Ann. Inst. Statist. Math.
Vol. 56, No. 2, 305–315 (2004)

A CLASS OF MULTIVARIATE SKEW-NORMAL MODELS

ARJUN K. GUPTA AND JOHN T. CHEN

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, U.S.A.

(Received October 11, 2002; revised July 4, 2003)

Abstract. The existing model for multivariate skew normal data does not cohere with the joint distribution of a random sample from a univariate skew normal distribution. This incoherence causes awkward interpretation for data analysis in practice, especially in the development of the sampling distribution theory. In this paper, we propose a refined model that is coherent with the joint distribution of the univariate skew normal random sample, for multivariate skew normal data. The proposed model extends and strengthens the multivariate skew model described in Azzalini (1985, *Scandinavian Journal of Statistics*, **12**, 171–178). We present a stochastic representation for the newly proposed model, and discuss a bivariate setting, which confirms that the newly proposed model is more plausible than the one given by Azzalini and Dalla Valle (1996, *Biometrika*, **83**, 715–726).

Key words and phrases: Moment generating function, skewness, stochastic representation, quadratic form, multivariate normal distribution, Helmert matrix.

1. Introduction

Let X and Y be two independent random variables following the standard normal distribution. For any real number λ, random variable $Z = \frac{\lambda}{\sqrt{1+\lambda^2}}|X| + \frac{1}{\sqrt{1+\lambda^2}}Y$ follows a distribution with density function

$$f(z) = 2\phi(z)\Phi(\lambda z), \qquad -\infty < x < \infty \tag{1.1}$$

where $\phi(\cdot)$, and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. The distribution family of Z with density (1.1), denoted as $Z \sim SN(\lambda)$, is called the skew normal distribution (see for example Azzalini (1985); Gupta *et al.* (2002)). The skew normal distribution family extends the widely employed family of normal distributions by introducing a skewness factor λ. The advantage of this distribution family is that it persists many statistical properties of the normal distribution family (Azzalini (1985), Zack (1981)). The study of skew normal distributions explores an approach for statistical analysis without the assumption of symmetry for the underlying population distribution. Such an extension is necessary because in practice, the underlying distribution may be well skew rather than being symmetric. Thus, the skew normal distribution family emerges to take into account the skewness property. For example, the application of skew normal distribution to time series and spatial statistics was discussed by Genton *et al.* (2001), among others.

Although random variables with density function (1.1) fragmentarily appear in earlier writings in econometrics and medical studies (see for example Aigner *et al.* (1977)

and Roberts (1966)), the primary work of Azzalini (1985) defines the univariate skew normal distribution family and systematically addresses its statistical properties.

When applying the skew normal distribution family in statistical inference, frequently we need to discuss the joint distribution of a random sample from the population. This consequently necessitates the study of multivariate skew normal distribution. Toward this end, Azzalini (1985) mentioned the following multivariate extension of the density in (1.1).

DEFINITION 1. (Azzalini (1985)) Random vector $\boldsymbol{y}$ follows a multivariate skew normal distribution if the joint density of $\boldsymbol{y}$ has the following form:

$$f(\boldsymbol{y}) = c\phi_k(\boldsymbol{y};\Omega)\prod_{j=1}^{k}\Phi(\beta_j y_j) \tag{1.2}$$

where $\phi_k(\cdot,\Omega)$ is the density of a multivariate normal distribution with $k\times k$ correlation matrix $\Omega = (\rho_{ij})$, $\boldsymbol{y} = (y_1,\ldots,y_k)' \in R^k$, $\beta_1,\ldots,\beta_k$ are k real numbers. Note that $\Omega = (\rho_{ij})$ is generally not the correlation matrix of $\boldsymbol{y}$, and c^{-1} is the orthant probability of a standardized normal random variable. The off-diagonal elements of the correlation matrix are of the form $\delta_i\delta_j\rho_{ij}$ with $\delta_i = \lambda_i/\sqrt{1+\lambda_i^2}$. We denote the set of random vectors following a joint distribution in Definition 1 as $SN_1(k,\Omega,\underline{\beta})$, where $\underline{\beta} = (\beta_1,\ldots,\beta_k)'$.

Obviously, Definition 1 is just a direct and formal extension of the univariate skew normal distribution. In 1996, Azzalini and Dalla Valle pointed out the disadvantage of Definition 1, and put forward the following version for multivariate skew normal distributions.

DEFINITION 2. (Azzalini and Dalla Valle (1996)) Random vector $\boldsymbol{y}$ follows a multivariate skew normal distribution if the joint density of $\boldsymbol{y}$ takes the following form:

$$f(\boldsymbol{y}) = 2\phi_k(\boldsymbol{y},\Omega)\Phi(\underline{\alpha}'\boldsymbol{y}) \quad \text{for} \quad \boldsymbol{y}\in R^k \tag{1.3}$$

where $\underline{\alpha}$ is the vector of k real numbers. We denote the set of random vectors following a distribution in Definition 2 as $SN_2(k,\Omega,\underline{\alpha})$.

As delineated in Azzalini and Dalla Valle (1996), and also in Azzalini and Capitanio (1999), Definition 2 is endowed with statistical properties such as the quadratic form of a skew normal random vector is χ_k^2, along with a probabilistic interpretation on the basis of the stochastic representation. However, in the definition, Ω is not the correlation matrix of $\boldsymbol{y}$, and $\Omega = I$ does not imply that all components of $\boldsymbol{y}$ are independent in Definition 2.

For Definition 2, Gupta and Chen (2001, 2003) observed that if $Y_1,\ldots,Y_k$ constitute a random sample from skew normal population $SN(\lambda)$, the distribution of $\boldsymbol{y} = (Y_1,\ldots,Y_k)'$ is not included in $SN_2(k,\Omega,\underline{\alpha})$. Consequently, the distribution of the mean of a random sample does not belong to $SN_2(k,\Omega,\underline{\alpha})$. This phenomenon casts a doubt on the appropriateness of Definition 2 being a statistical model for multivariate skew normal data.

In this paper, we propose a refined model for multivariate skew normal data. We shall show that the newly defined model retains many useful statistical properties, and it coherently embraces the joint distribution of identically and independently distributed skew normal random variables. To keep the paper concrete and directed, we focus on the definition of the new multivariate skew normal family in Section 2, together with the comparison between Definitions 1 and 3, as well as the comparison between Definitions 2 and 3. To provide a probability interpretation, we present a stochastic presentation for the proposed model in Section 3, which is then followed by a discussion on bivariate skew normal distributions in Section 4.

2. A new definition of multivariate skew normal distribution

Consider a random vector with skewness parameter $(\delta_1, \dots, \delta_k)'$ corresponding to its k random elements. Note that the individual shape parameter δ_j may affect the shape of the other random elements via the correlation coefficient matrix used in the multivariate normal density. We propose

DEFINITION 3. Let Ω be a $k \times k$ positive definitive matrix. A $k \times 1$ random vector $\boldsymbol{y}$ is called a multivariate skew normal random vector if the density of $\boldsymbol{y}$ is of the form

$$f(\boldsymbol{y}, \Omega, \boldsymbol{d}) = 2^k \phi_k(\boldsymbol{y}, \Omega) \prod_{j=1}^{k} \Phi(\underline{\lambda}_j' \boldsymbol{y}) \tag{2.1}$$

where $\boldsymbol{d} = (\delta_1, \dots, \delta_k)'$ for some real numbers $\delta_1, \dots, \delta_k$, and $\underline{\lambda}_1, \dots, \underline{\lambda}_k$ are k real vectors satisfying

$$\Lambda = (\underline{\lambda}_1, \dots, \underline{\lambda}_k) = \Omega^{-1/2} \operatorname{diag}(\delta_1, \dots, \delta_k). \tag{2.2}$$

In the sequel, we consider the set of random vectors following a distribution in the family defined in (2.1) as $SN_3(k, \Omega, \boldsymbol{d})$.

Remark 1. Note that in Definition 3, setting $\delta_1 = \cdots = \delta_k = 0$ in (2.1) and (2.2) yields the joint density of the multivariate normal distribution $\phi_k(\boldsymbol{y}, \Omega)$. And from (2.2), letting $\Lambda = (\underline{\lambda}_1, \dots, \underline{\lambda}_k)$ results in

$$\Lambda' \Omega \Lambda = \operatorname{diag}(\delta_1^2, \dots, \delta_k^2).$$

Thus, skew vector $\boldsymbol{d}$ affects the shape of the distribution via eigenvectors $\underline{\lambda}_1, \dots, \underline{\lambda}_k$.

Remark 2. Note that Definition 1 and Definition 3 are two different distribution families. Definition 1 is a special case of Definition 3. To see this, note that for any $\boldsymbol{x} \in SN_1(k, \Omega, \underline{\beta})$, if $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with $|\boldsymbol{A}| \neq 0$, when $\boldsymbol{A} \neq \boldsymbol{I}$, $\boldsymbol{y} \notin SN_1(k, \Omega, \underline{\beta})$, but $\boldsymbol{y} \in SN_3(k, \Omega, \boldsymbol{d})$. (We thank one of the referees for pointing out this connection). Thus $SN_1(k, \Omega, \underline{\beta})$ is not closed for linear transformations. However, for any $\boldsymbol{x} \in SN_3(k, \Omega, \boldsymbol{d})$, $\boldsymbol{A}\boldsymbol{x} \in SN_3(k, \Omega, \boldsymbol{d})$ for any matrix $\boldsymbol{A}$ with $|\boldsymbol{A}| \neq 0$.

To see the eligibility of Definition 3, we need to prove that the integral over the whole space for the density function given in (2.1) equals 1.

LEMMA 2.1. *With $f(\boldsymbol{y}, \Omega, \boldsymbol{d})$ defined in* (2.1), $\int_{R^k} f(\boldsymbol{y}, \Omega, \boldsymbol{d}) d\boldsymbol{y} = 1$.

PROOF. Note that

$$\begin{aligned}
&\int_{R^k} f(\boldsymbol{y}, \Omega, \boldsymbol{d}) d\boldsymbol{y} \\
&\quad = \int_{R^k} 2^k (2\pi)^{-k/2} |\Omega|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{y}'\Omega^{-1}\boldsymbol{y}\right) \prod_{j=1}^{k} \Phi(\underline{\lambda}_j' \boldsymbol{y}) d\boldsymbol{y} \\
&\quad = \int_{R^k} 2^k (2\pi)^{-k/2} \exp\left(-\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right) \prod_{j=1}^{k} \Phi(\underline{\lambda}_j' \Omega^{1/2} \boldsymbol{t}) d\boldsymbol{t}
\end{aligned}$$

Since $\Lambda'\Omega^{1/2} = \operatorname{diag}(\delta_1, \ldots, \delta_k)$, we have

$$\begin{aligned}
&= \int_{R^k} 2^k (2\pi)^{-k/2} \exp\left(-\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right) \prod_{j=1}^{k} \Phi(\delta_j t_j) d\boldsymbol{t} \\
&= 1. \qquad \square
\end{aligned}$$

Lemma 2.1 guarantees that (2.1) is eligible for being a density function. Next, we shall clarify the relation between Definition 2 and Definition 3. In this regard, we show that Definition 3 includes Definition 2 as a special case. The following theorem also describes the relation between Definition 3 and a random sample drawn from a skew normal population.

THEOREM 2.1. *Let $SN_3(k, \Omega, \boldsymbol{d})$ and $SN_2(k, \Omega, \underline{\alpha})$ denote the two distribution families defined in Definition 3 and Definition 2, respectively. For any univariate skew normal random sample $\boldsymbol{y}$, $\boldsymbol{y} = (Y_1, \ldots, Y_k)' \in SN_3(k, \Omega, \boldsymbol{d})$ but $\boldsymbol{y} \notin SN_2(k, \Omega, \underline{\alpha})$.*

To prove Theorem 2.1, we need the following lemma, which was given by Azzalini and Capitanio (1999).

LEMMA 2.2. (Proposition 6 of Azzalini and Capitanio (1999)) *If $\boldsymbol{z} \sim SN_2(n, \Omega, \underline{\alpha})$, and $\boldsymbol{A}'\Omega\boldsymbol{A}$ is a positive definite correlation matrix, $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, then the elements of random vector $\boldsymbol{y} = \boldsymbol{A}'\boldsymbol{z}$ are independent if and only if the following conditions hold simultaneously:*

(a) *$\boldsymbol{a}_i'\Omega\boldsymbol{a}_j = 0$ for $i \neq j$, with $i, j = 1, \ldots, n$ and*

(b) *$\boldsymbol{a}_i'\Omega\underline{\alpha} \neq 0$ for at most one i, with $i = 1, \ldots, n$.*

PROOF OF THEOREM 2.1. First, we notice that for any skew normal random sample, since $Y_1, \ldots, Y_k$ are i.i.d. $SN(\lambda)$ (the univariate skew normal distribution with skew parameter λ), the joint distribution of $\boldsymbol{y}$ is

$$\prod_{j=1}^{k} 2\phi(y_j)\Phi(\lambda y_j)$$

which can be written as

$$2^k \phi_k(\boldsymbol{y}, I) \prod_{j=1}^{k} \Phi(\lambda y_j).$$

Thus $\boldsymbol{y}$ is in $SN_3(k, \Omega, \boldsymbol{d})$, with $\Omega = \boldsymbol{I}$, and $\boldsymbol{d} = (\lambda, \dots, \lambda)'$.

Next let us examine whether $\boldsymbol{y}$ is in $SN_2(k, \Omega, \underline{\alpha})$. If $\boldsymbol{y} \sim SN_2(k, \Omega, \underline{\alpha})$, for some correlation matrix Ω and skewness vector $\underline{\alpha} = (\alpha_1, \dots, \alpha_n)'$, according to Azzalini and Capitanio (1999), Ω is not necessarily equal to the correlation matrix of $\boldsymbol{y}$. Now, let $\boldsymbol{A} = \boldsymbol{I}$ in Lemma 2.2, $\boldsymbol{a}_i$ is the vector with the i-th element (taking value 1) as the only non-zero element. And $\boldsymbol{A}'\Omega\boldsymbol{A}$ is still a positive definite correlation matrix. By Lemma 2.2 with $\Omega = \boldsymbol{I}$, $\underline{\alpha}$ can only have one non-zero element. Without loss of generality, we assume that $\alpha_1 \neq 0$ and $\alpha_j = 0$ for $j = 2, \dots, n$. Thus if $Y_1, \dots, Y_n$ constitute a random sample from a skew normal population, and if we assume that $\boldsymbol{y} = (Y_1, \dots, Y_n)'$ follows the multivariate skew normal distribution $SN_2(k, \Omega, \underline{\alpha})$, from the independence we conclude that $\Omega = \boldsymbol{I}$ and the skewness vector $\underline{\alpha} = (\alpha_1, 0, \dots, 0)'$, $\alpha_1 \neq 0$.

Now, by Proposition 5 of Azzalini and Capitanio (1999) in the case where $\boldsymbol{A} = \boldsymbol{I}$, the marginal distribution of $\boldsymbol{y}$ reads

$$Y_1 \sim SN(\sigma_1, \alpha_1), \qquad \text{and} \qquad Y_j \sim SN(\sigma_j, 0) \quad \text{for} \quad j = 2, \dots, n,$$

where $SN(\sigma_i, \alpha_i)$ is the skew normal distribution with density function, $2\phi(z, \sigma)\Phi(\alpha_i z, \sigma)$, with $\phi(z, \sigma)$ as the density of normal distribution with standard deviation σ_i, and $\Phi(z, \sigma)$ as the cdf corresponding to $\phi(z, \sigma)$. In another words, σ is the scale parameter and α_i becomes the skew parameter. The marginal distributions obtained here are in contradiction with the fact that $Y_1, \dots, Y_n$ constitute a random sample, which means that $Y_1, \dots, Y_n$ follow an identical distribution. Therefore $\boldsymbol{y} \notin SN_2(k, \Omega, \underline{\alpha})$, this completes the proof of Theorem 2.1. □

From Theorem 2.1, we know that given a set of skew normal random sample $\boldsymbol{y}$, it can not be in $SN_2(k, \Omega, \underline{\alpha})$, but $\boldsymbol{y} \in SN_3(k, \Omega, \boldsymbol{d})$, thus $SN_3(k, \Omega, \boldsymbol{d})$ is not the same as $SN_2(k, \Omega, \underline{\alpha})$.

3. Stochastic representation for distribution family $SN_3(k, \Omega, \boldsymbol{d})$

For Definition 3 proposed in Section 2, we shall illustrate the distribution family with a probability interpretation. To give a probabilistic interpretation of $SN_3(k, \Omega, \boldsymbol{d})$, we present a stochastic representation for the newly defined distribution family, $SN_3(k, \Omega, \boldsymbol{d})$. We first prove a theorem in the case where $\Omega = \boldsymbol{I}$, and then, on the basis of this theorem, we show in Corollary 3.1 that any member in $SN_3(k, \Omega, \boldsymbol{d})$ can have a stochastic representation.

THEOREM 3.1. *Let $\boldsymbol{x}$, $\boldsymbol{y}$ be two independent random vectors following $N_k(\boldsymbol{0}, \boldsymbol{I})$. Let*

$$\boldsymbol{z} = \mathrm{diag}\left(\frac{\delta_1}{\sqrt{1+\delta_1^2}}, \dots, \frac{\delta_k}{\sqrt{1+\delta_k^2}}\right)|\boldsymbol{x}| + \mathrm{diag}\left(\frac{1}{\sqrt{1+\delta_1^2}}, \dots, \frac{1}{\sqrt{1+\delta_k^2}}\right)\boldsymbol{y}, \tag{3.1}$$

where $|\boldsymbol{x}| = (|X_1|, \dots, |X_k|)'$. Then $\boldsymbol{z} \sim SN_3(k, \boldsymbol{I}, \boldsymbol{d})$ for any $\boldsymbol{d} = (\delta_1, \dots, \delta_k)' \in R^k$.

PROOF. Let

$$\boldsymbol{U} = \mathrm{diag}\left(\frac{\delta_1}{\sqrt{1+\delta_1^2}}, \dots, \frac{\delta_k}{\sqrt{1+\delta_k^2}}\right)$$

and

$$\boldsymbol{V} = \operatorname{diag}\left(\frac{1}{\sqrt{1+\delta_1^2}}, \ldots, \frac{1}{\sqrt{1+\delta_k^2}}\right),$$

then we have $\boldsymbol{z} = \boldsymbol{U}|\boldsymbol{x}| + \boldsymbol{V}\boldsymbol{y}$ and

$$\boldsymbol{V}\boldsymbol{y} \sim N_k(\boldsymbol{0}, \boldsymbol{V}'\boldsymbol{V}). \tag{3.2}$$

For any real vector $\boldsymbol{w} \in R^k$, we have

$$\begin{aligned}
P(\boldsymbol{z} \le \boldsymbol{w}) &= E_{|\boldsymbol{x}|}\{P(\boldsymbol{z} \le w \mid |\boldsymbol{x}|)\} \\
&= \int_{R_+^k} P(\boldsymbol{z} \le w) 2^k \phi_k(\boldsymbol{x}, \boldsymbol{I}) d\boldsymbol{x} \\
&\quad (\text{where} \quad R_+^k = (0, +\infty)^k) \\
&= \int_{R_+^k} P(\boldsymbol{V}\boldsymbol{y} \le \boldsymbol{w} - \boldsymbol{U}\boldsymbol{x}) 2^k \phi_k(\boldsymbol{x}, \boldsymbol{I}) d\boldsymbol{x} \\
&= \int_{R_+^k} \Phi_k(\boldsymbol{w} - \boldsymbol{U}\boldsymbol{x}, \boldsymbol{V}'\boldsymbol{V}) 2^k \phi_k(\boldsymbol{x}, \boldsymbol{I}) d\boldsymbol{x}
\end{aligned}$$

where $\Phi_k(\boldsymbol{x}, \Sigma) = \int_{R(\boldsymbol{x})} \phi_k(\boldsymbol{y}, \Sigma) d\boldsymbol{y}$. Thus, the joint density of $\boldsymbol{z}$ is

$$\begin{aligned}
f(w_1, \ldots, w_k) &= d(P(\boldsymbol{z} \le \boldsymbol{w}))/d\boldsymbol{w} \\
&= \int_{R_+^k} \{d(\Phi_k(\boldsymbol{w} - \boldsymbol{U}\boldsymbol{x}, \boldsymbol{V}'\boldsymbol{V}))/d\boldsymbol{w}\} 2^k \phi_k(\boldsymbol{x}, \boldsymbol{I}) d\boldsymbol{x} \\
&= \int_{R_+^k} (2\pi)^{-k} |\boldsymbol{V}|^{-1} 2^k \exp\left\{-\frac{1}{2}((\boldsymbol{w} - \boldsymbol{U}\boldsymbol{x})'(\boldsymbol{V}^{-1})'\boldsymbol{V}^{-1}(\boldsymbol{w} - \boldsymbol{U}\boldsymbol{x}))\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right\} d\boldsymbol{x} \\
&\quad (\text{note that } \boldsymbol{V} \text{ is a diagonal matrix and (3.2)}) \\
&= \int_{R_+^k} (2\pi)^{-k} |\boldsymbol{V}|^{-1} 2^k \exp\Big\{-\frac{1}{2}[\boldsymbol{w}'\boldsymbol{V}^{-2}\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{V}^{-2}\boldsymbol{U}\boldsymbol{x} \\
&\qquad - \boldsymbol{x}'\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{w} + \boldsymbol{x}'\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{U}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{x}]\Big\} d\boldsymbol{x}.
\end{aligned}$$

Since $\boldsymbol{U}$ and $\boldsymbol{V}$ are diagonal matrices, they are exchangeable. Also

$$\boldsymbol{x}'\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{U}\boldsymbol{x} + x'x = \boldsymbol{x}'\boldsymbol{V}^{-2}\boldsymbol{x},$$

thus we have

$$\begin{aligned}
&f(w_1, \ldots, w_k) \\
&= \int_{R_+^k} (2\pi)^{-k} |\boldsymbol{V}|^{-1} 2^k \exp\Big\{-\frac{1}{2}[\boldsymbol{x}'\boldsymbol{V}^{-2}\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{V}^{-2}\boldsymbol{U}\boldsymbol{x} \\
&\qquad + \boldsymbol{w}'\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{U}\boldsymbol{w} - \boldsymbol{w}'\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{U}\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{V}^{-2}\boldsymbol{w}]\Big\} d\boldsymbol{x}
\end{aligned}$$

$$= \int_{R_+^k} (2\pi)^{-k} |\boldsymbol{V}|^{-1} 2^k \exp\left\{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w})'\boldsymbol{V}^{-2}(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w}) + \boldsymbol{w}'(\boldsymbol{V}^{-2}-\boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{U})\boldsymbol{w}]\right\} d\boldsymbol{x}.$$

Note that $\boldsymbol{V}^{-2} - \boldsymbol{U}'\boldsymbol{V}^{-2}\boldsymbol{U} = \boldsymbol{I}$, we have

$$\begin{aligned}
&f(w_1,\dots,w_k)\\
&= \int_{R_+^k} (2\pi)^{-k/2} |\boldsymbol{V}|^{-1} 2^k \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w})'\boldsymbol{V}^{-2}(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w})\right\} \phi_k(\boldsymbol{w},\boldsymbol{I}) d\boldsymbol{x}\\
&= 2^k \phi_k(\boldsymbol{w},\boldsymbol{I}) \int_{R_+^k} (2\pi)^{-k/2} |\boldsymbol{V}|^{-1} 2^k \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w})'\boldsymbol{V}^{-2}(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w})\right\} d\boldsymbol{x}\\
&= 2^k \phi_k(\boldsymbol{w},\boldsymbol{I}) \prod_{j=1}^{k} \int_{R^*} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}t_j^2\right\} dt_j\\
&\qquad (\text{letting} \quad \boldsymbol{t} = \boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{U}\boldsymbol{w}))\\
&= 2^k \phi_k(\boldsymbol{w},\boldsymbol{I}) \prod_{j=1}^{k} (1-\Phi(-\delta_j w_j))\\
&= 2^k \phi_k(\boldsymbol{w},\boldsymbol{I}) \prod_{j=1}^{k} \Phi(\delta_j w_j),
\end{aligned}$$

where $R^* = (-\delta_j w_j, \infty)$. Therefore $\boldsymbol{z} \in SN_3(k,\Omega,\boldsymbol{d})$. □

From Theorem 3.1, one can construct a stochastic representation for any member in $SN_3(k,\Omega,\boldsymbol{d})$ as follows.

COROLLARY 3.1. *For any $\boldsymbol{y} \in SN_3(k,\Omega,\boldsymbol{d})$ with parameter matrix Ω and skewness factor $\boldsymbol{d} = (\delta_1,\dots,\delta_k)'$, $\boldsymbol{w} = \Omega^{1/2}\boldsymbol{z}$ has the same distribution as $\boldsymbol{y}$, where $\boldsymbol{z}$ is the random vector defined in* (3.1).

PROOF. First, let the δ_i's in Theorem 3.1 the same as the skewness factors δ_j's for random vector $\boldsymbol{y}$. By Theorem 3.1, we know that the density of $\boldsymbol{z}$ in (3.1) reads

$$2^k \phi_k(\boldsymbol{z},\boldsymbol{I}) \prod_{j=1}^{k} \Phi(\delta_j z_j).$$

Thus the distribution of random vector $\boldsymbol{w} = \Omega^{1/2}\boldsymbol{z}$ is

$$2^k \phi_k(\boldsymbol{w},\Omega) \prod_{j=1}^{k} \Phi((0,\dots,\delta_j,\dots,0)\Omega^{-1/2}\boldsymbol{w}).$$

Letting $\underline{\lambda}_j^* = \Omega^{-1/2}(0,\dots,\delta_j,\dots,0)'$, we know that $\boldsymbol{w}$ has the same distribution as $\boldsymbol{y}$ (since $\Lambda = (\underline{\lambda}_j^*)' = \Omega^{-1/2}\operatorname{diag}(\delta_1,\dots,\delta_k)$). □

4. The bivariate skew-normal distribution

After describing the model and its probability interpretation, in this section, we shall discuss a special case where $k = 2$, namely the bivariate distribution of the newly defined distribution family. First the density function in (2.1) reads

$$f(\boldsymbol{y}, \Omega, \boldsymbol{d}) = 2^2(2\pi)^{-1}\det(\Omega)^{-1/2}e^{(-1/2)\boldsymbol{y}'\Omega^{-1}\boldsymbol{y}}\Phi(\delta_1\boldsymbol{p}_1'\boldsymbol{y})\Phi(\delta_2\boldsymbol{p}_2'\boldsymbol{y}), \tag{4.1}$$

where the parameter matrix

$$\Omega = \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}.$$

Thus the inverse matrix of Ω reads

$$\Omega^{-1} = \frac{1}{1-\omega^2}\begin{pmatrix} 1 & -\omega \\ -\omega & 1 \end{pmatrix}$$

and

$$\Omega^{-1/2} = (1-\omega^2)^{-1/2}\begin{pmatrix} (\sqrt{1-\omega}+\sqrt{1+\omega})/2 & (\sqrt{1-\omega}-\sqrt{1+\omega})/2 \\ (\sqrt{1-\omega}-\sqrt{1+\omega})/2 & (\sqrt{1-\omega}+\sqrt{1+\omega})/2 \end{pmatrix}. \tag{4.2}$$

From (4.2), Equation (4.1) becomes

$$\begin{aligned} &f(y_1, y_2, \omega, \delta_1, \delta_2) \\ &\quad = 2^2(2\pi)^{-1}(1-\omega^2)^{-1/2}e^{(-(y_1^2-2\omega y_1 y_2+y_2^2)/2(1-\omega^2))}\Phi(\delta_1\boldsymbol{p}_1'\boldsymbol{y})\Phi(\delta_2\boldsymbol{p}_2'\boldsymbol{y}) \end{aligned} \tag{4.3}$$

with

$$\boldsymbol{p}_1 = \begin{pmatrix} (\sqrt{1-\omega}+\sqrt{1+\omega})/2 \\ (\sqrt{1-\omega}-\sqrt{1+\omega})/2 \end{pmatrix},$$
$$\boldsymbol{p}_2 = \begin{pmatrix} (\sqrt{1-\omega}-\sqrt{1+\omega})/2 \\ (\sqrt{1-\omega}+\sqrt{1+\omega})/2 \end{pmatrix}.$$

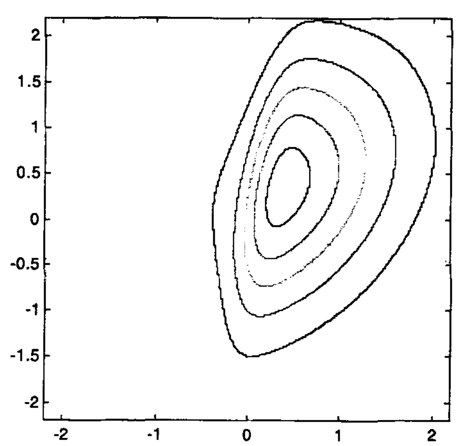

Fig. 1. Bivariate skew normal with $w = 0.3$.

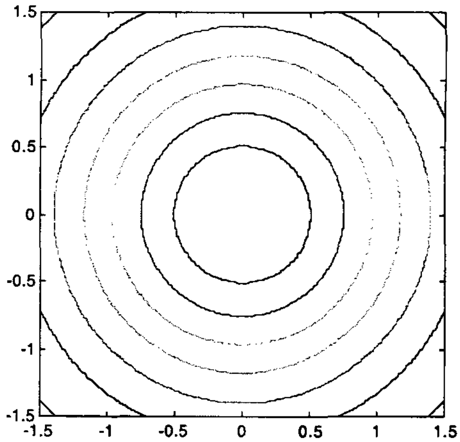

Fig. 2. Bivariate normal distribution.

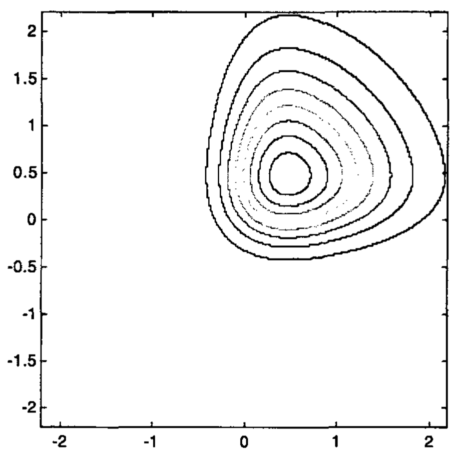

Fig. 3. Bivariate skew normal with $w = 0$.

The contour plot of density (4.3) when $\delta_1 = 4$, $\delta_2 = 0.4$ and $\omega = 0.3$ is given in Fig. 1. Comparing Fig. 1 with Fig. 2 that outlines a contour plot of the standard bivariate normal density, we can see the symmetry of the standard bivariate normal density is changed by the shape parameters ω, δ_1 and δ_2. These parameters make the density skew and distorted. In Fig. 3, we provide a contour plot of density (4.3) with $\delta_1 = \delta_2 = 3.0$ and $\omega = 0$. The plot in Fig. 3 suggests that in bivariate case, random variables y_1 and y_2 are independent if and only if $\omega = 0$. While this property holds for the bivariate normal distribution family, in the bivariate skew normal family defined by Azzalini and Dalla Valle (1996), this property is invalid, which causes the interpretation of ω somewhat awkward. In the following theorem, we prove that in the definition of $SN_3(k, \Omega, \boldsymbol{d})$, the property between independence and $\omega = 0$ remains valid. This makes the definition of $SN_3(k, \Omega, \boldsymbol{d})$ more plausible.

THEOREM 4.1. *Let $\boldsymbol{y}$ be a bivariate random vector in $SN_3(k, \Omega, \boldsymbol{d})$. Then Y_1 and*

Y_2 are independent if and only if $\omega = 0$.

PROOF. The moment generating function of $\boldsymbol{y} = \binom{Y_1}{Y_2}$ can be obtained as

$$M(\boldsymbol{t}) = 4\exp\{(t_1^2 + 2\omega t_1 t_2 + t_2^2)/2\}\Phi(h_1 p_{11} t_1 + h_1 p_{12} t_2)\Phi(h_2 p_{21} t_1 + h_2 p_{22} t_2)$$

where

$$h_1 = \frac{\delta_1}{\sqrt{1+\delta_1^2}},$$
$$h_2 = \frac{\delta_2}{\sqrt{1+\delta_2^2}},$$
$$p_{11} = p_{22} = (\sqrt{1-\omega} + \sqrt{1+\omega})/2,$$

and

$$p_{21} = p_{12} = (\sqrt{1-\omega} - \sqrt{1+\omega})/2.$$

Thus

$$\left.\frac{\partial M(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} = \omega + \frac{\pi}{2}(h_1 h_2 p_{11} p_{22} + h_1 h_2 p_{21} p_{12}). \tag{4.4}$$

By the fact that

$$E(Y_1) = \sqrt{\frac{2}{\pi}} h_1 p_{11} + \sqrt{\frac{2}{\pi}} h_2 p_{21}$$

and

$$E(Y_2) = \sqrt{\frac{2}{\pi}} h_1 p_{12} + \sqrt{\frac{2}{\pi}} h_2 p_{22},$$

we have, from (4.4), that

$$\mathrm{COV}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = \left(1 + \frac{1}{\pi}(h_1^2 + h_2^2)\right)\omega. \tag{4.5}$$

Therefore, if Y_1 and Y_2 are independent, by (4.5) we know immediately that $\omega = 0$; on the other hand, if $\omega = 0$, by the density in (4.3), one can see that the joint density is the product of two individual skew normal densities, which means that Y_1 and Y_2 are independent.

5. Concluding remarks

We present a distribution family of multivariate skew normal distributions that embraces the joint distribution of a random sample from a univariate skew normal distribution. The newly defined distribution family, $SN_3(\boldsymbol{k}, \Omega, \boldsymbol{d})$, is endowed with many statistical properties, including the coherency between a multivariate distribution family and the corresponding univariate skew normal samples. The persistence of statistical properties enriches the theory of skew normal distribution which can be employed when the symmetry of the underlying population cannot be plausibly assumed. Compared with existing multivariate models for populations with skew distributions, $SN_3(\boldsymbol{k}, \Omega, \boldsymbol{d})$

extends and strengthens the multivariate model proposed by Azzalini (1985), and it has greater flexibility with respect to the model proposed by Azzalini and Dalla Valle (1996). However, using the link between $SN_3(k, \Omega, \boldsymbol{d})$ and $SN_1(k, \Omega, \underline{\beta})$, statistical properties such as conditional and marginal distributions, moments of $SN_3(k, \Omega, \boldsymbol{d})$ may be derived indirectly from the properties of $SN_1(k, \Omega, \underline{\beta})$.

Acknowledgements

The Research of the authors was partially supported by Technology Innovation Enhancement Grant. We thank the editor, an associate editor and two referees for their constructive comments, which improved the original version of this article.

References

Aigner, D. J., Lovell, C. H. K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function model, *Journal of Econometrics*, **12**, 21–37.

Azzalini, A. (1985). A class of distribution which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **3**, 579–602.

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

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

Gupta, A. K. and Chen, T. (2001). Goodness of fit tests for the skew-normal distribution, *Communication in Statistics, Computation and Simulation*, **30**(4), 907–930.

Gupta, A. K. and Chen, J. T. (2003). On the sample characterization criterion for normal distributions, *Journal of Statistical Computation and Simulation*, **73**, 155–163.

Gupta, A. K., Chang, F. C. and Huang, W. J. (2002). Some skew-symmetric models, *Random Operations and Stochastic Equations*, **10**, 133–140.

Roberts, C. (1966). A correlation model useful in the study of twins, *Journal of the American Statistical Association*, **61**, 1184–1190.

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