Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 1, 110–115
DOI: 10.1214/11-BJPS159

A note on the parameterization of multivariate skewed-normal distributions

Luis M. Castro[a], **Ernesto San Martín**[b,c] **and Reinaldo B. Arellano-Valle**[b]

[a]*Department of Statistics, Universidad de Concepción*
[b]*Department of Statistics, Pontificia Universidad Católica de Chile*
[c]*Measurement Center MIDE UC, Pontificia Universidad Católica de Chile*

Abstract. Azzalini's skew-normal distribution is obtained through a conditional reduction of a multivariate normal distribution parameterized with a correlation matrix. It seems natural that when the parameterization of that multivariate normal distribution is complexified, more flexible skew-normal distributions could be obtained. In this note this specification strategy, previously explored by Azzalini [*Scand. J. Stat.* **33** (2006) 561–574] among many other authors, is formally analyzed through an identification analysis.

1 Introduction

Skewed-normal distributions can be obtained as a conditional reduction of a multivariate normal distribution as follows (see Capitanio et al., 2003; Arellano-Valle and Azzalini, 2006): let $U_0 \in \mathbb{R}$ and $\mathbf{U}_1 \in \mathbb{R}^d$ be two random vectors such that

$$\mathbf{U} = \begin{pmatrix} U_0 \\ \mathbf{U}_1 \end{pmatrix} \sim \mathcal{N}_{1+d}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \Omega^* = \begin{pmatrix} 1 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \bar{\Omega} \end{pmatrix}\right),$$

where $\boldsymbol{\delta} \in (-1, 1)^d$, $\bar{\Omega} \in \mathbb{R}^{d\times d}$ is a positive definite symmetric matrix and Ω^* is a correlation matrix. Let $\mathbf{Z} \stackrel{d}{=} (\mathbf{U}_1 | U_0 > -\gamma)$, where $\stackrel{d}{=}$ means *equal distribution*. Thus, the probability density function (p.d.f.) of $\mathbf{Z}$ is given by

$$f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}) = C(\gamma)\phi_d(\mathbf{z}|\bar{\Omega})\Phi\left(\frac{\gamma + \boldsymbol{\delta}^T\bar{\Omega}^{-1}\mathbf{z}}{\sqrt{1-\boldsymbol{\delta}^T\bar{\Omega}^{-1}\boldsymbol{\delta}}}\right), \qquad \mathbf{z} \in \mathbb{R}^d, \tag{1.1}$$

where $\phi_d(\mathbf{z}-\boldsymbol{\mu}|\Sigma)$ is the p.d.f. of the normal distribution $\mathcal{N}_d(\mathbf{0}, \Sigma)$, $\Phi(z)$ is the cumulative density function (c.d.f.) of the standard normal distribution evaluated at z, and $C^{-1}(\gamma) = \Phi(\gamma)$. Density (1.1) is parameterized by $\boldsymbol{\theta} = (\gamma, \boldsymbol{\delta}, \bar{\Omega}) \in \mathbb{R} \times (-1, 1)^d \times \mathbb{R}^{d\times d}$. In this context, a random variable with a density given by (1.1) and a parameter $\boldsymbol{\theta}$ is said to have a *standard extended skew-normal distribution*; it is denoted as $\mathbf{Z} \sim \mathcal{ESN}_d(\boldsymbol{\theta})$. Alternatively, along the paper, we will denote as $\mathbf{Z} \sim \mathcal{SN}_d(\boldsymbol{\theta})$ the extended skewed-normal distribution obtained from a reduction

Key words and phrases. Identification, minimal sufficient parameter, multivariate skewed-normal distributions.

Received November 2011; accepted June 2011.

of multivariate normal distribution parameterized with a covariance matrix instead of a correlation matrix and with a given parameter $\boldsymbol{\theta}$. In this case, we will call this distribution simply a *skewed-normal distribution*. It is important to stress that the skewed-normal distributions are closed under affine linear transformations of the form $\mathbf{Y} = \boldsymbol{\xi} + \omega\mathbf{Z}$, where $\boldsymbol{\xi} \in \mathbb{R}^d$ is a location parameter and $\omega \in \mathbb{R}^{d\times d}$ is a scale matrix. Thus, the skewed-normal distribution induced by $\mathbf{Y}$ will be called a *location-scale skewed-normal distribution*.

From time to time, some skewed-normal distributions have been introduced, all of them having a common genesis, namely, the reduction of a multivariate normal distribution with a covariance matrix instead of a correlation matrix. The covariance matrix which has been considered is of the following types:

$$\Omega = \begin{pmatrix} \mathbf{D}^T\Upsilon\mathbf{D} & -\mathbf{D}^T\Upsilon \\ -\Upsilon\mathbf{D} & \Upsilon+\Sigma \end{pmatrix}, \qquad \Omega = \begin{pmatrix} \Theta+\mathbf{D}^T\Sigma\mathbf{D} & \mathbf{D}^T\Sigma \\ \Sigma\mathbf{D} & \Sigma \end{pmatrix},$$

where $\mathbf{D} \in \mathbb{R}^d$, Υ and Σ are positive definite matrices in $\mathbb{R}^{d\times d}$, and $\Theta \in \mathbb{R}^+$. However, it is important to stress that the identifiability is typically lost when a multivariate normal distribution is reduced by conditioning (see Florens et al., 1990, Chapter 4). The objective of this note consists, therefore, in establishing an identified parameterization of the skewed-normal distribution when a general covariance matrix at the multivariate normal level is considered.

2 Identified parameterization for the skewed-normal distributions

Parameter identification can be defined in terms of parametric minimal sufficiency (see Kadane, 1974; Picci, 1977; Florens et al., 1990). It can be proved that if the mapping $\boldsymbol{\theta} \longmapsto P^{\boldsymbol{\theta}}$ is one-to-one, then $\boldsymbol{\theta}$ is the minimal sufficient parameter for the data $\mathbf{Z}$ in the sense that $\boldsymbol{\theta}$ fully characterizes the sampling process $p(\mathbf{Z}|\boldsymbol{\theta})$ and if $\boldsymbol{\phi}$ is another parameter characterizing the same sampling process, then $\boldsymbol{\theta}$ is an injective function of $\boldsymbol{\phi}$. Moreover, the minimal sufficient parameter is always a function of a countable number of sampling expectations (see Florens et al., 1990, Chapter 4). This means that a parameterization of interest $\boldsymbol{\psi} = g(\boldsymbol{\theta})$ for some function g is identified, when there exist measurable functions f and h such that $\boldsymbol{\psi} = h\{E(f(\mathbf{Z})|\boldsymbol{\theta})\}$. Consequently, only identified parameters have a statistical interpretation because they can be expressed in terms of the sampling process.

The identification of the skewed-normal distributions through the concept of minimal sufficient parameter is next established by assuming that the multivariate normal covariance matrix Ω is of the form

$$\Omega = \begin{pmatrix} \Gamma & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \bar{\Omega} \end{pmatrix}, \tag{2.1}$$

where $\boldsymbol{\delta} \in \mathbb{R}^d$, $\bar{\Omega} \in \mathbb{R}^{d \times d}$ is a variance–covariance matrix and $\Gamma \in \mathbb{R}^+$. In this case, the p.d.f. of $\mathbf{Z} \stackrel{d}{=} (\mathbf{U}_1 | U_0 > -\gamma)$ is given by

$$f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}) = C(\gamma, \Gamma)\phi_d(\mathbf{z}|\bar{\Omega})\Phi\left(\frac{\gamma + \boldsymbol{\delta}^T \bar{\Omega}^{-1}\mathbf{z}}{\sqrt{\Gamma - \boldsymbol{\delta}^T \bar{\Omega}^{-1}\boldsymbol{\delta}}}\right), \qquad \mathbf{z} \in \mathbb{R}^d, \tag{2.2}$$

where $C^{-1}(\gamma, \Gamma) = \Phi(\gamma\Gamma^{-1/2})$ and $\boldsymbol{\theta} = (\gamma, \boldsymbol{\delta}, \Gamma, \bar{\Omega}) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{R}^{d \times d}$, with $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_d)^T$ and $\bar{\omega}_{ij} = \bar{\Omega}[i, j]$ for $i, j = 1, \ldots, d$. Also, the rth cumulants κ^r for Z_i, Z_j and Z_k, $i, j, k = 1, \ldots, d$, are given by

$$\kappa_i^1 = \left(\frac{\delta_i}{\Gamma^{1/2}}\right)\zeta_1\left(\frac{\gamma}{\Gamma^{1/2}}\right), \qquad \kappa_{ij}^2 = \bar{\omega}_{ij} - \frac{\delta_i\delta_j}{\Gamma}\zeta_2\left(\frac{\gamma}{\Gamma^{1/2}}\right), \tag{2.3}$$

$$\kappa_{ijk}^3 = \left(\frac{\delta_i\delta_j\delta_k}{\Gamma^{3/2}}\right)\zeta_3\left(\frac{\gamma}{\Gamma^{1/2}}\right), \tag{2.4}$$

where $\zeta_r(x)$ is the rth derivative of $\zeta_0(x) = \log\{2\Phi(x)\}$. Considering $\eta = \Gamma^{-1/2}\gamma$ and $\alpha_i = \Gamma^{-1/2}\delta_i$, and using (2.3)–(2.4), it follows that, for $i, j, k = 1, \ldots, d$,

$$\alpha_i = \frac{\kappa_i^1}{\zeta_1(\eta)}, \tag{2.5}$$

$$\bar{\omega}_{ij} = \kappa_{ij}^2 - \kappa_i^1\kappa_j^1\frac{\zeta_2(\eta)}{\zeta_1^2(\eta)}, \tag{2.6}$$

$$\frac{\zeta_3(\eta)}{\zeta_1^3(\eta)} = \frac{\sum_i\sum_j\sum_k \kappa_{ijk}^3}{(\sum_i \kappa_i^1)(\sum_j \kappa_j^1)(\sum_k \kappa_k^1)}. \tag{2.7}$$

It is important to stress that there are as many well-defined α_i's as δ_i's are different from zero. Taking into account that κ^r, $r = 1, 2, 3$, are measurable functions of $E[f(\mathbf{Z})|\boldsymbol{\theta}]$, with f a measurable function, then, once η is identified from equation (2.7), the identification of both $\bar{\omega}_{ij}$ and α_i follows from equations (2.6) and (2.5), respectively. The identification of η follows after noticing that the function $\upsilon(\eta) = \zeta_3(\eta)/\zeta_1^3(\eta)$ is a strictly increasing function (for a proof, see the Appendix). Consequently, the equation (2.7) implies that

$$\eta = \upsilon^{-1}\left(\frac{\sum_i\sum_j\sum_k \kappa_{ijk}^3}{(\sum_i \kappa_i^1)(\sum_j \kappa_j^1)(\sum_k \kappa_k^1)}\right),$$

and, therefore, the minimal sufficient parameter of the skewed-normal sampling process is given by $\boldsymbol{\psi} = (\eta, \boldsymbol{\alpha}, \bar{\Omega}) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d \times d}$, where $\eta = \Gamma^{-1/2}\gamma$ and $\boldsymbol{\alpha} = \Gamma^{-1/2}\boldsymbol{\delta}$. In other words, $\boldsymbol{\psi}$ is identified and then the p.d.f. of the skewed-normal distribution can be written as a function of $\boldsymbol{\psi}$, namely, $f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\theta}) = C(\eta)\phi_d(\mathbf{z}|\bar{\Omega})\Phi(\frac{\eta + \boldsymbol{\alpha}^T\bar{\Omega}^{-1}\mathbf{z}}{\sqrt{1 - \boldsymbol{\alpha}^T\bar{\Omega}^{-1}\boldsymbol{\alpha}}}) \doteq f_{\mathbf{Z}}(\mathbf{z}|\boldsymbol{\psi})$ with $C^{-1}(\eta) = \Phi(\eta)$. Finally, note that, since Ω is a variance–covariance matrix, $1 - \boldsymbol{\alpha}^T\bar{\Omega}^{-1}\boldsymbol{\alpha} > 0$.

3 Discussion

Although the result proposed above may seem straightforward, the main objective of it is to establish the fact that the identified parameters at the multivariate normal level are quite different from those which can be identified at the skewed-normal level (after reduction by conditioning). Thus, for example, in the case of the extended skew-normal distribution (2.2), the parameterization $\boldsymbol{\theta} = (\gamma, \boldsymbol{\delta}, \Gamma, \bar{\Omega})$, indexing the extended skew-normal distribution (2.2), is not identified by the observations generated by the skew-normal process, although $\boldsymbol{\delta}$, Γ and $\bar{\Omega}$ are identified with respect to the underlying multivariate normal distribution. However, according to the result presented in Section 2, the parameterization $\boldsymbol{\psi} = (\eta = \Gamma^{-1/2}\gamma, \boldsymbol{\alpha} = \Gamma^{-1/2}\boldsymbol{\delta}, \bar{\Omega})$ is identified by the skewed-normal process (2.2); it can be noted that the identified parameter $\boldsymbol{\psi}$ is a noninjective function of the parameter $\boldsymbol{\theta}$.

The previous comment leads to focus our attention on a typical way used to construct a skewed-normal distribution, namely, through a reduction of a multivariate normal distribution in which Ω is a correlation matrix, instead of a variance–covariance matrix (see, e.g., Azzalini, 2005; Arellano-Valle and Azzalini, 2006). In fact, the Ω matrix (2.1) becomes a correlation matrix after imposing the following restrictions: $\Gamma = 1$, $\bar{\omega}_{ii} = 1$ and $\bar{\omega}_{ij} \in (-1, 1)$ $\forall i \neq j$ for $i, j = 1, \dots, d$. Under these restrictions, it follows that $\boldsymbol{\delta} \in (-1, 1)^d$ and, therefore, the parameter $\boldsymbol{\psi} = (\gamma, \boldsymbol{\delta}, \bar{\Omega}) \in \mathbb{R} \times (-1, 1)^d \times \mathbb{R}^{d \times d}$ is identified. When the parameter γ is equal to 0 and an affine linear transformation $\mathbf{Y} = \boldsymbol{\xi} + \omega\mathbf{Z}$ is considered, then a straightforward application of the result presented in Section 2 leads to prove that the parameter $\boldsymbol{\psi} = (\boldsymbol{\alpha}, \boldsymbol{\xi}, \Lambda) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d}$ is identified, with $\boldsymbol{\alpha} = \Gamma^{-1/2}\omega\boldsymbol{\delta}$ and $\Lambda = \omega\bar{\Omega}\omega$. It is important to stress that the parameter ω cannot be estimated unless an additional restriction is considered. A typically considered restriction is $\omega = \mathrm{diag}(\omega_{11}, \dots, \omega_{dd})$; in this case, the parameter $\boldsymbol{\psi} = (\boldsymbol{\delta}, \boldsymbol{\xi}, \omega, \bar{\Omega}) \in (-1, 1)^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d} \times \mathbb{R}^{d \times d}$ is identified.

The reader can recognize that this form of constructing skewed-normal distributions considers particular restrictions leading to identify the parameters of interest for the skewed-normal process, say, γ, $\boldsymbol{\delta}$, $\boldsymbol{\xi}$ and ω. In fact, the main advantage of that form of constructing skewed-normal distributions is that the scale parameter ω can be estimated from the skewed-normal sampling process and, when $\gamma = 0$, $\boldsymbol{\delta}$ is a function of the skewness index γ_1 as defined by Fisher (2003). In addition, the centered parameterization proposed by Azzalini and Dalla Valle (1996) allows to make interpretable the parameters $\boldsymbol{\xi}$ and ω in terms of expectation and variance of random variable $\mathbf{Y}$, respectively.

Summarizing, it can be concluded that an eventual extension of a skewed-normal distribution based on a parameterization of Ω more complex than (2.1) necessarily will fail to get a proper extension. This conclusion is based on the fact that an extension of a statistical model involves parameters which, on one hand, are not involved in the initial statistical model and, on the other hand, are capable of being estimated. However, if it is necessary to restrict the parameters of the extended model in order to get the parameter identifiability, then the extension would fail. It is important to remark that the loss of the identifiability is a consequence of the truncation and/or the reduction trough conditionalization or marginalization (see Hodoshima, 1988; Florens et al., 1990; Sapra, 2008). Finally, it should be remarked that the restrictions needed to obtain a skewed-normal distribution are not empirically justified because they are defined at the level of an unobservable process, namely, the multivariate normal process. In any case, it could be said that the assumed correlation structure is a way to fix the location and/or scale of the truncated process.

Appendix

To prove that the real-valued function $\upsilon(x) \doteq \zeta_3(x)/\zeta_1^3(x)$ is strictly increasing, we consider the following steps:

Step 0. By using $\zeta_0(x) = \log\{2\Phi(x)\}$, it is easy to prove that

$$\zeta_1(x) = \phi(x)/\Phi(x),$$
$$\zeta_2(x) = -\zeta_1(x)\big(\zeta_1(x) + x\big),$$
$$\zeta_3(x) = \zeta_1(x)\big\{\big(\zeta_1(x) + x\big)\big(2\zeta_1(x) + x\big) - 1\big\},$$

where $-1 < \zeta_2(x) < 0$ is an increasing function and $\zeta_3(x) > 0$ (see Sampford, 1953).

Step 1. Using the relations proposed in Step 0, we write $\upsilon(x) = \lambda(x) - \nu(x)$, where $\nu(x) = \zeta_2(x)/\zeta_1^2(x)$ and $\lambda(x) = \nu^2(x) - \frac{1}{\zeta_1^2(x)}$.

Step 2. It follows that $\upsilon'(x) = \lambda'(x) - \nu'(x)$, where $\lambda'(x) = 2\nu(x)\zeta_1(x)\{\zeta_3(x) \times \zeta_1(x) - 2\zeta_2^2(x)\} - 2\nu(x)\zeta_1^3(x) + 4\zeta_1(x)$ and $\nu(x)' = -\zeta_1(x)\{\nu(x) + \frac{1}{\zeta_1^2(x)} + \nu^2(x)\}$.

Step 4. $\lambda'(x) > 0$. In fact, taking into account that $\nu(x) < 0$ $\forall x \in \mathbb{R}$ and that $\zeta_1(x) > 0$ $\forall x \in \mathbb{R}$, the conclusion follows if $\zeta_3(x)\zeta_1(x) - 2\zeta_2^2(x) < 0$ $\forall x \in \mathbb{R}$. But $0 < \zeta_3(x)\zeta_1(x) = \zeta_2^2(x) - \zeta_1^2(x)[1 + \zeta_2(x)]$ because $\zeta_3(x) > 0$ $\forall x \in \mathbb{R}$. Taking into account that $\zeta_1^2(x)[1 + \zeta_2(x)] > 0$ (because $-1 < \zeta_2(x) < 0$), it follows that $\zeta_3(x)\zeta_1(x) < \zeta_2^2(x)$ $\forall x \in \mathbb{R}$, which implies that $\zeta_3(x)\zeta_1(x) - 2\zeta_2^2(x) < 0$.

Step 5. $\nu'(x) < 0$. In fact, it is enough to remark that $\nu(x) + \frac{1}{\zeta_1^2(x)} = \frac{1+\zeta_2(x)}{\zeta_1(x)} > 0$ $\forall x \in \mathbb{R}$.

Step 6. Finally, from Steps 4 and 5, it is concluded that $\upsilon'(x) > 0$ $\forall x \in \mathbb{R}$, thus the proof.

Acknowledgments

The first author was partially supported by the FONDECYT project N^{o} 11100076 from the Chilean government. The second author was partially supported by the PUENTE Project 08/2009 from the Pontificia Universidad Católica de Chile. The third author was partially supported by the FONDECYT Project N^{o} 1085241 from the Chilean government. The authors thank the editor and the referees for their helpful comments.

References

Arellano-Valle, R. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33**, 561–574. MR2298065

Azzalini, A. (2005). The skew normal distribution and related multivariate families. *Scandinavian Journal of Statistics* **32**, 159–188. MR2188669

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew normal distribution. *Biometrika* **83**, 715–726. MR1440039

Capitanio, A., Azzalini, A. and Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scandinavian Journal of Statistics* **30**, 129–144. MR1963897

Fisher, R. A. (1930). The moments of the distribution for normal samples of measures of departure from normality. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **130**, 16–28.

Florens, J.-P., Mouchart, M. and Rolin, J.-M. (1990). *Elements of Bayesian Statistics*. New York: Marcel and Dekker. MR1051656

Hodoshima, J. (1988). The effect of truncation on the identifiability of regression coefficients. *Journal of the American Statistical Association* **83**, 1055–1056. MR0997581

Kadane, J. (1974). The role of identification in Bayesian theory. In *In Studies in Bayesian Econometrics and Statistics* (S. Fienberg and A. Zellner, eds.) 175–191. Amsterdam: North Holland. MR0483124

Picci, G. (1977). Some connections between the theory of sufficient statistics and the identifiability problem. *SIAM Journal of Applied Mathematics* **33**, 383–398. MR0464455

Sampford, M. (1953). Some inequalities on Mill's ratio and related functions. *The Annals of Mathematical Statistics* **24**, 130–132. MR0054890

Sapra, S. (2008). Truncation and unidentifiability: the single sample case. *Bulletin of Statistics and Economics* **2**, 26–29. MR2427321

L. M. Castro
Department of Statistics
Universidad de Concepción
Av. Estaban Iturra s/n Barrio Universitario
Concepción
Chile
E-mail: luiscastroc@udec.cl
URL: http://stat.cfm.cl/

E. San Martín
R. B. Arellano-Valle
Department of Statistics
Pontificia Universidad Católica de Chile
Santiago 22
Chile
E-mail: esanmart@mat.puc.cl
reivalle@mat.puc.cl
URL: http://www.mat.puc.cl/