SORT 38 (1) January-June 2014, 43-52

Regression analysis using order statistics and their concomitants

Abdul Rasoul Ziaei[1], Ayyub Sheikhi[2] and Vahid Amirzadeh[2]

Abstract

In this work we derive the exact joint distribution of linear combinations of order statistics and linear combinations of their concomitants and some auxiliary variables in multivariate normal distribution. By extending the results of Sheikhi and Jamalizadeh we investigate some regression equations. Our results generalize those obtained in previous research by Viana, Lee and Loperfido.

MSC: 62G30, 62J05

Keywords: Unified skew normal, L-statistic, exchangeable multivariate normal distribution, order statistic, concomitant.

1. Introduction

The theory of order statistics and their concomitants plays an essential role in statistical inference. An excellent review of development in this field is available in David and Nagaraja (2003). There have been many studies with emphasis on distribution theory. Tsukibayashi (1998) found the moments and the joint distribution of an extreme value and its concomitant. Goel and Hall (1994) discussed the difference between concomitants and order statistics. Yang (1981) studied the linear functions of concomitants of order statistics. Arellano-Valle and Genton (2007) obtained the exact distribution of linear combinations of order statistics from dependent random variables. Viana and Lee (2006) studied the covariance structure of two random vectors $\mathbf{X}_{(n)}$ and $\mathbf{Y}_{[n]}$ in the presence of a random variable Z where $\mathbf{X}_{(n)} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})^{\mathsf{T}}$ is the vector of order statistics

[1] Islamic Azad University, Dashtestan Branch (IRN), ziaei85@yahoo.com

[2] Department of Statistics, Shahid Bahonar University, Kerman, Iran,
sheikhy.a@uk.ac.ir, v_amirzadeh@uk.ac.ir

Received: January 2013
Accepted: November 2013

obtained from $\mathbf{X}_{n\times 1}$ and $\mathbf{Y}_{[n]} = (Y_{[1]},\ Y_{[2]},\ldots,Y_{[n]})^{\mathsf{T}}$ is the vector of concomitants. They also discussed some regression equations between order statistics, concomitants and the covariate variable Z, while Olkin and Viana (1995) studied the covariance structure and several regression models when $(X,Y_1,Y_2)^{\mathsf{T}}$ has a trivariate exchangeable normal distribution. Loperfido (2008) determined the joint distribution of an auxiliary variable X and the maximum of Y_1 and Y_2, i.e. $(X,\ Y_{(2)})^{\mathsf{T}}$. Sheikhi and Jamalizadeh (2011) found the joint distribution of two linear combinations of order statistics in the presence of a covariate random variable and presented some regression analyses.

We assume that the joint distribution of a covariate p-dimensional random vector $\mathbf{Z}$ and two n-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ follows a $2n+p$ dimensional multivariate normal vector with positive definite covariance matrix, i.e.

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n+p}\left(\mu = \begin{pmatrix} \mu_z \\ \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{pmatrix}, \sum = \begin{pmatrix} \sum_{\mathbf{zz}} & \sum_{\mathbf{xz}}^{\mathsf{T}} & \sum_{\mathbf{yz}}^{\mathsf{T}} \\ \sum_{\mathbf{xz}} & \sum_{\mathbf{xx}} & \sum_{\mathbf{xy}}^{\mathsf{T}} \\ \sum_{\mathbf{yz}} & \sum_{\mathbf{xy}} & \sum_{\mathbf{yy}} \end{pmatrix} \right) \tag{1}$$

where $\boldsymbol{\mu}_{\mathbf{x}}$, $\boldsymbol{\mu}_{\mathbf{y}}$ and $\boldsymbol{\mu}_{\mathbf{z}}$ are the mean vectors of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ respectively and $\sum_{\mathbf{uv}}$ is the covariance matrix of two random vectors $\mathbf{U}$ and $\mathbf{V}$. We assume that all of these covariance matrices are positive definite. The aim of this article is to derive the exact joint distribution of a linear combination of order statistics $(\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)})$ and a linear combination of their concomitants $(\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]})$ in the presence of a p-dimensional random vector $\mathbf{Z}$, where $\mathbf{a} = (a_1,a_2,\ldots,a_n)^{\mathsf{T}}$ and $\mathbf{b} = (b_1,b_2,\ldots,b_n)^{\mathsf{T}}$ are arbitrary vectors in $\mathbb{R}^n$. We show that the joint distribution of $\left(\mathbf{Z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)},\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\right)^{\mathsf{T}}$ is a mixture of skew-normal distributions. Furthermore, we may explore some regression equations using order statistics, concomitants and covariate variables which generalizes those investigated in Viana and Lee (2006).

Following Arellano Valle and Azzalini (2006), we say that the random vector $\mathbf{X}$ follows a multivariate skew-normal distribution, denoted by $\mathbf{Y} \sim SUN_{d,m}(\boldsymbol{\xi},\ \boldsymbol{\delta},\ \boldsymbol{\Omega},\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda})$, if its density can be written as

$$f_{\mathbf{Y}}(\mathbf{y}) = \phi_d(\mathbf{y};\boldsymbol{\xi},\boldsymbol{\Omega})\,\frac{\Phi_m\left(\boldsymbol{\delta}+\boldsymbol{\Lambda}^{\mathsf{T}}\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})\,;\boldsymbol{\Gamma}-\boldsymbol{\Lambda}^{\mathsf{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\right)}{\Phi_m(\boldsymbol{\delta};\boldsymbol{\Gamma})} \qquad \mathbf{y}\in\mathbb{R}^d \tag{2}$$

where $\boldsymbol{\delta} \in \mathbb{R}^m$, $\boldsymbol{\xi} \in \mathbb{R}^d$, $\boldsymbol{\Gamma} \in \mathbb{R}^{m\times m}$, $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times d}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{d\times d}$ is a positive definite covariance matrix and $\varphi_d(.,\boldsymbol{\xi},\boldsymbol{\Omega})$ is the density function of a d-dimensional normal with mean vector $\boldsymbol{\xi}$ and covariance matrix $\boldsymbol{\Omega}$ and $\Phi_m(.;\ \sum)$ is the multivariate normal cumulative function with covariance matrix $\sum$.

Let

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim N_{m+d}\left(\begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^\top \\ \boldsymbol{\Lambda} & \boldsymbol{\Omega} \end{pmatrix}\right)$$

A d-dimensional random vector $\mathbf{Y}$ is said to have a unified multivariate skew-normal, $\mathbf{Y} \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ say, if

$$\mathbf{Y} \stackrel{d}{=} \mathbf{V} \mid \mathbf{U} > 0. \tag{3}$$

The density of this random vector can be written as (2).

For more information on multivariate skew-normal distributions and their applications we refer the reader to Azzalini and Dalla Valle (1996), González-Farías et al. (2003), etc. An overview of which can be found in the book edited by Genton (2004) and in the review paper by Azzalini (2005).

The remainder of this paper is organized as follows. In Section 2, we state and prove the main theorem of the paper and deduce some useful corollaries in regression analysis. Section 3 contains some numerical applications of our results.

2. Main results

Consider the following partition of $\mathbf{Y}$ and its corresponding parameters

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \ \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \ \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{21}^\top \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \ \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_2 \end{pmatrix} \tag{4}$$

where $\mathbf{Y}_1$ is a vector of dimension $d-1$ and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ have dimensions $(d-1)\times m$ and $1 \times m$ respectively. The two following lemmas are generalizations of those presented in Sheikhi and Jamalizadeh (2011).

Lemma 1 [9]. *If* $Y = \left(Y_1^\top, Y_2\right)^\top \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$, *then*

a) $\mathbf{Y}_1^\top \sim SUN_{(d-1),m}(\boldsymbol{\xi}_1, \boldsymbol{\delta}, \boldsymbol{\Omega}_{11}, \boldsymbol{\Upsilon}, \boldsymbol{\Lambda}_1)$

b) $\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1 \sim SUN_{1,m}(\xi_{2.1}, \boldsymbol{\delta}_{2.1}, \omega_{22.1}, \boldsymbol{\Gamma}_{2.1}, \boldsymbol{\Lambda}_{2.1})$

c) $M_{Y_2|\mathbf{Y}_1=\mathbf{y}_1}(t) = \exp(\xi_{2.1}t + \frac{1}{2}t^2\omega_{22.1})\frac{\Phi_m(\boldsymbol{\delta}_{2.1}+\boldsymbol{\Lambda}_{2.1}t;\boldsymbol{\Gamma}_{21.})}{\Phi_m(\boldsymbol{\delta}_{2.1};\boldsymbol{\Gamma}_{2.1})} \quad t \in \mathbb{R}$

where $\xi_{2.1} = \xi_2 + \boldsymbol{\Omega}_{2.1}\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\xi}_1)$, $\boldsymbol{\delta}_{2.1} = \boldsymbol{\delta} + \boldsymbol{\Lambda}_1^\top\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\xi}_1)$, $\omega_{22.1} = \omega_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}$ $\boldsymbol{\Omega}_{21}^\top$, $\boldsymbol{\Gamma}_{2.1} = \boldsymbol{\Gamma} - \boldsymbol{\Lambda}_1^\top\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Lambda}_1$ *and* $\boldsymbol{\Lambda}_{2.1} = \boldsymbol{\Lambda}_2 - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Lambda}_1$.

Lemma 2 [8]. *If* $\mathbf{Y} = \left(\mathbf{Y}_1^\top, Y_2\right)^\top \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$, *then the conditional expectation of* Y_2 *given* $\boldsymbol{Y}_1 = \boldsymbol{y}_1$ *in (4) is*

$$E(Y_2|\mathbf{Y}_1 = \mathbf{y}_1) = \xi_{2.1} + \frac{G_m(0;\ \boldsymbol{\delta}_{2.1}, \boldsymbol{\Lambda}_{2.1}, \boldsymbol{\Gamma}_{2.1})}{\Phi_m(\boldsymbol{\delta}_{2.1}; \boldsymbol{\Gamma}_{2.1})} \tag{5}$$

where $G_m(\boldsymbol{0};\ \boldsymbol{A}, \boldsymbol{B},\ \boldsymbol{\Sigma}) = \frac{\partial}{\partial t}\Phi_m\left(\boldsymbol{A}t + \boldsymbol{B}; \boldsymbol{\Sigma}\right)|_{t=0}$.

We now define $S(\mathbf{X})$ to be the class of all permutations of components of the random vector $\mathbf{X}$, i.e. $S(\mathbf{X})=\{\mathbf{X}^{(i)} = \mathbf{P}_i\mathbf{X}; i = 1, 2, \ldots, n!\}$ where $\mathbf{P}_i$ is an $n \times n$ permutation matrix. We also define the matrix $\boldsymbol{\Delta}$ to be the difference matrix of dimension $n-1$ by n such that its ith row is $\mathbf{e}_{n,i+1}^\top - \mathbf{e}_{n,i}^\top$, $i = 1, 2, \ldots, n-1$, where $\mathbf{e}_{n,1}, \mathbf{e}_{n,2}, \ldots, \mathbf{e}_{n,n-1}$ are n-dimensional unit basis vectors, i.e. $\boldsymbol{\Delta}X = (X_2 - X_1, X_3 - X_2, \ldots, X_n - X_{n-1})^\top$. Further, let $\mathbf{X}^{(i)}$ be the ith permutation of the random vector $\mathbf{X}$. We have $P\left(\boldsymbol{\Delta}X^{(i)} \geq \mathbf{0}\right) = 1 - \Phi_m(-(\boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\boldsymbol{\Delta}^\top)^{-1/2}\boldsymbol{\Delta}\,\boldsymbol{\mu}_x^{(i)})$ where $\boldsymbol{\mu}_x^{(i)}$ and $\sum_{\mathbf{xx}}^{(i)}$ are the mean vector and covariance matrix of the random vector $\mathbf{X}^{(i)}$, respectively. Hereafter we adopt the notation $G_i(\mathbf{t}, \boldsymbol{\xi}, \sum)$ for $P\left(\boldsymbol{\Delta}X^{(i)} \geq \mathbf{t}\right)$.

Theorem 1 *Under the assumption of model (1) The cdf of random vector* $\left(\mathbf{Z}, \mathbf{a}^\top\mathbf{X}_{(n)}, \mathbf{b}^\top\mathbf{Y}_{[n]}\right)^\top$ *is the mixture*

$$F_{\mathbf{Z},\mathbf{a}^\top\mathbf{X}_{(n)},\mathbf{b}^\top\mathbf{Y}_{[n]}}(\mathbf{z}, x, y) = \textstyle\sum_{i=1}^{n!} F_{SUN}(\mathbf{z}, x, y; \boldsymbol{\xi}_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i)G_i(\mathbf{t}, \boldsymbol{\xi}, \sigma)$$

where $F_{SUN}(\,.\,; \boldsymbol{\xi}_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i)$ *is the cdf of unified multivariate skew-normal with* $\boldsymbol{\xi}_i = \left(\boldsymbol{\mu}_z, \mathbf{a}^\top\boldsymbol{\mu}_{\mathbf{x}}^{(i)}, \mathbf{b}^\top\boldsymbol{\mu}_{\mathbf{y}}^{(i)}\right)^\top$, $\boldsymbol{\delta}_i = \boldsymbol{\Delta}\,\boldsymbol{\mu}_{\mathbf{x}}^{(i)}$, $\boldsymbol{\Gamma}_i = \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\boldsymbol{\Delta}^\top$, $\boldsymbol{\Lambda}_i = \left(\boldsymbol{\Delta}\sum_{\mathbf{xz}}^{(i)}, \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\mathbf{a}, \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(i)}\mathbf{b}\right)^\top$ *and* $\sum_{\mathbf{ux}}^{(i)}$ *is the covariance matrix of random vector* U *and the ith permutation of the random vector* X. *Moreover,* Ω_i *is the covariance matrix of* $\left(\mathbf{Z}, \mathbf{a}^\top\mathbf{X}^{(i)}, \mathbf{b}^\top\mathbf{Y}^{(i)}\right)^\top$.

Proof. We have

$$\begin{aligned}
F_{\mathbf{Z},\mathbf{a}^\top\mathbf{X}_{(2)},\mathbf{b}^\top\mathbf{Y}_{[2]}}(\mathbf{z}, x, y) &= P\left(\mathbf{Z} \leq \mathbf{z}, \mathbf{a}^\top\mathbf{X}_{(n)} \leq x, \mathbf{b}^\top\mathbf{Y}_{[n]} \leq y\right) \\
&= \sum_{i=1}^{n!} P\left(\mathbf{Z} \leq \mathbf{z}, \mathbf{a}^\top\mathbf{X}^{(i)} \leq x, \mathbf{b}^\top\mathbf{Y}^{(i)} \leq y | \boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) P\left(\boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) \\
&= \sum_{i=1}^{n!} P\left(\mathbf{Z} \leq \mathbf{z}, \mathbf{a}^\top\mathbf{X}^{(i)} \leq x, \mathbf{b}^\top\mathbf{Y}^{(i)} \leq y | \boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) G_i(\mathbf{t}, \boldsymbol{\xi}, \Sigma).
\end{aligned}$$

Furthermore,

$$\begin{pmatrix} \boldsymbol{\Delta}\mathbf{X}^{(i)} \\ \mathbf{Z} \\ \mathbf{a}^{\mathsf{T}}\mathbf{X}^{(i)} \\ \mathbf{b}^{\mathsf{T}}\mathbf{Y}^{(i)} \end{pmatrix} \sim N_{n+p+1}\left(\begin{pmatrix} \boldsymbol{\Delta}\,\mu_{\mathbf{x}}^{(i)} \\ \mu_{z} \\ \mathbf{a}^{\mathsf{T}}\mu_{\mathbf{x}}^{(i)} \\ \mathbf{b}^{\mathsf{T}}\mu_{\mathbf{y}}^{(i)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\boldsymbol{\Delta}^{\mathsf{T}} & \boldsymbol{\Delta}\sum_{\mathbf{xz}}^{(i)} & \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\mathbf{a} & \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(i)}\mathbf{b} \\ & \sum_{\mathbf{zz}} & \sum_{\mathbf{zx}}^{(i)}\mathbf{a} & \sum_{\mathbf{zy}}^{(i)}\mathbf{b} \\ & & \mathbf{a}^{\mathsf{T}}\sum_{\mathbf{xx}}^{(i)}\mathbf{a} & \mathbf{a}^{\mathsf{T}}\sum_{\mathbf{xy}}^{(i)}\mathbf{b} \\ & & & \mathbf{b}^{\mathsf{T}}\sum_{\mathbf{yy}}^{(i)}\mathbf{b} \end{pmatrix}\right).$$

Now using (3) we immediately conclude that

$$\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}^{(i)}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}^{(i)}\right)^{\mathsf{T}} |\boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0} \sim SUN_{p+2,n-1}(\boldsymbol{\xi}_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i).$$

This establishes the theorem.

Using the previous theorem and lemma 1, we find the conditional distribution of linear combinations of concomitants given linear combinations of order statistics and covariates.

Corollary 1 Under the assumptions of model (1) the cdf of the random variable $\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}$ condition on **Z=z** and $\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)} = x$ is the mixture

$$F_{\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}|\mathbf{Z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}}(y|\mathbf{z},x) = \sum_{i=1}^{n!} F_{SUN}(y\,|\,\mathbf{z},x;\boldsymbol{\xi}_i,\boldsymbol{\delta}_i,\boldsymbol{\Gamma}_i,\boldsymbol{\Omega}_i,\boldsymbol{\Lambda}_i)G_i(\mathbf{t},\boldsymbol{\xi},\textstyle\sum)$$

where $F_{SUN}(y\,|\,\mathbf{z},x;\boldsymbol{\xi}_i,\boldsymbol{\delta}_i,\boldsymbol{\Gamma}_i,\boldsymbol{\Omega}_i,\boldsymbol{\Lambda}_i)$ is the cdf of conditional unified skew-normal of $\mathbf{b}^{\mathsf{T}}\mathbf{Y}^{(i)} = y$ given **Z=z** and $\mathbf{a}^{\mathsf{T}}\mathbf{X}^{(i)} = x$ and the parameters are as in theorem 1.

The following corollary is obvious via lemma 2.

Corollary 2 Under the assumptions of model (1) the regression equation of $\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}$ on **Z** and $\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}$ is

$$E(\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\,|\,\mathbf{Z=z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)} = x) = \sum_{i=1}^{n!} \xi_{2.1}^{(i)} + \frac{G_{n-1}^{(i)}(0;\boldsymbol{\delta}_{2.1},\boldsymbol{\Lambda}_{2.1},\boldsymbol{\Gamma}_{2.1})}{\Phi_{n-1}(\boldsymbol{\delta}_{2.1};\boldsymbol{\Gamma}_{2.1})}$$

where the superscript (i) denotes the parameters based on the ith permutation of **X**.

In the remainder of this section, we shall focus on a special case of the multivariate normal distribution. Let the joint distribution of a p-dimensional random vector **Z**, and two random vectors **X** and **Y** follow a $2n+p$ dimensional exchangeable multivariate normal random distribution, i.e. its covariance matrix is equicorrelated. Hence we have

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n+p}\left(\mu = \begin{pmatrix} \mu_{\mathbf{z}} \\ \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{pmatrix}, \sum = \begin{pmatrix} \sum_{\mathbf{zz}} & \sum_{\mathbf{xz}}^{\mathsf{T}} & \sum_{\mathbf{yz}}^{\mathsf{T}} \\ \sum_{\mathbf{xz}} & \sum_{\mathbf{x}} & \sum_{\mathbf{xy}}^{\mathsf{T}} \\ \sum_{\mathbf{yz}} & \sum_{\mathbf{xy}} & \sum_{\mathbf{y}} \end{pmatrix}\right) \tag{6}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}} = \mu_x \mathbf{1}_n,\ \boldsymbol{\mu}_{\mathbf{y}} = \mu_y \mathbf{1}_n, \Sigma_{\mathbf{xx}} = \sigma_x^2 \left[\rho_x \mathbf{1}_n \mathbf{1}_n^{\mathsf{T}} + (1-\rho_x)\,\mathbf{I}_n\right],$$

$$\Sigma_{\mathbf{yy}} = \sigma_y^2 \left[\rho_y \mathbf{1}_n \mathbf{1}_n^{\mathsf{T}} + (1-\rho_y)\,\mathbf{I}_n\right], \Sigma_{\mathbf{xy}} = \rho_{xy}\sigma_x\sigma_y \mathbf{J},$$

where $\mathbf{1}_n\text{=}(1,\ldots,1)^{\mathsf{T}}$, $\mathbf{I}_n = diag(1,\ldots,1)$ and $\mathbf{J} = [1]_{n\times n}$. This model is the generalization of that assumed in Viana and Lee (2006).

Sheikhi and Jamalizadeh (2011) found the joint distribution of two linear combinations of order statistics in the presence of a covariate random variable under the exchangeable assumption and presented some regression analyses.

Theorem 2 Under the assumption of model (5),

$$\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\right)^{\mathsf{T}} \sim SUN_{p+2,n-1}\left(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}\right)$$

where $\boldsymbol{\xi} = \left(\boldsymbol{\mu}_z, \mathbf{a}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{y}}\right)^{\mathsf{T}}, \boldsymbol{\delta} = \mathbf{0}$, $\boldsymbol{\Gamma} = \boldsymbol{\Delta}\sum_{\mathbf{x}}\boldsymbol{\Delta}^{\mathsf{T}}, \boldsymbol{\Lambda} = \left(\boldsymbol{\Delta}\sum_{\mathbf{xz}}, \boldsymbol{\Delta}\sum_{\mathbf{xx}} a, \Delta\sum_{\mathbf{yy}} b\right)^{\mathsf{T}}$

and $\boldsymbol{\Omega}$ is the covariance matrix of $\left(\mathbf{Z},\ \mathbf{a}^{\mathsf{T}}\mathbf{X},\ \mathbf{b}^{\mathsf{T}}\mathbf{Y}\right)^{\mathsf{T}}$.

Proof. Since $P\left(\boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) = \frac{1}{n!}$, $i = 1,\ldots,n!$, by exchangeability we have

$$F_{\mathbf{Z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(2)},\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[2]}}\left(\mathbf{t}_1, t_2, t_3\right) = P\left(\mathbf{Z} \leq \mathbf{t}_1, \mathbf{a}^{\mathsf{T}}\mathbf{X} \leq t_2, \mathbf{b}^{\mathsf{T}}\mathbf{Y} \leq t_3 | \boldsymbol{\Delta}\mathbf{X} > 0\right).$$

Moreover,

$$\begin{pmatrix} \boldsymbol{\Delta}\mathbf{X} \\ \mathbf{Z} \\ \mathbf{a}^{\mathsf{T}}\mathbf{X} \\ \mathbf{b}^{\mathsf{T}}\mathbf{Y} \end{pmatrix} \sim N_{n+p+1}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu}_z \\ \mathbf{a}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}} \\ \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{y}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta}\sum_{\mathbf{xx}}\boldsymbol{\Delta}^{\mathsf{T}} & \boldsymbol{\Delta}\sum_{\mathbf{xz}} & \boldsymbol{\Delta}\sum_{\mathbf{xx}}\mathbf{a} & \boldsymbol{\Delta}\sum_{\mathbf{xy}}\mathbf{b} \\ & \sum_{\mathbf{zz}} & \sum_{\mathbf{xz}}^{\mathsf{T}}\mathbf{a} & \sum_{\mathbf{yz}}^{\mathsf{T}}\mathbf{b} \\ & & \mathbf{a}^{\mathsf{T}}\sum_{\mathbf{xx}}\mathbf{a} & \mathbf{a}^{\mathsf{T}}\sum_{\mathbf{xy}}\mathbf{b} \\ & & & \mathbf{b}^{\mathsf{T}}\sum_{\mathbf{yy}}\mathbf{b} \end{pmatrix}\right).$$

So $\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}\right)^{\mathsf{T}} | \left(\boldsymbol{\Delta}\mathbf{X} > 0\right) \sim SUN_{p+2,n-1}\left(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}\right)$, where the parameters are as given in theorem 2.

We may also interested in predicting the concomitants using order statistics and some covariates. The following corollary provides such a regression equation.

Corollary 3 Under the assumptions of model (1) the regression equation of $\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}$ on $\mathbf{Z}$ and $\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}$ is

$$E(\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\,|\,\mathbf{Z{=}z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)} = x) = \xi_{2.1} + \frac{G_{n-1}(0;\, A, \boldsymbol{\Omega})}{\Phi_{n-1}\,(\boldsymbol{\delta}; \boldsymbol{\Gamma})}$$

where

$$\xi_{2.1} = \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{y}} + \frac{1}{\mathbf{b}^{\mathsf{T}}\sum_{\mathbf{yy}}\mathbf{b}}\left[\mathbf{b}^{\mathsf{T}}\sum\nolimits_{\mathbf{yz}}(\mathbf{Z}-\mathbf{z}) + \mathbf{b}^{\mathsf{T}}\sum\nolimits_{\mathbf{xy}}\mathbf{a}(\mathbf{a}^{\mathsf{T}}\mathbf{X} - \mathbf{a}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}})\right]$$

and $G_{n-1}(0;\, A, \boldsymbol{\Omega})$ was defined in lemma 2.

The regression equation of $Y_{[i]}$ on $\mathbf{Z}$ and $X_{(i)}$ may be determined by letting the i-th component of the random vectors $\mathbf{a}$ and $\mathbf{b}$ equal to 1 and other all components equal to zero. Specifically, this regression equation and the regression equation of $Y_{[i]}$ on $\mathbf{Z}$ and X coincide and expressed as

$$Y_{[i]} = \mu_y + \frac{1}{\rho_y\sigma_y^2}\left[\boldsymbol{\sigma}_{y\mathbf{z}}(\mathbf{Z}-\mathbf{z}) + \rho_{xy}\sigma_x\sigma_y(X-\mu_x)\right] + \frac{G_{n-1}(0;\, A, \boldsymbol{\Omega})}{\Phi_{n-1}\,(\boldsymbol{\delta}; \boldsymbol{\Gamma})} \tag{7}$$

where $\boldsymbol{\sigma}_{y\mathbf{z}} = \mathrm{cov}(\mathbf{Z}, Y)$.

Olkin and Viana (1995) discussed that the linear regression of a random variable Z on $\mathbf{X}$ and Z on $\mathbf{X}_{(n)}$ coincide.

3. Numerical results

Viana and Lee (2006) considered the data from a pilot study in which a number of physiological parameters were measured at a specific site on the left and right brain hemispheres of subjects participating in a study conducted at the sleep centre of the University of Illinois at Chicago. They considered the following variables, jointly observed in a sample of $N = 30$ subjects:

(Z) : age;
(X_s) : tissue oxygenation on the right site;
(X_d) : tissue oxygenation on the left site;
(Y_s) : total hemoglobin on the right site;
(Y_d) : total hemoglobin on the left site.

They explored their results by assuming $\mathbf{X}^{\mathsf{T}} = (X_s, X_d)$ and $\mathbf{Y}^{\mathsf{T}} = (Y_s, Y_d)$. In particular, they estimated the covariance matrix of the random vector $(Z, \mathbf{X}_{(2)}^{\mathsf{T}}, \mathbf{Y}_{[2]}^{\mathsf{T}})^{\mathsf{T}}$. In this section we use these data to estimate the distributions obtained in the previous section. We assume that $(Z, \mathbf{X}^{\mathsf{T}}, \mathbf{Y}^{\mathsf{T}})^{\mathsf{T}}$ follows a 5 dimensional exchangeable multivariate normal distribution. The MLE of parameters are as follows:

$$\hat{\mu} = \begin{pmatrix} 50.000 \\ 57.275 \\ 57.275 \\ 39.695 \\ 39.695 \end{pmatrix}, \hat{\Sigma} = \begin{pmatrix} 128.410 & -28.110 & -28.110 & -35.996 & -35.996 \\ -28.110 & 40.016 & 26.090 & 42.909 & 42.909 \\ -28.110 & 26.090 & 40.016 & 42.909 & 42.909 \\ -35.996 & 42.909 & 42.909 & 159.58 & 113.06 \\ -35.996 & 42.909 & 42.909 & 113.06 & 159.58 \end{pmatrix}$$

Hence, $\hat{\rho}_x$=0.651**,** $\hat{\rho}_y$=0.722, $\hat{\rho}_{xz} = 0.392$**,** $\hat{\rho}_{yz} = 0.251$**,** $\hat{\rho}_{xy} = 0.536$. By considering **a = b = (**0, 1)$^{\mathsf{T}}$, theorem 2 implies that $(Z, X_{(2)}, Y_{[2]})^{\mathsf{T}} \sim SUN_{3,1}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$, where

$$\xi = \begin{pmatrix} 50.000 \\ 57.275 \\ 39.695 \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} 128.410 & -28.110 & -35.996 \\ -28.110 & 40.016 & 42.909 \\ -35.996 & 42.909 & 159.58 \end{pmatrix}, \Lambda = \begin{pmatrix} 0 \\ 13.965 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\delta} = 0, \boldsymbol{\Gamma} = 27.931.$$

Furthermore, the estimated regression equation of $Y_{[2]}$ on $X_{(2)}$ and Z follows from (6). We readily obtain

$$Y_{[2]} = -16.685 + 0.054Z + 1.034X_{(2)} \quad .$$

Also, the linear regression of $Y_{[2]}$ on Z is easily estimated as $Y_{[2]} = 59.909 - 0.287Z$, then as obtained by Viana and Lee (2006). In addition, the regression of $Y_{[2]}$ on $X_{(2)}$ can be expressed as $Y_{[2]} = 24.294 + 0.269X_{(2)}$.

Similarly, we may obtain the joint distribution of $(Z, X_{(1)}, Y_{[1]})$ as well as the regression equation of $Y_{[1]}$ on $X_{(1)}$ and Z by letting **a = b = (**1, 0**)**$^{\mathsf{T}}$.

4. Conclusion

In this work we find the joint distribution of a linear combination of order statistics and a linear combination of their concomitants in the presence of some covariate random variables as a member of the skew-normal distribution. Some useful special cases of this distribution are investigated as well as some conditional distributions.

The application of our results in density estimation and regression analysis is illustrated by a numerical data set. We hope to extend our results to elliptical distributions in the future.

References

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

Arellano-Valle, R. B. and Genton, M. G. (2007). On the exact distribution of linear combinations of order statistics from dependent random variables. *Journal of Multivariate Analysis*, 98, 1876–1894.

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

David, H. A. and Nagaraja, H.N. (2003). *Order Statistics*, third ed. John Wiley and Sons, New York.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Edited vol., Chapman and Hall/CRC, Boca Raton, FL, 416 pp.

Goel, P. K. and Hall, P. (1994). On the average difference between concomitants and order statistics. *Annals of Probability*, 22, 126–144.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A. K. (2004b). Additive properties of skew-normal random vectors. *Journal of Statistical Planning and Inference*, 126, 521–534.

Gupta, A. K., Gonzlez-Faras, G. and Domnguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89, 181–190.

Loperfido, N. (2008). Modeling maxima of longitudinal contralateral observations. *Test* 17, 370–380.

Olkin, I. and Viana, M. (1995). Correlation analysis of extreme observations from a multivariate normal distribution. *Journal of the American Statistical Association*, 90, 1373–1379.

Sheikhi, A. and Jamalizadeh, A. (2011). Regression analysis using order statistics. *Statistical Papers*, 52, 885–892.

Tsukibayashi, S. (1998). The joint distribution and moments of an extreme of the dependent variable and the concomitant of an extreme of the independent variable. *Communications in Statistics-Theory and Methods*, 27(7), 639–651.

Viana, M. A. G. and Lee, H. M. (2006). Correlation analysis of ordered symmetrically dependent observations and their concomitants of order statistics. *Scandinavian Journal of Statistics*, 34(2), 327–340.

Yang, S. (1981). Linear functions of concomitant of order statistics with application to nonparametric estimation of a regression function. *Journal of the American Statistical Association*, 76, 658–662.