doi: 10.1111/j.1467-9469.2006.00503.x
 Vol 33: 561–574, 2006

On the Unification of Families of Skew-normal Distributions

REINALDO B. ARELLANO-VALLE

Departamento de Estadística Pontificia Universidad Católica de Chile

ADELCHI AZZALINI

Dipartimento di Scienze Statistiche Università di Padova

ABSTRACT. The distribution theory literature connected to the multivariate skew-normal distribution has grown rapidly in recent years, and a number of extensions and alternative formulations have been put forward. Presently there are various coexisting proposals, similar but not identical, and with rather unclear connections. The aim of this paper is to unify these proposals under a new general formulation, clarifying at the same time their relationships. The final part sketches an extension of the argument to the skew-elliptical family.

Key words: skew-elliptical family, skew-normal distribution, skew-*t* distribution, stochastic representation

1. The skew-normal distribution and its descendants

1.1. General remarks

Recently, there has been a resumption of interest in the study of parametric classes of probability distributions for continuous multivariate random variables. A substantial fraction of this activity has resulted from the proposal of the multivariate skew-normal (SN) distribution, which represents an extension of the multivariate normal family to which an additional parameter is added to regulate skewness. The multivariate SN class has been studied by Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999), and subsequently by several authors, with intense developments in various directions, often considering non-normal symmetric families as the starting point in place of the normal class.

These further developments are now so numerous that is not feasible to recall here all of them; hence we shall restrict our references to specific contributions, as their relevance occurs in the exposition of the paper. For reviews of the research work produced in this area, see Arnold & Beaver (2002), Genton (2004) and Azzalini (2005). Here we shall confine ourselves to a single remark: while the original motivation of this approach was the introduction of skewness in the normal family of distributions, the current developments are of a much higher level of generality and the effect of these extensions to standard parametric families is much more substantial than simple 'insertion' of skewness.

This very active and stimulating context has seen the appearance of several variants or alternative proposals of the original SN distribution. Among these, we mention in particular: the closed skew-normal (CSN) of González-Farías *et al.* (2004), the hierarchical skew-normal (HSN) of Liseo & Loperfido (2003), the fundamental skew-normal (FUSN) of Arellano-Valle & Genton (2005), the multivariate skew-normal of Gupta *et al.* (2004), the skew-normal of Sahu *et al.* (2003); although on a slightly different direction, see also Arnold & Beaver (2000).

While the existence of so many proposals is a sign of the high vitality of this stream of literature, it inevitably poses problems, namely: is there one version which is the 'overall best' skew-normal distribution? are some of these proposals equivalent up to a reparametrization?

more generally, what are the connections among them? The purpose of this paper is to tackle the above questions, and to reconstitute a 'unified formulation' of the SN distribution. To achieve this target we shall introduce a further type of SN distribution, and this very fact seems to contradict our purpose. However, we show that this new formulation encompasses all previously mentioned proposals, once some redundancies in parametrization are removed, and this process also clarifies the connections among previous variants. In addition, reasons are given for preferring certain forms of parametrization in place of others.

1.2. The basic SN distribution

To establish notation, the density function at point x ($x \in \mathbb{R}^d$) of a $N_d(\mu, \Sigma)$ random variable will be denoted by $\phi_d(x-\mu;\Sigma)$; similarly, $\Phi_d(x-\mu;\Sigma)$ represents the corresponding distribution function. When $d=1$, we omit the subscript of ϕ and Φ.

The original version of the multivariate SN distribution refers to the d-dimensional density function whose value at x ($x \in \mathbb{R}^d$) is given by

$$2\phi_d(x-\xi;\Omega)\Phi\left(\alpha^\top \omega^{-1}(x-\xi)\right), \tag{1}$$

where ξ ($\xi \in \mathbb{R}^d$) is a location parameter, Ω is a positive definite covariance matrix, α ($\alpha \in \mathbb{R}^d$) is a parameter which regulates skewness, and ω is a diagonal matrix formed by the standard deviations of Ω; hence $\Omega = \omega\bar{\Omega}\omega$, where $\bar{\Omega}$ is a correlation matrix. If a random variable $Y=(Y_1,\ldots, Y_d)^\top$ has density (1), we write $Y \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$. As the symbols for the location parameter and the scale matrix, we have adopted ξ and Ω, which differ from those used for the normal distribution, to mark the fact that these quantities do not represent the mean vector and the variance matrix of distribution (1).

Clearly, setting $\alpha=0$ in (1) produces the $N_d(\xi,\Omega)$ distribution. Besides this simple fact, many other formal properties and analogies with the normal family of distributions support the adoption of the term 'skew-normal', but we do not review these aspects here, and refer the reader to the literature quoted above. What is more important for our development are two forms of stochastic representation for a random variable of the SN type.

The first of these representations is obtained via the following conditioning mechanism. Denote by U_0 and U_1 two random variables of dimension 1 and d, respectively, such that

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{1+d}(0, \Omega^*), \quad \Omega^* = \begin{pmatrix} 1 & \delta^\top \\ \delta & \bar{\Omega} \end{pmatrix}, \tag{2}$$

where $\bar{\Omega}$ is a correlation matrix and

$$\delta = (1+\alpha^\top\bar{\Omega}\alpha)^{-1/2}\bar{\Omega}\alpha, \tag{3}$$

which ensures that Ω^* is a correlation matrix for any $\alpha \in \mathbb{R}^d$. Then $Z=(U_1 | U_0>0)$ has density function

$$2\phi_d(x;\bar{\Omega})\,\Phi(\alpha^\top x) \tag{4}$$

and the affine transformation $Y=\xi+\omega Z$ has density function (1). For later use, we recall a form of dual expression of (3), given by

$$\alpha = (1-\delta^\top\bar{\Omega}^{-1}\delta)^{-1/2}\bar{\Omega}^{-1}\delta. \tag{5}$$

The other stochastic representation is via a convolution. Specifically, assume that V_0 and V_1 are independent variables with distribution $N(0,1)$ and $N_d(0,\Psi)$, respectively, where Ψ is a correlation matrix; also, let $\Delta = \mathrm{diag}(\delta_1,\ldots,\delta_d)$ where $\delta_j \in (-1,1)$ for all js, denote by I_d the identity matrix of order d and by 1_d the d-dimensional vector of all 1s. Then

$$Z = \Delta 1_d |V_0| + (I_d - \Delta^2)^{1/2} V_1 \tag{6}$$

has distribution of type (4), with a known relationship between the (Ψ, Δ) and the $(\bar{\Omega}, \alpha)$ sets of parameters; for the explicit expression of this relationship see Azzalini & Capitanio (1999, Appendix A1). This connection explains why we have adopted the same symbol δ is (3) and for the ingredients of Δ appearing in (6). Again, the affine transformation $Y = \xi + \omega Z$ leads to density (1).

1.3. Some extensions and variants

The basic form (1) lends itself to a number of extensions, which have been extensively studied in the papers quoted in section 1.1 and other publications mentioned therein. These extensions follow one, or possibly more, of the following directions of work.

1. Replace the 0 value of $\mathbb{E}\{U_0\}$ in (2) by a further parameter, γ say. This additional parameter changes the normalizing constant in (1) from 2 to $1/\Phi(\gamma)$, and it prevents the stochastic representation (6), at least in the exact form given above. Notice that another variant is equivalent, namely setting $Z = (U_1 | U_0 > \tau)$, but the two parameters τ and γ cannot be included simultaneously, as only $\tau - \gamma$ is identifiable.

A stochastic representation similar to (6) is given by

$$Z = \Delta 1_d V_{0\gamma} + (I_d - \Delta^2)^{1/2} V_1, \tag{7}$$

where $V_{0\gamma} \sim LTN(-\gamma; 0, 1)$, which represents the distribution obtained by truncation below $-\gamma$ of a $N(0, 1)$ variate.

This form of representation has been obtained by Arnold & Beaver (2004). It is also connected with a discussion, started by Weinstein (1964) and summarized by Nelson (1964), which contains in an implicit form the ingredients leading to the scalar SN distribution.

2. The scalar hidden variable U_0 in (2) can be replaced by a multivariate variable. Correspondingly, there is a set of constraints of type $U_0 > 0$, where it is understood that the inequality must hold for each of the components of U_0. A similar form of multivariate extension has been considered for the convolution representation (6). Various formulations of this type will be discussed more in detail in section 2.2 below.

3. Much work has been carried out replacing the normality assumption by a more general one, especially using densities of the elliptical families. In this paper, an extension of our main argument to the elliptical family is sketched in the final section, but we do not dwell on the connections with other proposals.

Some of the variants which have been put forward occur as reparametrizations of the above formulation, and some comments are in order.

It is algebraically simpler and apparently equivalent to drop the term ω inside the argument of $\Phi(\cdot)$ in (1), but in doing so α looses its interpretation as 'skewness parameter'. To see this, suppose that $Y \sim \mathrm{SN}_d(0, \Omega, \alpha)$, where we take the location parameter to be 0 for simplicity of argument, and consider the scaled version $\tilde{Y} = \tilde{\omega} Y$, for some diagonal matrix $\tilde{\omega}$. In the formulation given above, $\tilde{Y} \sim \mathrm{SN}_d(0, \tilde{\omega}\Omega\tilde{\omega}, \alpha)$ with α unchanged, but this is not the case in the variant parametrization, as its α parameter changes to $\tilde{\omega}^{-1}\alpha$. It is clearly unsatisfactory that a skewness parameter changes its value as an effect of change of scale.

It is true that even α does not reflect only the skewness of the components, as it depends on the shape of the marginals as well as the correlation matrix $\bar{\Omega}$, although at least it does not depend on the scale parameter ω. In this sense, the use of δ in place of α is preferable, as its generic component δ_j is directly related to the index of skewness of the corresponding component Y_j. However, there is the drawback that δ and $\bar{\Omega}$ are not variation independent, as α and $\bar{\Omega}$ are.

Another variant form of (1) starts replacing Ω in (2) with the identity matrix, and set $Z=(U_1 | U_0>0)$. To incorporate a covariance parameter, define further $Y'=\xi+\Omega^{1/2}Z$ for some square root $\Omega^{1/2}$ of a covariance matrix Ω. The resulting density involves then a term $\Omega^{-1/2}$ which is problematic when we come to state a 'definition' of SN distribution, as the square root is not unique, except when Ω is diagonal. Consequently, one must either suffer the fact that the 'definition' effectively defines several distributions or to impose somewhat arbitrarily a specific choice of square root, such as the Cholesky decomposition or the one connected to the spectral decomposition of Ω. The adoption of a specific type of square root has also the disadvantage of becoming hardly acceptable when the skewness parameter is 0, as it would be equivalent to parameterize the normal family by $(\xi, \Omega^{1/2})$ instead of (ξ, Ω).

2. Yet another SN distribution

2.1. Two equivalent representations

We introduce an extension of the basic SN distribution via two equivalent types of constructions. The first of these is based on a representation of type (2), but with U_0 of dimension m, and the new constraint $U_0+\gamma>0$ for some $\gamma\in\mathbb{R}^m$. As already indicated, a notation of type $X>c$, when X and c are vectors, means that each component of X exceeds the corresponding component of c.

Assume that (U_0, U_1) is jointly a multivariate normal variable of dimension $m+d$ with distribution

$$U=\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{m+d}\left(0, \Omega^*\right), \quad \Omega^*=\begin{pmatrix} \Gamma & \Delta^\top \\ \Delta & \bar{\Omega} \end{pmatrix}, \tag{8}$$

where Ω^* is a correlation matrix, and consider the distribution of $Z=(U_1 \mid U_0+\gamma>0)$. In the following development we assume that Ω^* is positive definite; the case where Ω^* has not full rank is of more technical nature and it is confined in appendix.

The density function of a variable of type $(X_1 \mid X_0>c)$, evaluated at point x, is most easily computed via the general relationship

$$f(x)=\frac{f_{X_1}(x)\mathbb{P}\{X_0>c \mid X_1=x\}}{\mathbb{P}\{X_0>c\}}, \tag{9}$$

in an obvious notation. The role of this simple yet important relationship, also clearly valid outside the normal context, has been stressed by Arellano-Valle *et al.* (2002); see their Theorem 5.1.

The case under consideration lends itself to simple calculation, since $(U_0+\gamma \mid U_1=y)$ still has a normal type distribution type with a well-known expression for the parameters. After simple algebra, one obtains that the density function of $Y=\xi+\omega Z$ is

$$f(y)=\phi_d(y-\xi;\Omega)\frac{\Phi_m(\gamma+\Delta^\top\bar{\Omega}^{-1}\omega^{-1}(y-\xi); \Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta)}{\Phi_m(\gamma;\Gamma)} \tag{10}$$

for $y\in\mathbb{R}^d$. For reasons explained in section 2.2, we shall call this expression the 'unified skew-normal' density and, to ease pronunciation, we adopt the acronym SUN. Hence we write $Y\sim \mathrm{SUN}_{d,m}(\xi,\gamma,\bar{\omega},\Omega^*)$, where $\bar{\omega}$ is the vector of the diagonal elements of ω, i.e. $\bar{\omega}=\omega 1_d$, and it is intended that the terms $\Gamma, \Delta, \bar{\Omega}$ required by (10) are recovered by appropriate partitioning of Ω^* as indicated in (8).

There exists another form of genesis for distribution (10), using a convolution mechanism similar to (7) instead of conditioning. Let V_0 and V_1 be independent variables such that

$$V_{0\gamma} \sim \mathrm{LTN}_m(-\gamma; 0, \Gamma), \quad V_1 \sim N_d(0, \Psi), \tag{11}$$

where Γ and Ψ are correlation matrices, and the notation $\mathrm{LTN}_m(c; \mu, \Sigma)$ denotes a multivariate normal variable with components truncated below c, and consider the transformation

$$Y' = \xi + \omega\{B_0 V_{0\gamma} + B_1 V_1\}, \tag{12}$$

where $B_0 = \Delta\Gamma^{-1}$ and B_1 is a $d \times d$ matrix such that

$$B_1 \Psi B_1^\top = \bar{\Omega} - \Delta\Gamma^{-1}\Delta^\top. \tag{13}$$

It is not difficult to show that the two forms of genesis, via conditioning and via (12), are equivalent. To see this, consider U_0 and U_1 with distribution (8) and construct the two independent sets of variables

$$V_0 = U_0, \quad V_1 = U_1 - \mathbb{E}\{U_1 \mid U_0\} = U_1 - \Delta\Gamma^{-1} U_0 \sim N_d(0, \Psi),$$

where $\Psi = \bar{\Omega} - \Delta\Gamma^{-1}\Delta^\top$. Now consider (11) and notice that the distribution of $V_{0\gamma}$ is the same of $(V_0 \mid V_0 + \gamma > 0)$, if $V_0 \sim N_m(0, \Gamma)$. Finally, set $B_0 = \Delta\Gamma^{-1}$, $B_1 = I_d$, and rewrite the convolution form in (12) as

$$\begin{aligned} B_0 V_{0\gamma} + B_1 V_1 &= (B_0 V_0 + B_1 V_1 \mid V_0 + \gamma > 0) \\ &= (B_0 U_0 + B_1(U_1 - \Delta\Gamma^{-1} U_0) \mid U_0 + \gamma > 0) = (U_1 | U_0 + \gamma > 0) \end{aligned} \tag{14}$$

which shows that the representation via convolution corresponds to a representation of conditioning type for another set of variables. On reading the above sequence of relationships backwards, we obtain the converse transformation of representation.

Representation (12) is the analogue of (6), with the main variant that the term $|V_0|$ of the latter is replaced by the $V_{0\gamma}$. The reason for the change is that taking the absolute value of the components of $V_0 \sim N_m(\gamma, \Gamma)$ does not produce any meaningful distribution. For $\gamma = 0$, an exception to this is the special case where Γ is proportional to I_m, a situation which will be discussed separately in section 2.2, in connection with the FUSN family.

One limitation of the basic SN distribution (1) is that, if we partition Z into two components, Z_1 and Z_2, say, then it allows independence between Z_1 and Z_2 only if one of the two has a symmetric normal distribution. This limitation is overcome by the SUN family, which removes that condition for independence; the required conditions on the structure of the matrix Ω^* in (8) are given in appendix. On applying these conditions recursively to the individual subvectors Z_1 and Z_2, it is seen that at most m independent non-symmetrical components are possible.

2.2. *Relationships with similar families*

The aim of this section is to show that the SUN family encompasses all extensions to the basic SN family mentioned in section 1.1, provided redundancies in the parametrization of some proposals are removed.

Basic SN family and a simple extension

It is easy to see that the basic $\mathrm{SN}(\xi, \Omega, \alpha)$ distribution is the same as the $\mathrm{SUN}_{d,1}(\xi, 0, \bar{\omega}, \Omega^*)$ distribution, except for a different form of representation; here $\Gamma = 1$ and Δ is given by (3). Conversely, given that a random variable has distribution $\mathrm{SUN}_{d,1}(\xi, 0, \bar{\omega}, \Omega^*)$, the corresponding $\mathrm{SN}_d(\xi, \Omega, \alpha)$ distribution is uniquely identified, taking into account (5).

A simple extension of the above case is given by $\mathrm{SUN}_{d,1}(\xi, \gamma, \bar{\omega}, \Omega^*)$ which corresponds to the distribution examined by Arnold & Beaver (2000) and by Capitanio *et al.* (2003), again up to change of parametrization.

Closed SN family

González-Farías *et al.* (2004) have examined a form of representation via conditioning similar to ours, and obtained a class of distributions called the CSN family. The main difference from our conditioning mechanism is that the analogue of the covariance matrix appearing in (8) is of the form

$$\begin{pmatrix} \Theta + D\Sigma D^\top & D\Sigma \\ \Sigma D^\top & \Sigma \end{pmatrix}$$

whose ingredients are not required to satisfy any condition, except that $\Sigma > 0$ and $\Theta > 0$. The CSN family is then obtained as the distribution of $Y = \xi + Z$, where $Z = (U_1 \mid U_0 + \gamma > 0)$, and parameterized as $\mathrm{CSN}_{d,m}(\xi, \Sigma, D, \gamma, \Theta)$, except that we have introduced a slight change of notation.

However, the lack of scale constraints on the above covariance matrix makes the parametric class not identifiable. To see this, consider a diagonal matrix G which can be chosen arbitrarily provided its diagonal terms are all positive, and notice that the condition $U_0 + \gamma > 0$ is equivalent to $GU_0 + G\gamma > 0$. Therefore, $\mathrm{CSN}_{d,m}(\xi, \Sigma, D, \gamma, \Theta)$ and $\mathrm{CSN}_{d,m}(\xi, \Sigma, GD, G\gamma, G\Theta G)$ refer to the same probability distribution.

Another critical aspect of the adopted parametrization is D, which is described as a skewness parameter. In fact, D is not invariant to changes of scale, for reasons analogous to those discussed in section 1.3.

Once the above aspects of the parametrization are adjusted, the SUN and CSN classes are equivalent, on setting $\bar{\Omega} = \Sigma$, $\Gamma = \Theta + D\Sigma D^\top$ and $\Delta = \Sigma D^\top$.

CSN-2 family

Another extension of the basic SN family is given by the density

$$f(y) = \phi_d(y - \xi; \Psi + D^\top \Gamma D) \frac{\Phi_m(\gamma + \Gamma D(\Psi + D^\top \Gamma D)^{-1}(y - \xi); (\Gamma^{-1} + D\Psi^{-1}D^\top)^{-1})}{\Phi_m(\gamma; \Gamma)} \tag{15}$$

which is an extension of a distribution considered by Arellano-Valle & Genton (2005); when $\xi = 0$, $\gamma = 0$ and $\Gamma = I_m$, this density reduces to their (1.9); that density is in turn an extension of the family introduced by Sahu *et al.* (2003).

As it can be shown that, similarly to the earlier CSN family, the family of densities (15) is closed by marginalization and conditioning, then we refer to it as the CSN-2 class. If a random variable Y has density function (15), we write $Y \sim \mathrm{CSNII}_{d,m}(\xi, \gamma, \Psi, \Gamma, D)$.

The above density function can be obtained by a conditioning mechanism, similar to those discussed earlier, on setting

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} \sim N_{m+d}(0, \Omega^{**}), \quad \Omega^{**} = \begin{pmatrix} \Gamma & \Gamma D \\ D^\top \Gamma & \Psi + D^\top \Gamma D \end{pmatrix}$$

and considering $Y = \xi + Z$, where $Z = (U_1 \mid U_0 + \gamma > 0)$. An alternative construction of this distribution will be discussed later. A peculiar aspect of this density is that the skewness matrix D enters into the density ϕ_d which represents the 'symmetric part' of the density.

Even for this class there is a problem of overparametrization if the covariance matrices are not restricted. In fact $\mathrm{CSNII}_{d,m}(\xi, G\gamma, \Psi, G\Gamma, G^{-1}D)$, which corresponds to the distribution of $(U_1 \mid GU_0 + G\gamma > 0)$ for any diagonal matrix $G > 0$, and $\mathrm{CSNII}_{d,m}(\xi, \gamma, \Psi, \Gamma, D)$ refers to the same density. The coincidence with the SUN family is achieved on setting $\bar{\Omega} = \Psi + D\Gamma D^\top$ and $\Delta = D^\top \Gamma$.

Hierarchical SN family

The HSN distribution was obtained by Liseo & Loperfido (2003) within a Bayesian context, but the purely probabilistic argument can be linked to our construction as follows. If $\theta_0 \sim N_d(0, \Sigma)$ and $\theta_1 \sim N_d(0, \Upsilon)$ are independent variables, and C is a full-rank $m \times d$ matrix $(1 \le m \le d)$, set

$$\begin{pmatrix} U_0 \\ U_1 \end{pmatrix} = \begin{pmatrix} -C\theta_0 \\ \theta_0 + \theta_1 \end{pmatrix} \sim N_{m+d}\left(0, \begin{pmatrix} C\Upsilon C^\top & -C\Upsilon \\ -\Upsilon C^\top & \Upsilon + \Sigma \end{pmatrix}\right)$$

and define the HSN distribution is the one for the variable $\mu + (U_1 | U_0 > c)$, parameterized as $\mathrm{HSN}_d(\mu, c, \Sigma, \Upsilon, C)$.

However, there is a problem of over-parametrization, very similar to the one of the CSN family, since even here it originates by the lack of suitable scale constraints. The condition $U_0 > c$ is the same of $GU_0 > Gc$ if G is as above. Hence the notations $\mathrm{HSN}_d(\mu, c, \Sigma, \Upsilon, C)$ and $\mathrm{HSN}_d(\mu, Gc, \Sigma, \Upsilon, GC)$ refer to the same distribution.

Fundamental SN family

The FUSN family studied by Arellano-Valle & Genton (2005) was obtained via a convolution mechanism which generalizes (6), rather than via the use of conditioning. An analogous representation using conditioning is however not difficult to construct. In its canonical form, $\mathrm{CFUSN}_{d,m}(\Delta)$ say, the distribution coincides with $\mathrm{SUN}_{d,m}(0, 0, 1_d, \Omega^*)$, where

$$\Omega^* = \begin{pmatrix} I_m & \Delta^\top \\ \Delta & I_d \end{pmatrix}.$$

This form does not include location and scale parameter, and a dependence structure separate from the one induced by the skewness parameter Δ. To insert these ingredients starting from a variable Z with the distribution of above type, one sets $Y = \mu + \Sigma^{1/2} Z$, for some positive definite matrix $\Sigma^{1/2}$. This choice involves the use of a square root matrix $\Sigma^{1/2}$, whose critical aspects have been already discussed in section 1.3.

The following table summarizes the correspondence between the parametrization of the SUN and those of the other families, assuming for simplicity that all scale factors are 1s and the location parameters are 0:

SUN	$\bar{\Omega}$	Γ	Δ
CSN	Σ	$\Theta + D\Sigma D^\top$	$(D\Sigma)^\top$
CSN-2	$\Psi + D^\top \Gamma D$	Γ	$(\Gamma D)^\top$
HSN	$\Upsilon + \Sigma$	$C\Upsilon C^\top$	$-(C\Upsilon)^\top$
CFUSN	I_d	I_m	Δ

2.3. *Some properties of the SUN*

We do not discuss in detail the formal properties of SUN family, as these coincide with those of the related families discussed earlier, once the parameterizations are properly selected. However, a brief account of the main facts seems in order.

Taking into account that the moment generating function of $V_{0\gamma}$ evaluated at s is

$$\exp(\tfrac{1}{2} s^\top \Gamma s) \frac{\Phi_m(\gamma + \Gamma s; \Gamma)}{\Phi_m(\gamma; \Gamma)}, \quad (s \in \mathbb{R}^m),$$

it follows that, under conditions (13), the moment generating function of (10) is

$$M(t) = \exp(\xi^\top t + \tfrac{1}{2} t^\top \Omega t) \frac{\Phi_m(\gamma + \Delta^\top \omega t; \Gamma)}{\Phi_m(\gamma; \Gamma)} \quad (t \in \mathbb{R}^d). \tag{16}$$

Moments and cumulants can be obtained directly from (16) or from suitable adaption of expressions given by Gupta *et al.* (2004). The derivation of moments is simplified when $\Gamma=\mathrm{diag}(\tau_1^2,\dots,\tau_m^2)$, as the cumulant generating function reduces to

$$K(t)=\log M(t)=\xi^\top t+\tfrac{1}{2}t^\top\Omega t+\sum_{j=1}^{m}\log\Phi(\tau_j^{-1}\gamma_j+\tau_j^{-1}\delta_{\cdot j}^\top\omega t)-\log\Phi(\gamma;\Gamma),$$

where $\delta_{\cdot 1},\dots,\delta_{\cdot m}$ are the columns of Δ. From this expression, we obtain

$$\mathbb{E}\{Y\}=K'(0)=\xi+\sum_{j=1}^{m}\zeta_1(\tau_j^{-1}\gamma_j)\tau_j^{-1}\omega\delta_{\cdot j}$$

and

$$\mathrm{var}\{Y\}=K''(0)=\Omega+\sum_{j=1}^{m}\zeta_2(\tau_j^{-1}\gamma_j)\tau_j^{-2}\omega\delta_{\cdot j}\delta_{\cdot j}^\top\omega,$$

where $\zeta_r(x)$ is the rth derivative of $\zeta_0(x)=\log\{2\Phi(x)\}$.

It is easy to see from the construction of section 2.1, in particular via the conditioning mechanism, that the SUN family is closed under marginalization and under conditioning, similarly to the related families discussed in section 2.2. Explicit expressions of the parameters of the marginal and conditional distributions are given in appendix.

Consider next the distribution of a quadratic form of type $Q(Z)=Z^\top AZ$, where $Z\sim\mathrm{SUN}_{d,m}(0,\gamma,1_d,\Omega^*)$ and A is a $d\times d$ symmetric matrix of rank p. It can be shown that the moment generating function of $Q(Z)$ is

$$M_Q(t)=|I_d-2tA\bar\Omega|^{-1/2}\frac{\Phi_m(\gamma;\Gamma+2t\Delta^\top(I_d-2tA\bar\Omega)^{-1}A\Delta)}{\Phi_m(\gamma;\Gamma)}.$$

An important special case is when $A=\bar\Omega^{-1}$. To obtain $Q(Z)\sim\chi_d^2$ similarly to the case when Z has normal or SN distribution, we need two conditions: (a) that $|I_d-2tA\bar\Omega|^{-1/2}=(1-2t)^{-d/2}$, which holds true if $A=\bar\Omega^{-1}$; (b) that the numerator and denominator of the fraction in $M_Q(t)$ cancel out. The latter condition is satisfied when $\gamma=0$ if $m=1$, and when $A\Delta=0$ if $m\geq 1$.

3. Extension to the skew-elliptical family

The aim of this section is to extend the results of the previous sections from the normal framework to the one of elliptical distributions, at least to some extent. In the development, we shall make use of some standard results on elliptical distributions, which we do not reproduce here in detail, as they can be found in standard references (see, e.g. Fang *et al.*, 1990).

To start with, replace the normality assumption in (8) by the assumption for an elliptical density; hence write

$$U=\begin{pmatrix}U_0\\U_1\end{pmatrix}\sim\mathrm{El}_{m+d}(0,\Omega^*;h_{m+d}),\qquad \Omega^*=\begin{pmatrix}\Gamma & \Delta^\top\\ \Delta & \bar\Omega\end{pmatrix},$$

where again Ω^* is a correlation matrix. Here h_{m+d} is the so-called density generator, such that the density of U is

$$f_{m+d}(u;\Omega^*)=|\Omega^*|^{-1/2}h_{m+d}\left(u^\top(\Omega^*)^{-1}u\right), \tag{17}$$

for $u\in\mathbb{R}^{m+d}$; we shall denote its distribution function by $F_{m+d}(u;\Omega^*)$. The two components of U have distribution of type

$$U_0\sim\mathrm{El}_m(0,\Gamma;h_m),\qquad U_1\sim\mathrm{El}_d(0,\bar\Omega;h_d),$$

whose density generators h_m and h_d are computed from h_{m+d} using formula (2.23) of Fang *et al.* (1990); replacing these generators in (17) in place of h_{m+d} provides the expressions for their densities. The conditional distribution of U_0 given $U_1=y$ is

$$(U_0 \mid U_1=y)=(U_0 \mid q(U_1)=q(y)) \sim \mathrm{El}_m(\Delta^\top \bar{\Omega}^{-1} y, \Gamma - \Delta^\top \bar{\Omega}^{-1}\Delta, h_{m;q(y)}),$$

where $q(y)=y^\top \bar{\Omega}^{-1} y$ and $h_{m;a}(u)=h_{m+d}(u+a)/h_d(a)$ denotes the density generator.

Proceeding in the same fashion as in the normal case, we consider the variable $Y=\xi+\omega Z$, where $Z=(U_1 \mid U_0+\gamma>0)$, and use the notation $Y\sim \mathrm{SUEl}_{d,m}(\xi,\gamma,\bar{\omega},\Omega^*;h_{m+d})$. Its density function can be computed inserting the above ingredients in (9), leading to

$$f(y)=f_d(y-\xi;\Omega)\frac{F_{m;q[\omega^{-1}(y-\xi)]}(\gamma+\Delta^\top\bar{\Omega}^{-1}\omega^{-1}(y-\xi);\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta)}{F_m(\gamma;\Gamma)}, \tag{18}$$

where $F_{m;a}(x;\Theta)$ denotes the distribution function of the $\mathrm{El}_m(0,\Theta,h_{m;a})$ distribution, namely

$$F_{m;a}(x;\Theta)=|\Theta|^{-1/2}\int_{v\le x}\frac{h_{m+d}(v^\top\Theta^{-1}v+a)}{h_d(a)}\,\mathrm{d}v.$$

Various variant forms can now be obtained from (18) similarly to the normal context above. We focus attention on the restriction $\gamma=0$, which has played an important role also for the normality case. Under this assumption, the simplification $F_m(0;\Gamma)=\Phi_m(0;\Gamma)$ holds (see, e.g. Fang *et al.*, 1990, pp. 53–54).

An important class of elliptical distributions is generated by the scale mixtures of the normal variables. For any elliptical distribution in this class, there is a non-negative random variable S such that $U \mid S=s\sim N_{m+d}(0,s^{-1}\Omega^*)$, with S having distribution function H. In this case

$$(Z\mid S=s)=(U_1 \mid U_0+\gamma>0, S=s)\sim \mathrm{SUN}_{d,m}(0,\gamma,1_d,s^{-1}\Omega^*)$$

and, using (10), the unconditional density of Z is

$$g(z)=\int_0^\infty s^{d/2}\phi_d(\sqrt{s}\,z;\bar{\Omega})\frac{\Phi_m(\sqrt{s}\gamma+\sqrt{s}\Delta^\top\bar{\Omega}^{-1}z;\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta)}{\Phi_m(\sqrt{s}\gamma;\Gamma)}\,\mathrm{d}H(s),$$

which for $\gamma=0$ reduces to

$$g(z)=\frac{1}{\Phi_m(0;\Gamma)}\int_{x\le\Delta^\top\bar{\Omega}^{-1}z}\int_0^\infty s^{(m+d)/2}\phi_d(\sqrt{s}z;\bar{\Omega})\phi_m(\sqrt{s}x;\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta)\,\mathrm{d}H(s)\,\mathrm{d}x,$$

from which one can obtain the density of $Y=\xi+\omega Z$, at least for some convenient forms of $H(s)$.

An important special case of the above density $g(z)$ occurs when $S\sim\mathrm{Gamma}(\nu/2,\nu/2)$, as it is well known that this mixing distribution applied to the normal variables produces the multivariate t distribution. After some algebraic manipulations, similar to those employed by Arellano-Valle & Bolfarine (1995), in this case $g(z)$ takes the special form

$$\begin{aligned}
g_T(z)&=\frac{c_{m+d,\,\nu}}{\Phi_m(0;\Gamma)|\bar{\Omega}|^{1/2}|\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta|^{1/2}}\\
&\quad\times\int_{x\le\Delta^\top\bar{\Omega}^{-1}z}\{\nu+q(z)+x^\top(\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta)^{-1}x\}^{-(m+d+\nu)/2}\,\mathrm{d}x\\
&=\frac{1}{\Phi_m(0;\Gamma)}t_d(z;\bar{\Omega},\nu)\int_{x\le\Delta^\top\bar{\Omega}^{-1}z}t_m\left(x;\frac{\nu+q(z)}{\nu+d}(\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta),\nu+d\right)\mathrm{d}x\\
&=\frac{1}{\Phi_m(0;\Gamma)}t_d(z;\bar{\Omega},\nu)T_m\left(\Delta^\top\bar{\Omega}^{-1}z;\frac{\nu+q(z)}{\nu+d}(\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta),\nu+d\right),
\end{aligned}$$

where $q(z)$ is defined above,

$$c_{k,\upsilon} = \frac{\Gamma[(1/2)(k+\upsilon)]\upsilon^{\upsilon/2}}{\Gamma((1/2)\,\upsilon)\pi^{k/2}},$$

and $t_k(x;\Theta,\upsilon)$, $T_k(x;\Theta,\upsilon)$ denote the density and distribution functions, respectively, of a k-dimensional t-distribution with dispersion matrix Θ and υ degrees of freedom.

We have then obtained a d-dimensional density function of the skew-t type. When $m=1$, this $g_T(z)$ reduces to a form of skew-t already proposed in the literature. This connection is more directly visible if one compares the density $g_T(z)$ with formula (26) of Azzalini & Capitanio (2003) and formula (2.3) of Gupta (2003); these densities are in turn equivalent to the skew-t density of Branco & Dey (2001), although expressed in a different form.

Under the assumption $\gamma=0$, it is possible to establish a connection between the representations of $\mathrm{SUEl}_{d,m}$ distribution via conditioning examined above and a representation via convolution. We omit the algebraic details.

Acknowledgements

We are grateful to Antonella Capitanio, Marc Genton, two referees and an associate editor for several remarks to the first draft of this paper, leading to a much improved version of the material. A substantial part of this work has been developed while the second author was visiting the Departamento de Estadística de la Pontificia Universidad Católica de Chile, whose kind hospitality is gratefully acknowledged. The work of the first author has been supported by Fondecyt, Chile.

References

Arellano-Valle, R. B. & Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statist. Probab. Lett.* **25**, 79–85.

Arellano-Valle, R. B. & Genton, M. G. (2005). On fundamental skew distributions. *J. Multivariate Anal.* **96**, 93–116.

Arellano-Valle, R. B., del Pino, G. & San Martín, E. (2002). Definition and probabilistic properties of skew-distribution. *Statist. Probab. Lett.* **58**, 111–121.

Arnold, B. C. & Beaver, R. J. (2000). Hidden truncation models. *Sankhyā Ser. A* **62**, 23–25.

Arnold, B. C. & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation. *Test* **11**, 7–54.

Arnold, B. C. & Beaver, R. J. (2004). Alternative construction of skewed multivariate distributions. *Acta Comment. Univ. Tartu. Math.* **8**, 1–9.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.* **32**, 159–188 (CR: 189–200).

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

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

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

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

Capitanio, A. C., Azzalini, A. & Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scand. J. Statist.* **30**, 129–144.

Fang, K.-T., Kotz, S. & Ng, K. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall, London.

Genton, M. G. (ed.) (2004). *Skew-elliptical distributions and their applications: a journey beyond normality*. Chapman & Hall/CRC, Boca Raton.

González-Farías, G., Domínguez-Molina, A. & Gupta, A. K. (2004). Additive properties of skew normal random vectors. *J. Statist. Plann. Inference* **126**, 521–534.

Gupta, A. K. (2003). Multivariate skew t-distribution. *Statistics* **37**, 359–363.

Gupta, A. K., González-Farías, G. & Domínguez-Molina, A. (2004). A multivariate skew normal distribution. *J. Multivariate Anal.* **89**, 181–190.
Liseo, B. & Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statist. Probab. Lett.* **61**, 395–401.
Nelson, L. S. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics* **6**, 469–471.
Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Statist.* **31**, 129–150.
Weinstein, M. A. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics* **6**, 104–105.

Received October 2004, in final form January 2006

Adelchi Azzalini, Dipartimento di Scienze Statistiche, Università di Padova, via C. Battisti 241, 35121 Padova, Italy.
E-mail: azzalini@stat.unipd.it

Appendix A: conditional and marginal distribution

Consider a variable $Y \sim \mathrm{SUN}_{d,m}(\xi, \gamma, \bar{\omega}, \Omega^*)$, where Ω^* is as in (2), and partition it as $Y = (Y_1^\top, Y_2^\top)^\top$ where Y_i has dimension $d_i \times 1$, for $i = 1, 2$ with $d_1 + d_2 = d$. We want to obtain expressions of the marginal distribution of Y_1 and the conditional distribution of Y_1 given that $Y_2 = y_2$, for any $y_2 \in \mathbb{R}^{d_2}$.

The above partition of Y induces corresponding partitions of $\xi = (\xi_1^\top, \xi_2^\top)^\top$, $\bar{\omega} = (\bar{\omega}_1^\top, \bar{\omega}_2^\top)^\top$,

$$\Delta = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}, \quad \bar{\Omega} = \begin{pmatrix} \bar{\Omega}_{11} & \bar{\Omega}_{12} \\ \bar{\Omega}_{21} & \bar{\Omega}_{22} \end{pmatrix}, \tag{19}$$

where ξ_i and $\bar{\omega}_i$ are of dimension $d_i \times 1$, Δ_i has dimension $d_i \times m$, and $\bar{\Omega}_{ij}$ is of dimension $d_i \times d_j$, for $i, j = 1, 2$. Keeping in mind the representation of Y via conditioning, the distribution of Y_1 is of the same type of (10) provided only the first d_1 components of U_1 in (2) are considered. It then follows immediately that $Y_1 \sim \mathrm{SUN}_{d_1,m}(\xi_1, \gamma, \bar{\omega}_1, \Omega_1^*)$, where

$$\Omega_1^* = \begin{pmatrix} \Gamma & \Delta_1^\top \\ \Delta_1 & \bar{\Omega}_{11} \end{pmatrix}.$$

Computation of the conditional density of Y_1 given $Y_2 = y_2$ can be accomplished by direct calculation which leads to

$$\begin{aligned} f_{Y_1|Y_2=y_2}(y_1) &= \phi_{d_1}(y_1 - \xi_{1.2}; \Omega_{11.2}) \frac{\Phi_m(\gamma + \Delta^\top \bar{\Omega}^{-1} \omega^{-1}(y - \xi); \Gamma - \Delta^\top \bar{\Omega}^{-1} \Delta)}{\Phi_m(\gamma_{1.2}; \Gamma_{1.2})} \\ &= \phi_{d_1}(y_1 - \xi_{1.2}; \Omega_{11.2}) \frac{\Phi_m(\gamma_{1.2} + \Delta_{1.2}^\top \bar{\Omega}_{11.2}^{-1} \omega_1^{-1}(y_1 - \xi_{1.2}); \Gamma_{1.2} - \Delta_{1.2}^\top \bar{\Omega}_{11.2}^{-1} \Delta_{1.2})}{\Phi_m(\gamma_{1.2}; \Gamma_{1.2})}, \end{aligned}$$

where $\xi_{1.2} = \xi_1 + \Omega_{12}\Omega_{22}^{-1}(y_2 - \xi_2)$, $\gamma_{1.2} = \gamma + \Delta_2^\top \bar{\Omega}_{22}^{-1} \omega_2^{-1}(y_2 - \xi_2)$, $\bar{\Omega}_{11.2} = \bar{\Omega}_{11} - \bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\bar{\Omega}_{21}$, $\Omega_{11.2} = \omega_1 \bar{\Omega}_{11.2} \omega_1 = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$, with $\Omega_{ij} = \omega_i \bar{\Omega}_{ij} \omega_j$, $i, j = 1, 2$, $\Delta_{1.2} = \Delta_1 - \bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\Delta_2$ and $\Gamma_{1.2} = \Gamma - \Delta_2^\top \bar{\Omega}_{22}^{-1} \Delta_2$. We then obtain that $Y_1 | Y_2 = y_2 \sim \mathrm{SUN}_{d_1, m}(\xi_{1.2}, \gamma_{1.2}, \bar{\omega}_1, \Omega_{11.2}^*)$, where

$$\Omega_{1.2}^* = \begin{pmatrix} \Gamma_{1.2} & \Delta_{1.2}^\top \\ \Delta_{1.2} & \bar{\Omega}_{11.2} \end{pmatrix}.$$

Appendix B: conditions for independence

We examine conditions under which partition of a random variable $\mathrm{SUN}_d(0, \gamma, 1_d, \Omega^*)$ into two components attains independence of these two variables. The question is elementary if we allow either of these variables to be regular normal, as the conditions are the same for independence between multivariate normal variables. We then discuss the case where both subcomponents are non-symmetric.

For (U_0, U_1) distributed as in (8), set $Z=(U_1 \mid U_0+\gamma>0)$, and consider the partition of Z in components of size d_1 and d_2, and the associated partition of U_1 of type $Z^\top=(Z_1^\top, Z_2^\top)$ and $U_1^\top=(U_{11}^\top, U_{12}^\top)$ and the corresponding partition of $\bar{\Omega}$ as (19). From the results of appendix A,

$$Z_1=(U_{11}|U_0+\gamma>0)\sim \mathrm{SUN}_{d_1,m}(0,\gamma,1_{d_1},\Omega_1^*)$$

and a similar fact holds for Z_2.

Consider further a partition of U_0 in blocks of size m_1 and m_2, respectively, for some integers such that $m_1+m_2=m$, and write

$$U_0=\begin{pmatrix}U_{01}\\ U_{02}\end{pmatrix},\quad \gamma=\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix},\quad \Gamma=\begin{pmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{pmatrix},\quad \begin{pmatrix}\Delta_1\\ \Delta_2\end{pmatrix}=\begin{pmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{pmatrix},$$

where U_{0i} and γ_i have dimension $m_i\times 1$, Γ_{ij} has dimension $m_i\times m_j$, and Δ_{ij} has dimension $d_i\times m_j$ $(i,j=1,2)$. As

$$Z_1=(U_{11}|U_{01}+\gamma_1>0, U_{02}+\gamma_2>0),\quad Z_2=(U_{12}|U_{01}+\gamma_1>0, U_{02}+\gamma_2>0),$$

then there are two situations under which Z_1 and Z_2 are independent:

(i) if (U_{11}, U_{01}) and (U_{12}, U_{02}) are independent, i.e.

$$Z_1\stackrel{d}{=}(U_{11}|U_{01}+\gamma_1>0),\quad Z_2\stackrel{d}{=}(U_{12}|U_{02}+\gamma_2>0),$$

which occurs if

$$\Gamma_{12}=0,\quad \bar{\Omega}_{12}=0,\quad \Delta_{12}=0,\quad \Delta_{21}=0; \tag{20}$$

(ii) if (U_{11}, U_{02}) and (U_{12}, U_{01}) are independent, i.e.

$$Z_1\stackrel{d}{=}(U_{11}|U_{02}+\gamma_2>0),\quad Z_2\stackrel{d}{=}(U_{12}|U_{01}+\gamma_1>0),$$

which occurs if

$$\Gamma_{12}=0,\quad \bar{\Omega}_{12}=0,\quad \Delta_{11}=0,\quad \Delta_{22}=0. \tag{21}$$

The above conditions are sufficient for independence; we now turn to necessary conditions. It is clear that the conditions $\bar{\Omega}_{12}=\mathrm{cov}\{U_{11}, U_{12}\}=0$, $\Gamma_{12}=\mathrm{cov}\{U_{01}, U_{02}\}=0$ are necessary, as we need U_{11} to be independent of U_{12} and U_{01} independent of U_{02}, respectively, in order that $\phi_d(z;\bar{\Omega})$ factorizes as $\phi_{d_1}(z_1;\bar{\Omega}_{11})\phi_{d_2}(z_2;\bar{\Omega}_{22})$ and $\Phi_m(\gamma;\Gamma)$ factorizes as $\Phi_{m_1}(\gamma_1;\Gamma_{11})\times\Phi_{m_2}(\gamma_2;\Gamma_{22})$.

In order to examine the conditions on the matrix Δ, let

$$\bar{\Omega}^{-1}=\begin{pmatrix}\bar{\Omega}^{11} & \bar{\Omega}^{12}\\ \bar{\Omega}^{21} & \bar{\Omega}^{22}\end{pmatrix}=\begin{pmatrix}\bar{\Omega}_{11.2}^{-1} & -\bar{\Omega}_{11.2}^{-1}\bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\\ -\bar{\Omega}_{22}^{-1}\bar{\Omega}_{21}\bar{\Omega}_{11.2}^{-1} & \bar{\Omega}_{22}^{-1}+\bar{\Omega}_{22}^{-1}\bar{\Omega}_{21}\bar{\Omega}_{11.2}^{-1}\bar{\Omega}_{12}\bar{\Omega}_{22}^{-1}\end{pmatrix}$$

and, under the conditions $\bar{\Omega}_{12}=0$, $\Gamma_{12}=0$, we have that

$$\gamma+\Delta^\top\bar{\Omega}^{-1}z=\begin{pmatrix}\gamma_1+\Delta_{11}^\top\bar{\Omega}_{11}^{-1}z_1+\Delta_{21}^\top\bar{\Omega}_{22}^{-1}z_2\\ \gamma_2+\Delta_{12}^\top\bar{\Omega}_{11}^{-1}z_1+\Delta_{22}^\top\bar{\Omega}_{22}^{-1}z_2\end{pmatrix},$$

and

$$\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta=\begin{pmatrix}\Gamma_{11}-(\Delta_{11}^\top\bar{\Omega}_{11}^{-1}\Delta_{11}+\Delta_{21}^\top\bar{\Omega}_{22}^{-1}\Delta_{21}) & -(\Delta_{11}^\top\bar{\Omega}_{11}^{-1}\Delta_{12}+\Delta_{21}^\top\bar{\Omega}_{22}^{-1}\Delta_{22})\\ -(\Delta_{12}^\top\bar{\Omega}_{11}^{-1}\Delta_{11}+\Delta_{22}^\top\bar{\Omega}_{22}^{-1}\Delta_{21}) & \Gamma_{22}-(\Delta_{12}^\top\bar{\Omega}_{11}^{-1}\Delta_{12}+\Delta_{22}^\top\bar{\Omega}_{22}^{-1}\Delta_{22})\end{pmatrix}.$$

Therefore, $\Phi_m(\gamma+\Delta^\top\bar{\Omega}^{-1}z;\Gamma-\Delta^\top\bar{\Omega}^{-1}\Delta)$ can be factorized in either of the following two ways:

(i) when $\Delta_{21}^\top\bar{\Omega}_{22}^{-1}=0$ and $\Delta_{12}^\top\bar{\Omega}_{11}^{-1}=0$, as

$$\Phi_{m_1}(\gamma_1+\Delta_{11}^\top\bar{\Omega}_{11}^{-1}z_1;\Gamma_{11}-\Delta_{11}^\top\bar{\Omega}_{11}^{-1}\Delta_{11})\Phi_{m_2}(\gamma_2+\Delta_{22}^\top\bar{\Omega}_{22}^{-1}z_2;\Gamma_{22}-\Delta_{22}^\top\bar{\Omega}_{22}^{-1}\Delta_{22});$$

(ii) or when $\Delta_{11}^\top \bar{\Omega}_{11}^{-1} = 0$ and $\Delta_{22}^\top \bar{\Omega}_{22}^{-1} = 0$, as

$$\Phi_{m_1}(\gamma_1 + \Delta_{21}^\top \bar{\Omega}_{22}^{-1} z_1; \Gamma_{11} - \Delta_{21}^\top \bar{\Omega}_{22}^{-1} \Delta_{21}) \Phi_{m_2}(\gamma_2 + \Delta_{12}^\top \bar{\Omega}_{11}^{-1} z_2; \Gamma_{22} - \Delta_{12}^\top \bar{\Omega}_{11}^{-1} \Delta_{12}).$$

Consequently, the conditions for independence are either of the following set:

(i) $\bar{\Omega}_{12} = 0$, $\Gamma_{12} = 0$, $\Delta_{21}^\top \bar{\Omega}_{22}^{-1} = 0$ and $\Delta_{12}^\top \bar{\Omega}_{11}^{-1} = 0$;
(ii) $\bar{\Omega}_{12} = 0$, $\Gamma_{12} = 0$, $\Delta_{11}^\top \bar{\Omega}_{11}^{-1} = 0$ and $\Delta_{22}^\top \bar{\Omega}_{22}^{-1} = 0$.

It is immediate to check that conditions (i) are equivalent to the sufficient conditions (20); this formal equivalence does not imply that they are necessary and sufficient conditions, since what is necessary is that either of the sets (i) and (ii) holds true. An analogous remark holds for conditions (ii) and (21).

Notice that, for the above conditions refer to a given partition of Z and a given partition of U_0. However, the actual question of independence between Z_1 and Z_2 is not linked to a specific partition of U_0. As, if $m > 2$, the partition of U_0 in two subcomponents is not unique, the above conditions must be checked for all possible partitions of U_0. Luckily, m can be expected to be small in most cases, as otherwise the mere computation of the density (10) becomes burdensome; hence the number of cases to be scanned is likely to be limited.

Appendix C: the singular SUN distribution

The development of section 2.1 has assumed that Ω^* in (8) is non-singular. Here, we discuss the similar problem when this condition is removed. The development follows the pattern of Rao (1973) for the singular normal distribution; see his section 8a.

Assume then that $\operatorname{rank}(\Omega^*) = r < m + d$, hence the distribution of U in (8) is singular normal. Recall that, in this case, then one can write $U = \mathrm{C}W$, where $\mathrm{CC}^\top = \Omega^*$ and $W \sim N_r(0, I_r)$. We discuss separately some possible cases.

Consider first that $\operatorname{rank}(\Gamma) = r < m$ and $\operatorname{rank}(\bar{\Omega}) = d$, and define $V_1 = U_1 - E(U_1 | U_0) = U_1 - \Delta\Gamma^- U_0$, where Γ^- is a generalized inverse of Γ according to Rao (1973, p. 24) definition. It is immediate to obtain

$$\mathbb{E}\{V_1\} = 0, \quad \operatorname{var}\{V_1\} = \bar{\Omega} + \Delta\Gamma^-\Gamma(\Gamma^-)^\top\Delta^\top - \Delta(\Gamma^-)^\top\Delta^\top - \Delta\Gamma^-\Delta^\top = \bar{\Omega} - \Delta\Gamma^-\Delta^\top$$

and $\operatorname{cov}\{V_1, U_0\} = \Delta - \Delta\Gamma^-\Gamma = 0$, since $\Delta\Gamma^-\Gamma = \Delta$, on using Rao (1973) formula (8a.2.12). Hence, $V_1 \sim N_d(0, \bar{\Omega} - \Delta\Gamma^-\Delta^\top)$ and is independent of U_0. Similarly to the non-singular case, we define

$$Z = (U_1 | U_0 + \gamma > 0) = \Delta\Gamma^- U_{0\gamma} + V_1$$

as having a SUN distribution. To show that its moment generating function $M(t)$ is of type (16), write $U_0 = C_0 W$, where $W \sim N_r(0, I_r)$ and $C_0 C_0^\top = \Gamma$. Hence

$$M(t) = M_{0\gamma}((\Gamma^-)^\top \Delta^\top t) M_1(t), \tag{22}$$

where $M_1(t) = \mathbb{E}\{e^{t^\top V_1}\} = \exp(\frac{1}{2} t^\top (\bar{\Omega} - \Delta\Gamma^-\Delta^\top) t)$, and

$$\begin{aligned} M_{0\gamma}(s) &= E[e^{s^\top U_0} | U_0 + \gamma > 0] = \frac{\int_{\{w: C_0 w + \gamma > 0\}} e^{s^\top C_0 w} \phi_r(w; I_r)\, dw}{\Phi_m(\gamma; \Gamma)} \\ &= e^{(1/2) s^\top \Gamma s} \frac{\int_{\{w: C_0 w + \gamma > 0\}} \phi_r(w - C_0^\top s; I_r)\, dw}{\Phi_m(\gamma; \Gamma)} = e^{(1/2) s^\top \Gamma s} \frac{\int_{\{u: C_0 u + \Gamma s + \gamma > 0\}} \phi_r(u; I_r)\, du}{\Phi_m(\gamma; \Gamma)} \\ &= e^{(1/2) s^\top \Gamma s} \frac{\Phi_m(\Gamma s + \gamma; \Gamma)}{\Phi_m(\gamma; \Gamma)}. \end{aligned}$$

Now, on using again that $\Delta\Gamma^{-}\Gamma=\Delta$, we have

$$M_{0\gamma}((\Gamma^{-})^{\top}\Delta^{\top}t)=\mathrm{e}^{(1/2)t^{\top}\Delta\Gamma^{-}\Delta^{\top}t}\frac{\Phi_m(\Delta^{\top}t+\gamma;\Gamma)}{\Phi_m(\gamma;\Gamma)},$$

which, when replaced in (22), gives for $M(t)$ the expression (16), except that here we have not included location and scale parameters. Therefore, the form of $M(t)$ does not depend on whether Γ has full rank or not. The only impact of rank deficiency is of computational nature, as in this case

$$\Phi_m(a;\Gamma)=\mathbb{P}\{C_0W\le a\}=\int_{\{w:C_0w\le a\}}\phi_r(w;I_r)\,\mathrm{d}w. \tag{23}$$

Therefore, if $\mathrm{rank}(\bar{\Omega})=d$, the density of the SUN distribution is given by (10), irrespectively of the fact that Γ is singular or not. The only difference is on the computation of $\Phi_m(a;\Gamma)$ according to (23). In fact, this conclusion is a direct consequence of the fact that, if $\mathrm{rank}(\Gamma)<m$, we can reduce the number of constraints of type $U_0+\gamma>0$ accordingly, leading to a new matrix Γ of full rank.

Consider next the case when $\mathrm{rank}(\bar{\Omega})=r<d$. Without loss of generality we assume that Γ has full rank m, for the reason explained above. Arguing similarly to Rao (1973, pp. 527–528), consider the transformation $U_{11}=BU_1$ and $U_{21}=NU_1$, where B is a $r\times d$ matrix of orthonormal row vectors belonging to the linear space spanned by $\bar{\Omega}$, and N is a $(d-r)\times d$ matrix such that $N\bar{\Omega}=0$. Thus, $U_{21}=0$ with probability 1 and by $U_{11}\sim N_r(0,B\bar{\Omega}B^{\top})$, where $|B\bar{\Omega}B^{\top}|$ is positive as it is equal to the product $\lambda_1\times\cdots\times\lambda_r$ of the non-zero eigenvalues of $\bar{\Omega}$. Now, by letting $Z_1=(U_{11}|U_0+\gamma>0)=BZ$ and $Z_2=(U_{21}|U_0+\gamma>0)=NZ$, where $Z=(U_1|U_0+\gamma>0)$, we have then that $Z_2=0$ with probability 1 and that $Z_1\sim\mathrm{SUN}_{r,m}(0,\gamma,1_r,\Omega_B^*)$ is a non-singular SUN distribution, as

$$\Omega_B^*=\begin{pmatrix}\Gamma & \Delta^{\top}B^{\top}\\ B\Delta & B\bar{\Omega}B^{\top}\end{pmatrix}.$$

Therefore, $Z_1=BZ$ has a SUN density given by

$$\phi_r(z_1;B\bar{\Omega}B^{\top})\frac{\Phi_m(\gamma+\Delta^{\top}B^{\top}(B\bar{\Omega}B^{\top})^{-1}z_1;\Gamma-\Delta^{\top}B^{\top}(B\bar{\Omega}B^{\top})^{-1}B\Delta)}{\Phi_m(\gamma;\Gamma)}.$$

After some algebraic manipulation, this expression can be rewritten as

$$\frac{\exp(-(1/2)z^{\top}\bar{\Omega}^{-}z)}{(2\pi)^{r/2}(\prod_{j=1}^{r}\lambda_j)^{1/2}}\frac{\Phi_m(\gamma+\Delta^{\top}\bar{\Omega}^{-}z;\Gamma-\Delta^{\top}\bar{\Omega}^{-}\Delta)}{\Phi_m(\gamma;\Gamma)}$$

which can be interpreted as the density at a point z lying on the hyperplane $Nz=0$; this hyperplane is such that $\mathbb{P}\{NZ=0\}=1$.

The final case to consider is when $\mathrm{rank}(\Gamma)=m$, $\mathrm{rank}(\bar{\Omega})=d$, but $\mathrm{rank}(\Omega^*)<m+d$. Under these assumptions we have

$$0=|\Omega^*|=|\Gamma||\bar{\Omega}-\Delta\Gamma^{-1}\Delta^{\top}|=|\bar{\Omega}||\Gamma-\Delta^{\top}\bar{\Omega}^{-1}\Delta|,$$

implying that $\bar{\Omega}-\Delta\Gamma^{-1}\Delta^{\top}$ and $\Gamma-\Delta^{\top}\bar{\Omega}^{-1}\Delta$ are singular positive semidefinite matrices. However, as $U_1\sim N_d(0,\bar{\Omega})$ is non-singular, then $Z=(U_1|U_0+\gamma>0)\sim\mathrm{SUN}_{d,m}(0,\gamma,1_d,\bar{\Omega}^*)$ is also non-singular, i.e. its density is again of type (10). The only computational difference is that the integral in the numerator of (10) is computed via (23), where in this case we use a matrix C_0 such that $C_0C_0^{\top}=\Gamma-\Delta^{\top}\bar{\Omega}^{-1}\Delta$.