Accepted Manuscript

Scale and shape mixtures of multivariate skew-normal distributions

Reinaldo B. Arellano-Valle, Clécio S. Ferreira, Marc G. Genton

PII: S0047-259X(17)30393-7
DOI: https://doi.org/10.1016/j.jmva.2018.02.007
Reference: YJMVA 4329

To appear in: *Journal of Multivariate Analysis*

Received date : 24 June 2017

Scale and shape mixtures of multivariate skew-normal distributions

Reinaldo B. Arellano-Valle[a], Clécio S. Ferreira[b], Marc G. Genton[c,*]

[a]*Department of Statistics, Pontificia Universidad Católica de Chile, Santiago 22, Chile*
[b]*Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora 36036-900, Brazil*
[c]*Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*

Abstract

We introduce a broad and flexible class of multivariate distributions obtained by both scale and shape mixtures of multivariate skew-normal distributions. We present the probabilistic properties of this family of distributions in detail and lay down the theoretical foundations for subsequent inference with this model. In particular, we study linear transformations, marginal distributions, selection representations, stochastic representations and hierarchical representations. We also describe an EM-type algorithm for maximum likelihood estimation of the parameters of the model and demonstrate its implementation on a wind dataset. Our family of multivariate distributions unifies and extends many existing models of the literature that can be seen as submodels of our proposal.

Keywords: EM-algorithm, Scale mixtures of normal distributions, Scale mixtures of skew-normal distributions, Shape mixtures of skew-normal distributions, Skew-normal distribution, Skew scale mixtures of normal distributions.

1. Introduction

In recent years, the use of the multivariate skew-normal (SN) distribution [13, 14] in both theoretical and applied studies has been increasingly popular. In several of these studies, the location-scale version of the multivariate SN distribution has been implemented using different parametrizations but equivalent to the original one [3]; see also the books by Genton [22] and Azzalini and Capitanio [16]. In this paper, we consider the multivariate SN version with the parametrization proposed by Arellano-Valle and Genton [9]. According to these authors, a p-dimensional random vector $\mathbf{Y}$ follows a multivariate SN distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale parameter $\boldsymbol{\Sigma} > 0$ (a $p \times p$ positive definite matrix) and skewness/shape parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$, denoted by $\mathbf{Y} \sim \mathcal{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its probability density function (pdf) is given, for all $\mathbf{y} \in \mathbb{R}^p$, by

$$f_{\rm SN}(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1\{\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\}, \tag{1}$$

where $\phi_p(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2}\phi_p(\mathbf{z})$, with $\phi_p(\mathbf{z}) = (2\pi)^{-p/2}\exp(-\mathbf{z}^\top\mathbf{z}/2)$ and $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$, is the pdf of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, denoted by $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Φ_1 denotes the cumulative distribution function (cdf) of the standard normal distribution $\mathcal{N}_1(0, 1)$, and $\boldsymbol{\Sigma}^{-1/2}$ is the symmetric square root matrix of $\boldsymbol{\Sigma}^{-1}$. When $\boldsymbol{\lambda} = \mathbf{0}$, the SN distribution reduces to the multivariate normal distribution, viz. $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Also, in terms of the "standardized" random vector $\mathbf{Z} \sim \mathcal{SN}_p(\boldsymbol{\lambda}) \equiv \mathcal{SN}_p(\mathbf{0}, \mathbf{I}_p, \boldsymbol{\lambda})$, where $\mathbf{I}_p$ is the identity matrix of dimension $p \times p$, the multivariate SN distribution can be represented stochastically as follows:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{Z}, \quad \text{with} \quad \mathbf{Z} \stackrel{d}{=} \boldsymbol{\delta}|Z_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{Z}_1, \tag{2}$$

where "$\stackrel{d}{=}$" means "equal in distribution", $\boldsymbol{\delta} = \boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$, $|Z_0|$ denotes the absolute value of Z_0, and $Z_0 \sim \mathcal{N}_1(0, 1)$ is independent of $\mathbf{Z}_1 \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$. This representation is very useful to derive most of the main properties of the multivariate SN distribution. It is equivalent to the following hierarchical representation as location mixture of the multivariate normal distribution, which is of great utility in the formulation of the SN statistical model

$$\mathbf{Y} \mid U = u \sim \mathcal{N}_p(\boldsymbol{\mu} + \bar{\boldsymbol{\lambda}}u, \boldsymbol{\Psi}), \tag{3}$$

*Corresponding author
Email address: marc.genton@kaust.edu.sa (Marc G. Genton)