (It would be of interest to dig deeper and to discover —if possible— intuitive reasons for these results. This may shed additional light on the properties and geneses of exponential and gamma distributions which has been extensively —perhaps even ravenously— studied by researchers, both theoretical and applied).

Concluding remarks

We have provided highlight points in the development of skewed continuous distributions during the last 120 years as reflected mainly in statistical periodical literature. Applications —for the most part— are not discussed. They seem to require additional invigoration. Uneven (chronologically) developments are certainly due to unexpected but welcome changes in technology related to statistical calculations. The earlier results in the late 19th century were prompted by the so-called "English Biometric School" and almost independently by its —less visible— Scandinavian counterparts. Advances in the last 20 years spearheaded by the Italian investigators have been better coordinated with the parallel research in the U.S., Canada, and more recently in South America. Indeed, it may be of interest to contrast the lack of coordination (and even antagonism) in the late 19th century and early 20th between the leaders of the English Biometric School and prominent economists and statisticians in Italy and Scandinavian countries with the cordial cooperation between the European scholars and those in the Western Hemisphere working at present in the area of skewed continuous distributions. No doubt that ease of communication is an important (but presumably not the only) factor in these encouraging developments.

The field of skewed distributions has become —in our opinion— one of the most fruitful and promising areas in the development of statistical distribution theory and applications, during the last 20 years which does not so far require using advanced mathematical tools.

The current year 2005 is the 20th Anniversary of the modern era of continuous skewed distributions initiated by Azzalini (1985). The simple idea of reallocating the probability mass of symmetric density f by defining the univariate "Azzalini skew density" g such that $g(x) + g(-x) = 2f(x)$, for any x, was the basic starting point. This lends itself quite naturally to multitude of extensions and generalizations described in this paper. Jones (2004) points out that there are two complementary approaches to introducing skewness. One tries to directly employ a skewness parameter that in some sense does not depend on the weights of the tails of distribution (Ferreira and Steel, 2004).

The second method, propagated by Azzalini among others, regards skewness as an implicit consequence of different left and right tails. Overabundance of parameters as we develop new flexible generalizations is an inevitable but unfortunate consequence of this activity (in Arnold's words: "risking a chorus of hisses").

A justification for the introduction of models involving four parameters (or more) is that these include as special cases many models which proved to be useful to model data. However, estimation of four or more parameters is usually a daunting and challenging task.

To utilize the available generalizations in practice would require substantial work on problems of parameter estimation, hypothesis testing and model fitting among others. In our opinion this activity should perhaps be a priority for further research putting, at least temporarily, further generalizations on the back burner.

We are thus confronted with two complementary possible avenues for further research (which are *not* mutually exclusive):

1) continuation of the mostly theoretical work related to further generalizations in effort to develop quite a general well structured and possibly elegant framework;
2) concentration of already available models and results in effort to tackle real-world problems using these models.

As Azzalini and Capitanio (2003) observe: success in tackling real problems is "the ultimate test to decide about the actual usefulness of all this work".

ACKNOWLEDGMENTS

To the memory of professor N. L.Johnson (1917-2004), an innovator, scholar and friend. We are thankful to Prof. A. Azzalini (Univ. of Padova) for drawing our attention to the historically important paper by De Helguero (1909).

REFERENCES

AIGNER, D. J., LOVELL, C. A. K., and SCHMIDT, P. (1977) Formulation and estimation of stochastic frontier production function model, *J. Econometrics*, 12, 21–37.

AITCHISON, J. (1986) *The Statistical Analysis of Compositional Data*, Chapman and Hall, London.

ALDRICH, J. (2003) The Language of the English Biometric School, *Inter. Stat. Review*, 71, 109–130.

ALI, M. M., and WOO, J. (2005) Skew-symmetric reflected distributions, *Soochow J. Math*, under review.

ARELLANO-VALLE, R. B., and AZZALINI, A. (2004) On the unification of families of skew-normal distributions, submitted.

ARELLANO-VALLE, R. B., BOLFARINE, H., and OZÁN, S. (2003) Skew-normal measurement error model, Technical Report, PUC-Chile/IME-USP-Chile/Brazil.

Arellano-Valle, R. B., del Pino, G., and Iglesias, P. (2003) Bayesian analysis for spherical linear models, Technical Report, PUC-Chile.

Arellano-Valle, R. B., del Pino, G., and San Martin, E. (2002) Definition and probabilistic properties of skew-distributions, *Stat. and Prob. Lett.*, 58, 111–121.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. (2003a) Statistical inference for a general class of skew distributions, Technical Report, PUC-Chile.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. (2003b) A new class of skew-normal distributions, Technical Report, PUC-Chile.

Arellano-Valle, R. B., and Iglesias, P. (2003) Bayesian inference for elliptical linear models: Conjugate analysis and model comparison (with discussion), *Bayesian Statistics* 7, 3–24. J. M. Bernardo, *et al.* (eds.). Oxford, UK; Oxford University Press.

Arnold, B. C. (2004) Discussion of M. C. Jones, Families of distributions arising from distributions of order statistics, *Test*, 13, 23–25.

Arnold, B. C., and Beaver, R. J. (2000a) Hidden truncation models, *Sankhyā*, A 62, 22–35.

Arnold, B. C., and Beaver, R. J. (2000b) Some skewed multivariate distributions, *Amer. J. Math. Mgmt. Sci.*, 20, 27–38.

Arnold, B. C., and Beaver, R. J. (2002) Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion), *Test*, 11, 7–54.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q. (1993) The non-truncated marginal of a truncated bivariate normal distribution, *Psychometrika*, 58, 471–478.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999) *Conditional Specification and Statistical Models*, New York: Springer Verlag.

Arnold, B. C., Castillo, E., and Sarabia, J. M. (2002) Conditionally specified multivariate skewed distributions, *Sankhyā*, Ser. A. 64, 1-21.

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

Azzalini, A. (1986) Further results on a class of distributions which includes the normal ones, *Statistica*, 46, 199–208.

Azzalini, A. (2001) A note on regions of given probability of the skew-normal distribution, *Metron*, 59, 27–34.

Azzalini, A. (2003) Multivariate Skew Normal Distributions, in *Encyclopedia of Statistical Sciences*, 2nd edition, New York: Wiley.

Azzalini, A., and Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distributions, *J. Roy. Stat. Soc.*, B 61, 579–602.

Azzalini, A., and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution, *J. Roy. Stat. Soc.*, B 65, 367–389.

Azzalini, A., Dal Cappello, T., and Kotz, S. (2003) Log-skew-normal and log-skew-*t* distributions as models of family income, *J. Income Distributions*, 11, 12–20.

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

Balakrishnan, N., and Ambagaspitiya, R. S. (1994) On skew-Laplace distributions, Technical Report, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada.

Barnard, G. (1989) Sophisticated theory and practive in quality improvement, *Phil. Trans. R. Soc. London*, A, 327, 581–589.

Bentler, P. M., and Berkane, M. (1985) Greatest lower bound to the elliptical kurtosis parameter, *Biometrika*, 73, 240–241.

BERGER, J. O., and BERNARDO, J. M. (1992) On the development of the reference prior method, in *Bayesian Statistics IV*, (J. M. Bernardo, Berger, J. O., Dawid, A. P. and Smith A. F. M., eds.), 35–60, Oxford University Press, London.

BIRNBAUM, Z. W. (1950) Effect of linear truncation on a multinormal population, *J. Amer. Stat. Assn.*, 21, 272–279.

BRANCO, M. D., and DEY, D. K. (2001) A general class of multivariate skew elliptical distributions, *J. Mult. Anal.*, 79, 99–113.

BRUNS, H. (1897) Über die Darstellung von Fehlergesetzen, *Astron. Nachr.*, 143, 329–340.

BRUNS, H. (1906) *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Leipzig: Teubner.

CAMBANIS, S., HUANG, S., and SIMONS, G. (1981) On the theory of elliptical contoured distributions, *J. Mult. Anal.*, 11, 368–385.

CAPITANIO, A., AZZALINI, A., and STANGHELLINI, E. (2003) Graphical models for skew-normal variates, *Scand. Journ. of Stat.*, 30, 129–144.

CHARLIER, C. V. L. (1905A) Über das Fehlergesetzes, *Ark. Mat. Astro. Fys.*, 2(8), 1–9.

CHARLIER, C. V. L. (1905B) Die zweite Form des Fehlergesetzes, *Ark. Mat. Astro. Fys.*, 2(15), 1–8.

CHARLIER, C. V. L. (1906) *Researches into the Theory of Probability*, Lund Univ. Arsskrift. N. F. Afd. 2, Bd. 2, Nr. 5, Lund.

CHAUVENET, W. (1863) *A Manual of Spherical and Practical Astronomy*, Vol. 1, Philadelphia: Lippincott.

CHIOGNA, M. (1998) Some results on the scalar skew-normal distribution, *J. Ital. Stat. Soc.*, 7, 1–13.

CHMIELEWSKI, M. (1981) Elliptically symmetric distributions: A review and bibliography, *Int. Stat. Review*, 49, 67–74.

CHOU, Y. M., and OWEN, D. B. (1984) An approximation to the percentiles of a variable of the bivariate normal distribution when the other variable is truncated, with applications, *Comm. Stat.*, A 13, 2535–2547.

COPAS, J. B., and LI, H.G. (1997) Inference for non-random samples, *J. Roy. Stat. Soc.*, B 59, 55–95.

CZUBER, E. (1903) *Wahrscheinlichkeitsrechnung, I*, Leipzig: Teubner.

DE FOREST, E. (1882–83) On an unsymmetrical probability curve, *Analyst*, 9, 135–143; 10, 67–74.

DE HELGUERO, F. (1909) Sulla rappresentazione analitica delle statistiche abnormali, in *Atti del IV Congresso Internazionale dei matematici* (Roma, 6-11 Aprile 1908), (G. Castelnuovo, ed.), vol. III, 288-299, Roma, Accademia dei Lincei.

DE MOIVRE, A. (1733) Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem Expansi, *Supplementum II to Miscellanae Analytica*, 1–7.

DUNAJEVA, O., KOLLO, T., and TRAAT, I. (2003) Bias correction for the shape parameter of the skew normal distribution, *Tatra Mt. Math. Publ.*, 26, 281–289.

EDGEWORTH, F. Y. (1883) The law of error, *Philos. Mag.* 5^{th} *Series*, 14, 300–309.

EDGEWORTH, F. Y. (1896) The compound law of error, *Philos. Mag.* 5^{th} *Series*, 41, 207–215.

EDGEWORTH, F. Y. (1917) On the mathematical representation of statistical data, *J. Royal Stat. Soc. Series A*, 80, 266–288.

ELDERTON, W. P., and JOHNSON, N. L. (1969) *Systems of Frequency Curves*, London: Cambridge University Press.

FANG, B. Q. (2003) The skew elliptical distributions and their quadratic forms, *J. Multiv. Anal.*, 87, 298–314.

FANG, K. T., KOTZ, S., and NG, K. W. (1990) *Symmetric Multivariate and Related Distributions*, London: Chapman and Hall.

FECHNER, G. T. (1897) *Kollektivmasslehre*, (G. F. Lipps, ed.), Engelmann, Leipzig.

FERNÁNDEZ, C., and STEEL, M. F. J. (1998) On Bayesian modelling of fat tails and skewness, *J. Amer. Stat. Assn.*, 93, 359–371.

FERREIRA, J. T. A. S., and STEEL, M. F. J. (2004) Discussion of M. C. Jones, Families of Distributions Arising from Distributions of Order Statistics, *Test*, 13, 30-33.

FIRTH, D. (1993) Bias reduction of maximum likelihood estimates, *Biometrika*, 80, 27–38.

GAUSS, C. F. (1809) *Theoria Motus Corporum Coelestium*, Hamburg: Perthes and Besser.

GAUSS, C. F. (1816) Bestimmung der Genauigkeit der Beobachtungen, *Z. Astronom. Verwandte Wiss.*, 1, 185–196.

GENTON, M. G. (2005) *Skew-elliptical distributions and their applications: a journey beyond normality*, Edited Volume. Chapman and Hall/CRC, Boca Raton, FL.

GENTON, M. G., HE, L., and LIU, X. (2001) Moments of skew normal random vectors and their quadratic forms, *Stat. Prob. Lett.*, 51, 319–325.

GENTON, M. G., and LOPERFIDO, N. (2005) Generalized skew-elliptical distributions and their quadratic forms, *Ann. Inst. Stat. Math.*, to appear.

GIARD, A. (1894) Sur certains cas de dedoublement des courbes de Galton dus au parasitisme et sur le dimorphisme d'origine parasitaire, *Comp. Rend. Acad. Sci. Paris*, 108.

GONZÁLEZ-FARÍAS, G., DOMINGUEZ-MOLINA, A., and GUPTA, A. K. (2004) Additive properties of skew normal random vector, *J. Stat. Plann. Inf.*, 126, 521–534.

GRAM, J. P. (1879) *Om Rakkeudviklinger, bestemte ved Hjalp af de mindeste. Kvadraters Methode*, [On Series Expansions Determined by Means of the Method of Least Squares], Host, Kjobenhavn.

GRAM, J. P. (1883) Püber die Entwickelung reeller Funktionen in Reihen mittelst der Methode der kleinsten Quadrate, *J. Reine Angew. Math.*, 94, 41–73.

GUPTA, A. K., and CHEN, J. T. (2004) A class of multivariate skew-normal models, *Ann. Inst. Stat. Math.*, 56, 305–315.

GUPTA, A. K., GONZÁLEZ-FARIÁS, G., and DOMÍNGUEZ-MOLINA, J. A. (2001) A Multivariate Skew Normal Distribution, Dept. of Mathematics and Statistics, Bowling Green State University, Tech. Rep. No. 01–10.

GUPTA, A. K., GONZÁLEZ-FARIÁS, G., and DOMÍNGUEZ-MOLINA, J. A. (2004) A multivariate skew normal distribution, *J. Mult. Anal.*, 89, 181–190.

GUPTA, A. K., and KOLLO, T. (2000) Multivariate skew normal distribution: some properties and density expansions, Dept. of Mathematics and Statistics, Bowling Green State University, Tech. Rep. No. 00–05.

GUPTA, A. K., NGUYEN, T. T., and SANQUI, J. A. T. (2004). Characterization of the skew-normal distribution, *Ann. Inst. Stat. Math.*, 56, 351–360.

GUPTA, A. K., and VARGA, T. (1993) *Elliptically Contoured Models in Statistics*, Dordrecht: Kluwer Academic.

HALD, A. (1981) T. N. Thiele's contributions to statistics, *Inter. Stat. Rev.*, 49, 1–20.

HARRIS, R. R., and KANJI, G. K. (1983) On the use of minimum chi-square estimation, *The Statistician*, 32, 379–394.

HARTMAN, P., and WINTNER, A. (1940) On the spherical approach to the normal distribution law, *Amer. J. Math.*, 62, 759–779.

HENZE, N. (1986) A probabilistic representation of the skew normal distribution, *Scand. J. Stat.*, 13, 271–275.

JOHNSON, N. L. (1949) Systems of frequency curves generated by methods of translation, *Biometrika*, 36, 149–176.

JOHNSON, N. L., KOTZ, S., and BALAKRISHNAN, N. (1994) *Continuous Univariate Distribution, 1*, Second Edition, New York: Wiley.

JONES, M. C. (2004) Families of Distributions Arising from Distributions of Order Statistics (with discussion), *Test*, 13, 1–43.

JONES, M. C., and FADDY, M. J. (2003) A skew extension of the t-distribution, with applications, *J. Roy. Stat. Soc.*, B 65, 159–174.

KAPTEYN, J. C. (1903) *Skew Frequency Curves in Biology and Statistics*, Groningen: P. Noordhoff.

KELKER, D. (1970) Distribution theory of spherical distributions and location-scale parameter, *Sankhyā A*, 32, 319–430.

KNOEBEL, B. R., and BURKHART, H. E. (1991) A bivariate distribution approach to modeling forest diameter distributions at two points in time, *Biometrics*, 47, 241–253.

KOLLO, T., and TRAAT, I. (2001) On the Multivariate skew-normal distribution, in Revista de Estatistica, *Proceedings of the 23rd European Meeting of Statisticians*, vol. II, 231–232, Instituto Nacional de Estatistica, Portugal.

KOTZ, S., JOHNSON, N. L.,and BALAKRISHNAN, N. (2000) *Continuous Multivariate Distributions*, Second Edition, New York: Wiley.

KOTZ, S., KOZUBOWSKI, J., and PODGÓRSKI, K. (2001) *The Laplace Distribution and Generalizations*, Boston, USA: Birkhäuser.

KUBOKAWA, T. (2000) Estimation of variance and covariance components in elliptically contoured distribution, *J. Japanese Stat. Soc.*, 30, 143–176.

LISEO, B. (1990) La classe delle densità normali sghembe: aspetti inferenziali da un punto di vista bayesiano, *Statistica*, 50, 71–82.

LISEO, B., and LOPERFIDO, N. (2003) A Bayesian interpretation of the multivariate skew-normal distribution, *Stat. Prob. Lett.*, 61, 395–401.

LISEO, B., and LOPERFIDO, N. (2004) Default Bayesian analysis of the skew normal distribution, *J. Stat. Plann. Inf.*, to appear.

LOPERFIDO, N. (2001) Quadratic forms of skew-normal random vectors, *Stat. Prob. Lett.*, 54, 381–387.

MA, Y., and GENTON, M. G. (2004) Flexible class of skew-symmetric distributions, *Scand. J. Statist.*, 31, 459–468.

MATEU-FIGUERAS, G., BARCELO-VIDAL, C., and PAWLOWSKY-GLAHN, V. (1998) Modelling compositional data with multivariate skew-normal distributions, *Proceedings of the IAMG'98.* The Fourth Annual Conference of the International Association for Mathematical Geology, (A. Buccianti, G. Nardi and R. Potenza, eds.) vol. II, 532–537, Napoli, De Frede Editore.

MCALISTER, D. (1879) The law of the geometric mean, *Proc. R. Soc. London*, 29, 367-376.

MCGRAW, D. K., and WAGNER, J. F. (1968) Elliptically symmetric distributions, *IEEE Trans. Inf. Theory*, IT14, 110-120.

MONTI, A. C. (2003) A note on the estimation of the skew normal and the skew exponential power distributions, *Metron*, LXI, 2, 205–210.

MUDHOLKAR, G. S., and HUTSON, A. D. (2000) The epsilon-skew-normal distribution for analyzing near-normal data, *J. Stat. Plann. Inf.*, 83, 291–309.

MUKHOPADHYAH, S., and VIDAKOVIC, B. (1995) Efficiency of linear Bayes rules for a normal mean: skewed prior class, *The Statistician*, 44, 389–397.

NADARAJAH, S., and KOTZ, S. (2003) Skewed distributions generated by the normal kernel, *Stat. Prob. Lett.*, 65, 269–277.

NADARAJAH, S., and KOTZ, S. (2004) Skewed distributions generated by the Laplace kernel, *Amer. J. Math. Manag. Sci.*, to appear.

NARUMI, S. (1923) On the general forms of bivariate frequency distributions which are mathematically possible when regression and variation are subject to limiting conditions, *Biometrika*, 15, 77–88, 201–221.

NELSON, L. S. (1964) The sum of values from a normal and a truncated normal distribution, *Technometrics*, 6, 469–470.

NEYMAN, J. (1949) Contributions to the theory of the z test, *First Berkley Symposium on Math. Stat. and Prob.*, 239–273.

O'HAGAN, A., and LEONARD, T. (1976) Bayes estimation subject to uncertainty about parameter constraints, *Biometrika*, 63, 201–203.

PEARSON, K. (1893) Asymmetrical frequency curves, *Nature*, 8, 615–616, correction: *ibid.*, 49, 6.

PEARSON, K. (1895) Contributions to the mathematical theory of evolution, II, *Phil. Trans. R. Soc. London, A*, 185, 71–110.

PEARSON, K. (1923) Notes on skew frequency surfaces, *Biometrika*, 15, 222–230.

PEARSON, K. (1923) On non-skew frequency surfaces, *Biometrika*, 15, 231–244.

PEARSON, K., and FILON, L. N. G. (1989) Mathematical contributions to the theory of evolution, IV, *Phil. Trans. R. Soc. London, A*, III, 229–311.

PEWSEY, A. (2000) The wrapped skew-normal distribution on the circle, *Comm. Stat.*, 29, 2459–2479.

QUETELET, A. (1846) *Lettres a S.A.R. le Duc Regnaut de Saxe-Cobourg et Gotha*, Brussels: Hazez.

RHODES, E. C. (1923) On a certain skew correlation surface, *Biometrika*, 14, 355–377.

RIECK, J. R., and NEDELMAN, J. (1991) A log-linear model for the Birnbaum-Saundres distribution, *Technometrics*, 33, 57–60.

RISSER, R. (1947) D'un certain caracteristique des surfaces de probabilité, *Comp. Rend. Acad. Sci. Paris*, 225, 1266–1268.

ROBERTS, C. (1966) A correlation model used in the study of twins, *J. Amer. Stat. Assn.*, 61, 1184–1190.

RUBEN, H. (1954) On moments of order statistics in samples from normal population, *Biometrika*, 41, 200–227.

SAGRISTA, S. N. (1952) On a generalization of Pearson's curves to the two-dimensional case, *Trab. Estadist.*, 3, 273–314.

SAHU, S. K., DEY, D. K., and BRANCO, M. D. (2003) A new class of multivariate skew distributions with applications to Bayesian regression models, *Canad. J. Statist.*, 31, 129–150.

SARTORI, N. (2003) Bias prevention of maximum likelihood estimates: Skew normal and skew distributions, Working paper 2003-1, Department of Statistical Science, University of Padua, Padua, Italy.

SCHREUDER, H. T., and HAFLEY, W. L. (1977) A useful bivariate distribution for describing stand structure of tree heights and diameters, *Biometrika*, 65, 650–653.

SERFLING, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley: New York.

STEYN, H. S. (1960) On regression properties of multivariate probability functions of Pearson's types, *Proc. Roy. Acad. Sci. Amsterdam*, 63, 302–311.

STIGLER, S. M. (1986) *The History of Statistics. The Measurement of Uncertainty before 1900*, Cambridge, MA: Belknap Press, Harvard University Press.

THIELE, T. N. (1897) *Elementar Iagttagelsesiare*, [Elementary Theory of Observations], Kobenhavn: Gyldendal.

THIELE, T. N. (1903) Theory of Observations, *London: Layton. Reprinted in Ann. Math. Stat.*, 2 (1931), 165–307.

THOMAS, K. J. (1993) *Fechner distributions and the assumption of normality in statistical inference*, M. S. Dissertation, University of Essex, England.

VAN UVEN, M. J. (1925–1926) On treating skew correlation, *Proc. Roy. Acad. Sci. Amsterdam*, 28, 797–811, 919–935; 29, 580–590.

VAN UVEN, M. J. (1929) Skew correlation between three and more variables, I-III, *Proc. Roy. Acad. Sci. Amsterdam*, 32, 793–807, 995–1007; 1085–1103.

VAN UVEN, M. J. (1947–1948) Extension of Pearson's probability distributions to two variables, I-IV, *Proc. Roy. Acad. Sci. Amsterdam*, 50, 1063–1070, 1252–1264; 51, 41–52, 191–196.

WANG, J., BOYER, J., and GENTON, M. G. (2004A) A skew-symmetric representation of multivariate distributions, *Stat. Sinica*, 14, 1259–1270.

WANG, J., BOYER, J., and GENTON, M. G. (2004B) A note on equivalence between chi-square and generalized skew-normal distributions, *Stat. Prob. Lett.*, 66, 395–398.

WEINSTEIN, M. A. (1964) The sum of values from a normal and a truncated normal distribution, *Technometrics*, 6, 104–105.

WICKSELL, S. D. (1917) On the genetic theory of frequency, *Arkiv Mate. Astron. Fysik*, 12, no. 20.

WICKSELL, S. D. (1923) Contributions to the analytical theory of sampling, *Arkiv Mate. Astron. Fysik*, 17, no. 19.

SAMUEL KOTZ
Department of Engineering Management
and Systems Engineering
The George Washington University
1776 G Street NW
Washington, D.C. 20052 (U.S.A)
kotz@gwu.edu

DONATELLA VICARI
Dipartimento di Statistica, Probabilità
e Statistiche Applicate
Università di Roma "La Sapienza"
Piazzale Aldo Moro, 5
00185 Rome (Italy)
donatella.vicari@uniroma1.it