Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 1758–1771
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

A class of weighted multivariate normal distributions and its properties

Hea-Jung Kim

Department of Statistics, Dongguk University, Pil-Dong 3Ga, Chung-Gu, Seoul 100-715, Republic of Korea

Received 25 September 2006
Available online 2 February 2008

Abstract

This article proposes a class of weighted multivariate normal distributions whose probability density function has the form of a product of a multivariate normal density and a weighting function. The class is obtained from marginal distributions of various doubly truncated multivariate normal distributions. The class strictly includes the multivariate normal and multivariate skew-normal. It is useful for selection modeling and inequality constrained normal mean vector analysis. We report on a study of some distributional properties and the Bayesian perspective of the class. A probabilistic representation of the distributions is also given. The representation is shown to be straightforward to specify the distribution and to implement computation, with output readily adapted for the required analysis. Necessary theories and illustrative examples are provided.

AMS 2000 subject classifications: 62E10; 62F30

Keywords: Weighted multivariate normal distribution; Doubly truncated normal; Selection model; Constrained normal mean vector; Probabilistic representation

1. Introduction

There have recently been many studies on the family of weighted distributions that occupy a central place in the development of applications (see Arellano-Valle and Azzalini [1] and the references therein). The weighted distribution arises when the density $g(x; \theta_1)$ of the potential observation x gets distorted so that it is multiplied by some non-negative weight function

E-mail address: kim3hj@dongguk.edu.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.01.008

$w(x; \theta_1, \theta_2)$ involving an additional parameter vector θ_2. Then, the observed data is a random sample from a weighted distribution with density

$$f(x; \theta_1, \theta_2) = w(x; \theta_1, \theta_2) g(x; \theta_1)/E_{\theta_1}[w(X; \theta_1, \theta_2)], \tag{1.1}$$

where the expectation in the denominator is just a normalizing constant.

Interesting classes of weighted distributions are discussed in Rao [23], Bayarri and DeGroot [8], Arnold and Beaver [3], Branco and Dey [10], Ma et al. [20], Azzalini [6], and Kim [17]. As elaborated in the articles by Arnold et al. [4], Ma and Genton [19] as well as in the book edited by Genton [13], the application of the weighted distributions extends to the areas of econometrics, astronomy, engineering, medicine as well as psychology in scenarios where the observed random phenomena can be described by (1.1). In particular, if the potential observation x is obtained only from a selected portion of the population of interest, then (1.1) is called a selection model. Weighted distributions, establishing links with selection models obtained from various forms of selection mechanisms, are well addressed in the literature; see Genton [13,14], Arellano-Valle et al. [2] and the references therein.

The main objective of this study, described here, is to investigate various properties of a class of weighted distributions arising from a selection mechanism where the underlying distribution is a doubly truncated multivariate normal. Although the class can be taken as a special case of selection distributions developed by Arellano-Valle et al. [2], we are not aware of any detailed exposition of the distributional properties. This lack of detailed exposition motivates the investigation described in this article. This class is interesting both from a theoretical and an applied point of view. On the theoretical side, the class enjoys a number of formal properties which resemble those of multivariate skew-normal distribution by Azzalini and Dalla Valle [7] and is a superset of the multivariate skew-normal family (which in turn is a superset of the multivariate normal one). In addition, the class provides distribution for the sum of multivariate normal and a doubly truncated multivariate normal random vectors. From a Bayesian perspective, the class of distributions leads to yet other conjugate family of priors. The latter is especially useful for a Bayesian subjective methodology for an inequality constrained multivariate normal mean vector problem. In the applied view point, the class provides new models that enable us to analyze the inequality constrained linear models, the screened multivariate data sets, and the skewed multivariate data sets.

The material in this article is arranged as follows. In Section 2, we set the scene and define a class of weighted multivariate normal distributions whose underlying distribution is a doubly truncated multivariate normal. This is followed by discussion of the properties of the distributions mainly based on their probabilistic representations. Moment calculations for the class of distributions are presented in Section 3. In Section 4, we provide an important property useful for the Bayesian approach. Section 5 contains illustrative examples that have been selected for the purpose of validating and motivating the contents of this article. In Section 6, we provide a discussion and some conclusions.

2. The class of weighted distributions

According to (1.1), we may define a class of weighted multivariate normal distributions: For $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \Omega)$, if the p.d.f. (probability density function) of $\mathbf{Y}$ is distorted by a multiplicative nonnegative weight function $w(\mathbf{y}; \Theta_1, \Theta_2)$ such that

$$f_{\mathbf{Y}}(\mathbf{y}; \Theta_1, \Theta_2) = \phi_k(\mathbf{y}; \Theta_1) \frac{w(\mathbf{y}; \Theta_1, \Theta_2)}{E_{\Theta_1}[w(\mathbf{y}; \Theta_1, \Theta_2)]}, \quad \mathbf{y} \in \mathbb{R}^k, \tag{2.1}$$

we say that the distribution is a weighted multivariate normal distribution, where $\phi_k(\mathbf{y};\, \Theta_1)$ is the p.d.f. of the $\mathbf{Y}$ random vector, $\Theta_1 = \{\boldsymbol{\mu}, \Omega\}$, and Θ_2 is an additional parameter vector involved in the weight function.

2.1. The derivation from a doubly truncated multivariate normal

A class of weighted distributions of the form (2.1) is obtained from the marginal joint distributions of doubly truncated multivariate normal distributions: Let $\mathbf{X}^* = (X_0, \mathbf{X}^\top)^\top$ be a $(k+1)$-variate normal random vector such that $\mathbf{X}^* \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma)$, where $\boldsymbol{\mu}^* = (\mu_0, \boldsymbol{\mu})^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^\top$, and the covariance matrix Σ has the form

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_0\boldsymbol{\delta}^\top \\ \sigma_0\boldsymbol{\delta} & \Omega \end{pmatrix}$$

with $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)^\top$. $\Omega = (\omega_{ij})$ is the covariance matrix associated to the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)^\top$ so that $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Omega)$. From now on, the notation for the parameters will remain the same.

Under the distribution of $\mathbf{X}^*$, suppose that the random variable X_0 is truncated at the lower truncation point a and the upper point b, then we see that the marginal joint p.d.f. of $\mathbf{X}$ is the same as that of $\mathbf{Y} = [\mathbf{X} \mid a < X_0 < b]$ and it is given by

$$f_{\mathbf{Y}}(\mathbf{y};\, \boldsymbol{\mu}^*, \Sigma) = \phi_k(\mathbf{y};\, \Theta_1)\frac{P(a < X_0 < b \mid \mathbf{y})}{E_{\Theta_1}[P(a < X_0 < b \mid \mathbf{y})]}, \quad \mathbf{y} \in \mathbb{R}^k, \tag{2.2}$$

where $\Theta_1 = \{\boldsymbol{\mu}, \Omega\}$, $\Theta_2 = \{\mu_0, \sigma_0^2, \boldsymbol{\delta}\}$, and $E_{\Theta_1}[P(a < X_0 < b \mid \mathbf{y})] = P(a < X_0 < b)$. Comparing (2.2) with (2.1), we see that (2.2) is a special case of the class of weighted multivariate normal distributions having the nonnegative weight function $w(\mathbf{y};\, \Theta_1, \Theta_2) = P(a < X_0 < b \mid \mathbf{y})$. The exact expression for $f_{\mathbf{Y}}(\mathbf{y};\, \boldsymbol{\mu}^*, \Sigma)$ in (2.2) is

$$\phi_k(\mathbf{y};\, \Theta_1)\frac{\Phi\left(\xi u(b) - \boldsymbol{\lambda}^\top(\mathbf{y}-\boldsymbol{\mu})\right) - \Phi\left(\xi u(a) - \boldsymbol{\lambda}^\top(\mathbf{y}-\boldsymbol{\mu})\right)}{\Phi(u(b)) - \Phi(u(a))}, \quad \mathbf{y} \in \mathbb{R}^k, \tag{2.3}$$

where $\Phi(\cdot)$ is the d.f. of the univariate standard normal variate, $u(a) = (a-\mu_0)/\sigma_0$, $u(b) = (b-\mu_0)/\sigma_0$, $\xi = (1-\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta})^{-1/2}$, and $\boldsymbol{\lambda}^\top = \xi\boldsymbol{\delta}^\top\Omega^{-1}$. This is obtained from noticing that

$$X_0 \mid \mathbf{y} \sim N(\mu_0 + \sigma_0\boldsymbol{\delta}^\top\Omega^{-1}(\mathbf{y}-\boldsymbol{\mu}), \sigma_0^2(1-\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta})).$$

Thus considering

$$Z_0 = \{X_0 - (\mu_0 + \sigma_0\boldsymbol{\delta}^\top\Omega^{-1}(\mathbf{y}-\boldsymbol{\mu}))\}/\{\sigma_0(1-\boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta})^{1/2}\}$$

gives

$$\begin{aligned} P(a < X_0 < b \mid \mathbf{y}) &= P\left(\xi u(a) - \boldsymbol{\lambda}^\top(\mathbf{y}-\boldsymbol{\mu}) < Z_0 < \xi u(b) - \boldsymbol{\lambda}^\top(\mathbf{y}-\boldsymbol{\mu}) \mid \mathbf{y}\right) \\ &= \Phi\left(\xi u(b) - \boldsymbol{\lambda}^\top(\mathbf{y}-\boldsymbol{\mu})\right) - \Phi\left(\xi u(a) - \boldsymbol{\lambda}^\top(\mathbf{y}-\boldsymbol{\mu})\right). \end{aligned}$$

Definition 1. For $\mathbf{X}^* \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma)$, $\mathbf{X}^* = (X_0, \mathbf{X}^\top)^\top$, the random vector $\mathbf{Y} = [\mathbf{X} \mid a < X_0 < b]$ with p.d.f. given by (2.3) is said to have a weighted multivariate two-sided conditioning normal (WTN) distribution. This is written by the notation $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$.

The distribution of $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ is the same as the conditional distribution of $\mathbf{X}$ given $a < X_0 < b$, where $(X_0, \mathbf{X}^\top)^\top \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma)$. Therefore, the cumulative distribution function of $\mathbf{Y}$ is

$$\begin{aligned} F_k(\mathbf{y}) &= \mathrm{pr}(Y_1 \le y_1, \ldots, Y_k \le y_k) \\ &= \mathrm{pr}(a \le X_0 \le b, X_1 \le y_1, \ldots, X_k \le y_k)/\{\Phi(u(b)) - \Phi(u(a))\} \end{aligned} \tag{2.4}$$

for $\mathbf{y} = (y_1, \ldots, y_k)^\top \in \mathbb{R}^k$. To conclude, the distribution function of the k-dimensional variable $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ can be obtained from the distribution function of a $(k+1)$-dimensional random variable with distribution $(X_0, \mathbf{X}^\top)^\top \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma)$ and that of $Z \sim N(0, 1)$.

Note that both the $WN_k^{(-\infty,\infty)}(\boldsymbol{\mu}^*, \Sigma)$ distribution and the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distribution with $\boldsymbol{\delta} = \mathbf{0}$ are equivalent to the $N_k(\boldsymbol{\mu}, \Omega)$ distribution. On the other hand, the $WN_k^{(\mu_0,\infty)}(\boldsymbol{\mu}^*, \Sigma)$ distribution is equal to the multivariate skew-normal distribution by Azzalini and Dalla Valle [7]. Further note that the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distribution is a special case of the selection normal distribution discussed by Arellano-Valle et al. [2]. Azzalini and Dalla Valle [7] gave a representation of the $WN_k^{(\mu_0,\infty)}(\boldsymbol{\mu}^*, \Sigma)$ distribution by using a transformation method. Later, the representation was extended to a more general context (the selection model distribution) by Arellano-Valle et al. [2]. Utilizing the same method, we can derive the following stochastic representation for the two side-conditioned normal vector $\mathbf{Y} = [\mathbf{X} \mid a < X_0 < b]$.

Theorem 1. *Let $U \sim N(\mu_0, \sigma_0^2)$ be independent of the U_i's, where $U_i \sim N(\theta_i, \tau_i^2)$ and $\mathrm{Cov}(U_i, U_j) = \tau_{ij}$, $i, j = 1, \ldots, k, i \neq j$. Define*

$$Y_i = \alpha_i U_{(a,b)} + \beta_i U_i, \quad i = 1, \ldots, k, \tag{2.5}$$

then the distribution of $\mathbf{Y} = (Y_1, \ldots, Y_k)^\top$ is $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ for any real values $\alpha_i (\alpha_i \neq 0)$ and β_i, provided that $\alpha_i = \delta_i/\sigma_0$, $\beta_i\beta_j\tau_{ij} = \omega_{ij} - \alpha_i\alpha_j\sigma_0^2$, $\beta_i\theta_i = \mu_i - \alpha_i\mu_0$, and $\beta_i^2\tau_i^2 = \omega_{ii} - \alpha_i^2\sigma_0^2$, where $U_{(a,b)}$ indicates that the distribution of U is doubly truncated with respect to lower and upper truncation points a and b.

Proof. Note that for $\mathbf{X}^* \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma)$, $\mathbf{X}^* = (X_0, \mathbf{X}^\top)^\top$, conditional distribution of random vector $\mathbf{X}$ given that $a < X_0 < b$ is the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distribution. Let $X_0 = U$ and $X_i = \alpha_i U + \beta_i U_i$ for $i = 1, \ldots, k$. Then $\mathbf{X}^* = (X_0, X_1, \ldots, X_k)^\top$ is a $(k+1)$-variate normal with the mean vector $\boldsymbol{\mu}^*$ and covariance matrix Σ, provided that $\alpha_i = \delta_i/\sigma_0$, $\beta_i\beta_j\tau_{ij} = \omega_{ij} - \alpha_i\alpha_j\sigma_0^2$, $\beta_i\theta_i = \mu_i - \alpha_i\mu_0$, and $\beta_i^2\tau_i^2 = \omega_{ii} - \alpha_i^2\sigma_0^2$. Since U is independent of the U_i's and the conditional distribution of U given $a < U < b$ equals to the distribution of $U_{(a,b)}$ appearing in the statement of Theorem 1, we are done. □

This representation indicates a kind of departure from the multivariate normal law and intrinsic structure of the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distribution. The representation (2.5) is also useful for studying properties of the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distributions and for computations of their moments. Upon setting $\mu_0 = 0, \theta_1 = \cdots = \theta_k = 0$, $\tau_i^2 = 1$, and $\tau_{ij} = \psi_{ij}$ in Theorem 1, the stochastic representation of a standard version of the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distribution (i.e. the $WN_k^{(a,b)}(\mathbf{0}, R)$ distribution) is immediate, where

$$R = D_\Sigma^{-1/2} \Sigma D_\Sigma^{-1/2} = \begin{pmatrix} 1 & \boldsymbol{\rho}^\top \\ \boldsymbol{\rho} & R_{11} \end{pmatrix},$$

the $(k+1)\times(k+1)$ correlation matrix obtained from Σ, $D_{\Sigma} = \mathrm{diag}\{\sigma_0^2, \omega_{11}, \ldots, \omega_{kk}\}$, $R_{11} = (r_{ij})$, and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_k)^\top$ with $\rho_i = \delta_i/\sqrt{\omega_{ii}}, i = 1, \ldots, k$.

Corollary 1. *If $r_{ij} = \rho_i\rho_j + \psi_{ij}(1-\rho_i^2)^{1/2}(1-\rho_j^2)^{1/2}$ for $i, j = 1, \ldots, k$, then the stochastic representation of $\mathbf{W} \sim WN_k^{(a,b)}(\mathbf{0}, R)$ is*

$$\mathbf{W} = V_{(a,b)}\boldsymbol{\rho} + D_{\boldsymbol{\rho}}\mathbf{V}, \tag{2.6}$$

where $V \sim N(0,1)$ and $\mathbf{V} \sim N_k(\mathbf{0}, \Psi)$ are independent, $\Psi = \{\psi_{ij}\} = Corr(\mathbf{V})$, and $D_{\boldsymbol{\rho}} = \mathrm{diag}\left\{\sqrt{1-\rho_1^2}, \ldots, \sqrt{1-\rho_k^2}\right\}$.

2.2. Some formal properties

The theorems described above provide probabilistic proofs for the following properties of the class of $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distributions.

Property 1. *Let $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ and $\mathbf{Z} = D_{\Omega}^{-1/2}(\mathbf{Y} - \mu)$. Then*

$$\mathbf{Z} \sim WN_k^{(u(a),u(b))}(\mathbf{0}, R), \tag{2.7}$$

where $D_{\Omega} = \mathrm{diag}\{\omega_{11}, \ldots, \omega_{kk}\}$, $u(a) = (a-\mu_0)/\sigma_0$ and $u(b) = (b-\mu_0)/\sigma_0$.

Corollary 1 and Property 1 enable us to implement a one-for-one method for generating a random vector with the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ distribution. For generating the doubly truncated standard normal variable $V_{(a,b)}$, the method by Devroye [11, sec. 14.1] may be used.

Property 2. *Let $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ and $\mathbf{Y}_1 \sim N_k(\boldsymbol{\mu}_1, \Omega_1)$ be independent. Then*

$$\mathbf{Y} + \mathbf{Y}_1 \sim WN_k^{(a,b)}(\boldsymbol{\mu}_1^*, \Sigma_1), \tag{2.8}$$

where $\boldsymbol{\mu}_1^* = \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu} + \boldsymbol{\mu}_1 \end{pmatrix}$ and $\Sigma_1 = \begin{pmatrix} \sigma_0^2 & \sigma_0\boldsymbol{\delta}^\top \\ \sigma_0\boldsymbol{\delta} & \Omega + \Omega_1 \end{pmatrix}$.

Theorem 1 and Property 2 imply that the linear sum of independent k-dimensional normals and the truncated k-dimensional normal follows the WTN distribution law. A linear transformation of the variable $\mathbf{Y}$ with joint p.d.f. (2.3) yields the following property.

Property 3. *Let $\mathbf{C}$ be a non-singular matrix with dimension k, $\mathbf{d}$ be a $k \times 1$ vector, and $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$. Then*

$$\mathbf{C}\mathbf{Y} + \mathbf{d} \sim WN_k^{(a,b)}(\boldsymbol{\mu}_2^*, \Sigma_2), \tag{2.9}$$

where $\boldsymbol{\mu}_2^* = \begin{pmatrix} \mu_0 \\ \mathbf{C}\boldsymbol{\mu} + \mathbf{d} \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} \sigma_0^2 & \sigma_0\boldsymbol{\delta}^\top\mathbf{C}^\top \\ \sigma_0\mathbf{C}\boldsymbol{\delta} & \mathbf{C}\Omega\mathbf{C}^\top \end{pmatrix}$.

In the next property, we obtain the marginal distribution of the weighted multivariate normal distribution. For that we consider the partitions

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y_A} \\ \mathbf{Y_B} \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X_A} \\ \mathbf{X_B} \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_\mathbf{A} \\ \boldsymbol{\mu}_\mathbf{B} \end{pmatrix}$$

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_0\boldsymbol{\delta}_\mathbf{A}^\top & \sigma_0\boldsymbol{\delta}_\mathbf{B}^\top \\ \sigma_0\boldsymbol{\delta}_\mathbf{A} & \Omega_{11} & \Omega_{12} \\ \sigma_0\boldsymbol{\delta}_\mathbf{B} & \Omega_{21} & \Omega_{22} \end{pmatrix},$$

where $\mathbf{Y_A}$, $\mathbf{X_A}$, $\boldsymbol{\mu}_\mathbf{A}$, and $\boldsymbol{\delta}_\mathbf{A}$ ($\mathbf{Y_B}$, $\mathbf{X_B}$, $\boldsymbol{\mu}_\mathbf{B}$, and $\boldsymbol{\delta}_\mathbf{B}$) are vectors with the first m (the last $k-m$) elements of the $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\delta}$, respectively. Further Ω_{11} is a $m \times m$ matrix, $\Omega_{12} = \Omega_{21}^\top$ is a $m \times (k-m)$ matrix, and Ω_{22} is a $(k-m) \times (k-m)$ matrix. Then Definition 1 of the weighted multivariate normal distribution yields the following distribution of $\mathbf{Y}_A = [\mathbf{X_A} \mid a < X_0 < b]$.

Property 4. *If* $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ *then* $\mathbf{Y_A} \sim WN_m^{(a,b)}(\boldsymbol{\mu}_1^*, \Sigma_{11})$, *where*

$$\boldsymbol{\mu}_1^* = \begin{pmatrix} \mu_0 \\ \boldsymbol{\mu}_\mathbf{A} \end{pmatrix}, \quad \textit{and} \quad \Sigma_{11} = \begin{pmatrix} \sigma_0^2 & \sigma_0 \boldsymbol{\delta}_\mathbf{A}^\top \\ \sigma_0 \boldsymbol{\delta}_\mathbf{A} & \Omega_{11} \end{pmatrix}.$$

Properties 3 and 4 immediately give a distribution of a linear combination of $\mathbf{Y}_A$. The marginal distribution of $\mathbf{Y}_B$ can be defined similarly.

3. Moments

To compute the moments of a random vector $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$, it suffices to compute the moments of $\mathbf{Z} = D_\Omega^{-1/2}(\mathbf{Y} - \mu)$ where $\mathbf{Z} \sim WN_k^{(u(a),u(b))}(\mathbf{0}, R)$. From (2.3), we see that $\mathbf{Z}$ has density

$$f_\mathbf{Z}(\mathbf{z}; R) = \phi_k(\mathbf{z}; R_{11}) \frac{\Phi\left(\xi^* u(b) - \boldsymbol{\lambda}^{*\top}\mathbf{z}\right) - \Phi\left(\xi^* u(a) - \boldsymbol{\lambda}^{*\top}\mathbf{z}\right)}{\Phi(u(b)) - \Phi(u(a))}, \mathbf{z} \in \mathbb{R}^k, \tag{3.1}$$

where $\phi_k(\mathbf{z}; R_{11})$ denotes the density of a standard k-variate normal having the correlation matrix R_{11}, $\xi^* = (1 - \boldsymbol{\rho}^\top R_{11}^{-1}\boldsymbol{\rho})^{-1/2}$, and $\boldsymbol{\lambda}^{*\top} = \xi^* \boldsymbol{\rho}^\top R_{11}^{-1}$.

Theorem 2. *Let* $\mathbf{Z} \sim WN_k^{(u(a),u(b))}(\mathbf{0}, R)$ *then the moment generating function of the* $\mathbf{Z}$ *distribution is*

$$M_\mathbf{Z}(\mathbf{t}) = \exp\left\{\frac{\mathbf{t}^\top R_{11}\mathbf{t}}{2}\right\}\left\{\frac{\Phi(u(b) - \boldsymbol{\rho}^\top\mathbf{t}) - \Phi(u(a) - \boldsymbol{\rho}^\top\mathbf{t})}{\Phi(u(b)) - \Phi(u(a))}\right\}, \quad \mathbf{t} \in \mathbb{R}^k. \tag{3.2}$$

Proof. An extension of the analogous function as given by Azzalini and Dalla Valle [7, pp. 719] yields

$$\begin{aligned} E[\exp\{\mathbf{t}^\top\mathbf{Z}\}] = \exp\left\{\frac{\mathbf{t}^\top R_{11}\mathbf{t}}{2}\right\}\Big(&E_{R_{11}}\left[\Phi\left(\xi^* u(b) - \boldsymbol{\lambda}^{*\top}R_{11}\mathbf{t} - \boldsymbol{\lambda}^{*\top}\mathbf{Z}^*\right)\right] \\ &- E_{R_{11}}\left[\Phi\left(\xi^* u(a) - \boldsymbol{\lambda}^{*\top}R_{11}\mathbf{t} - \boldsymbol{\lambda}^{*\top}\mathbf{Z}^*\right)\right]\Big) / \{\Phi(u(b)) - \Phi(u(a))\}, \end{aligned}$$

where $\mathbf{Z}^* = \mathbf{Z} - R_{11}\mathbf{t}$. Here $E_{R_{11}}$ means that the expectation is taken with respect to the distribution of $\mathbf{Z}^* \sim N_k(\mathbf{0}, R_{11})$. Now, applying Proposition 4 of Azzalini and Dalla Valle [7] to the expectations, we have the result. □

Naturally, the moments of $\mathbf{Z}$ can be obtained by using moment generating function differentiation. This differentiation yields

$$E[\mathbf{Z}] = -\frac{\phi(u(b)) - \phi(u(a))}{\Phi(u(b)) - \Phi(u(a))}\boldsymbol{\rho}, \quad \text{and} \quad \mathrm{Var}(\mathbf{Z}) = \mathbf{H}, \tag{3.3}$$

where $\mathbf{H} = (h_{ij})$, $i, j = 1, \ldots, k$, with

$$h_{ii} = \mathrm{Var}(Z_i) = 1 - \rho_i^2 \left[\frac{u(b)\phi(u(b)) - u(a)\phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} + \left\{ \frac{\phi(u(b)) - \phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} \right\}^2 \right]$$

and

$$\begin{aligned} h_{ij} &= \mathrm{Cov}(Z_i, Z_j) \\ &= r_{ij} - \rho_i \rho_j \left[\frac{u(b)\phi(u(b)) - u(a)\phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} + \left\{ \frac{\phi(u(b)) - \phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} \right\}^2 \right]. \end{aligned}$$

Thus the mean vector and the covariance matrix of $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ can be obtained from using the relation $\mathbf{Z} = D_\Omega^{-1/2}(\mathbf{Y} - \mu)$ where $\mathbf{Z} \sim WN_k^{(u(a),u(b))}(\mathbf{0}, R)$.

We recall the functions $\alpha_3(k_1, k_2)$ and $\alpha_4(k_1, k_2)$ studied by Sugiura and Gomi [24]. For real values of k_1 and $k_2 (k_1 < k_2)$, $\alpha_3(k_1, k_2)$ and $\alpha_4(k_1, k_2)$ give respective skewness and kurtosis of a doubly truncated standard normal distribution. Here k_1 and k_2 denote the lower and upper truncation points of the distribution. Using these functions, we have the following result.

Theorem 3. *Let* $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$, $\mathbf{Y} = (Y_1, \ldots, Y_k)^\top$, *and let* $\alpha_3(u(a), u(b))$ *and* $\alpha_4(u(a), u(b))$ *be respective skewness and kurtosis of a doubly truncated standard normal distribution, where* $u(a)$ *and* $u(b)$ *denote the lower and upper truncation points of the distribution. Then the skewness* $\alpha_3(i)$ *and kurtosis* $\alpha_4(i)$ *of the marginal distribution of* Y_i, $i = 1, \ldots, k$, *are*

$$\begin{aligned} &\alpha_3(i) = \rho_i^3 (\beta/\gamma_i)^{3/2} \alpha_3(u(a), u(b)), \\ &\alpha_4(i) = \left\{ 3(1-\rho_i^2)^2 + \rho_i^4 \beta^2 \alpha_4(u(a), u(b)) + 6\rho_i^2(1-\rho_i^2)\beta \right\} / \gamma_i^2, \end{aligned}$$

where

$$\begin{aligned} &\beta = \left[1 - \frac{u(b)\phi(u(b)) - u(a)\phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} - \left\{ \frac{\phi(u(b)) - \phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} \right\}^2 \right], \\ &\gamma_i = 1 - \rho_i^2 \left[\frac{u(b)\phi(u(b)) - u(a)\phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} + \left\{ \frac{\phi(u(b)) - \phi(u(a))}{\Phi(u(b)) - \Phi(u(a))} \right\}^2 \right]. \end{aligned}$$

Proof. Note that $\mathbf{Z} \sim WN_k^{(u(a),u(b))}(\mathbf{0}, R)$ where $\mathbf{Z} = D_\Omega^{-1/2}(\mathbf{Y} - \mu)$. Since $\alpha_3(i)$ and $\alpha_4(i)$ for the Y_i and $Z_i = (Y_i - \mu_i)/(\omega_{ii})^{1/2}$ distributions are the same, it suffices to obtain $\alpha_3(i)$ and $\alpha_4(i)$ for the distribution of Z_i random variables. From Corollary 1, we see that $Z_i = \rho_i V_{(u(a),u(b))} + \sqrt{1-\rho_i^2} V_i$, where V and V_i are independent standard normal variables. $\mathrm{Var}(Z_i) = \gamma_i$ by (3.3) and

$$\gamma_i^{3/2} \alpha_3(i) = \rho_i^3 E[V_{(u(a),u(b))} - EV_{(u(a),u(b))}]^3 = \rho_i^3 \alpha_3(u(a), u(b)) \mathrm{Var}(V_{(u(a),u(b))})^{3/2}.$$

Therefore, some algebra using the fact that

$$\begin{aligned} &E[V_{(u(a),u(b))}] = \{\phi(u(a)) - \phi(u(b))\}/\{\Phi(u(b)) - \Phi(u(a))\} \quad \text{and} \\ &\mathrm{Var}(V_{(u(a),u(b))}) = \beta, \end{aligned}$$

we have the expression for $\alpha_3(i)$. A similar proof holds for the derivation of $\alpha_4(i)$. See, for example, Johnson et al. [16] for the moments of the doubly truncated standard normal variable $V_{(u(a),u(b))}$. □

Consider a special case when $(u(a), u(b)) = (0, \infty)$. Using the results of Sugiura and Gomi [24], we see that $\alpha_3(0, \infty) = (4/\pi - 1)(2/\pi)^{1/2}(1 - 2/\pi)^{-3/2}$ and $\alpha_4(0, \infty) = (3 - 4/\pi - 12/\pi^2)(1 - 2/\pi)^2$. These are the skewness and kurtosis of the left truncated half normal distribution. For the case the expressions of the indices of skewness and kurtosis given by Theorem 3 reduce to

$$\alpha_3(i) = \left(\frac{4}{\pi} - 1\right)\left(\frac{2}{\pi}\right)^{1/2}\left(1 - \frac{2\rho_i^2}{\pi}\right)^{-3/2}\rho_i^3,$$

$$\alpha_4(i) = \left(3 - \frac{12\rho_i^2}{\pi} - \frac{4\rho_i^4}{\pi}\right)\left(1 - \frac{2\rho_i^2}{\pi}\right)^{-2}.$$

The value of the skewness of Y_i agrees with that of the skew-normal given in Azzalin and Dalla Valle [7].

Corollary 2. *For* $\mathbf{Y} \sim WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$, $\mathbf{Y} = (Y_1, \ldots, Y_k)^\top$, *with* $a < b$, *the skewness of the marginal distribution of* Y_i, $i = 1, \ldots, k$ *is:* (i) *If* $|u(a)| < u(b)$, *the distribution of the* Y_i *variable is skewed to the right (left) when* $\rho_i > 0$ $(\rho_i < 0)$. (ii) *If* $|u(a)| > u(b)$, *the distribution of the* Y_i *variable is skewed to the right (left) when* $\rho_i < 0$ $(\rho_i > 0)$. (iii) *If* $|u(a)| = u(b)$ *for* $u(a) \neq 0$, *the distribution of the* Y_i *variable is symmetric, where* $u(a) = (a - \mu_0)/\sigma_0$ *and* $u(b) = (b - \mu_0)/\sigma_0$.

Proof. It is obvious that $\alpha_3(u(a), u(b)) > 0$ for $|u(a)| < u(b)$, $\alpha_3(u(a), u(b)) < 0$ for $|u(a)| > u(b)$, and $\alpha_3(u(a), u(b)) = 0$ for $|u(a)| = u(b)$, for $u(a) \neq 0$. This fact and the formula of $\alpha_3(i)$ in Theorem 3 immediately give the result. □

4. A Bayesian perspective

A well-known property of the multivariate normal distribution is that, if $\mathbf{X}$ is $N_k(\mathbf{Y}, \Lambda)$ where a priori $\mathbf{Y}$ is a k-variate normal random vector variable, then the posterior distribution $\mathbf{Y}$ is still k-variate normal. An analogous fact is true for the weighted normal distributions. See Azzalini [5] in a scalar case and Liseo [18] and the references therein for the multivariate cases. The property of closure also applies for the $WN_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma)$ prior. If a priori $\mathbf{Y}$ has the probability density function in (2.3), the posterior density function of $\mathbf{Y}$ given that $\mathbf{X} = \mathbf{x}$ is

$$p(\mathbf{y} \mid \mathbf{x}) \propto \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{x})^\top \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{x})\right\} \times \frac{\Phi\left(\xi u(b) - \boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu})\right) - \Phi\left(\xi u(a) - \boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu})\right)}{\Phi(u(b)) - \Phi(u(a))}, \quad \mathbf{y} \in \mathbb{R}^k,$$

where $\boldsymbol{\mu}_\mathbf{x} = (\Omega^{-1} + \Lambda^{-1})^{-1}(\Omega^{-1}\boldsymbol{\mu} + \Lambda^{-1}\mathbf{x})$, $\mathbf{K} = (\Omega^{-1} + \Lambda^{-1})^{-1}$, $u(a) = (a - \mu_0)/\sigma_0$, $u(b) = (b - \mu_0)/\sigma_0$, $\xi = (1 - \boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta})^{-1/2}$, and $\boldsymbol{\lambda}^\top = \xi\boldsymbol{\delta}^\top\Omega^{-1}$.

Some algebra, including the binomial inverse theorem (see, for example, Press [22, pp. 49]) for $(\Omega^{-1} + \Lambda^{-1})^{-1}$, shows that

$$p(\mathbf{y} \mid \mathbf{x}) = \phi_k(\mathbf{y}; \boldsymbol{\mu}_\mathbf{x}, \mathbf{K})\frac{\Phi\left(\xi_* u(b)^* - \boldsymbol{\lambda}_*^\top(\mathbf{y} - \boldsymbol{\mu}_\mathbf{x})\right) - \Phi\left(\xi_* u(a)^* - \boldsymbol{\lambda}_*^\top(\mathbf{y} - \boldsymbol{\mu}_\mathbf{x})\right)}{\Phi(u(b)^*) - \Phi(u(a)^*)} \tag{4.1}$$

for $\mathbf{y} \in \mathbb{R}^k$, where $\phi_k(\mathbf{y}; \boldsymbol{\mu}_{\mathbf{x}}, \mathbf{K})$ denotes the p.d.f. of $N_k(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{K})$ distribution,

$$\xi_* = (1 - \boldsymbol{\delta}^{*T}\mathbf{K}^{-1}\boldsymbol{\delta}^*)^{-1/2}, \qquad \boldsymbol{\lambda}_* = \xi_*\boldsymbol{\delta}^{*T}\mathbf{K}^{-1},$$

$$\boldsymbol{\delta}^* = \frac{\xi \mathbf{K}\Omega^{-1}\boldsymbol{\delta}}{\left\{1 + \xi^2 \left(\mathbf{K}\Omega^{-1}\boldsymbol{\delta}\right)^\top \mathbf{K}^{-1} \left(\mathbf{K}\Omega^{-1}\boldsymbol{\delta}\right)\right\}^{1/2}},$$

$$u(a)^* = \frac{\xi}{\xi_*}\left\{u(a) + \boldsymbol{\delta}^\top \Lambda^{-1}(\Lambda^{-1} + \Lambda^{-1}\Omega\Lambda^{-1})^{-1}\Lambda^{-1}\Omega(\Lambda^{-1}\mathbf{x} + \Omega^{-1}\boldsymbol{\mu}) - \boldsymbol{\delta}^\top\Lambda^{-1}\mathbf{x}\right\},$$

$$u(b)^* = \frac{\xi}{\xi_*}\left\{u(b) + \boldsymbol{\delta}^\top \Lambda^{-1}(\Lambda^{-1} + \Lambda^{-1}\Omega\Lambda^{-1})^{-1}\Lambda^{-1}\Omega(\Lambda^{-1}\mathbf{x} + \Omega^{-1}\boldsymbol{\mu}) - \boldsymbol{\delta}^\top\Lambda^{-1}\mathbf{x}\right\}.$$

We note that (4.1) is the p.d.f. of the $WN_k^{(u(a)^*, u(b)^*)}(\boldsymbol{\mu}^*_{\mathbf{x}}, \Sigma_{\mathbf{x}})$, where

$$\boldsymbol{\mu}^*_{\mathbf{x}} = \begin{pmatrix} 0 \\ \boldsymbol{\mu}_{\mathbf{x}} \end{pmatrix} \quad \text{and} \quad \Sigma_{\mathbf{x}} = \begin{pmatrix} 1 & \boldsymbol{\delta}^{*T} \\ \boldsymbol{\delta}^* & \mathbf{K} \end{pmatrix}.$$

To conclude, the class of WTN prior distributions forms a conjugate family for the k-variate normal mean vector. In other words, if $\mathbf{X}$ is $N_k(\mathbf{Y}, \Lambda)$, then the posterior density of $\mathbf{Y}$ is in the class of WTN distributions for all $\mathbf{x}$ whenever the prior distribution of $\mathbf{Y}$ is in the class of WTN distributions.

5. Applications

5.1. A constrained estimation of the normal mean vector

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with $n > k$, be an independent sample of size n from $N_k(\boldsymbol{\theta}, \Lambda)$. Suppose that $\boldsymbol{\theta}$ satisfies a constrained inequality of the form

$$a \le \mathbf{c}^\top\boldsymbol{\theta} \le b \tag{5.1}$$

for any constant vector $\mathbf{c} \neq \mathbf{0}$. This framework is slightly different from that discussed by Liseo [18] where a one-sided inequality constraint is considered. Suppose that we want to estimate $\boldsymbol{\theta}$ under the constraint (5.1) by using a Bayesian approach. As a joint prior density in the analysis of the inequality constrained multivariate normal model, we assume that our prior information about $\boldsymbol{\theta}$ and Λ are a priori independent. Accordingly, the hierarchical model considered by Liseo [18] yields

$$p(\boldsymbol{\theta}, \Lambda \mid a \le \mathbf{c}^\top\boldsymbol{\theta}_0 \le b) = p(\boldsymbol{\theta} \mid a \le \mathbf{c}^\top\boldsymbol{\theta}_0 \le b)p(\Lambda). \tag{5.2}$$

Suppose that the prior information is readily accessible by the distributions about the parameters where $\boldsymbol{\theta} \mid \boldsymbol{\theta}_0 \sim N_k(\boldsymbol{\theta}_0, \Sigma), \boldsymbol{\theta}_0 \sim N_k(\boldsymbol{\mu}, \Psi)$, and $p(\Lambda) \sim |\Lambda|^{-(k+1)/2}$, a diffuse prior for Λ, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^\top$ and $\Omega = (\Sigma + \Psi)$ are arbitrary parameters to be predetermined by an analyst. The joint prior may be written as

$$p(\boldsymbol{\theta}, \Lambda \mid a \le \mathbf{c}^\top\boldsymbol{\theta} \le b) \propto |\Lambda|^{-(k+1)/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^\top\Omega^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})\right\}$$
$$\times \frac{\Phi\left(\xi u(b) - \boldsymbol{\lambda}^\top(\boldsymbol{\theta} - \boldsymbol{\mu})\right) - \Phi\left(\xi u(a) - \boldsymbol{\lambda}^\top(\boldsymbol{\theta} - \boldsymbol{\mu})\right)}{\Phi(u(b)) - \Phi(u(a))},$$

where $u(a) = (a - \mathbf{c}^\top\boldsymbol{\mu})(\mathbf{c}^\top\Omega\mathbf{c})^{-1/2}$, $u(b) = (b - \mathbf{c}^\top\boldsymbol{\mu})(\mathbf{c}^\top\Omega\mathbf{c})^{-1/2}$, $\xi = (1 - \boldsymbol{\delta}^\top\Omega^{-1}\boldsymbol{\delta})^{-1/2}$, and $\boldsymbol{\lambda} = \xi\Omega^{-1}\boldsymbol{\delta}$ with $\boldsymbol{\delta} = \Omega\mathbf{c}(\mathbf{c}^\top\Omega\mathbf{c})^{-1/2}$.

The joint posterior density for $\boldsymbol{\theta}$ and Λ is found by multiplying the likelihood function of $\mathbf{X_1}, \ldots, \mathbf{X}_n$. It is

$$p(\boldsymbol{\theta}, \Lambda \mid a \le \mathbf{c}^\top \boldsymbol{\theta} \le b, \text{Data}) \propto |\Lambda|^{-(n+k+1)/2} \exp\left\{-\frac{1}{2}\text{tr}\left[\Lambda^{-1} V\right]\right\} \exp\left\{-\frac{\mathbf{Q}}{2}\right\}$$
$$\times \frac{\Phi\left(\xi u(b) - \boldsymbol{\lambda}^\top(\boldsymbol{\theta} - \boldsymbol{\mu})\right) - \Phi\left(\xi u(a) - \boldsymbol{\lambda}^\top(\boldsymbol{\theta} - \boldsymbol{\mu})\right)}{\Phi(u(b)) - \Phi(u(a))},$$

where

$$\mathbf{Q} = (\boldsymbol{\theta} - \boldsymbol{\mu}_\mathbf{P})^\top \mathbf{P}^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_\mathbf{P}), \qquad \mathbf{P} = \left(n\Lambda^{-1} + \Omega^{-1}\right)^{-1},$$
$$\boldsymbol{\mu}_\mathbf{P} = \mathbf{P}\left[n\Lambda^{-1}\bar{\mathbf{x}} + \Omega^{-1}\boldsymbol{\mu}\right],$$

and $V = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$.

Marginal posterior densities of $\boldsymbol{\theta}$ and Λ are complicated. Instead, the Gibbs sampler can be used for inference of the constrained normal model. To apply the Gibbs sampler, we need full conditional posterior distributions of $\boldsymbol{\theta}$ and Λ. The full conditional distribution of $\boldsymbol{\theta}$ is obtained from algebra using the conjugate property of the WTN distribution. The conjugate property shows that the full conditional distribution of $\boldsymbol{\theta}$ has the following p.d.f.:

$$p(\boldsymbol{\theta} \mid \Lambda, a \le \mathbf{c}^\top \boldsymbol{\theta} \le b, \text{Data})$$
$$= \phi_k(\boldsymbol{\theta}; \boldsymbol{\mu}_\mathbf{P}, \mathbf{P}) \frac{\Phi\left(\xi_0 u_0(b) - \boldsymbol{\lambda}_0^\top(\mathbf{y} - \boldsymbol{\mu}_\mathbf{P})\right) - \Phi\left(\xi_0 u_0(a) - \boldsymbol{\lambda}_0^\top(\mathbf{y} - \boldsymbol{\mu}_\mathbf{P})\right)}{\Phi(u_0(b)) - \Phi(u_0(a))}, \tag{5.3}$$

where $\phi_k(\boldsymbol{\theta}; \boldsymbol{\mu}_\mathbf{P}, \mathbf{P})$ denotes the p.d.f. of $N_k(\boldsymbol{\mu}_\mathbf{P}, \mathbf{P})$ distribution,

$$\xi_0 = (1 - \boldsymbol{\delta}_0^\top \mathbf{P}^{-1} \boldsymbol{\delta}_0)^{-1/2}, \qquad \boldsymbol{\lambda}_0 = \xi_0 \boldsymbol{\delta}_0^\top \mathbf{P}^{-1},$$
$$\boldsymbol{\delta}_0 = \frac{\xi \mathbf{P}\Omega^{-1}\boldsymbol{\delta}}{\left\{1 + \xi^2 \left(\mathbf{P}\Omega^{-1}\boldsymbol{\delta}\right)^\top \mathbf{P}^{-1} \left(\mathbf{P}\Omega^{-1}\boldsymbol{\delta}\right)\right\}^{1/2}},$$
$$u_0(a) = \frac{\xi}{\xi_0}\left\{u(a) + n\boldsymbol{\delta}^\top \Lambda^{-1}(\Lambda^{-1} + n\Lambda^{-1}\Omega\Lambda^{-1})^{-1}\Lambda^{-1}\Omega(n\Lambda^{-1}\bar{\mathbf{x}} + \Omega^{-1}\boldsymbol{\mu}) \right.$$
$$\left. - n\boldsymbol{\delta}^\top \Lambda^{-1}\bar{\mathbf{x}}\right\},$$
$$u_0(b) = \frac{\xi}{\xi_0}\left\{u(b) + n\boldsymbol{\delta}^\top \Lambda^{-1}(\Lambda^{-1} + n\Lambda^{-1}\Omega\Lambda^{-1})^{-1}\Lambda^{-1}\Omega(n\Lambda^{-1}\bar{\mathbf{x}} + \Omega^{-1}\boldsymbol{\mu}) \right.$$
$$\left. - n\boldsymbol{\delta}^\top \Lambda^{-1}\bar{\mathbf{x}}\right\}.$$

Therefore, the full conditional distributions are

$$\boldsymbol{\theta} \mid (\Lambda, a \le \mathbf{c}^\top \boldsymbol{\theta} \le b, \text{Data}) \sim WN_k^{(u_0(a), u_0(b))}(\boldsymbol{\mu}_0^*, \Sigma_0), \tag{5.4}$$

and

$$\Lambda \mid (\boldsymbol{\theta}, a \le \mathbf{c}^\top \boldsymbol{\theta} \le b, \text{Data}) \sim W_k^{-1}(n + k + 1, V + n(\boldsymbol{\theta} - \bar{\mathbf{x}})(\boldsymbol{\theta} - \bar{\mathbf{x}})^\top), \tag{5.5}$$

an inverted Wishart distribution with scale matrix $V + n(\boldsymbol{\theta} - \bar{\mathbf{x}})(\boldsymbol{\theta} - \bar{\mathbf{x}})^\top$ and $n + k + 1$ degrees of freedom. Here

$$\boldsymbol{\mu}_0^* = \begin{pmatrix} 0 \\ \boldsymbol{\mu}_{\mathbf{P}} \end{pmatrix} \quad \text{and} \quad \Sigma_0 = \begin{pmatrix} 1 & \boldsymbol{\delta}_0^\top \\ \boldsymbol{\delta}_0 & \mathbf{P} \end{pmatrix}.$$

When Λ is assumed to be known, the distribution (5.4) becomes the exact marginal distribution of $\boldsymbol{\theta}$. The Gibbs sampler proceeds by alternatively sampling from the full conditional distributions. For the sampling from the $WN_k^{(u_0(a),u_0(b))}(\boldsymbol{\mu}_0^*, \Sigma_0)$ distribution, we may use the one-for-one method described in Section 2.2. Odell and Feiveson [21] provide an easy and efficient algorithm for sampling from the inverted Wishart distribution.

5.2. Screening in a normal model

The problem of screening has received considerable attention in the literature (see, Boys and Dunsmore [9] and the references therein). Suppose that certain measurements on an individual are described by a k-variate normal random vector $\mathbf{X}$. We assume that we are given a specification region $R_{\mathbf{X}}$ such that if $\mathbf{c}^\top\mathbf{X} \in R_{\mathbf{X}}$ the individual is considered success, where $\mathbf{c}$ is a known $k \times 1$ constant vector. Then $\gamma = P(\mathbf{c}^\top\mathbf{X} \in R_{\mathbf{X}})$ is the probability of success or the proportion of successful individuals in the population. Typically, $R_{\mathbf{X}}$ is of the form $[L^*, U^*]$. The motivation behind screening is to try to increase the proportion of successes by eliminating some individuals. This is done by recording a second measurement X_0 on each individual, where X_0 and $\mathbf{X}$ are correlated and a normal variable X_0 is easier to measure than $\mathbf{X}$.

Let $\mathbf{X}^* = (X_0, \mathbf{X}^\top)^\top$ be a $(k+1)$-variate normal random vector such that $\mathbf{X}^* \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma)$, where $\boldsymbol{\mu}^* = (\mu_0, \boldsymbol{\mu})^\top$ and the covariance matrix Σ has the form

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_0\boldsymbol{\delta}^\top \\ \sigma_0\boldsymbol{\delta} & \Omega \end{pmatrix}.$$

Suppose that $R_{X_0} = \{a \le X_0 \le b\}$ denotes the specification region for the X_0 values so that the individual will be retained or screened out if $X_0 \in R_{X_0}$ or not, respectively. Then our problem of screening is to specify the shortest length $R_{\mathbf{X}}$ so that $P(\mathbf{c}^\top\mathbf{X} \in R_{\mathbf{X}} \mid X_0 \in R_{X_0}) = \alpha$, for some specified $\alpha > \gamma$.

A simple method may be followed if we use the d.f. (2.4) of the WTN distribution: Although at the screening stage only the dichotomy on $\mathbf{c}^\top\mathbf{X}$ is important. We assume throughout that there is an underlying variable which can be measured and that we have available data in the form of a random sample $(\mathbf{x}_1, x_{01}), \ldots, (\mathbf{x}_n, x_{0n})$ from the unscreened data.

The shortest region of the form $R_{\mathbf{X}} = [L^*, U^*]$ might require a two-sided region such that

$$P(L^* \le \mathbf{c}^\top\mathbf{X} \le U^* \mid a \le X_0 \le b) = \alpha. \tag{5.6}$$

This is equivalent to

$$\frac{L(z_{L^*}, u(a); \rho) - L(z_{L^*}, u(b); \rho)}{\Phi(u(b)) - \Phi(u(a))} - \frac{L(z_{U^*}, u(a); \rho) - L(z_{U^*}, u(b); \rho)}{\Phi(u(b)) - \Phi(u(a))} = \alpha \tag{5.7}$$

by the d.f. (2.4), where $z_{L^*} = (L^* - \mathbf{c}^\top\boldsymbol{\mu})(\mathbf{c}^\top\Omega\mathbf{c})^{-1/2}$, $z_{U^*} = (U^* - \mathbf{c}^\top\boldsymbol{\mu})(\mathbf{c}^\top\Omega\mathbf{c})^{-1/2}$, $u(a) = (a - \mu_0)/\sigma_0$, and $u(b) = (b - \mu_0)/\sigma_0$, and $\rho = \mathbf{c}^\top\boldsymbol{\delta}(\mathbf{c}^\top\Omega\mathbf{c})^{-1/2}$. Here $L(\alpha, \beta; \rho) = P(V_1 > \alpha, V_2 > \beta)$ is the orthant probability of a bivariate standard normal variable (V_1, V_2) with correlation ρ. Computing methods that evaluate $L(\alpha, \beta; \rho)$ have been given by Donnelly [12] and Joe [15]. When $\boldsymbol{\mu}^*$ and Σ are unknown this can be estimated by using the sample.

Our first problem is to find values of L^* and U^* that minimizes the length $U^* - L^*$ under the restriction

$$F_Y(U^*) - F_Y(L^*) = \int_{L^*}^{U^*} f_Y(y)\mathrm{d}y = \alpha,$$

where $F_Y(y)$ and $f_Y(y)$ are the d.f. and p.d.f. of the $WN^{(a,b)}(\boldsymbol{\mu}_3^*, \Sigma_3)$ distribution defined by (2.3) and (2.4), respectively, where

$$\boldsymbol{\mu}_3^* = \begin{pmatrix} \mu_0 \\ \mathbf{c}^\top \boldsymbol{\mu} \end{pmatrix} \quad \text{and} \quad \Sigma_3 = \begin{pmatrix} \sigma_0^2 & \sigma_0 \mathbf{c}^\top \boldsymbol{\delta} \\ \sigma_0 \mathbf{c}^\top \boldsymbol{\delta} & \mathbf{c}^\top \Omega \mathbf{c} \end{pmatrix}$$

by Definition 1. By taking derivatives of the restricting equation with respect to U^*, we see that $\mathrm{d}L^*/\mathrm{d}U^* = f_Y(U^*)/f_Y(L^*)$. Upon setting $\mathrm{d}(U^* - L^*)/\mathrm{d}U^* = 0$, we see that the constrained minimization yields a condition $f_Y(U^*) = f_Y(L^*)$, i.e.

$$\frac{\phi(z_{U^*})}{\phi(z_{L^*})} = \frac{\Phi(\delta u(b) - \lambda z_{L^*}) - \Phi(\delta u(a) - \lambda z_{L^*})}{\Phi(\delta u(b) - \lambda z_{U^*}) - \Phi(\delta u(a) - \lambda z_U^*)} \tag{5.8}$$

by (2.3), where $\delta = (1-\rho^2)^{-1/2}$, $\lambda = \rho\delta$, and $\phi(\cdot)$ denotes the p.d.f. of the $N(0, 1)$ distribution. This condition (5.8) along with (5.7) will be used to calculate the values U^* and L^*.

Let our second problem of screening be to determine a specification region R_{X_0} for a given specification region $R_{\mathbf{X}}$. Then the aim is to specify R_{X_0} so that $P(\mathbf{c}^\top \mathbf{X} \in R_{\mathbf{X}} \mid X_0 \in R_{X_0}) = \alpha$, for some specified $\alpha > \gamma$. In this case, there are many pairs of (a, b), $a < b$, which satisfy (5.7), so some additional criteria become necessary. One such criterion is obtained by considering the probability attached to the possible mistake that an individual screened out by the process would have been a success. It seems sensible therefore to choose that pair (a, b) which maximizes $\beta = P(X_0 \in R_{X_0}) = \Phi(u(b)) - \Phi(u(a))$. A Lagrangian constrained maximization analysis (maximizing β subject to (5.7)) yields the additional criterion

$$Q(u(a)) = Q(u(b)), \tag{5.9}$$

where $Q(c) = \int_{z_{L^*}}^{z_{U^*}} f(v_1 \mid v_2 = c)dv_1$ and $f(v_1, v_2)$ is the p.d.f. of the bivariate standard normal random variables (V_1, V_2) with correlation ρ. Thus

$$Q(c) = \Phi\left(\frac{z_{U^*} - \rho c}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{z_{L^*} - \rho c}{\sqrt{1-\rho^2}}\right).$$

The solution satisfying (5.7) and (5.9) gives an optimal (a, b) provided that $a < b$ and $\alpha > \gamma$.

6. Concluding remarks

The primary focus of this study is to propose and study a class of weighted multivariate normal distributions that is associated with a doubly truncated multivariate normal distribution. From an applied viewpoint, the distribution is useful for inequality constrained analysis and selection modeling. Procedures for solving those problems by using the distribution are illustrated in Section 5. On the theoretical side, the class enables us to have at hand a family of densities with the following properties: (i) strict inclusion of the multivariate normal and multivariate skew-normal by Azzalini and Dalla Valle [7]; (ii) mathematical tractability; (iii) rich distributional properties; (iv) an important Bayesian property (providing a family of conjugate priors to an inequality constrained multivariate normal mean vector); and (v) availability for various scale mixture distributions.

For a class $\mathfrak{F}$ of mixtures of the WTN distributions may be defined by

$$\mathfrak{F} = \left\{ F : WN_k^{(a,b)}(\boldsymbol{\mu}^*, \kappa(\eta)\Sigma), \eta \sim \pi(\eta) \right.$$
$$\left. \text{with } \kappa(\eta) > 0, \eta > 0 \text{ and } \int_0^\infty \pi(\eta)\mathrm{d}\eta = 1 \right\}. \tag{6.1}$$

Then, from (2.3), the density $f_{\mathbf{Y}}(\mathbf{y})$ of the class is

$$D_\eta^{-1} \int_0^\infty \phi_k(\mathbf{y}; \boldsymbol{\mu}, \kappa(\eta)\Omega) \left\{ \Phi\left(\frac{\xi u(b) - \boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu})}{\kappa(\eta)^{1/2}} \right) \right.$$
$$\left. - \Phi\left(\frac{\xi u(a) - \boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu})}{\kappa(\eta)^{1/2}} \right) \right\} \mathrm{d}H(\eta), \quad \mathbf{y} \in \mathbb{R}^k, \tag{6.2}$$

where $\phi_k(\mathbf{y}; \boldsymbol{\mu}, \kappa(\eta)\Omega)$ is the p.d.f. of $N_k(\boldsymbol{\mu}, \kappa(\eta)\Omega)$, η is a mixing variable with the d.f. $H(\eta)$, $\kappa(\eta)$ is a mixing function, and

$$D_\eta = \int_0^\infty \Phi\left(\frac{u(b)}{\kappa(\eta)^{1/2}} \right) \mathrm{d}H(\eta) - \int_0^\infty \Phi\left(\frac{u(a)}{\kappa(\eta)^{1/2}} \right) \mathrm{d}H(\eta).$$

As in the two special cases, (6.2) gives the p.d.f. of weighted multivariate normal and weighted multivariate t_ν when $H(\cdot)$ degenerate, with $\kappa(\eta) = 1$ and $\kappa(\eta) = 1/\eta$ with $H(\eta)$ as the d.f. of a Gamma distribution, i.e., $\eta \sim G(\nu/2, 2/\nu)$, where $G(\alpha, \beta)$ denotes a Gamma distribution with mean $\alpha\beta$. See Branco and Dey [10] for various cases of mixing variable and mixing function.

In this regard, the weighted multivariate t_ν distribution can be defined as follows. Let $\mathbf{X}^* \sim N_{k+1}(\boldsymbol{\mu}^*, \Sigma/\eta)$ with $\eta \sim G(\nu/2, 2/\nu)$, where $\mathbf{X}^* = (X_0, \mathbf{X}^\top)^\top$. Then the scale mixture distribution of $\mathbf{Y} = [\mathbf{X} \mid a < X_0 < b]$ is a weighted k-variate t_ν distribution. The final expression of its density is

$$f_{\mathbf{Y}}(\mathbf{y}) = f_\nu(\mathbf{y}; \boldsymbol{\mu}, \Omega) \left\{ F_{\nu+k}\left(\frac{(\nu + k)^{1/2}(\xi u(b) - \boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu}))}{(\nu + (\mathbf{y} - \boldsymbol{\mu})^\top \Omega^{-1}(\mathbf{y} - \boldsymbol{\mu}))^{1/2}} \right) \right.$$
$$\left. - F_{\nu+k}\left(\frac{(\nu + k)^{1/2}(\xi u(a) - \boldsymbol{\lambda}^\top(\mathbf{y} - \boldsymbol{\mu}))}{(\nu + (\mathbf{y} - \boldsymbol{\mu})^\top \Omega^{-1}(\mathbf{y} - \boldsymbol{\mu}))^{1/2}} \right) \right\} \Big/ \{F_\nu(u(b)) - F_\nu(u(a))\}, \tag{6.3}$$

where $f_\nu(\cdot; \boldsymbol{\mu}, \Omega)$ is the p.d.f. of a k-variate generalized Student's t_ν distribution with location and scale parameters $\boldsymbol{\mu}$ and Ω, respectively. $F_{\nu^*}(\cdot)$ is the c.d.f. of an univariate standard t distribution with degrees of freedom ν^*. The distribution of $\mathbf{Y}$ may be abbreviated by saying $\mathbf{Y}$ is $Wt_k^{(a,b)}(\boldsymbol{\mu}^*, \Sigma, \nu)$.

Acknowledgments

This research was supported by the Korean Research Foundation Grant funded by the Korean Government (MOEHRD) KRF-2007-313-C00126.

References

[1] R.B. Arellano-Valle, A. Azzalini, On unification of families of skew-normal distributions, Scand. J. Statist. 33 (2006) 561–574.

[2] R.B. Arellano-Valle, M.D. Branco, M.G. Genton, A unified view on skewed distributions arising from selections, Canad. J. Statist. 34 (2006) 581–601.

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

[4] B.C. Arnold, R.J. Beaver, R.A. Groeneveld, W.Q. Meeker, The nontruncated marginal of a truncated bivariate normal distribution, Psychometrica 58 (1993) 471–478.

[5] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Statist. 12 (1985) 171–178.

[6] A. Azzalini, The skew-normal distribution and related multivariate families (with discussion), Scand. J. Statist. 32 (2005) 159–188.

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

[8] M.J. Bayarri, M. DeGroot, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), A BAD View of Weighted Distributions and Selection Models, in: Bayesian Statistics, vol. 4, Oxford University Press, Oxford, 1992.

[9] R.J. Boys, I.R. Dunsmore, Screening in a normal model, J. Roy. Statist. Soc. Ser. B 48 (1986) 60–69.

[10] M.D. Branco, D.K. Dey, A general class of multivariate skew-normal distributions, J. Multivariate Anal. 79 (2001) 99–113.

[11] L. Devroye, Non-Uniform Random Variate Generation, Springer Verlag, New York, 1986.

[12] T.G. Donnelly, Algorithm 426: Bivariate normal distribution, Communications in the Association for Computing Machinery 16 (1973) 638.

[13] M.G. Genton, Skew-Elliptical Distributions and their Applications; A Journey beyond Normality, Chapman & Hall, New York, 2004.

[14] M.G. Genton, Discussion of "the skew-normal", Scand. J. Statist. 32 (2005) 189–198.

[15] H. Joe, Approximation to multivariate rectangle probabilities based on conditional expectations, J. Amer. Statist. Assoc. 90 (1995) 957–966.

[16] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, vol. 1, John Wiley & Sons, New York, 1994.

[17] H.J. Kim, On a class of two-piece skew-normal distributions, Statistics 39 (2005) 537–553.

[18] B. Liseo, Skew-elliptical distributions in Bayesian inference, in: M.G. Genton (Ed.), Skew-Elliptical Distributions and their Applications, Chapman & Hall, New York, 2004, pp. 153–171.

[19] Y. Ma, M.G. Genton, A flexible class of skew-symmetric distributions, Scand. J. Statist. 31 (2004) 459–468.

[20] Y. Ma, M.G. Genton, A.A. Tsiatis, Locally efficient semiparametric estimators for generalized skew-elliptical distributions, J. Amer. Statist. Assoc. 100 (2005) 980–989.

[21] P.L. Odell, A.H. Feiveson, A numerical procedure to generate a sample covariance matrix, J. Amer. Statist. Assoc. 61 (1996) 199–203.

[22] S.J. Press, Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference, Krieger, Florida, 1982.

[23] C.R. Rao, Weighted distributions arising out of methods of ascertainment, in: A.G. Aitkinson, S.E. Feinberg (Eds.), A Celebration of Statistics: The ISI Centenary Volume, Springer, New York, 1985.

[24] N. Sugiura, A. Gomi, Pearson diagrams for truncated normal and truncated Weibull distributions, Biometrika 72 (1985) 219–222.