Chilean Journal of Statistics
Vol. 4, No. 1, April 2013, 69–85

Distribution Theory
Research Paper

Distributional results for L-statistics based on random vectors with multivariate elliptical distributions

Alexander R. de Leon[1], Hamed Mahmoodian[2], and Ahad Jamalizadeh[3,*]

[1]Department of Mathematics and Statistics, University of Calgary, Calgary, Canada
[2]Department of Statistics, Yazd University, Yazd, Iran
[3]Department of Statistics, Shahid Bahonar University, Kerman, Iran

(Received: 13 September 2012 · Accepted in final form: 03 December 2012)

Abstract

We consider random vectors $\mathbf{X}_{K\times 1}$ and $\mathbf{Y}_{N\times 1}$ having a multivariate elliptical joint distribution, and derive the exact joint distribution of $\mathbf{X}$ and L-statistics from $\mathbf{Y}$, as a mixture of multivariate unified skew-elliptical distributions. This mixture representation enables us to predict $\mathbf{X}$ based on L-statistics from $\mathbf{Y}$, and vice versa, when $K = 1$ and with normal and t-distributions. Our results extend and generalize previous ones in two ways: first, we consider a general multivariate set-up for which $K \geq 1$ and $N \geq 2$, and second, we adopt the multivariate elliptical distribution to include previous multivariate normal and t-formulations as special cases.

Keywords: Linear combination · Mixture distribution · Multivariate unified skew-elliptical distribution · Order statistics · Squared-error loss.

Mathematics Subject Classification: Primary 62E15 · Secondary 62H05.

1. Introduction

A motivation for the paper is the following problem: it is usual practice to use the quiz scores to predict a student's final test mark. More formally, if X is the final test mark and $\mathbf{Y} = (Y_1, \ldots, Y_N)^\top$ are the quiz scores, we wish to study X in terms of some linear combination $\sum_{i=1}^N a_i Y_{(i)}$, where $Y_{(1)} < \cdots < Y_{(N)}$ are the ordered quiz scores, with varying weights a_i. The results in this paper enable us to study X based on $\sum_{i=1}^N a_i Y_{(i)}$, when $(X, \mathbf{Y}^\top)^\top$ follows a $(N + 1)$-dimensional multivariate elliptical distribution.

The above problem is akin to the following that arises in electrical engineering and discussed by Wiens et al. (2006). In processing cellular phone signals from several antennae, receivers normally select only the strongest signals to reduce signal fading. Specifically, if a receiver receives N signals, only the $n \leq N$ strongest signals will be processed, which are then combined and used to analyze and predict the transmission system's performance, as measured by $\mathbf{X} = (X_1, \ldots, X_K)^\top$, say.

Various incarnations of the above problem, simplified in some way, have been studied previously by several authors. Viana (1998) and Olkin and Viana (1995) obtained the best

*Corresponding author. Email: A.Jamalizadeh@uk.ac.ir

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

linear predictors in the trivariate normal distribution case with $N = 2$ and $K = 1$, where Y_1 and Y_2 are exchangeable such that $(X, Y_1)^\top$ and $(X, Y_2)^\top$ share a common correlation. Loperfido (2008b) considered the same set-up and derived the exact joint distribution of X and $Y_{(2)} = \max(Y_1, Y_2)$. Jamalizadeh and Balakrishnan (2009a) similarly derived the exact joint distribution of X and $\sum_{i=1}^2 a_i Y_{(i)}$ in the case of a trivariate normal distribution for X, Y_1, Y_2, with arbitrary covariance structure. They showed that this joint distribution is a mixture of bivariate unified skew-normal distributions and obtained a predictor for X using linear combinations of order statistics from Y_1, Y_2; see also Balakrishnan et al. (2012) for the case of elliptical distributions.

Order statistics, their linear combinations (i.e., L-statistics), and their corresponding distributions have also been similarly widely studied. Early results are provided by Gupta and Pillai (1965), Basu and Ghosh (1978), Nagaraja (1982), and Balakrishnan (1993). More recent work include Genc (2006), who derived the exact distribution of L-statistics from the bivariate normal distribution; Arellano-Valle and Genton (2007, 2008), who considered multivariate elliptical distributions; and Jamalizadeh and Balakrishnan (2008), who worked with bivariate skew-normal and skew-t distributions. Additional results are given by Jamalizadeh et al. (2009a,b), Jamalizadeh and Balakrishnan (2009b, 2010), and Loperfido (2008a).

In this paper, we consider the general case of $N > 2$ and $K > 1$, and assume an elliptical joint distribution for $\mathbf{X} = (X_1, \ldots, X_K)^\top$ and $\mathbf{Y} = (Y_1, \ldots, Y_N)^\top$, i.e., let

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim \mathrm{EC}_{K+N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h^{(K+N)}\right), \tag{1}$$

the $(K + N)$-dimensional elliptical distribution with density generator function $h^{(K+N)}$, and respective location parameter and dispersion-shape matrix

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{yx}^\top \\ & \boldsymbol{\Sigma}_{yy} \end{pmatrix}, \tag{2}$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are respective $K \times 1$ and $N \times 1$ location vectors of $\mathbf{X}$ and $\mathbf{Y}$, $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{yy}$ are their respective $K \times K$ and $N \times N$ dispersion matrices, and $\boldsymbol{\Sigma}_{yx}$ is a $N \times K$ shape matrix. Note that this specification includes the multivariate normal and t-distributions, among others, and generalizes previous cases studied in the literature. With $\mathbf{Y}_{(N)} = (Y_{(1)}, \ldots, Y_{(N)})^\top$ as the vector of order statistics from $\mathbf{Y}$, we derive the exact joint distribution of $\mathbf{X}$ and $\mathbf{LY}_{(N)}$, where $\mathbf{L}$ is a $P \times N$ matrix of $\mathrm{rank}(\mathbf{L}) = P$. We show that this joint distribution is a mixture of multivariate unified skew-elliptical distributions, and obtain in the process, in the special case when $K = 1$ and $\mathbf{L} = \mathbf{a}^\top = (a_1, \ldots, a_N)$, the best (nonlinear) predictors of X based on $\mathbf{a}^\top \mathbf{Y}_{(N)}$, and of $\mathbf{a}^\top \mathbf{Y}_{(N)}$ based on X, under square loss function in the case of normal and t-distributions. We also present a mixture representation for the joint cumulative distribution function (CDF) of X and $Y_{(r)}$, $r = 1, \ldots, N$, in terms of bivariate unified skew-elliptical distributions.

We organize the paper as follows. Section 2 presents a brief review of skew-elliptical distribution theory and presents specialized results for normal and t-distributions in the univariate and bivariate cases. The main results of the paper are then obtained in Section 3. Section 4 then concludes the paper.

2. Skew-Elliptical Distributions: Preliminaries

Consider the random vectors $\mathbf{X}_{K\times 1}$ and $\mathbf{Y}_{N\times 1}$ in Equation (1) above. A random vector $\mathbf{U}_{N\times 1}$ is defined to have a multivariate unified skew-elliptical (SUE) distribution if and only if

$$\mathbf{U} \stackrel{d}{=} (\mathbf{Y}|\mathbf{X} > \mathbf{0}), \tag{3}$$

where "$\stackrel{d}{=}$" denotes equality in distribution, and it is understood that the inequality "$\mathbf{X} > \mathbf{0}$" must hold for each of the components of $\mathbf{X}$; we write $\mathbf{U} \sim \mathrm{SUE}_{N,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx}, h^{(K+N)})$. A closed-form expression for the corresponding probability density function (PDF) is given by Arellano-Valle and Azzalini (2006) and Arellano-Valle and Genton (2005, 2010a); see also Branco and Dey's (2001) formulation. Note that our construction of skew-elliptical distributions in Equation (3) relies on dispersion matrices which may generate an identifiability problem; see for example, Arellano-Valle and Azzalini (2006) and Arellano-Valle and Genton (2010b), for the construction of skew-normal and skew-t distributions. In such cases, the dispersion matrices should be replaced with correlation matrices. Taking $h^{(K+N)}(u) = \phi^{(K+N)}(u) = (2\pi)^{-(K+N)/2}\exp(-u/2)$, $u > 0$, we obtain the multivariate unified skew-normal distribution (SUN) (Arellano-Valle and Genton, 2010b). Given $\mathbf{U} \sim \mathrm{SUN}_{N,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx})$, its respective PDF $f_{\mathbf{U}}(\cdot)$ and moment-generating function (MGF) $M_{\mathbf{U}}(\cdot)$ are given by

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{u}) &= \frac{\phi_N(\mathbf{u}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})\Phi_K(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{u} - \boldsymbol{\mu}_y); \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})}{\Phi_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx})}, \\ M_{\mathbf{U}}(\mathbf{s}) &= \frac{\exp\left(\boldsymbol{\mu}_y^\top \mathbf{s} + \frac{1}{2}\mathbf{s}^\top \boldsymbol{\Sigma}_{yy}\mathbf{s}\right)\Phi_K(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx}^\top \mathbf{s}; \boldsymbol{\Sigma}_{xx})}{\Phi_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx})}, \end{aligned} \tag{4}$$

where $\phi_Q(\cdot; \boldsymbol{\Sigma})$ and $\Phi_Q(\cdot; \boldsymbol{\Sigma})$ are the PDF and CDF of the centered Q-dimensional normal distribution with dispersion matrix $\boldsymbol{\Sigma}$, respectively. In the case of the t_ν-kernel distribution with generator

$$h^{(K+N)}(u) \;=\; t_\nu^{(K+N)}(u) \;=\; \frac{\Gamma\left(\frac{\nu+K+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{K+N}{2}}}\left(1 + \frac{u}{\nu}\right)^{-(\nu+K+N)/2},$$

for $u \geq 0$, $\nu > 0$, where $\Gamma(\cdot)$ is the gamma function, we generate the multivariate unified skew-t (SUT) distribution, with PDF

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{u}) = T_K &\left(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{u} - \boldsymbol{\mu}_y); \tfrac{\nu + (\mathbf{u}-\boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{u}-\boldsymbol{\mu}_y)}{\nu+N}(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{yx}^\top \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}), \nu + N\right) \\ &\times \frac{t_N(\mathbf{u}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}, \nu)}{T_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx}, \nu)}, \end{aligned}$$

where $t_Q(\cdot; \boldsymbol{\Sigma}, \nu)$ and $T_Q(\cdot; \boldsymbol{\Sigma}, \nu)$ are the respective PDF and CDF of the centered Q-dimensional t-distribution with scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν, which we write as $\mathbf{U} \sim \mathrm{SUT}_{N,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx}, \nu)$. These distributions were developed recently in Jamalizadeh and Balakrishnan (2012), who obtained marginal and conditional distributions of SUE distributions. In what follows, we study univariate and bivariate SUN and SUT distributions and derive moment expressions, among others, for later use in Section 3.

2.1 Univariate case

We now consider the univariate class of SUE distributions that arises from Equation (1) with $K \geq 1$ and $N = 1$. In this case, Equation (2) becomes

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\sigma}_{xy} \\ & \sigma_{yy} \end{pmatrix}.$$

For $U \sim \mathrm{SUN}_{1,K}(\mu_y, \boldsymbol{\mu}_x, \sigma_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{xy})$, we get

$$\begin{aligned} \mathrm{E}[U] = \mu_y + \frac{1}{\Phi_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx})} \sum_{i=1}^{K} \frac{\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}} \phi\left(\frac{\mu_{x,i}}{\sqrt{\sigma_{xx,ii}}}\right) \\ \times \Phi_{K-1}\left(\boldsymbol{\mu}_{x,-i} - \frac{\mu_{x,i}}{\sigma_{xx,ii}} \boldsymbol{\sigma}_{xx,-ii}; \boldsymbol{\Sigma}_{xx,-i|i}\right), \end{aligned} \tag{5}$$

where, for some i,

$$\boldsymbol{\sigma}_{xy} = \begin{pmatrix} \sigma_{xy,i} \\ \boldsymbol{\sigma}_{xy,-i} \end{pmatrix}, \quad \boldsymbol{\mu}_x = \begin{pmatrix} \mu_{x,i} \\ \boldsymbol{\mu}_{x,-i} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{xx} = \begin{pmatrix} \sigma_{xx,ii} & \boldsymbol{\sigma}_{xx,-ii}^\top \\ & \boldsymbol{\Sigma}_{xx,-i-i} \end{pmatrix},$$

with $\boldsymbol{\Sigma}_{xx,-i|i} = \boldsymbol{\Sigma}_{xx,-i-i} - \boldsymbol{\sigma}_{xx,-ii}\boldsymbol{\sigma}_{xx,-ii}^\top / \sigma_{xx,ii}$, and with $\phi(\cdot)$ the standard normal PDF. Equation (5) is easily obtained by differentiating Equation (4) and using the following:

$$\begin{aligned} \frac{\partial}{\partial s} \Phi_K(\boldsymbol{\mu}_x + s\boldsymbol{\sigma}_{xy}; \boldsymbol{\Sigma}_{xx}) = \sum_{i=1}^{K} \frac{\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}} \phi\left(\frac{\mu_{x,i} + s\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}}\right) \Phi_{K-1}\left(\left\{\boldsymbol{\sigma}_{xy,-i} - \frac{\sigma_{xy,i}}{\sigma_{xx,ii}} \boldsymbol{\sigma}_{xx,-ii}\right\} s \right. \\ \left. + \boldsymbol{\mu}_{x,-i} - \frac{\mu_{x,i}}{\sigma_{xx,ii}} \boldsymbol{\sigma}_{xx,-ii}; \boldsymbol{\Sigma}_{xx,-i|i}\right). \end{aligned}$$

Next, if $U \sim \mathrm{SUT}_{1,K}(\mu_y, \boldsymbol{\mu}_x, \sigma_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{xy}, \nu)$, we similarly obtain

$$\begin{aligned} \mathrm{E}[U] = \mu_y + \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right) T_K(\boldsymbol{\mu}_x; \boldsymbol{\Sigma}_{xx}, \nu)} \sum_{i=1}^{K} \frac{\sigma_{xy,i}}{\sqrt{\sigma_{xx,ii}}} \left(\nu + \frac{\mu_{x,i}^2}{\sigma_{xx,ii}}\right)^{-(\nu-1)/2} \\ \times T_{K-1}\left(\frac{\sqrt{\nu-1}}{\sqrt{\nu + \frac{\mu_{x,i}^2}{\sigma_{xx,ii}}}} \left(\boldsymbol{\mu}_{x,-i} - \frac{\mu_{x,i}}{\sigma_{xx,ii}} \boldsymbol{\sigma}_{xx,-ii}\right); \boldsymbol{\Sigma}_{xx,-i|i}, \nu - 1\right). \end{aligned}$$

This follows from Equation (5) and from the result that for a χ^2_ν random variable νV with ν degrees of freedom,

$$\begin{aligned} \mathrm{E}\left[V^{-1/2}\phi(aV^{1/2})\Phi_Q(V^{1/2}\mathbf{b}; \boldsymbol{\Sigma})\right] = \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} (\nu + a^2)^{-(\nu-1)/2} \\ \times T_Q\left(\frac{\sqrt{\nu-1}}{\sqrt{\nu + a^2}} \mathbf{b}; \boldsymbol{\Sigma}, \nu - 1\right), \end{aligned}$$

for any real number a, any $Q \times 1$ real vector $\mathbf{b}$, and any positive definite $Q \times Q$ matrix $\boldsymbol{\Sigma}$.

2.2 Bivariate case

Next, we consider the case $N = 2$ and $K \geq 1$. Let

$$\boldsymbol{\Sigma}_{yy} = \begin{pmatrix} \sigma_{yy,11} & \sigma_{yy,12} \\ & \sigma_{yy,22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{yx} = \begin{pmatrix} \boldsymbol{\sigma}_{yx1}^\top \\ \boldsymbol{\sigma}_{yx2}^\top \end{pmatrix}.$$

For $\mathbf{U} = (U_1, U_2)^\top \sim \mathrm{SUN}_{2,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx})$, it follows from Arellano-Valle and Genton (2010a) that $U_1 \sim \mathrm{SUN}_{1,K}(\mu_{y,1}, \boldsymbol{\mu}_x, \sigma_{yy,11}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{yx1})$. In addition, we get

$$U_2|U_1 = u_1 \sim \mathrm{SUN}_{1,K}(\mu_y^{2\cdot 1}(u_1), \boldsymbol{\mu}_x^{2\cdot 1}(u_1), \sigma_{yy}^{22\cdot 1}, \boldsymbol{\Sigma}_{xx}^{2\cdot 1}, \boldsymbol{\sigma}_{yx}^{2\cdot 1}),$$

where $\mu_y^{2\cdot 1}(u_1) = \mu_{y,2} + \sigma_{yy,12}(u_1 - \mu_{y,1})/\sigma_{yy,11}$, $\boldsymbol{\mu}_x^{2\cdot 1}(u_1) = \boldsymbol{\mu}_x + \boldsymbol{\sigma}_{yx1}(u_1 - \mu_{y,1})/\sigma_{yy,11}$, $\sigma_{yy}^{22\cdot 1} = \sigma_{yy,22} - \sigma_{yy,12}^2/\sigma_{yy,11}$, $\boldsymbol{\Sigma}_{xx}^{2\cdot 1} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\sigma}_{yx1}\boldsymbol{\sigma}_{yx1}^\top/\sigma_{yy,11}$, and $\boldsymbol{\sigma}_{yx}^{2\cdot 1} = \boldsymbol{\sigma}_{yx2} - \sigma_{yy,12}\boldsymbol{\sigma}_{yx1}/\sigma_{yy,11}$. It can also be shown that

$$\mathrm{E}[U_2|U_1 = u_1] = \mu_y^{2\cdot 1}(u_1) + \frac{1}{\Phi_K(\boldsymbol{\mu}_x^{2\cdot 1}(u_1); \boldsymbol{\Sigma}_{xx}^{2\cdot 1})} \sum_{i=1}^{K} \frac{\sigma_{yx,i}^{2\cdot 1}}{\sqrt{\sigma_{xx,ii}^{2\cdot 1}}} \phi\left(\frac{\mu_{x,i}^{2\cdot 1}(u_1)}{\sqrt{\sigma_{xx,ii}^{2\cdot 1}}}\right)$$
$$\times \Phi_{K-1}\left(\boldsymbol{\mu}_{x,-1}^{2\cdot 1}(u_1) - \frac{\mu_{x,i}^{2\cdot 1}(u_1)}{\sigma_{xx,ii}^{2\cdot 1}}\boldsymbol{\sigma}_{xx,-ii}^{2\cdot 1}; \boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot 1}\right), \tag{6}$$

where $\sigma_{xx,ii}^{2\cdot 1} = \sigma_{xx,ii} - \sigma_{yx1,i}/\sigma_{yy,11}$, $\boldsymbol{\mu}_{x,-i}^{2\cdot 1}(u_1) = \boldsymbol{\mu}_{x,-i} + \boldsymbol{\sigma}_{yx1,-i}(u_1 - \mu_{y,1})/\sigma_{yy,11}$, $\boldsymbol{\sigma}_{xx,-ii}^{2\cdot 1} = \boldsymbol{\sigma}_{xx,-ii} - \sigma_{yx1,i}\boldsymbol{\sigma}_{yx1,-i}/\sigma_{yy,11}$,

$$\boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot 1} = \boldsymbol{\Sigma}_{xx,-i-i} - \frac{\boldsymbol{\sigma}_{yx1,-i}\boldsymbol{\sigma}_{yx1,-i}^\top}{\sigma_{yx,11}} - \frac{\left(\boldsymbol{\sigma}_{xx,-ii} - \frac{\sigma_{yx1,i}\boldsymbol{\sigma}_{yx1,-i}}{\sigma_{yy,11}}\right)\left(\boldsymbol{\sigma}_{xx,-ii} - \frac{\sigma_{yx1,i}\boldsymbol{\sigma}_{yx1,-i}}{\sigma_{yy,11}}\right)^\top}{\sigma_{xx,ii} - \frac{\sigma_{yx1,i}}{\sigma_{yy,11}}},$$

and where $\boldsymbol{\sigma}_{yx1}$, $\boldsymbol{\sigma}_{yx2}$, $\boldsymbol{\mu}_x$, and $\boldsymbol{\Sigma}_{xx}$ are similarly partitioned as in Section 2.1. Analogous results may be similarly obtained for $\mathbf{U} = (U_1, U_2)^\top \sim \mathrm{SUT}_{2,K}(\boldsymbol{\mu}_y, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yx}, \nu)$:

$$U_1 \sim \mathrm{SUT}_{1,K}(\mu_{y,1}, \boldsymbol{\mu}_x, \sigma_{yy,11}, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\sigma}_{yx1}, \nu), \tag{7}$$

$$U_2|U_1 = u_1 \sim \mathrm{SUT}_{1,K}\left(\mu_y^{2\cdot 1}(u_1), \boldsymbol{\mu}_x^{2\cdot 1}(u_1), q_1(u_1,\nu)\sigma_{yy}^{22\cdot 1}, q_1(u_1,\nu)\boldsymbol{\Sigma}_{xx}^{2\cdot 1}, q_1(u_1,\nu)\boldsymbol{\sigma}_{yx}^{2\cdot 1}, \nu+1\right),$$

$$\mathrm{E}[U_2|U_1 = u_1] = \mu_y^{2\cdot 1}(u_1) + \frac{(\nu+1)^{(\nu+1)/2} q_1^{1/2}(u_1,\nu)\Gamma\left(\frac{\nu}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right) T_K(\boldsymbol{\mu}_x^{2\cdot 1}(u_1); q_1(u_1,\nu)\boldsymbol{\Sigma}_x^{2\cdot 1}, \nu+1)} \sum_{i=1}^{K} \frac{\sigma_{yx,i}^{2\cdot 1}}{\sqrt{\sigma_{xx,ii}^{2\cdot 1}}}$$
$$\times T_{K-1}\left(\frac{\sqrt{\nu}\left(\boldsymbol{\sigma}_{x,-i}^{2\cdot 1}(u_1) - \frac{\sigma_{xi}^{2\cdot 1}(u_1)}{\sigma_{xx,ii}^{2\cdot 1}}\boldsymbol{\sigma}_{xx,-ii}^{2\cdot 1}\right)}{\sqrt{\nu+1+\frac{\{\mu_{x,i}^{2\cdot 1}(u_1)\}^2}{q_1(u_1,\nu)\sigma_{xx,ii}^{2\cdot 1}}}}; q_1(u_1,\nu)\boldsymbol{\Sigma}_{xx,-i|i}^{2\cdot 1}, \nu\right)$$
$$\times \left(\nu+1+\frac{\left(\mu_{x,i}^{2\cdot 1}(u_1)\right)^2}{q_1(u_1,\nu)\sigma_{xx,ii}^{2\cdot 1}}\right)^{-\nu/2}, \tag{8}$$

where $\sigma^{2\cdot 1}_{yx,i}$, $\sigma^{2\cdot 1}_{xx,ii}$, $\mu^{2\cdot 1}_{x,i}(u_1)$, $\boldsymbol{\mu}^{2\cdot 1}_{x,-1}(u_1)$, $\sigma^{2\cdot 1}_{yx,-ii}$, and $\boldsymbol{\Sigma}^{2\cdot 1}_{xx,-i|i}$ are as defined previously, and

$$q_1(u_1,\nu) = \frac{1}{\nu+1}\left\{\nu + \frac{(u_1-\mu_{x,1})^2}{\sigma_{yy,11}}\right\}.$$

Note that the above show that the class of SUN and SUT distributions are conveniently closed under marginalization and conditionalization.

3. Main Results

Assume Equation (1) holds, where $\boldsymbol{\Sigma}$ is positive definite. In this section, we show that $\mathbf{X}$ and $\mathbf{LY}_{(N)}$ are jointly distributed according to a mixture of SUE distributions, where $\mathbf{Y}_{(N)} = (Y_{(1)},\ldots,Y_{(N)})^\top$ is the vector of order statistics from $\mathbf{Y}$, and $\mathbf{L}$ is a $P\times N$ matrix of rank$(\mathbf{L}) = P$. To this end, let $\mathbf{Y}_{(N)} \in \mathcal{P}(\mathbf{Y})$, where $\mathcal{P}(\mathbf{Y}) = \{\mathbf{Y}_i : \mathbf{Y}_i = \mathbf{P}_i\mathbf{Y}, i = 1,\ldots,N!\}$ is the collection of vectors $\mathbf{Y}_i$ corresponding to the $N!$ different permutations of the components of $\mathbf{Y}$, with $\mathbf{P}_i$ a $N\times N$ permutation matrix such that $\mathbf{P}_i \neq \mathbf{P}_{i'}$, for all $i \neq i'$. Further, let $\mathbf{D}$ be an $(N-1)\times N$ difference matrix such that $\mathbf{DY} = (Y_2-Y_1, Y_3-Y_2,\ldots,Y_N-Y_{N-1})$, i.e., row i of $\mathbf{D}$ is given by $\mathbf{e}_{i+1}^\top - \mathbf{e}_i^\top$, $i = 1,\ldots,N$, where $\mathbf{e}_1,\ldots,\mathbf{e}_N$ are N-dimensional unit basis vectors. We give below our main result on the joint, marginal and conditional distributions of $\mathbf{X}$ and $\mathbf{LY}_{(N)}$.

Proposition 3.1 Assume Equation (1) holds. Then the following are true:

(i) the joint CDF $F_{\mathbf{X},\mathbf{LY}_{(N)}}(\cdot)$ and joint PDF $f_{\mathbf{X},\mathbf{LY}_{(N)}}(\cdot)$ of $\mathbf{X}$ and $\mathbf{LY}_{(N)}$ is then given by

$$F_{\mathbf{X},\mathbf{LY}_{(N)}}(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N!} \pi_i F^{h^{(K+N+P-1)}}_{K+P,N-1}(\mathbf{x},\mathbf{w};\boldsymbol{\Theta}_i), \tag{9}$$

$$f_{\mathbf{X},\mathbf{LY}_{(N)}}(\mathbf{x},\mathbf{w}) = \sum_{i=1}^{N!} \pi_i f^{h^{(K+N+P-1)}}_{K+P,N-1}(\mathbf{x},\mathbf{w};\boldsymbol{\Theta}_i), \tag{10}$$

where $F^{h^{(K+N+P-1)}}_{K+P,N-1}(\cdot;\boldsymbol{\Theta}_i)$ and $f^{h^{(K+N+P-1)}}_{K+P,N-1}(\cdot;\boldsymbol{\Theta}_i)$ are the CDF and PDF of $\mathrm{SUE}_{K+P,N-1}(\boldsymbol{\Theta}_i,\, h^{(K+N+P-1)})$, and

$$\pi_i = G^{h^{(N-1)}}_{N-1}(\boldsymbol{\eta}_i;\boldsymbol{\Gamma}_i),$$

with $G^{h^{(N-1)}}_{N-1}(\cdot;\boldsymbol{\Gamma}_i)$ the CDF of $\mathrm{EC}_{N-1}(\mathbf{0},\boldsymbol{\Gamma}_i,h^{(N-1)})$, $\boldsymbol{\Theta}_i = \{\boldsymbol{\xi}_i,\boldsymbol{\eta}_i,\boldsymbol{\Omega}_i,\boldsymbol{\Gamma}_i,\boldsymbol{\Lambda}_i\}$,

$$\boldsymbol{\xi}_i = \begin{pmatrix}\boldsymbol{\mu}_x\\ \mathbf{L}\boldsymbol{\mu}_{y,i}\end{pmatrix},\quad \boldsymbol{\eta}_i = \mathbf{D}\boldsymbol{\mu}_{y,i},\quad \boldsymbol{\Omega}_i = \begin{pmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{yx,i}^\top\mathbf{L}^\top\\ & \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top\end{pmatrix},$$

$$\boldsymbol{\Gamma}_i = \mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}^\top,\quad \boldsymbol{\Lambda}_i = \begin{pmatrix}\boldsymbol{\Sigma}_{yx,i}^\top\mathbf{D}^\top\\ \mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top\end{pmatrix},$$

$\boldsymbol{\mu}_{y,i} = \mathbf{P}_i\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_{yy,i} = \mathbf{P}_i\boldsymbol{\Sigma}_{yy}\mathbf{P}_i^\top$, and $\boldsymbol{\Sigma}_{yx,i} = \mathbf{P}_i\boldsymbol{\Sigma}_{yx}$;

(ii) the marginal CDF of $\mathbf{LY}_{(N)}$ is given by

$$F_{\mathbf{LY}_{(N)}}(\mathbf{w}) = \sum_{i=1}^{N!} \pi_i F_{P,N-1}^{h^{(N+P-1)}}(\mathbf{w}; \mathbf{\Theta}_i^1),$$

where $\mathbf{\Theta}_i^1 = \{\mathbf{L}\boldsymbol{\mu}_{y,i}, \mathbf{D}\boldsymbol{\mu}_{y,i}, \mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top, \mathbf{D}\mathbf{\Sigma}_{yy,i}\mathbf{D}^\top, \mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{D}^\top\}$;

(iii) the conditional CDF of $\mathbf{X}$ given $\mathbf{LY}_{(N)} = \mathbf{w}$ is given by

$$F_{\mathbf{X}|\mathbf{LY}_{(N)}}(\mathbf{x}|\mathbf{w}) = \sum_{i=1}^{N!} \pi_i F_{K,N-1}^{h_{q_{2i}(\mathbf{w})}^{(K+N-1)}}(\mathbf{x}; \mathbf{\Theta}_i^{1\cdot 2}),$$

where $F_{K,N-1}^{h_{q_{2i}(\mathbf{w})}^{(K+N-1)}}(\cdot; \mathbf{\Theta}_i^{1\cdot 2})$ is the CDF of $\mathrm{SUE}_{K,N-1}(\mathbf{\Theta}_i^{1\cdot 2}, h_{q_{2i}(\mathbf{w})}^{(K+N-1)})$ with conditional density generator function $h_a^{(m)}$ given by

$$h_a^{(m)}(u) = \frac{h^{(m+1)}(u+a)}{h^{(1)}(a)}, \quad a, u \geq 0,$$

and $\mathbf{\Theta}_i^{1\cdot 2} = \{\boldsymbol{\xi}_i^{1\cdot 2}(\mathbf{w}), \boldsymbol{\eta}_i^{1\cdot 2}(\mathbf{w}), \mathbf{\Omega}_i^{11\cdot 2}, \mathbf{\Gamma}_i^{1\cdot 2}, \mathbf{\Lambda}_i^{1\cdot 2}\}$, with $\boldsymbol{\xi}_i^{1\cdot 2}(\mathbf{w}) = \boldsymbol{\mu}_x + \mathbf{\Sigma}_{yx,i}^\top \mathbf{L}^\top(\mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}(\mathbf{w} - \mathbf{L}\boldsymbol{\mu}_{y,i})$, $\boldsymbol{\eta}_i^{1\cdot 2}(\mathbf{w}) = \mathbf{D}\boldsymbol{\mu}_{y,i} + \mathbf{D}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top(\mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}(\mathbf{w} - \mathbf{L}\boldsymbol{\mu}_{y,i})$, $\mathbf{\Omega}_i^{11\cdot 2} = \mathbf{\Sigma}_{xx} - \mathbf{\Sigma}_{yx,i}^\top \mathbf{L}^\top(\mathbf{L}\mathbf{\Sigma}_{yy,i}\ \mathbf{L}^\top)^{-1}\mathbf{L}\mathbf{\Sigma}_{yx,i}$, $\mathbf{\Gamma}_i^{1\cdot 2} = \mathbf{\Gamma}_i - \mathbf{D}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top(\mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}\mathbf{L}\mathbf{\Sigma}_{yy,i}\ \mathbf{D}^\top$, $\mathbf{\Lambda}_i^{1\cdot 2} = \mathbf{\Sigma}_{yx,i}^\top \mathbf{D}^\top - \mathbf{\Sigma}_{yx,i}^\top \mathbf{L}^\top(\mathbf{L}\mathbf{\Sigma}_{yy,i}\ \mathbf{L}^\top)^{-1}\mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{D}^\top$, and $q_{2i}(\mathbf{w}) = (\mathbf{w} - \mathbf{L}\boldsymbol{\mu}_{y,i})^\top(\mathbf{L}\mathbf{\Sigma}_{yy,i}\ \mathbf{L}^\top)^{-1}(\mathbf{w} - \mathbf{L}\boldsymbol{\mu}_{y,i})$;

(iv) the conditional CDF of $\mathbf{LY}_{(N)}$ given $\mathbf{X} = \mathbf{x}$ is given by

$$F_{\mathbf{LY}_{(N)}|\mathbf{X}}(\mathbf{w}|\mathbf{x}) = \sum_{i=1}^{N!} \pi_i F_{P,N-1}^{h_{q_1(\mathbf{x})}^{(N+P-1)}}(\mathbf{w}; \mathbf{\Theta}_i^{2\cdot 1}),$$

where $F_{P,N-1}^{h_{q_1(\mathbf{x})}^{(N+P-1)}}(\cdot; \mathbf{\Theta}_i^{2\cdot 1})$ is the CDF of $\mathrm{SUE}_{P,N-1}(\mathbf{\Theta}_i^{2\cdot 1}, h_{q_1(\mathbf{x})}^{(N+P-1)})$ with conditional density generator function $h_{q_1(\mathbf{x})}^{(N+P-1)}$, and $\mathbf{\Theta}_i^{2\cdot 1} = \{\boldsymbol{\xi}_i^{2\cdot 1}(\mathbf{x}), \boldsymbol{\eta}_i^{2\cdot 1}(\mathbf{x}), \mathbf{\Omega}_i^{22\cdot 1}, \mathbf{\Gamma}_i^{2\cdot 1}, \mathbf{\Lambda}_i^{2\cdot 1}\}$, with $\boldsymbol{\xi}_i^{2\cdot 1}(\mathbf{x}) = \boldsymbol{\mu}_{y,i} + \mathbf{L}\mathbf{\Sigma}_{yx,i}\mathbf{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$, $\boldsymbol{\eta}_i^{2\cdot 1}(\mathbf{x}) = \mathbf{D}\boldsymbol{\mu}_{y,i} + \mathbf{D}\mathbf{\Sigma}_{yx,i}\mathbf{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$, $\mathbf{\Omega}_i^{22\cdot 1} = \mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{L}^\top - \mathbf{L}\mathbf{\Sigma}_{yx,i}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{yx,i}^\top \mathbf{L}^\top$, $\mathbf{\Gamma}_i^{2\cdot 1} = \mathbf{\Gamma}_i - \mathbf{D}\mathbf{\Sigma}_{yx,i}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{yx,i}^\top \mathbf{D}^\top$, $\mathbf{\Lambda}_i^{2\cdot 1} = \mathbf{L}\mathbf{\Sigma}_{yy,i}\mathbf{D}^\top - \mathbf{L}\mathbf{\Sigma}_{yx,i}^\top \mathbf{\Sigma}_{xx}^{-1}\ \mathbf{\Sigma}_{yx,i}^\top \mathbf{D}^\top$, and $q_1(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_x)^\top \mathbf{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$.

Proof For (i), note first that

$$F_{\mathbf{X},\mathbf{LY}_{(N)}}(\mathbf{x}, \mathbf{w}) \ = \ \sum_{i=1}^{N!} \mathrm{P}(\mathbf{DY}_i \geq \mathbf{0})\mathrm{P}(\mathbf{X} \leq \mathbf{x}, \mathbf{LY}_i \leq \mathbf{w}|\mathbf{DY}_i \geq \mathbf{0}), \qquad (11)$$

where the inequalities hold componentwise. Next, note that

$$\begin{pmatrix} \mathbf{DY}_i \\ \mathbf{X} \\ \mathbf{LY}_i \end{pmatrix} \sim \mathrm{EC}_{K+N+P-1}\left(\begin{pmatrix} \boldsymbol{\eta}_i \\ \boldsymbol{\xi}_i \end{pmatrix}, \begin{pmatrix} \mathbf{\Gamma}_i & \mathbf{\Lambda}_i^\top \\ & \mathbf{\Omega}_i \end{pmatrix}, h^{(K+N+P-1)} \right),$$

for $i = 1, \ldots, N!$. For the ith term of Equation (11), we have $\mathrm{P}(\mathbf{DY}_i \geq \mathbf{0}) = \pi_i$ and by

Equation (3), we get

$$\mathrm{P}(\mathbf{X} \le \mathbf{x}, \mathbf{L}\mathbf{Y}_i \le \mathbf{w} | \mathbf{D}\mathbf{Y}_i \ge \mathbf{0}) = F_{K+P,N-1}^{h^{(K+N+P-1)}}(\mathbf{x}, \mathbf{w}; \boldsymbol{\Theta}_i),$$

which proves (i). Parts (ii)-(iv) follow in a straightforward manner from the mixture representation given in Equation (9) and results in Jamalizadeh and Balakrishnan (2012) on marginal and conditional distributions of SUE distributions. ∎

Proposition 3.1 is a generalization of Jamalizadeh and Balakrishnan (2009b) and Balakrishnan et al. (2012), and gives the relevant joint, marginal, and conditional distributions as mixtures of SUE distributions. It can also be specialized to the case of normal and t-distributions; hence, it can be viewed as extensions of Viana (1998) and Olkin and Viana (1995). In either case, we need only replace $F_{K+P,N-1}^{h^{(K+N+P-1)}}(\cdot; \boldsymbol{\Theta}_i)$, the CDF of a SUE distribution with parameter $\boldsymbol{\Theta}_i$ and density generator function $h^{(K+N+P-1)}$ with the CDF $F_{K+P,N-1}^{\phi^{(K+N+P-1)}}(\cdot; \boldsymbol{\Theta}_i)$ of a SUN distribution (with density generator function $\phi^{(K+N+P-1)}$) in the normal case, or with the CDF $F_{K+P,N-1}^{t_\nu^{(K+N+P-1)}}(\cdot; \boldsymbol{\Theta}_i)$ of a SUT distribution (with density generator function $t_\nu^{(K+N+P-1)}$) in the case of t. It is also important to mention that the density generator function should be replaced by the characteristic generator function when $\boldsymbol{\Sigma}$ is singular.

3.1 Special case: results for X and $\mathbf{a}^\top \mathbf{Y}_{(N)}$

In this section, we consider the special case $K = P = 1$ with $\mathbf{L} = \mathbf{a}^\top = (a_1, \ldots, a_N)$. Provided $\sum_{i=1}^N a_i \neq 0$, the joint distribution of X and $\mathbf{a}^\top \mathbf{Y}_{(N)}$ is given by Equation (9) in Proposition 3.1. If $\sum_{i=1}^N a_i = 0$, then the bivariate SUE distributions in Equation (9) are singular, in which case the density generator function $h^{(N+1)}$ is replaced by the characteristic generator function $\varphi^{(N+1)}$. Corresponding marginal and conditional distributions likewise follow from Proposition 3.1. For example, consider the exchangeable case

$$\begin{pmatrix} X \\ \mathbf{Y} \end{pmatrix} \sim \mathrm{EC}_{N+1}\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu \mathbf{1}_N \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \delta\tau\sigma\mathbf{1}_N^\top \\ & \tau^2\{(1-\rho)\mathbf{I}_N + \rho\mathbf{1}_N\mathbf{1}_N^\top\} \end{pmatrix}, h^{(N+1)}\right),$$

where $\tau > 0$, $|\rho| < 1$, $\sqrt{N}|\delta| \le \sqrt{1+\rho(N-1)}$, $\mathbf{1}_N = (1, \ldots, 1)^\top$ is the $N \times 1$ summing vector, and $\mathbf{I}_N = \mathrm{diag}(1, \ldots, 1)$ is the $N \times N$ identity matrix. This equicorrelation structure for $\mathbf{Y}$ is found most frequently in familial studies in genetics, for example, and in animal teratology, where data arise in clusters from litters. Then, it follows from Proposition 3.1 that $X | \mathbf{a}^\top \mathbf{Y}_{(N)} = w \sim \mathrm{SUE}_{1,N-1}(\boldsymbol{\Theta}^{1\cdot 2}, h_{q_2(w)}^{(N)})$, where $\boldsymbol{\Theta}^{1\cdot 2} = \{\xi^{1\cdot 2}(w), \boldsymbol{\eta}^{1\cdot 2}(w), \omega^{11\cdot 2}, \boldsymbol{\Gamma}^{1\cdot 2}, \boldsymbol{\lambda}^{1\cdot 2}\}$, with

$$\xi^{1\cdot 2}(w) = \mu_x + \frac{\delta\sigma \sum_{i=1}^N a_i}{\tau\left\{(1-\rho)\sum_{i=1}^N a_i^2 + \rho\left(\sum_{i=1}^N a_i\right)^2\right\}}\left[w - \mu\sum_{i=1}^N a_i\right],$$

$$\boldsymbol{\eta}^{1\cdot 2}(w) = \frac{(1-\rho)\mathbf{D}\mathbf{a}}{(1-\rho)\sum_{i=1}^N a_i^2 + \rho\left(\sum_{i=1}^N a_i\right)^2}\left[w - \mu\sum_{i=1}^N a_i\right],$$

$$\omega^{11\cdot 2} = \sigma^2 - \frac{\delta^2\sigma^2\left(\sum_{i=1}^N a_i\right)^2}{(1-\rho)\sum_{i=1}^N a_i^2 + \rho\left(\sum_{i=1}^N a_i\right)^2},$$

$$\boldsymbol{\Gamma}^{1\cdot 2} = \tau^2(1-\rho)\mathbf{D}\mathbf{D}^\top - \frac{\tau^2(1-\rho)^2\mathbf{D}\mathbf{a}\mathbf{a}^\top\mathbf{D}^\top}{(1-\rho)\sum_{i=1}^N a_i^2 + \rho\left(\sum_{i=1}^N a_i\right)^2},$$

$$\boldsymbol{\lambda}^{1\cdot 2} = -\frac{\delta\tau\sigma(1-\rho)\mathbf{D}\mathbf{a}\sum_{i=1}^N a_i}{(1-\rho)\sum_{i=1}^N a_i^2 + \rho\left(\sum_{i=1}^N a_i\right)^2},$$

$$q_2(w) = \frac{\left(w - \mu\sum_{i=1}^N a_i\right)^2}{\tau\left\{(1-\rho)\sum_{i=1}^N a_i^2 + \rho\left(\sum_{i=1}^N a_i\right)^2\right\}}.$$

These results may be specialized as well to the normal and t-cases as generalizations of previous results by Jamalizadeh and Balakrishnan (2009b), Balakrishnan et al. (2012), Viana (1998), and Olkin and Viana (1995), among others.

We now give the best (non-linear) predictors $\mathrm{E}[X|\mathbf{a}^\top\mathbf{Y}_{(N)} = w]$ and $\mathrm{E}[\mathbf{a}^\top\mathbf{Y}_{(N)}|X = x]$ of X and $\mathbf{a}^\top\mathbf{Y}_{(N)}$, respectively, in the normal and t-cases under squared error loss. Let

$$\mathbf{D} = \begin{pmatrix} \mathbf{d}_j^\top \\ \mathbf{D}_{-j} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{j+1}^\top - \mathbf{e}_j^\top \\ \mathbf{D}_{-j} \end{pmatrix}, \tag{12}$$

where $\mathbf{D}_{-j}$ is the matrix obtained from $\mathbf{D}$ by deleting row j, where $\mathbf{e}_1, \ldots, \mathbf{e}_N$ are N-dimensional unit basis vectors.

Proposition 3.2 Suppose X and $\mathbf{Y}$ have a $(N+1)$-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, under squared error loss, the best (non-linear) predictor of X based on $\mathbf{a}^\top\mathbf{Y}_{(N)} = w$ is

$$\begin{aligned}\mathrm{E}\left[X|\mathbf{a}^\top\mathbf{Y}_{(N)} = w\right] &= \sum_{i=1}^{N!}\pi_i\xi_i^{1\cdot 2}(w) + \sum_{i=1}^{N!}\frac{\pi_i}{\Phi_{N-1}(\boldsymbol{\eta}_i^{1\cdot 2}(w);\boldsymbol{\Gamma}_i^{1\cdot 2})}\sum_{j=1}^{N-1}\frac{\lambda_{i,j}^{1\cdot 2}}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}} \\ &\quad\times\phi\left(\frac{\eta_{i,j}^{1\cdot 2}(w)}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}}\right)\Phi_{N-2}\left(\boldsymbol{\eta}_{i,-j}^{1\cdot 2}(w) - \frac{\eta_{i,j}^{1\cdot 2}(w)}{\gamma_{i,jj}^{1\cdot 2}}\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2};\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2}\right);\end{aligned} \tag{13}$$

if, on the other hand, X and $\mathbf{Y}$ have a $(N+1)$-dimensional t_ν-distribution with mean $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$, then the best (non-linear) predictor of X based on $\mathbf{a}^\top\mathbf{Y}_{(N)} = w$

becomes

$$\mathrm{E}\left[X|\mathbf{a}^\top\mathbf{Y}_{(N)}=w\right]=\sum_{i=1}^{N!}\pi_i\xi_i^{1\cdot 2}(w)+\frac{(\nu+1)^{(\nu+1)/2}\Gamma\left(\frac{\nu}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right)}\sum_{i=1}^{N!}\frac{\pi_i q_{2i}^{1/2}(w,\nu)}{T_{N-1}\left(\boldsymbol{\eta}_i^{1\cdot 2}(w);\ \begin{matrix} q_{2i}(w,\nu)\boldsymbol{\Gamma}_i^{1\cdot 2}, \\ \nu+1\end{matrix}\right)}$$
$$\times\sum_{j=1}^{N-1}\frac{\lambda_{i,j}^{1\cdot 2}}{\sqrt{\gamma_{i,jj}^{1\cdot 2}}}T_{N-2}\left(\begin{matrix}\frac{\sqrt{\nu}}{\sqrt{\nu+1+\frac{\{\eta_{i,j}^{1\cdot 2}(w)\}^2}{q_{2i}(w,\nu)\gamma_{i,jj}^{1\cdot 2}}}}\left(\boldsymbol{\eta}_{i,-j}^{1\cdot 2}(w)-\frac{\eta_{i,j}^{1\cdot 2}(w)}{\gamma_{i,jj}^{1\cdot 2}}\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\right); \\ q_{2i}(w,\nu)\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2},\nu\end{matrix}\right)$$
$$\times\left(\nu+1+\frac{\{\eta_{i,j}^{1\cdot 2}(w)\}^2}{q_{2i}(w,\nu)\gamma_{i,jj}^{1\cdot 2}}\right)^{-\nu/2}. \tag{14}$$

All relevant quantities in Equations (13)–(14) are as defined in Proposition 3.1 and in Section 2.2; for example,

$$\boldsymbol{\eta}_i^{1\cdot 2}(\mathbf{w})=\begin{pmatrix}\eta_{i,j}^{1\cdot 2}(w)\\ \boldsymbol{\eta}_{i,-j}^{1\cdot 2}(w)\end{pmatrix}\ =\ \mathbf{D}\boldsymbol{\mu}_{y,i}+\mathbf{D}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top(\mathbf{L}\boldsymbol{\Sigma}_{yy,i}\mathbf{L}^\top)^{-1}(\mathbf{w}-\mathbf{L}\boldsymbol{\mu}_{y,i}), \tag{15}$$

$$\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot 2}=\mathbf{D}_{-j}\boldsymbol{\Sigma}_{yy,i}\mathbf{D}_{-j}^\top-\frac{\mathbf{D}_{-j}\boldsymbol{\Sigma}_{yy,i}\mathbf{a}\mathbf{a}^\top\boldsymbol{\Sigma}_{yy,i}\mathbf{D}_{-j}^\top}{\mathbf{a}^\top\boldsymbol{\Sigma}_{yy,i}\mathbf{a}}-\frac{\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\left(\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\right)^\top}{\gamma_{i,jj}^{1\cdot 2}}$$
$$=\boldsymbol{\Gamma}_{i,-j-j}^{1\cdot 2}-\frac{\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\left(\boldsymbol{\gamma}_{i,-jj}^{1\cdot 2}\right)^\top}{\gamma_{i,jj}^{1\cdot 2}},$$

with $\mathbf{D}$ given by Equation (12). The best (non-linear) predictor $\mathrm{E}[\mathbf{a}^\top\mathbf{Y}_{(N)}|X=x]$ of $\mathbf{a}^\top\mathbf{Y}_{(N)}$ based on $X=x$ in both the normal and t-cases is obtained by replacing "w" with "x", changing the superscript "$1\cdot 2$" to "$2\cdot 1$", and replacing "$q_{2i}(w,\nu)$" in Equation (14) with "$q_1(x,\nu)$".

Proof Expressions (13) and (14) are straightforward from Equations (6) and (8). ■

Proposition 3.2 extends results in Loperfido (2008b), Jamalizadeh and Balakrishnan (2009a), and Balakrishnan et al. (2012), to the case $N\geq 2$.

3.2 Special case: results for X and $Y_{(r)}$

The joint CDF of X and $Y_{(r)}$ can be obtained from Proposition 3.1 or from Section 3.1, by taking $\mathbf{a}=\mathbf{e}_r$, $r=1,\ldots,N$. However, note that the resulting CDF involves a mixture consisting of $N!$ components. Evaluating the CDF may thus be a problem in practice when N is large.

In this section, we present an alternative approach for deriving this joint CDF with only $N\binom{N-1}{r-1}<N!$ terms. To do this, let $1\leq r\leq N$ be an integer, and for integers $1\leq j_1<\cdots<j_{r-1}\leq N-1$, let $\mathbf{S}_{j_1\ldots j_{r-1}}=\mathrm{diag}\,(s_1,\ldots,s_{N-1})$ be a $(N-1)\times(N-1)$ diagonal matrix such that if $j_{r-1}=N$, then

$$s_i=\begin{cases}1, & \text{for } i=j_1,\ldots,j_{r-2}, \text{ and } i=N-1;\\ -1, & \text{otherwise}\end{cases}$$

and otherwise

$$s_i = \begin{cases} 1, & \text{for } i = j_1, \ldots, j_{r-1}; \\ -1, & \text{otherwise.} \end{cases}$$

In particular, $\mathbf{S}_{j_1 \ldots j_{N-1}} = \mathbf{I}_{N-1}$ and $\mathbf{S}_{j_0} = -\mathbf{I}_{N-1}$. Further, let $\mathbf{Y}_{j_1 \ldots j_{r-1}} = (Y_{j_1}, \ldots, Y_{j_{r-1}})^\top$, and for $i = 1, \ldots, N$, let the vector $\mathbf{Y}_{-i-j_1-\cdots-j_{r-1}}$ $(j_k \neq i, k = 1, \ldots, r-1)$ be obtained from $\mathbf{Y}$ by deleting $Y_i, Y_{j_1}, \ldots, Y_{j_{r-1}}$. Consider also the partitions

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{-i} \\ Y_i \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{y,-i} \\ \mu_{y,i} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{yy} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy,-i-i} & \boldsymbol{\sigma}_{yy,-ii} \\ & \sigma_{yy,ii} \end{pmatrix},$$

so that

$$\begin{pmatrix} X \\ \mathbf{Y}_{-i} \\ Y_i \end{pmatrix} \sim \mathrm{EC}_{m+N}\left(\begin{pmatrix} \mu_x \\ \boldsymbol{\mu}_{y,-i} \\ \mu_{y,i} \end{pmatrix}, \begin{pmatrix} \sigma_{xx} & \boldsymbol{\sigma}_{yx,-i}^\top & \sigma_{yx,i} \\ & \boldsymbol{\Sigma}_{yy,-i-i} & \boldsymbol{\sigma}_{yy,-ii} \\ & & \sigma_{yy,i} \end{pmatrix}, h^{(N+1)} \right).$$

The following proposition gives the joint CDF of X and $Y_{(r)}$ as a mixture of SUE distributions with only $N\binom{N-1}{r-1} < N!$ terms. The resulting CDF is thus computationally simpler to evaluate and hence, more useful in practice than the corresponding CDF obtained from Proposition 3.1. Moreover, the result is technically elegant and should be of theoretical interest in its own right.

Proposition 3.3 For $r = 1, \ldots, N$, and $j_k \neq i$, the joint CDF of X and $Y_{(r)}$ is given by

$$F_{X,Y_{(r)}}(x, y) = \sum_{i=1}^{N} \sum_{\substack{j_1 < \cdots < j_{r-1} \\ 1 \leq j_k \leq N, j_k \neq i}} \pi_{i,j_1 \ldots j_{r-1}} F_{2,N-1}^{h^{(N+1)}}(x, y; \boldsymbol{\Theta}_{i,j_1 \ldots j_{r-1}}), \tag{16}$$

where $\boldsymbol{\Theta}_{i,j_1 \ldots j_{r-1}} = \{\boldsymbol{\xi}_{i,j_1 \ldots j_{r-1}}, \boldsymbol{\eta}_{i,j_1 \ldots j_{r-1}}, \boldsymbol{\Omega}_{i,j_1 \ldots j_{r-1}}, \boldsymbol{\Gamma}_{i,j_1 \ldots j_{r-1}}, \boldsymbol{\Lambda}_{i,j_1 \ldots j_{r-1}}\}$, $F_{2,N-1}^{h^{(N+1)}}(\cdot; \boldsymbol{\Theta}_{i,j_1 \ldots j_{r-1}})$ is the CDF of $\mathrm{SUE}_{2,N-1}(\boldsymbol{\Theta}_{i,j_1 \ldots j_{r-1}}, h^{(N+1)})$, and

$$\pi_{i,j_1 \ldots j_{r-1}} = G_{N-1}^{h^{(N-1)}}(\mathbf{S}_{j_1 \ldots j_{r-1}}(\mu_{y,i} \mathbf{1}_{N-1} - \boldsymbol{\mu}_{y,-i}); \boldsymbol{\Gamma}_{i,j_1 \ldots j_{r-1}}), \tag{17}$$

with $G_{N-1}^{h^{(N-1)}}(\cdot; \boldsymbol{\Gamma}_{i,j_1 \ldots j_{r-1}})$ the CDF of $\mathrm{EC}_{N-1}(\mathbf{0}, \boldsymbol{\Gamma}_{i,j_1 \ldots j_{r-1}}, h^{(N+1)})$, $\boldsymbol{\xi}_{i,j_1 \ldots j_{r-1}} = (\mu_x, \mu_{y,i})^\top$, $\boldsymbol{\eta}_{i,j_1 \ldots j_{r-1}} = \mathbf{S}_{j_1 \ldots j_{r-1}}(\mu_{y,i} \mathbf{1}_{N-1} - \boldsymbol{\mu}_{y,-i})$,

$$\begin{aligned}
\boldsymbol{\Omega}_{i,j_1 \ldots j_{r-1}} &= \begin{pmatrix} \sigma_{xx} & \sigma_{yx,i} \\ & \sigma_{yy,ii} \end{pmatrix}, \\
\boldsymbol{\Gamma}_{i,j_1 \ldots j_{r-1}} &= \mathbf{S}_{j_1 \ldots j_{r-1}}(\sigma_{yy,ii} \mathbf{1}_{N-1} \mathbf{1}_{N-1}^\top + \boldsymbol{\Sigma}_{yy,-i-i} - \mathbf{1}_{N-1} \boldsymbol{\sigma}_{yy,-ii}^\top - \boldsymbol{\sigma}_{yy,-ii} \mathbf{1}_{N-1}^\top) \mathbf{S}_{j_1 \ldots j_{r-1}}, \\
\boldsymbol{\Lambda}_{i,j_1 \ldots j_{r-1}} &= \begin{pmatrix} (\sigma_{yx,i} \mathbf{1}_{N-1}^\top - \boldsymbol{\sigma}_{yx,-i})^\top \mathbf{S}_{j_1 \ldots j_{r-1}} \\ (\sigma_{yy,ii} \mathbf{1}_{N-1}^\top - \boldsymbol{\sigma}_{yy,-ii})^\top \mathbf{S}_{j_1 \ldots j_{r-1}} \end{pmatrix} = \begin{pmatrix} (\sigma_{yx,i} \mathbf{1}_{N-1}^\top - \boldsymbol{\sigma}_{yx,-i})^\top \mathbf{S}_{j_1 \ldots j_{r-1}} \\ \boldsymbol{\lambda}_{i,j_1 \ldots j_{r-1}}^\top \end{pmatrix}.
\end{aligned}$$

The marginal CDF of $Y_{(r)}$ is readily obtained from Equation (16) as

$$F_{Y_{(r)}}(y) = \sum_{i=1}^{N} \sum_{\substack{j_1 < \cdots < j_{r-1} \\ 1 \leq j_k \leq N, j_k \neq i}} \pi_{i,j_1 \ldots j_{r-1}} F_{1,N-1}^{h^{(N+1)}}(y; \boldsymbol{\Theta}_{i,j_1 \ldots j_{r-1}}^{(r)}),$$

where $\boldsymbol{\Theta}^{(r)}_{i,j_1\cdots j_{r-1}} = \{\mu_{y,i}, \boldsymbol{\eta}_{i,j_1,\ldots,j_{r-1}}, \sigma_{yy,ii}, \boldsymbol{\Gamma}_{i,j_1,\ldots,j_{r-1}}, \boldsymbol{\lambda}_{i,j_1,\ldots,j_{r-1}}\}$.

Proof We have

$$F_{X,Y_{(r)}}(x,y) = \sum_{i=1}^{N} \mathrm{P}\left(X \le x, Y_i \le y, Y_i = Y_{(r)}\right). \tag{18}$$

The ith term of the RHS of Equation (18) is

$$\begin{aligned}
\mathrm{P}\left(X \le x, Y_i \le y, Y_i = Y_{(r)}\right) &= \sum_{\substack{j_1<\cdots<j_{r-1}\\ 1\le j_k\le N, j_k\ne i}} \mathrm{P}\begin{pmatrix} X < x, Y_i \le y, \\ \max(\mathbf{Y}_{j_1\cdots j_{r-1}}) < Y_i < \min(\mathbf{Y}_{-i-j_1-\cdots-j_{r-1}}) \end{pmatrix} \\
&= \sum_{\substack{j_1<\cdots<j_{r-1}\\ 1\le j_k\le N, j_k\ne i}} \mathrm{P}(X < x, Y_i \le y | \mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}) \\
&\quad \times \mathrm{P}(\mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}),
\end{aligned}$$

where $\max(\mathbf{Y}_{j_1\cdots j_{r-1}}) = \max(Y_{j_1}, \ldots, Y_{j_{r-1}})$ and $\min(\mathbf{Y}_{-i-j_1-\cdots-j_{r-1}})$ is the minimum of the elements of $\mathbf{Y}_{-i-j_1-\cdots-j_{r-1}}$, which is defined at the beginning of Section 3.2. Now, we have for $i = 1, \ldots, N$,

$$\begin{pmatrix} \mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} \\ (X, Y_i)^\top \end{pmatrix} \sim \mathrm{EC}_{N+1}\left(\begin{pmatrix} \boldsymbol{\eta}_{i,j_1\cdots j_{r-1}} \\ \boldsymbol{\xi}_{i,j_1\cdots j_{r-1}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma}_{i,j_1\cdots j_{r-1}} & \boldsymbol{\Lambda}^\top_{i,j_1\cdots j_{r-1}} \\ & \boldsymbol{\Omega}_{i,j_1\cdots j_{r-1}} \end{pmatrix}, h^{(N+1)}\right),$$

so that

$$\mathrm{P}(X < x, Y_i \le y | \mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}) = F^{h^{(N+1)}}_{2,N-1}(x, y; \boldsymbol{\Theta}_{i,j_1\cdots j_{r-1}}),$$

and since $\mathrm{P}(\mathbf{S}_{j_1\cdots j_{r-1}}\{\mathbf{1}_{N-1}Y_i - \mathbf{Y}_{-i}\} > \mathbf{0}) = \pi_{i,j_1\cdots j_{r-1}}$, the proof is complete. ■

Proposition 3.3 extends Loperfido's (2008b) result to the general contralateral data set-up and for an arbitrary multivariate skew-elliptical distribution with an arbitrary correlation structure. Best nonlinear predictors of X and of $Y_{(r)}$ based on X and on $Y_{(r)}$, respectively, are also conveniently obtained from Proposition 3.3. To do this for the case $r = N$, consider the following partitions for $j \ne i$:

$$\begin{aligned}
\boldsymbol{\mu}_{y,-i} &= \begin{pmatrix} \mu_{y,j} \\ \boldsymbol{\mu}_{y,-i-j} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{yy,-i-i} = \begin{pmatrix} \sigma_{yy,jj} & \boldsymbol{\sigma}^\top_{yy;-i-j,j} \\ & \boldsymbol{\Sigma}_{yy;-i-j,-i-j} \end{pmatrix}, \\
\boldsymbol{\sigma}_{yy,-ii} &= \begin{pmatrix} \sigma_{yy,ji} \\ \boldsymbol{\sigma}_{yy;-i-j,i} \end{pmatrix}, \quad \boldsymbol{\sigma}_{yx,-i} = \begin{pmatrix} \sigma_{yx,j} \\ \boldsymbol{\sigma}_{yx,-i-j} \end{pmatrix}.
\end{aligned}$$

Assuming joint normality for X and $\mathbf{Y}$, we get

$$\begin{aligned}
\mathrm{E}\left[X | Y_{(N)} = y\right] = \sum_{i=1}^{N} \pi_i \Bigg\{ & \xi^{1\cdot 2}_i(y) + \frac{1}{\Phi_{N-1}\left(\boldsymbol{\eta}^{1\cdot 2}_i(y); \boldsymbol{\Gamma}^{1\cdot 2}_i\right)} \sum_{j\ne i} \frac{\lambda^{1\cdot 2}_{i,j} \phi\left(\frac{\eta^{1\cdot 2}_{i,j}(y)}{\sqrt{\gamma^{1\cdot 2}_{i,jj}}}\right)}{\sqrt{\gamma^{1\cdot 2}_{i,jj}}} \\
& \times \Phi_{N-2}\left(\boldsymbol{\eta}^{1\cdot 2}_{i,-j}(y) - \frac{\eta^{1\cdot 2}_{i,j}(y)}{\gamma^{1\cdot 2}_{i,jj}} \boldsymbol{\gamma}^{1\cdot 2}_{i,-jj}; \boldsymbol{\Gamma}^{1\cdot 2}_{i,-j|j}\right) \Bigg\},
\end{aligned} \tag{19}$$

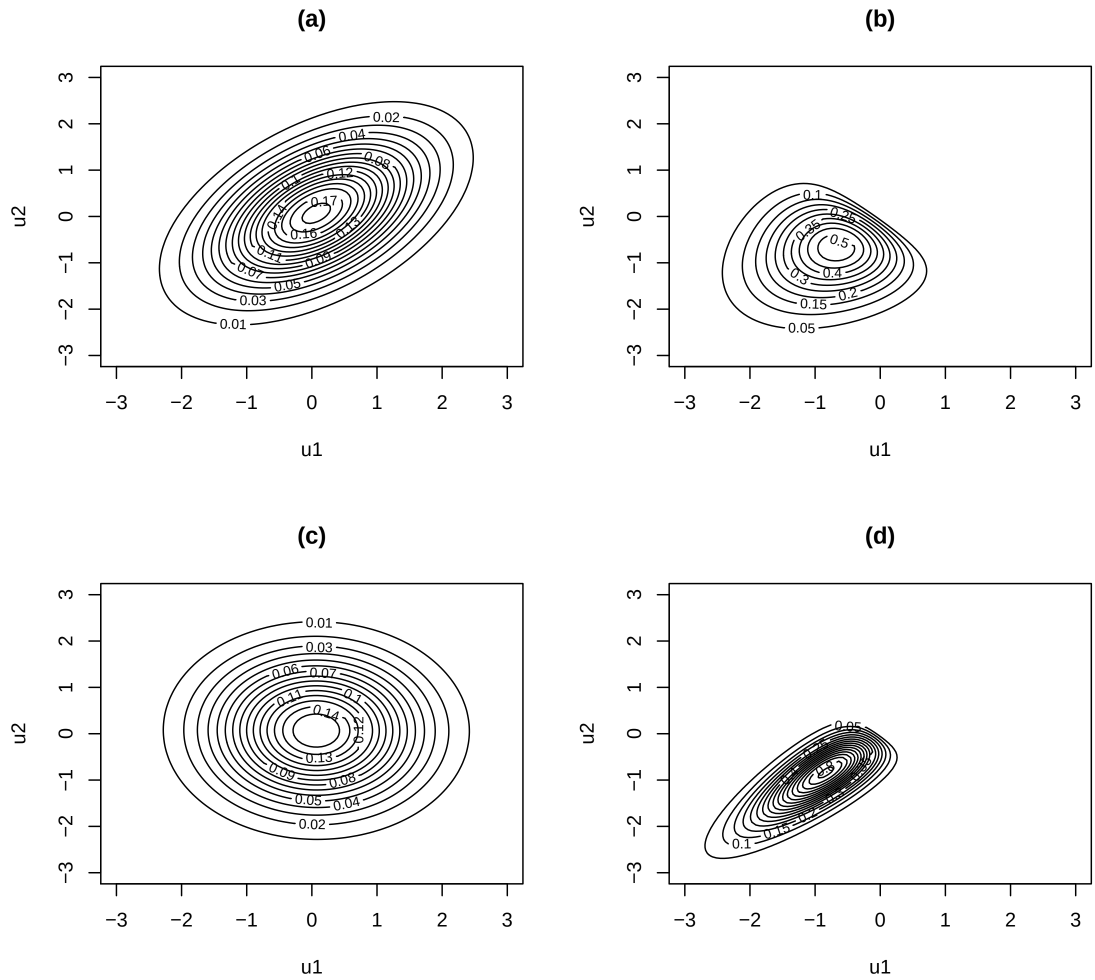

Figure 1. Contour plots of $f_{X,Y_{(N)}}(x,y)$, for $\mu_x = \mu = 0$ and $\sigma^2 = \tau^2 = 1$, with (a) $\rho = 0.1, \delta = 0.5$, (b) $\rho = -0.45, \delta = 0.5$, (c) $\rho = 0.1, \delta = 0$, and (d) $\rho = -0.45, \delta = 0.9$.

where $\pi_i = G_{N-1}^{h^{(N-1)}}(\boldsymbol{\eta}_i;\boldsymbol{\Gamma}_i)$, $\xi_i^{1\cdot2}(y) = \mu_x + \sigma_{yx,i}(y-\mu_{y,i})/\sigma_{yy,ii}$, $\lambda_{i,j}^{1\cdot2} = \sigma_{yx,i} - \sigma_{yx,j} - \sigma_{yx,i}(\sigma_{yy,ii}-\sigma_{yy,ji})/\sigma_{yy,ii}$, $\gamma_{i,jj}^{1\cdot2} = \sigma_{yy,ii}+\sigma_{yy,jj} - 2\sigma_{yy,ji} - (\sigma_{yy,ii}-\sigma_{yy,ji})^2/\sigma_{yy,ii}$, $\boldsymbol{\gamma}_{i,-jj}^{1\cdot2} = \sigma_{yy,ii}\mathbf{1}_{N-2}^\top + \boldsymbol{\sigma}_{yy;-i-j,j}^\top - \boldsymbol{\sigma}_{yy;-i-j,i}^\top - \sigma_{yy,ji}\mathbf{1}_{N-2}^\top - (\sigma_{yy,ii}-\sigma_{yy,ji})(\sigma_{yy,ii}\mathbf{1}_{N-2} - \boldsymbol{\sigma}_{yy;-i-j,i})^\top/\sigma_{yy,ii}$,

$$\boldsymbol{\eta}_i^{1\cdot2}(y) = \begin{pmatrix} \eta_{i,j}^{1\cdot2}(y) \\ \boldsymbol{\eta}_{i,-j}^{1\cdot2}(y) \end{pmatrix} = \mu_{y,i}\mathbf{1}_{N-1} - \boldsymbol{\mu}_{y,-i} + \frac{\sigma_{yy,ii}\mathbf{1}_{N-1} - \boldsymbol{\sigma}_{yy,-ii}}{\sigma_{yy,ii}}(y-\mu_{y,i}),$$

$$\boldsymbol{\Gamma}_i^{1\cdot2} = \boldsymbol{\Gamma}_i - \frac{(\sigma_{yy,ii}\mathbf{1}_{N-1} - \boldsymbol{\sigma}_{yy,-ii})(\sigma_{yy,ii}\mathbf{1}_{N-1} - \boldsymbol{\sigma}_{yy,-ii})^\top}{\sigma_{yy,ii}},$$

$$\boldsymbol{\Gamma}_{i,-j-j}^{1\cdot2} = \sigma_{yy,ii}\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top + \boldsymbol{\Sigma}_{yy;-i-j,-i-j} - \mathbf{1}_{N-2}\boldsymbol{\sigma}_{yy;-i-j,i} - \boldsymbol{\sigma}_{yy;-i-j,i}\mathbf{1}_{N-2}^\top - \frac{(\sigma_{yy,ii}\mathbf{1}_{N-2} - \boldsymbol{\sigma}_{yy;-i-j,i})(\sigma_{yy,ii}\mathbf{1}_{N-2} - \boldsymbol{\sigma}_{yy;-i-j,i})^\top}{\sigma_{yy,ii}},$$

with $\boldsymbol{\Gamma}_{i,-j|j}^{1\cdot2}$ given in Equation (15). Similarly, the predictor $\mathrm{E}[Y_{(N)}|X = x]$ can be analogously obtained from Equation (19) by replacing "y" with "x" and changing the super-

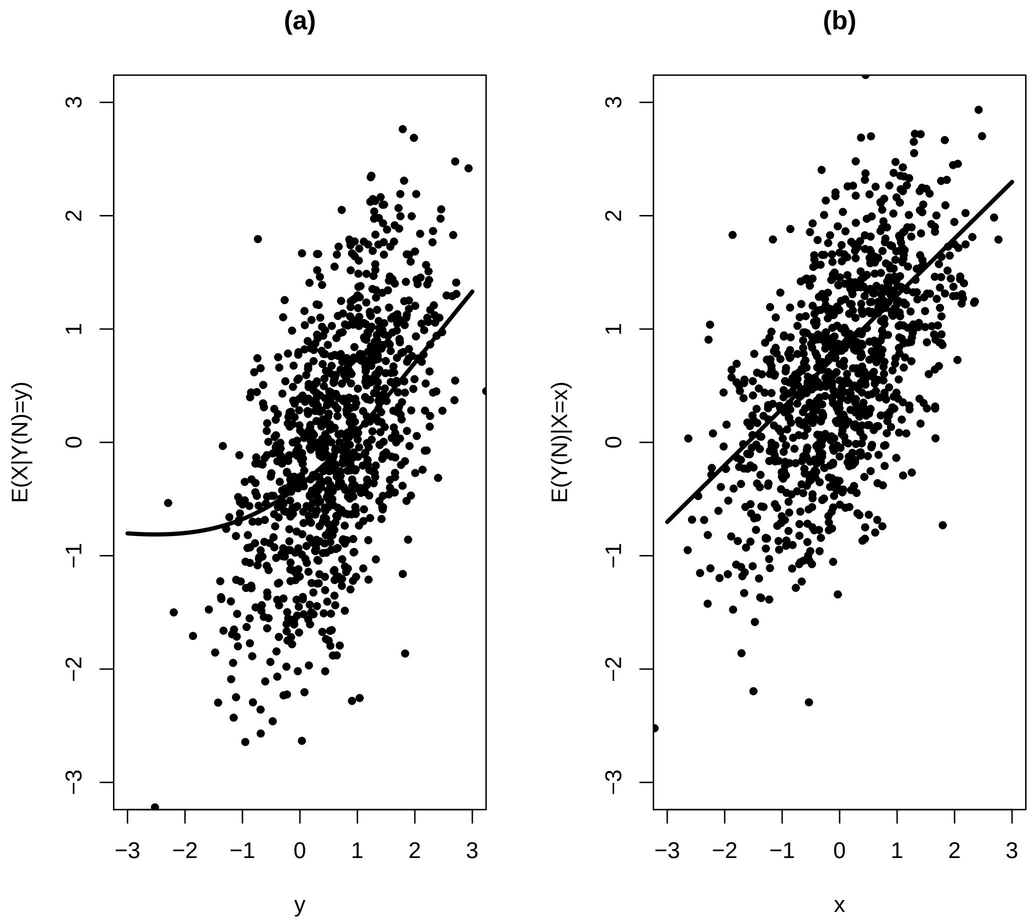

Figure 2. Plots of (a) $\mathrm{E}[X|Y_{(N)}=y]$ in Equation (21) and of (b) $\mathrm{E}[Y_{(N)}|X=x]$ in Equation (20), with $N=3$, $\mu_x=\mu=0$, $\sigma^2=\tau^2=1$, and $\rho=\delta=0.5$. Superimposed on (a) and (b) are scatterplots of x vs. $y=\max(y_1,y_2,y_3)$, and of $y=\max(y_1,y_2,y_3)$ vs. x, for sample size 1000.

script "$1\cdot 2$" to "$2\cdot 1$", where $\xi_i^{2\cdot1}(x)=\mu_{y,i}+\sigma_{yx,i}(x-\mu_x)/\sigma_{xx}$, $\lambda_{i,j}^{2\cdot1}=\sigma_{yy,ii}-\sigma_{yy,ji}-\sigma_{yx,i}(\sigma_{yx,i}-\sigma_{yx,j})/\sigma_{xx}$, $\gamma_{i,jj}^{2\cdot1}=\sigma_{yy,ii}+\sigma_{yy,jj}-2\sigma_{yy,ji}-(\sigma_{yx,i}-\sigma_{yx,j})^2/\sigma_{xx}$, $\boldsymbol{\gamma}_{i,-jj}^{2\cdot1}=\sigma_{yy,ii}\mathbf{1}_{N-2}^\top+\boldsymbol{\sigma}_{yy;-i-j,j}^\top-\boldsymbol{\sigma}_{yy;-i-j,i}^\top-\sigma_{yy,ji}\mathbf{1}_{N-2}^\top-(\sigma_{yx,i}-\sigma_{yx,j})(\sigma_{yx,i}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yx,-i-j})^\top/\sigma_{xx}$,

$$\boldsymbol{\eta}_i^{2\cdot1}(x)=\begin{pmatrix}\eta_{i,j}^{2\cdot1}(x)\\ \boldsymbol{\eta}_{i,-j}^{2\cdot1}(x)\end{pmatrix}=\mu_{y,i}\mathbf{1}_{N-1}-\boldsymbol{\mu}_{y,-i}+\frac{\sigma_{yx,i}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yx,-i}}{\sigma_{xx}}(x-\mu_x),$$

$$\boldsymbol{\Gamma}_i^{2\cdot1}=\boldsymbol{\Gamma}_i-\frac{(\sigma_{yx,i}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yx,-i})(\sigma_{yx,i}\mathbf{1}_{N-1}-\boldsymbol{\sigma}_{yx,-i})^\top}{\sigma_{xx}},$$

$$\begin{aligned}\boldsymbol{\Gamma}_{i,-j-j}^{2\cdot1}&=\sigma_{yy,ii}\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top+\boldsymbol{\Sigma}_{yy;-i-j,-i-j}-\mathbf{1}_{N-2}\boldsymbol{\sigma}_{yy;-i-j,i}-\boldsymbol{\sigma}_{yy;-i-j,i}\mathbf{1}_{N-2}^\top\\&\quad-\frac{(\sigma_{yx,i}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yx,-i-j})(\sigma_{yx,i}\mathbf{1}_{N-2}-\boldsymbol{\sigma}_{yx,-i-j})^\top}{\sigma_{xx}}.\end{aligned}$$

Note that these predictors are computationally simpler than the corresponding ones given in Proposition 3.2.

In the exchangeable case, it is easy to see that X and $Y_{(r)}$ are jointly $\mathrm{SUE}_{2,N-1}(\boldsymbol{\Theta}, h^{(N+1)})$, where

$$\boldsymbol{\Theta} = \left\{ \begin{pmatrix} \mu_x \\ \mu \end{pmatrix}, \mathbf{0}, \begin{pmatrix} \sigma^2 & \delta\tau\sigma \\ & \tau^2 \end{pmatrix}, \tau^2(1-\rho)(\mathbf{I}_{N-1} + \rho \mathbf{1}_{N-1}\mathbf{1}_{N-1}^\top), \begin{pmatrix} \mathbf{0}^\top \\ \tau^2(1-\rho)\mathbf{1}_{N-1}^\top \end{pmatrix} \right\}.$$

That is, X and $Y_{(r)}$ have an exact bivariate SUE joint distribution under exchangeability. In the case with $r = N$ and assuming X and $\mathbf{Y}$ are jointly normal, their exact joint distribution is SUN, the contour plots of which are shown in Figure 1, with $\mu_x = \mu = 0$, $\sigma^2 = \tau^2 = 1$, for $(\rho, \delta) = (0.1, 0.5), (-0.45, 0.5), (0.1, 0), (-0.45, 0.9)$. In this case, the best predictors above reduce to

$$\mathrm{E}\left[Y_{(N)} | X = x\right] = \mu + \delta\tau \left(\frac{x - \mu_x}{\sigma}\right) + \frac{(N-1)\tau\sqrt{1-\rho}}{\sqrt{\pi}}, \tag{20}$$

$$\mathrm{E}\left[X | Y_{(N)} = y\right] = \mu_x + \delta\sigma \left(\left(\frac{y-\mu}{\tau}\right) - \frac{(N-1)\sqrt{\frac{1-\rho}{1+\rho}}\,\phi\left(\sqrt{\frac{1-\rho}{1+\rho}}\frac{(y-\mu)}{\tau}\right)}{\Phi_{N-1}\begin{pmatrix} (1-\rho)\tau^2(y-\mu)\mathbf{1}_{N-1}; \\ (1-\rho)\tau^2\left\{\mathbf{I}_{N-1} + \rho\mathbf{1}_{N-1}\mathbf{1}_{N-1}^\top\right\} \end{pmatrix}} \right)$$
$$\times \Phi_{N-2}\begin{pmatrix} \frac{(1-\rho)(y-\mu)}{1+\rho}\mathbf{1}_{N-2}; \\ (1-\rho)\tau^2\left\{\mathbf{I}_{N-2} + \frac{\rho}{1+\rho}\mathbf{1}_{N-2}\mathbf{1}_{N-2}^\top\right\} \end{pmatrix}. \tag{21}$$

We plot Equation (20) as a function of $x \in (-3, 3)$ and Equation (21) as a function of $y \in (-3, 3)$ in Figures 2(b) and 2(a), respectively, with $N = 3$, $\mu_x = \mu = 0$, $\sigma^2 = \tau^2 = 1$, and $\rho = \delta = 0.5$. Also shown in Figures 2(b) and 2(a) are scatterplots of x_k vs. $y_k = \max(y_{1k}, y_{2k}, y_{3k})$ and of $y_k = \max(y_{1k}, y_{2k}, y_{3k})$ vs. x_k, respectively, from a sample $(x_k, y_{1k}, y_{k2}, y_{3k})^\top$, $k = 1, \ldots, 1000$, from the 4-dimensional normal distribution with mean vector $\boldsymbol{\mu} = (\mu_x, \mu, \mu, \mu)^\top = \mathbf{0}$ and exchangeable covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & \delta\tau\sigma & \delta\tau\sigma & \delta\tau\sigma \\ & \tau^2 & \rho & \rho \\ & & \tau^2 & \rho \\ & & & \tau^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ & 1 & 0.5 & 0.5 \\ & & 1 & 0.5 \\ & & & 1 \end{pmatrix}.$$

Figure 2 shows that the regression functions in Equations (20) and (21) both provide relatively good fits for the data.

Note that Equations (20) and (21) reproduce earlier results for $N = 2$ given, for example, by Viana (1998) and Olkin and Viana (1995). Proposition 3.3 may also be used to obtain the best predictors of X and of $Y_{(N)}$ based on $Y_{(N)}$ and on X, respectively, in the case of a joint t-distribution for X and $\mathbf{Y}$.

4. Conclusion

In this paper, we derived general results on joint distributions and prediction for a $K \times 1$ vector $\mathbf{X}$ and a vector of L-statistics $\mathbf{L}\mathbf{Y}_{(N)}$ (i.e., an affine transformation of the vector of order statistics $\mathbf{Y}_{(N)}$) of an $N \times 1$ vector $\mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are assumed to have a joint multivariate elliptical distribution. The results involving multivariate skew-elliptical distributions are general enough to include several previous results as special cases; see, e.g., Jamalizadeh and Balakrishnan (2009b) and Balakrishnan et al. (2012). By considering elliptical distributions, which include normal as well as t-distributions, we provide a robust alternative to conventional formulations based on normality. Note that while we paid particular attention to the special cases of normal and t-distributions, other elliptical distributions such as the Laplace and slash distributions may be considered as well. We however defer these other cases for future work. In addition, the results in this paper can be easily extended to SUE distributions. For example, the results can be easily extended to multivariate skew-elliptical distributions introduced by Branco and Dey (2001), as a referee suggested (since the multivariate skew-elliptical distributions presented by these authors are special cases of SUN distributions). Note that our results give the joint distribution of $\mathbf{X}$ and $\mathbf{LY}_{(N)}$ as $N!$-component mixtures. In practice, evaluating such distributions can become computationally infeasible, especially when N is large. Proposition 3.3 gives a computationally more efficient alternative form that involves a smaller number of mixture components but only for the special case $K = 1$ and $\mathbf{LY}_{(N)} = Y_{(r)}$, $r = 1, \ldots, N$. An analogous result in the general case should be useful in practice. Alternatively, accurate approximations may be obtained to alleviate the computational demands of evaluating the exact distributions. This would be the subject of future work as well.

Acknowledgements

This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. We thank Beilei Wu for computational assistance. We are also grateful to an Associate Editor and a referee for their constructive comments that helped improve the paper.

References

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

Arellano-Valle, R.B., Genton, M.G., 2005. On fundamental skew distributions. Journal of Multivariate Analysis, 96, 93–116.

Arellano-Valle, R.B., 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.

Arellano-Valle, R.B., Genton, M.G., 2008. On the exact distribution of the maximum of absolutely continuous dependent random variables. Statistics and Probability Letters, 78, 27–35.

Arellano-Valle, R.B., Genton, M.G., 2010a. Multivariate unified skew-elliptical distributions. Chilean Journal of Statistics, 1, 17–33.

Arellano-Valle, R.B., Genton, M.G., 2010b. Multivariate extended t distributions and related families. Metron, 68, 201–234.

Balakrishnan, N., 1993. Multivariate normal distributions and multivariate order statistics induced by ordering linear combinations. Statistics and Probability Letters, 17, 343–350.

Balakrishnan, N., Ong, S.H., Habibi, M., Jamalizadeh, A., 2012. Nonlinear regression in a trivariate elliptical distribution via a linear combination of order statistics. Statistics, 46, 605–619.

Basu, A.P., Ghosh, J.K., 1978. Identifiability of the multinormal and other distributions under competing risks model. Journal of Multivariate Analysis, 8, 413–429.

Branco, M.D., Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis, 79, 99–113.

Genc, A., 2006. A note on the distribution of linear functions of two ordered correlated random variables. Communications in Statistics - Theory and Methods, 35, 989–995.

Gupta, S.S., Pillai, K.C.S., 1965. On linear functions of ordered correlated normal random variables. Biometrika, 52, 367–379.

Jamalizadeh, A., Balakrishnan, N., 2008. On order statistics from bivariate skew-normal and skew-t distributions. Journal of Statistical Planning and Inference, 138, 4187–4197.

Jamalizadeh, A., Balakrishnan, N., 2009a. Prediction in a trivariate normal distribution via a linear combination of order statistics. Statistics and Probability Letters, 79, 2289–2296.

Jamalizadeh, A., Balakrishnan, N., 2009b. Order statistics from trivariate normal and t_ν-distributions in terms of generalized skew-normal and skew-t_ν distributions. Journal of Statistical Planning and Inference, 139, 3799–3819.

Jamalizadeh, A., Balakrishnan, N., 2010. Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of unified skew-elliptical distributions. Journal of Multivariate Analysis, 101, 1412–1427.

Jamalizadeh, A., Balakrishnan, N., 2012. Concomitants of order statistics from multivariate elliptical distributions. Journal of Statistical Planning and Inference, 142, 397–409.

Jamalizadeh, A., Khosravi, M., Balakrishnan, N., 2009a. Recurrence relations for distributions of a skew-t and a linear combination of order statistics from a bivariate-t. Computational Statistics and Data Analysis, 53, 847–852.

Jamalizadeh, A., Mehrali, Y., Balakrishnan, N., 2009b. Recurrence relations for bivariate t and extended skew-t distributions and an application to order statistics from bivariate-t. Computational Statistics and Data Analysis, 53, 4018–4027.

Loperfido, N., 2008a. A note on skew-elliptical distributions and linear functions of order statistics. Statistics and Probability Letters, 78, 3184–3186.

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

Nagaraja, H.N., 1982. A note on linear functions of ordered correlated normal random variables. Biometrika, 69, 284–285.

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

Viana, M., 1998. Linear combinations of ordered symmetric observations with applications to visual acuity. In Balakrishnan, N., Rao, C.R., (eds.). Handbook of Statistics–Order Statistics: Applications. North Holland, Amsterdam, pp. 513–524.

Wiens, D.P., Beaulieu, N.C., Loskot, P., 2006. On the exact distribution of the sum of the largest $n-k$ out of n normal random variables with differing means. Statistics, 40, 165–173.