Biometrika (1996), **83**, 4, *pp.* 715–726
Printed in Great Britain

The multivariate skew-normal distribution

BY A. AZZALINI AND A. DALLA VALLE
Department of Statistical Sciences, University of Padua, Via S. Francesco 33, 35121 Padova, Italy

SUMMARY

The paper extends earlier work on the so-called skew-normal distribution, a family of distributions including the normal, but with an extra parameter to regulate skewness. The present work introduces a multivariate parametric family such that the marginal densities are scalar skew-normal, and studies its properties, with special emphasis on the bivariate case.

Some key words: Bivariate distribution; Multivariate normal distribution; Specified marginal; Skewness.

1. THE SKEW-NORMAL DISTRIBUTION

1·1. *Introduction*

The term skew-normal ($\mathcal{SN}$) refers to a parametric class of probability distributions which includes the standard normal as a special case. A random variable Z is said to be skew-normal with parameter λ, written $Z \sim \mathcal{SN}(\lambda)$, if its density function is

$$\phi(z; \lambda) := 2\phi(z)\Phi(\lambda z) \quad (z \in \mathbf{R}), \tag{1·1}$$

where $\phi(z)$ and $\Phi(z)$ denote the $\mathcal{N}(0, 1)$ density and distribution function, respectively; the parameter λ which regulates the skewness varies in $(-\infty, \infty)$, and $\lambda = 0$ corresponds to the $\mathcal{N}(0, 1)$ density.

The density (1·1) has appeared at various places in the literature, sometimes in a somewhat casual manner. A systematic treatment of this distribution, developed independently from earlier work, has been given by Azzalini (1985, 1986).

The interest in density (1·1) comes from two directions. On the theoretical side, it enjoys a number of formal properties which reproduce or resemble those of the normal distribution and appear to justify its name skew-normal; in particular, $Z^2 \sim \chi_1^2$. From the applied viewpoint, (1·1) is suitable for the analysis of data exhibiting a unimodal empirical distribution but with some skewness present, a situation often occurring in practical problems. See Hill & Dixon (1982) for a discussion and numerical evidence of the presence of skewness in real data. Arnold et al. (1993) include an application of the $\mathcal{SN}$ distribution to real data.

The purpose of the present paper is to introduce a multivariate version of the skew-normal density. Such an extension is potentially relevant for practical applications, since in the multivariate case there are far fewer distributions available for dealing with non-normal data than in the univariate case, especially for the problem of moderate skewness of the marginals.

A multivariate version of (1·1) has been discussed briefly in Azzalini (1985), but it has the disadvantage of being a purely formal extension of the mathematical expression (1·1), and its marginals do not enjoy the property of being skew-normal densities, as remarked by Ana Quiroga in her 1992 doctoral dissertation at the University of Uppsala.

To achieve a multivariate distribution with specified marginals, a standard method is to use the Farlie–Gumbel–Morgenstern formula, which in the bivariate case takes the form

$$F(x_1, x_2) = F_1(x_1)F_2(x_2)[1 + \alpha\{1 - F_1(x_1)\}\{1 - F_2(x_2)\}],$$

where F_1 and F_2 denote the marginal distributions and α is a parameter varying in $(-1, 1)$. One limitation of this approach is that again the extension from the univariate to the bivariate case is of purely formal nature, and it is not linked to the intrinsic properties of the F_j's. In particular, the above formula does not produce the bivariate normal distribution when $F_1(x) = F_2(x) = \Phi(x)$. Another disadvantage is that the range of achievable correlations obtained by varying α is restricted to the interval $(-\frac{1}{3}, \frac{1}{3})$ (Schucany, Parr & Boyer, 1978).

In the present paper, we consider a different approach, more directly linked to the nature of the scalar $\mathcal{SN}$ distribution, which ensures that the marginals are scalar $\mathcal{SN}$ variates. After presenting the general k-dimensional version, we focus on the bivariate case, including an illustration with real data.

1·2. *Some formal preliminaries*

Before tackling the multivariate case, it is useful to recall some properties of the scalar $\mathcal{SN}$ distribution, for later use.

The moment generating function and the first moments of Z are given in Azzalini (1985). In particular,

$$E(Z) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \delta, \quad \text{var}(Z) = 1 - \frac{2}{\pi}\delta^2, \tag{1·2}$$

where δ is related to λ via the relationships

$$\delta(\lambda) = \frac{\lambda}{(1 + \lambda^2)^{\frac{1}{2}}}, \quad \lambda(\delta) = \frac{\delta}{(1 - \delta^2)^{\frac{1}{2}}}. \tag{1·3}$$

Also,

$$\gamma_1 := \tfrac{1}{2}(4 - \pi)\frac{E(Z)^3}{\text{var}(Z)^{3/2}},$$

the index of skewness, varies in the interval $(-0·995, 0·995)$. Henze (1986) has given the general expression of the odd moments in closed form; the even moments coincide with the normal ones, because $Z^2 \sim \chi_1^2$.

Proposition 1. *If Y and W are independent $\mathcal{N}(0, 1)$ variates, and Z is set equal to Y conditionally on $\lambda Y > W$, for some real λ, then $Z \sim \mathcal{SN}(\lambda)$.*

This result is essentially equivalent to the proof that (1·1) is a density function, given in Azzalini (1985). For random number generation, it is more efficient to use a variant of this result, namely to put

$$Z = \begin{cases} Y & \text{if } \lambda Y > W, \\ -Y & \text{if } \lambda Y \leqslant W, \end{cases}$$

hence avoiding rejection of samples.

The next result follows immediately from Proposition 1 on setting $X = (\lambda Y - W)/(1+\lambda^2)^{\frac{1}{2}}$.

PROPOSITION 2. *If (X, Y) is a bivariate normal random variable with standardised marginals and correlation δ, then the conditional distribution of Y given $X > 0$ is $\mathscr{SN}\{\lambda(\delta)\}$.*

This representation of the $\mathscr{SN}$ distribution, derived in a different way by Arnold et al. (1993), is interesting since it links the $\mathscr{SN}$ distribution to a censoring operation on normal variates, a situation naturally occurring in a large number of practical cases. A similar conditioning mechanism can be used to obtain the skew-normal distribution as a prior distribution for the mean of a normal variable, in a Bayesian framework (O'Hagan & Leonard, 1976).

The following further representation of Z, included in Azzalini (1986), is familiar in the econometric literature on the stochastic frontier model, starting from Aigner, Lovell & Schmidt (1977). Andel, Netuka & Zvara (1984) obtain (1·1) as the stationary distribution of a nonlinear autoregressive process, with transition law related to the next result.

PROPOSITION 3. *If Y_0 and Y_1 are independent $\mathscr{N}(0, 1)$ variables and $\delta \in (-1, 1)$, then*

$$Z = \delta |Y_0| + (1-\delta^2)^{\frac{1}{2}} Y_1$$

is $\mathscr{SN}\{\lambda(\delta)\}$.

2. A multivariate extension

2·1. *Transformation method*

For the k-dimensional extension of (1·1), we consider a multivariate variable Z such that each component is skew-normal. It is then natural to define the joint distribution of Z as being multivariate $\mathscr{SN}$.

Consider a k-dimensional normal random variable $Y = (Y_1, \ldots, Y_k)^{\mathrm{T}}$ with standardised marginals, independent of $Y_0 \sim \mathscr{N}(0, 1)$; thus

$$\begin{pmatrix} Y_0 \\ Y \end{pmatrix} \sim \mathscr{N}_{k+1}\left\{0, \begin{pmatrix} 1 & 0 \\ 0 & \Psi \end{pmatrix}\right\}, \tag{2·1}$$

where Ψ is a $k \times k$ correlation matrix. If $(\delta_1, \ldots, \delta_k)$ are in $(-1, 1)$, define

$$Z_j = \delta_j |Y_0| + (1-\delta_j^2)^{\frac{1}{2}} Y_j \quad (j = 1, \ldots, k), \tag{2·2}$$

so that $Z_j \sim \mathscr{SN}\{\lambda(\delta_j)\}$, by Proposition 3. Computation of the distribution of $Z = (Z_1, \ldots, Z_k)^{\mathrm{T}}$ is trivial but lengthy; the basic steps are reported in the Appendix. The final expression for its density function is

$$f_k(z) = 2\phi_k(z; \Omega)\Phi(\alpha^{\mathrm{T}} z) \quad (z \in \mathbb{R}^k), \tag{2·3}$$

where

$$\alpha^{\mathrm{T}} = \frac{\lambda^{\mathrm{T}}\Psi^{-1}\Delta^{-1}}{(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)^{\frac{1}{2}}}, \tag{2·4}$$

$$\Delta = \mathrm{diag}((1-\delta_1^2)^{\frac{1}{2}}, \ldots, (1-\delta_k^2)^{\frac{1}{2}}),$$

$$\lambda = (\lambda(\delta_1), \ldots, \lambda(\delta_k))^{\mathrm{T}}, \tag{2·5}$$

$$\Omega = \Delta(\Psi + \lambda\lambda^{\mathrm{T}})\Delta, \tag{2·6}$$

and $\phi_k(z; \Omega)$ denotes the density function of the k-dimensional multivariate normal distribution with standardised marginals and correlation matrix Ω.

We shall say that a random variable Z with density function (2·3) is a k-dimensional skew-normal variable, with vector λ of shape parameters and dependence parameter Ψ. For brevity, we shall write

$$Z \sim \mathscr{SN}_k(\lambda, \Psi). \tag{2·7}$$

2·2. *Conditioning method*

An immediate question is whether we would obtain the same class of density functions if we start from another property of the scalar $\mathscr{SN}$ distribution, instead of Proposition 3.

Proposition 2 suggests the following scheme. Let $X = (X_0, X_1, \ldots, X_k)^{\mathrm{T}}$ be a $(k+1)$-dimensional multivariate normal random variable such that

$$X \sim \mathscr{N}_{k+1}(0, \Omega^*), \tag{2·8}$$

with standardised marginals and correlation matrix

$$\Omega^* = \begin{pmatrix} 1 & \delta_1 & \ldots & \delta_k \\ \delta_1 & & & \\ \vdots & & \Omega & \\ \delta_k & & & \end{pmatrix}, \tag{2·9}$$

and consider the distribution of $(X_1, \ldots, X_k)$ given $X_0 > 0$. By Proposition 2, each of these conditional distributions is a scalar $\mathscr{SN}$, specifically $\mathscr{SN}\{\lambda(\delta_j)\}$, for $j = 1, \ldots, k$, and it would be reasonable to say that $(X_1, \ldots, X_k)$ conditionally on $X_0 > 0$ is a multivariate skew-normal random variable. However, it can be shown that the class of distributions obtained by this conditioning method is the same as that obtained by the transformation method of § 2·1.

To prove this, consider the following transformation of the components of the Y's, defined in (2·1):

$$X_0' = Y_0, \quad X_j' = \delta_j Y_0 + (1 - \delta_j^2)^{\frac{1}{2}} Y_j \quad (j = 1, \ldots, k).$$

Clearly, $X' = (X_0', \ldots, X_k')^{\mathrm{T}}$ is still a multivariate normal random variable with standardised marginals, and the resulting correlations are

$$\begin{aligned} &\operatorname{corr}(X_0', X_j') = \delta_j \quad (j = 1, \ldots, k), \\ &\operatorname{corr}(X_i', X_j') = \omega_{ij} = \delta_i \delta_j + \psi_{ij}(1 - \delta_i^2)^{\frac{1}{2}}(1 - \delta_j^2)^{\frac{1}{2}} \quad (i > 0, j > 0), \end{aligned} \tag{2·10}$$

where ψ_{ij} is the corresponding element of Ψ. Therefore X' has the same distribution as X introduced for the conditioning method.

The transformation method and the conditioning method differ in that the former replaces Y_0 by $|Y_0|$, while the latter requires $Y_0 > 0$. The symmetry of the $\mathscr{N}(0, 1)$ density renders these two operations equivalent, and then the distribution of $(X_1', \ldots, X_k')$ conditionally on $X_0' > 0$ is the same as for (2·2), provided that the generic element ω_{ij} of Ω and the corresponding element ψ_{ij} of Ψ are related as in (2·10).

The requirement of positive definiteness of Ω^* in (2·9) imposes some restrictions on the element of Ω, once the δ_j's are fixed. This is why in (2·7) we preferred to refer to Ψ which does not have such a restriction. Obviously, if the ω_{ij}'s are obtained via (2·10), the constraints are automatically satisfied. These aspects will be considered in greater detail in § 3, for the case $k = 2$.

2·3. *Some formal properties*

If Z has density function (2·3), then its cumulative distribution function is

$$F_k(z_1, \ldots, z_k) = \mathrm{pr}\{Z_1 \leqslant z_1, \ldots, Z_k \leqslant z_k\}$$
$$= 2 \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_k} \int_{-\infty}^{\alpha^{\mathrm{T}} v} \phi_k(v; \Omega)\phi(u)\, du\, dv_1 \ldots dv_k$$

for $z = (z_1, \ldots, z_k)^{\mathrm{T}} \in \mathbb{R}^k$, with α defined by (2·4). Then

$$F_k(z_1, \ldots, z_k) = 2\, \mathrm{pr}\{\tilde{Y}_0 \leqslant 0,\, Y_1 \leqslant z_1, \ldots, Y_k \leqslant z_k\}, \tag{2·11}$$

where $(Y_0, Y_1, \ldots, Y_k)^{\mathrm{T}}$ has distribution (2·1) and

$$\tilde{Y}_0 = Y_0 - \alpha^{\mathrm{T}} Y,$$

so that

$$\begin{pmatrix} \tilde{Y}_0 \\ Y \end{pmatrix} \sim \mathscr{N}_{k+1}\left\{0, \begin{pmatrix} 1 + \alpha^{\mathrm{T}}\Psi\alpha & -\alpha^{\mathrm{T}}\Psi \\ -\Psi\alpha & \Psi \end{pmatrix}\right\}. \tag{2·12}$$

To conclude, the distribution function of the k-dimensional variable $Z \sim \mathscr{SN}_k(\lambda, \Psi)$ can be obtained from (2·11) by computing the distribution function of a $(k+1)$-dimensional normal variable with distribution (2·12).

For computing the moment generating function of Z, we need the following well-known result.

PROPOSITION 4. *If $U \sim \mathscr{N}_k(0, \Omega)$, then*

$$E\{\Phi(u + v^{\mathrm{T}} U)\} = \Phi\left\{\frac{u}{(1 + v^{\mathrm{T}}\Omega v)^{\frac{1}{2}}}\right\}$$

for any scalar u and $v \in \mathbb{R}^k$.

This result is usually stated for a scalar random variable; see for instance Zacks (1981, pp. 53–4). The present formulation is a trivial extension, since $v^{\mathrm{T}} U \sim \mathscr{N}(0, v^{\mathrm{T}}\Omega v)$.

Computation of the moment generating function of Z is now immediate:

$$M(t) = 2 \int_{\mathbb{R}^k} \exp(t^{\mathrm{T}} z)\phi_k(z; \Omega)\Phi(\alpha^{\mathrm{T}} z)\, dz$$
$$= \frac{2}{(2\pi)^{\frac{1}{2}k}|\Omega|^{\frac{1}{2}}} \int_{\mathbb{R}^k} \exp\{-\tfrac{1}{2}(z^{\mathrm{T}}\Omega^{-1}z - 2t^{\mathrm{T}} z)\}\, \Phi(\alpha^{\mathrm{T}} z)\, dz$$
$$= 2 \exp\{\tfrac{1}{2}(t^{\mathrm{T}}\Omega t)\} E[\Phi\{\alpha^{\mathrm{T}}(U + \Omega t)\}]$$
$$= 2 \exp\{\tfrac{1}{2}(t^{\mathrm{T}}\Omega t)\}\, \Phi\left\{\frac{\alpha^{\mathrm{T}}\Omega t}{(1 + \alpha^{\mathrm{T}}\Omega\alpha)^{\frac{1}{2}}}\right\}, \tag{2·13}$$

where $U \sim \mathscr{N}_k(0, \Omega)$.

For computing the correlation matrix of Z, it is best to refer to the transformation method of its generation. Simple algebra leads to

$$\rho_{ij} = \mathrm{corr}(Z_i, Z_j) = \frac{\psi_{ij}\{(1 - \delta_i^2)(1 - \delta_j^2)\}^{\frac{1}{2}} + \delta_i\delta_j(1 - 2\pi^{-1})}{\{(1 - 2\pi^{-1}\delta_i^2)(1 - 2\pi^{-1}\delta_j^2)\}^{\frac{1}{2}}} \quad (i, j = 1, \ldots, k), \tag{2·14}$$

taking into account (1·2). Substituting (2·10) in this expression, we can also write

$$\rho_{ij} = \frac{\omega_{ij} - 2\pi^{-1}\delta_i\delta_j}{\{(1 - 2\pi^{-1}\delta_i^2)(1 - 2\pi^{-1}\delta_j^2)\}^{\frac{1}{2}}}.$$

In the scalar case, changing the sign of a $\mathscr{S}\mathscr{N}(\lambda)$ variable produces a $\mathscr{S}\mathscr{N}(-\lambda)$ variable. A similar property holds in the multivariate case.

PROPOSITION 5. *If* $Z \sim \mathscr{S}\mathscr{N}_k(\lambda, \Psi)$ *and* $S = \mathrm{diag}(s_1, \ldots, s_k)$, *where* $s_1, \ldots, s_k$ *is a sequence whose terms are either* $+1$ *or* -1, *then*

$$SZ \sim \mathscr{S}\mathscr{N}_k(S\lambda, S\Psi S).$$

The proof of this statement follows easily by applying the SZ transformation to the Y_j's in (2·2). When all s_j's are equal to -1, we obtain the simple case

$$-Z \sim \mathscr{S}\mathscr{N}_k(-\lambda, \Psi). \tag{2·15}$$

If the conditioning described in § 2·2 applies when $X_0 < 0$ instead of $X_0 > 0$, then the resulting distribution has the δ and λ vectors of reversed sign. Combining this fact with (2·15), we obtain the next result.

PROPOSITION 6. *If* X_0 *is a scalar random variable and* X *is k-dimensional, such that*

$$\begin{pmatrix} X_0 \\ X \end{pmatrix} \sim \mathscr{N}_{k+1}\left\{0, \begin{pmatrix} 1 & \delta^{\mathrm{T}} \\ \delta & \Omega \end{pmatrix}\right\},$$

and Z *is defined by*

$$Z = \begin{cases} X & \text{if } X_0 > 0, \\ -X & \text{otherwise}, \end{cases}$$

then $Z \sim \mathscr{S}\mathscr{N}_k\{\lambda(\delta), \Psi\}$, *where* $\lambda(\delta)$ *is given by* (1·3) *and the entries of* Ψ *are computed from those of* Ω *by solving* (2·10).

A corollary of this fact, which generalises a property of the scalar skew-normal distribution, is obtained by noticing that $Z^{\mathrm{T}}\Omega^{-1}Z$ is equal to $X^{\mathrm{T}}\Omega^{-1}X$ in distribution.

PROPOSITION 7. *If* $Z \sim \mathscr{S}\mathscr{N}_k(\lambda, \Psi)$ *and* Ω *is given by* (2·6), *then*

$$Z^{\mathrm{T}}\Omega^{-1}Z \sim \chi_k^2.$$

3. THE BIVARIATE SKEW-NORMAL DISTRIBUTION

3·1. *Generalities*

For the rest of the paper, we shall concentrate on the bivariate case. Setting $k = 2$ in (2·3), the density function of $Z = (Z_1, Z_2)$ is

$$f_2(z_1, z_2) = 2\phi_2(z_1, z_2; \omega)\Phi(\alpha_1 z_1 + \alpha_2 z_2), \tag{3·1}$$

where ω is the off-diagonal element of Ω and, after some algebra,

$$\alpha_1 = \frac{\delta_1 - \delta_2\omega}{\{(1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1\delta_2\omega)\}^{\frac{1}{2}}},$$

$$\alpha_2 = \frac{\delta_2 - \delta_1\omega}{\{(1 - \omega^2)(1 - \omega^2 - \delta_1^2 - \delta_2^2 + 2\delta_1\delta_2\omega)\}^{\frac{1}{2}}}.$$

An alternative way to obtain (3·1) is to consider the joint distribution of (X_1, X_2), components of the vector $X = (X_0, X_1, X_2)^{\mathrm{T}}$ distributed as (2·8), conditioned on $X_0 > 0$. This computation can be tackled via a technique similar to that of Cartinhour (1990). The resulting algebra is messy and is omitted.

Particularising (2·13) gives, after substantial reduction,

$$M(t_1, t_2) = 2\exp\{\tfrac{1}{2}(t_1^2 + 2\omega t_1 t_2 + t_2^2)\}\,\Phi(t_1\delta_1 + t_2\delta_2).$$

We have already noted the existence of restrictions on the elements of Ω. In the bivariate case, ω must satisfy

$$\delta_1\delta_2 - \{(1-\delta_1^2)(1-\delta_2^2)\}^{\frac{1}{2}} < \omega < \delta_1\delta_2 + \{(1-\delta_1^2)(1-\delta_2^2)\}^{\frac{1}{2}}. \tag{3·2}$$

Figure 1 shows the achievable correlation ρ as a function of ψ and δ, when $\delta_1 = \delta_2 = \delta$. In the general case $\delta_1 \neq \delta_2$, the upper bound for ρ is less than 1; Figure 2 shows the upper and lower limits for ρ for a specific value of δ_1. The overall message contained in these figures is that (3·1) spans a reasonably wide range of ρ values, much wider than that achieved by the Farlie–Gumbel–Morgenstern formula.

Outside the normal distribution context, the correlation is no longer the obvious measure of association between two variables. A recent proposal for an alternative measure, especially relevant in the context of nonlinear dependence among continuous variables, is the 'correlation curve', discussed by Bjerve & Doksum (1993), which is essentially a local analogue of the usual correlation, computed for each given value of the conditioning variable. This correlation curve can be readily obtained for the bivariate skew-normal distribution, by using the expressions given in § 3·2.

Figures 3(a) and (b) provide contour plots of the bivariate density (3·1) when $\delta_1 = \delta_2 = 0·8$ and ω is equal to 0·3 and 0·4074, respectively. The value $\omega = 0·4074$ corresponds to uncorrelated components. Additional plots of similar type are presented in the final section.

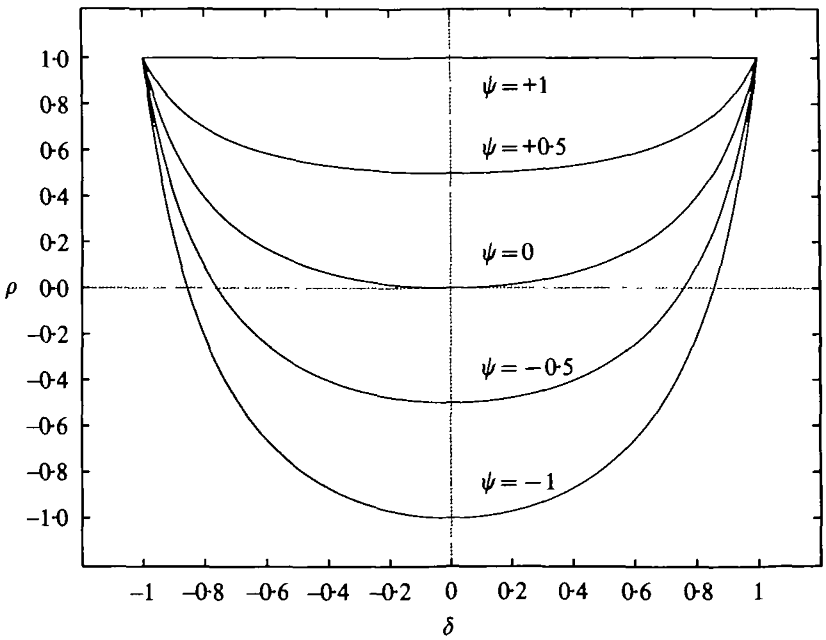

Fig. 1. Relation between ρ and δ for varying ψ.

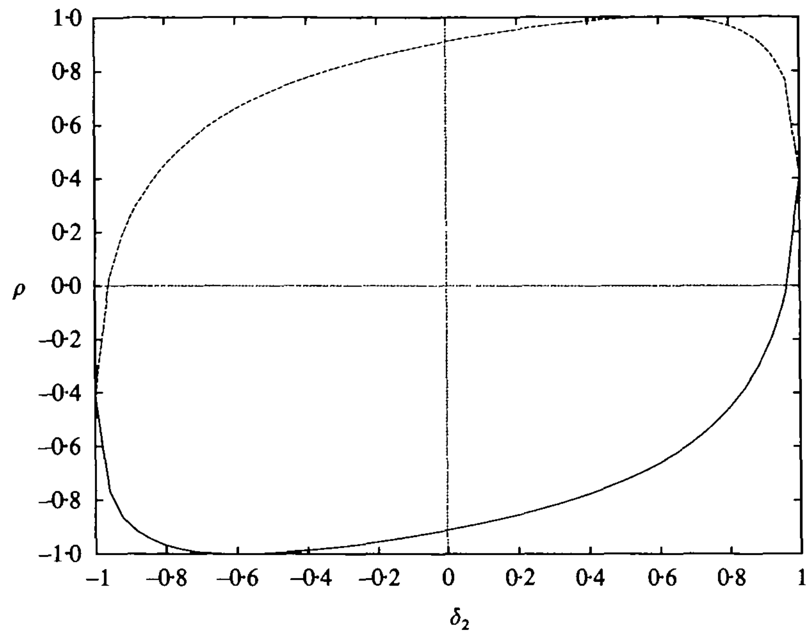

Fig. 2. Upper limit (dashed curve) and lower limit (continuous curve) of ρ as a function of δ_2 when $\delta_1 = 0{\cdot}6$.

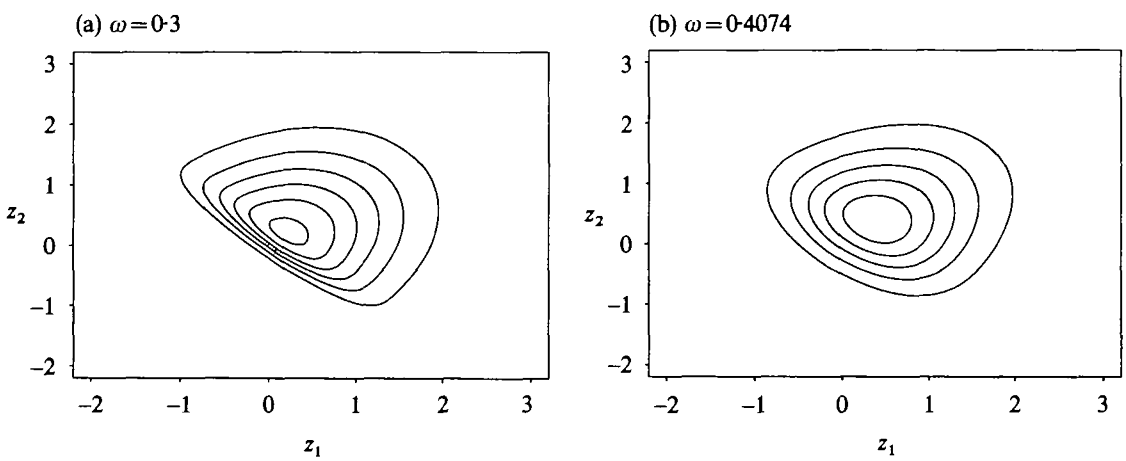

Fig. 3. Contour plot $\mathcal{SN}_2$ for $\delta_1=\delta_2=0{\cdot}8$ and $\omega=0{\cdot}3$, and $\delta_1=\delta_2=0{\cdot}8$ and $\omega=0{\cdot}4074$.

These contour plots are clearly different from ellipses, and their shape is substantially influenced by the value of ω.

3·2. *The conditional distribution*

The conditional density of Z_2 given $Z_1=z_1$ is

$$f(z_2|z_1)=\phi_c(z_2|z_1;\,\omega)\,\frac{\Phi(\alpha_1 z_1+\alpha_2 z_2)}{\Phi(\lambda_1 z_1)}, \tag{3·3}$$

where $\phi_c(z_2|z_1;\,\omega)$ denotes the conditional density associated with a bivariate normal variable with standardised marginals and correlation ω. We observe that (3·3) is a member of an extended skew-normal class of densities, given by equation (9) of Azzalini (1985).

The moment generating function of Z_2 given Z_1 is

$$M_{z_2|z_1}(t_2) = \exp[\tfrac{1}{2}\{(1-\omega^2)t_2^2 + 2\omega t_2 z_1\}]\Phi\left\{\frac{\delta_1 z_1 + t_2(\delta_2 - \omega\delta_1)}{(1-\delta_1^2)^{\frac{1}{2}}}\right\}\Bigg/\Phi(\lambda_1 z_1). \quad (3{\cdot}4)$$

By differentiation of (3·4), we obtain, after some algebra, the expressions for the conditional mean and variance:

$$E(Z_2|Z_1 = z_1) = \omega z_1 + \left\{\frac{\delta_2 - \omega\delta_1}{(1-\delta_1^2)^{\frac{1}{2}}}\right\} H(-\lambda_1 z_1),$$

$$\operatorname{var}(Z_2|Z_1 = z_1) = 1 - \omega^2 - \left\{\frac{\delta_2 - \omega\delta_1}{(1-\delta_1^2)^{\frac{1}{2}}}\right\}^2 H(-\lambda_1 z_1)\{\lambda_1 z_1 + H(-\lambda_1 z_1)\},$$

where $H(x) = \phi(x)/\Phi(-x)$ denotes the hazard function of the standard normal density.

4. Some examples with real data

The aim of this section is to provide simple numerical illustration of the material presented so far, focusing on the bivariate case to ease numerical work and graphical presentation. Neither a full discussion of the statistical aspects involved nor a thorough exploration of the practical relevance of the skew-normal distribution is attempted here.

To fit real data, a location and a scale parameter should be included. In the bivariate case, the observed variable $Y = (Y_1, Y_2)^{\mathrm{T}}$ would be naturally expressed as

$$Y_j = \lambda_{1j} + \lambda_{2j} Z_j \quad (j = 1, 2), \quad (4{\cdot}1)$$

where $Z = (Z_1, Z_2)^{\mathrm{T}} \sim \mathcal{SN}_2(\lambda\ \psi)$, involving seven components to be estimated.

The likelihood function is readily computed from (3·1) and, for these exploratory examples, the corresponding estimates have been computed by direct maximisation of the log-likelihood itself. It appears however preferable to reparametrise (4·1) to

$$Y_j = \theta_{1j} + \theta_{2j}\left[\frac{Z_j - E(Z_j)}{\{\operatorname{var}(Z_j)\}^{\frac{1}{2}}}\right] \quad (j = 1, 2), \quad (4{\cdot}2)$$

as suggested in Azzalini (1985) for the scalar case. The appropriateness of this parametrisation, at least for the scalar case, is confirmed by numerical work by Bruno Genetti, in his 1993 dissertation presented at the University of Padua, where it is shown that the log likelihood associated with (4·2) has a much more regular shape than that associated with (4·1), which has long narrow ridges, often curved. This sort of behaviour is also found by Arnold et al. (1993). After the likelihood function has been maximised, it seems however preferable to convert back the estimates from (4·2) to the simpler form (4·1).

We shall make use of a data set, collected by the Australian Institute of Sport and reported by Cook & Weisberg (1994), containing several variables measured on $n = 202$ Australian athletes. Specifically, we shall consider the pairs of variables (Wt, Ht) and (LBM, BMI) where the meaning of the names is: Wt, weight (Kg); Ht, height (cm); BMI, body mass index $=$ Wt/(Ht)^2; LBM, lean body mass.

Figure 4 displays the scatter-plot of (Wt, Ht) with superimposed contours of the fitted $\mathcal{SN}_2$ distribution. Both the observed points and the fitted density exhibit moderate skewness for each of the components.

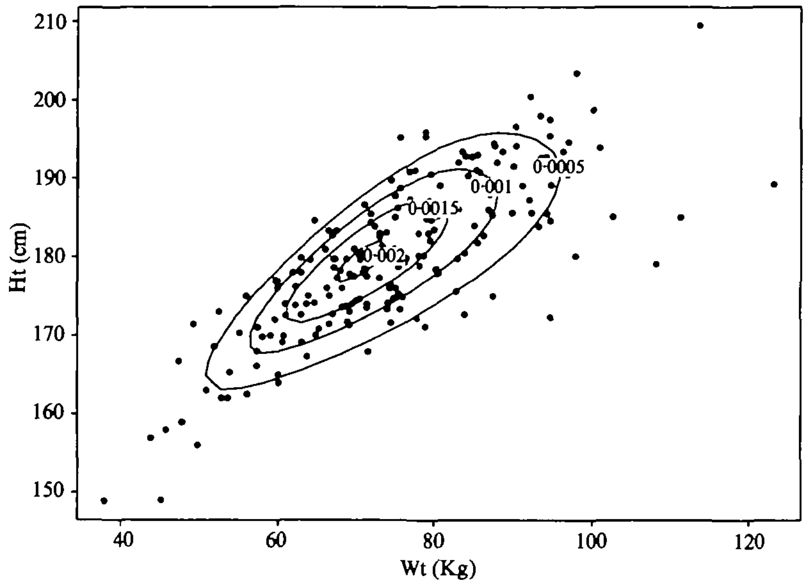

Fig. 4. Australian athletes data: scatter plot (Wt, Ht) and fitted bivariate skew-normal distribution.

An 'obvious' way of tackling these data, and the other components of the same data set, would be via some monotonic transformation of the individual components, a device which often cures nonnormality of the data; log-transformation is the most commonly employed transformation for this purpose. It can however be argued that retaining the variables on the original scale allows a more meaningful interpretation of the quantities involved. Moreover, transformation of the individual components can improve normality of the components, but it is of no value when the failure of normality refers to the joint behaviour of the variables.

To examine somewhat more formally whether the skewness is significantly different from 0, we use the sample index of skewness $\hat{\gamma}_1$. This has been shown by Salvan (1986) to be the locally most powerful location-scale invariant statistic, for testing departure from normality within the scalar skew normal densities. Standardisation of $\hat{\gamma}_1$ with its asymptotic standard deviation $(6/n)^{\frac{1}{2}}$ gives 1·39 and $-1{\cdot}15$, for Wt and Ht, respectively. Neither of them is significant at 5%, but they are both fairly high in absolute value. This suggests to consider a combined measure of asymmetry, such as the index $b_{1,2}$ proposed by Mardia (1970) as a test statistic of bivariate normality. For our data, the observed value of $nb_{1,2}/6$ is 56·8, which is markedly significant when compared with the χ_4^2 percentage points.

The summary conclusion is that some departure from normality is present, although it is not evident in a single-component analysis. This is in conflict with the common belief that anthropometric measurements are 'normal', and is essentially in agreement with the general findings of Hill & Dixon (1982).

Figure 5 displays the scatter plot and the fitted density for the (BMI, LBM) pair. In this case, obvious skewness is present in both components, and the observed significance level of $\hat{\gamma}_1$, after standardisation, is nearly 0 for both components. The contour level curves appear to follow satisfactorily the scatter plot, and then the fitted $\mathcal{SN}_2$ density seems to provide an adequate summary of the data.

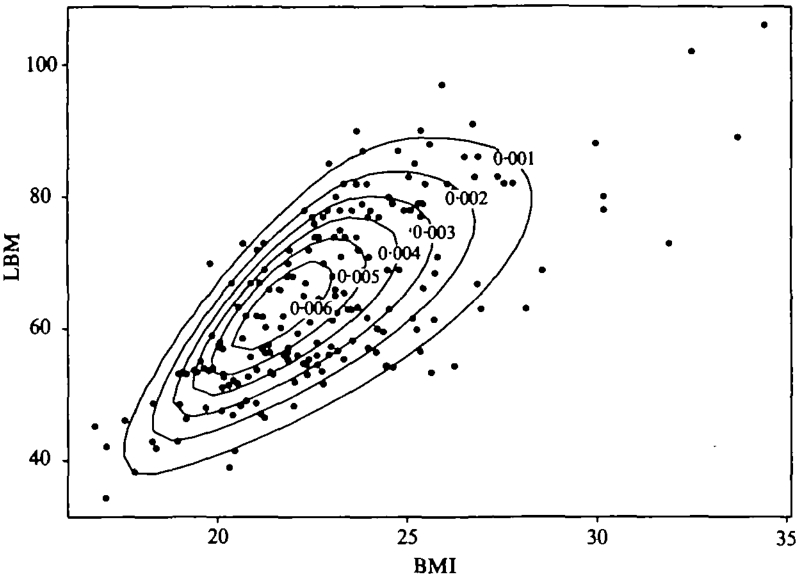

Fig. 5. Australian athletes data: scatter plot (BMI, LBM) and fitted bivariate skew-normal distribution.

Acknowledgement

We thank David Cox, Tom DiCiccio, an Associate Editor and a referee for a number of suggestions which led to a better presentation of the material. This research has been supported by Ministero dell'Università e della Ricerca Scientifica e Tecnologica, Italy.

Appendix

Computation of the density

We compute the density function of Z, whose generic component is defined by (2·2), at a point $z=(z_1,\ldots,z_k)^{\mathrm{T}}\in\mathbb{R}^k$. For subsequent use, we define

$$w_j=\frac{z_j}{(1-\delta_j^2)^{\frac{1}{2}}}\quad(j=1,\ldots,k)$$

and $w=(w_1,\ldots,w_k)^{\mathrm{T}}$. Using standard methods for transformations of random variables, the density function of Z at point $z\in\mathbb{R}^k$ is

$$\begin{aligned}
f_k(z)&=\frac{2}{(2\pi)^{\frac{1}{2}k}\{|\Psi|\prod_{j=1}^{k}(1-\delta_j^2)\}^{\frac{1}{2}}}\times\int_0^\infty\phi(v)\exp\{-\tfrac{1}{2}(w-\lambda v)^{\mathrm{T}}\Psi^{-1}(w-\lambda v)\}\,dv\\
&=\frac{2\exp\{-\frac{1}{2}(w^{\mathrm{T}}\Psi^{-1}w)\}(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}k}\{|\Psi|\prod_{j=1}^{k}(1-\delta_j^2)\}^{\frac{1}{2}}}\exp\left[\frac{1}{2}\left\{\frac{(\lambda^{\mathrm{T}}\Psi^{-1}w)^2}{1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda}\right\}\right]\\
&\quad\times\int_0^\infty\frac{(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\phi\left\{(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)^{-\frac{1}{2}}\left(v-\frac{\lambda^{\mathrm{T}}\Psi^{-1}w}{1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda}\right)\right\}dv\\
&=\frac{2\exp\left[-\dfrac{1}{2}\left\{(w^{\mathrm{T}}\Psi^{-1}w)-\dfrac{(\lambda^{\mathrm{T}}\Psi^{-1}w)^2}{1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda}\right\}\right]}{(2\pi)^{\frac{1}{2}k}\{|\Psi|\prod_{j=1}^{k}(1-\delta_j^2)\}^{\frac{1}{2}}(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)^{\frac{1}{2}}}\Phi\left\{\frac{\lambda^{\mathrm{T}}\Psi^{-1}w}{(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)^{\frac{1}{2}}}\right\}.
\end{aligned}\qquad(\mathrm{A}{\cdot}1)$$

We recall the binomial inverse theorem, which states

$$(A+UBV)^{-1}=A^{-1}-A^{-1}UB(B+BVA^{-1}UB)^{-1}BVA^{-1}$$

for any conformable matrices, provided the inverses exist; see for instance Mardia, Kent & Bibby (1979, p. 459). Using this result, we obtain

$$(\Psi+\lambda\lambda^{\mathrm{T}})^{-1}=\Psi^{-1}-\frac{\Psi^{-1}\lambda\lambda^{\mathrm{T}}\Psi^{-1}}{1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda},$$

$$w^{\mathrm{T}}\Psi^{-1}w-\frac{(\lambda^{\mathrm{T}}\Psi^{-1}w)^2}{1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda}=(\Delta^{-1}z)^{\mathrm{T}}(\Psi+\lambda\lambda^{\mathrm{T}})^{-1}(\Delta^{-1}z)=z^{\mathrm{T}}\Omega^{-1}z, \tag{A·2}$$

where Ω, defined as in (2·6), is such that

$$|\Omega|=|\Psi|(1+\lambda^{\mathrm{T}}\Psi^{-1}\lambda)\left\{\prod_{j=1}^{k}(1-\delta_j^2)^{\frac{1}{2}}\right\}^2. \tag{A·3}$$

Replacing (A·2) and (A·3) in (A·1), we obtain (2·3).

REFERENCES

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

ANDEL, J., NETUKA, I. & ZVARA, K. (1984). On threshold autoregressive processes. *Kybernetika* **20**, 89–106.

ARNOLD, B. C., BEAVER, R. J., GROENEVELD, R. A. & MEEKER, W. Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrica* **58**, 471–8.

AZZALINI, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–8.

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

BJERVE, S. & DOKSUM, K. (1993). Correlation curves: measures of association as functions of covariate values. *Ann. Statist.* **21**, 890–902.

CARTINHOUR, J. (1990). One dimensional marginal density function of a truncated multivariate Normal density function. *Commun. Statist.* A **19**, 197–203.

COOK, R. D. & WEISBERG, S. (1994). *An Introduction to Regression Graphics.* New York: Wiley.

HENZE, N. (1986). A probabilistic representation of the 'skew-normal' distribution. *Scand. J. Statist.* **13**, 271–5.

HILL, M. A. & DIXON, W. J. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrics* **38**, 377–96.

MARDIA, K. V. (1970). Measures of multivariates skewness and kurtosis with applications. *Biometrika* **57**, 519–30.

MARDIA, K. V., KENT, J. T. & BIBBY, J. M. (1979). *Multivariate Analysis.* London: Academic Press.

O'HAGAN, A. & LEONARD, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* **63**, 201–3.

SALVAN, A. (1986). Locally most powerful invariant tests of normality (in Italian). In *Atti della XXXIII Riunione Scientifica della Società Italiana di Statistica*, **2**, pp.173–9. Bari: Cacucci.

SCHUCANY, W. R., PARR, W. C. & BOYER, J. E. (1978). Correlation structure in Farlie–Gumbel–Morgenstern distributions. *Biometrika* **65**, 650–3.

ZACKS, S. (1981). *Parametric Statistical Inference.* Oxford: Pergamon.

[*Received September* 1995. *Revised March* 1996]