Statistics and Probability Letters 117 (2016) 40–45

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A derivation of the multivariate singular skew-normal density function

Phil D. Young[a,*], Jane L. Harvill[b], Dean M. Young[b]

[a] One Bear Place #98005, Department of Management and Information Systems, Baylor University, Waco, TX 76798-7140, United States

[b] One Bear Place #97140, Department of Statistical Science, Baylor University, Waco, TX 76798-7140, United States

ARTICLE INFO

Article history:
Received 4 January 2016
Received in revised form 29 April 2016
Accepted 30 April 2016
Available online 6 May 2016

Keywords:
Moment generating function
Pseudoinverse
Affine subspace
Lebesgue measure

ABSTRACT

We prove the existence of a multivariate singular skew-normal density function, derive its moment generating function, and demonstrate that the skewness parameter-vector is confined to the column space of the singular dispersion matrix.

1. Introduction

Multivariate data that do not follow a normal distribution arise naturally in a wide variety of fields. Disciplines such as economics, oceanography, engineering, and the biomedical sciences all produce such data. The skew-normal distribution has statistical relevance in GARCH modelling, spatial statistics, and Bayesian statistics, to name a few.

Here, we derive the multivariate singular skew-normal (*MSSN*) probability density function (*PDF*). We also establish the moment generating function (*MGF*) of the *MSSN* random variable and utilize it to find the mean vector and covariance matrix of a *MSSN* random vector. Furthermore, we show that the skewness parameter must be contained in the column space of the dispersion parameter. To our knowledge, this fact has not been previously stated or verified for the singular case.

We use the following notation throughout the remainder of the paper. Here, $\mathbb{R}_{m\times n}$ represents the vector space of all $m \times n$ matrices over the real field $\mathbb{R}$, $\mathbb{R}_n$ represents the space of all $n \times n$ vectors over $\mathbb{R}$, $\mathbb{R}^n$ represents the vector space of all $n \times 1$ matrices over $\mathbb{R}$, $\mathbb{R}_n^S$ represents the cone of all $n \times n$ symmetric matrices of $\mathbb{R}_n$, and $\mathbb{R}_n^{\geq}$ and $\mathbb{R}_n^{>}$ represent the cone of all symmetric nonnegative-definite and positive-definite matrices in $\mathbb{R}_n$, respectively. For the matrix $\mathbf{A} \in \mathbb{R}_{m\times n}$, $\mathcal{N}(\mathbf{A})$ denotes the null space and $\mathcal{C}(\mathbf{A})$ the column space. The Moore–Penrose pseudoinverse of $\mathbf{A}$ is represented by $\mathbf{A}^+$. Finally, letting $\lambda_i(\mathbf{A})$ denote the ith eigenvalue of $\mathbf{A} \in \mathbb{R}_n^S$, we denote the product of the r non-zero eigenvalues by

$$\det{}_k\mathbf{A} \equiv \prod_{i=1}^{k} \lambda_i(\mathbf{A}), \tag{1}$$

where $\operatorname{rank}(\mathbf{A}) = k < n$.

* Corresponding author.
E-mail address: philyoung36@gmail.com (P.D. Young).

http://dx.doi.org/10.1016/j.spl.2016.04.024
0167-7152/

The remainder of the paper is organized as follows. In Section 2, we discuss the multivariate skew-normal (*MSN*) and multivariate singular normal (*SMN*) distributions. In Section 3, we determine the *MGF* of a multivariate singular skew-normal random vector $\mathbf{x} \in \mathbb{R}^p$. We then utilize the *MSSN MGF* to determine the *MSSN PDF* as well as the mean and covariance matrix. Furthermore, we demonstrate that the skewness parameter vector must be contained in the column space of the dispersion matrix.

2. Preliminaries

2.1. The multivariate skew-normal distribution

The extension of the univariate normal density function family to one that incorporates truncation through a population selection mechanism was first brought about by de Helguero (1908). One can find a discussion concerning the importance of this research in Azzalini and Regoli (2012). However, the first appearance of a density function equivalent to the univariate skew-normal density function appeared in Birnbaum (1950), though the objective of this development was not as an extension to the normal density function. Other early occurrences of density functions that are equivalent to the skew-normal density include (Roberts, 1966; O'Hagan and Leonard, 1976; Aigner et al., 1977). The present-day univariate skew-normal distribution was developed by Azzalini (1985) as an extension of the normal density function family and further developed by Azzalini (1986) to incorporate an additional shape parameter.

Azzalini and Dalla Valle (1996) first introduced the *MSN* distribution. Numerous authors have studied the *MSN* and its applications, including Arnold et al. (2002), Liseo and Loperfido (2003), Azzalini (2005), Vernic (2005), Vernic (2006), Arellano-Valle and Azzalini (2006) and Arellano-Valle and Azzalini (2008).

In particular, Azzalini and Capitanio (1999) have defined the *MSN* distribution, provided linear and quadratic forms, discussed the location and scale parameters, and identified several applications of the multivariate skew-normal distribution. Loperfido (2010) has determined linear functions that maximize skewness and kurtosis and proposed measures of shape and non-normality, which are functions of the skewness parameter. Also, Loperfido (2004) has proposed the generalized skew-normal distribution, which generalizes the *MSN* distribution. A nice set of applications of the skew-normal and *MSN* distributions is provided in Azzalini and Capitanio (1999) and Genton (2004) as well as other papers.

Here, we use the following definition of the *MSN PDF*, given by Vernic (2006).

Definition 1. A random vector $\mathbf{x} \in \mathbb{R}^p$ has a *multivariate skew-normal* distribution with skewness parameter $\boldsymbol{\gamma} \in \mathbb{R}^p$, written $\mathbf{x} \sim MSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \delta_0, \boldsymbol{\gamma})$, if its density function is

$$p(\mathbf{x}) = \frac{1}{\Phi(\delta_0)} \phi_p(\mathbf{x};\ \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \Phi\left(\frac{\delta_0 + \boldsymbol{\gamma}' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{1 - \boldsymbol{\gamma}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}}} \right), \tag{2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathbb{R}_p^{>}$ such that $\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} < 1$, δ_0 is a real number,

$$\phi_p(\mathbf{x};\ \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \tag{3}$$

is the multivariate normal *PDF*, and $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function.

One attractive property of the *MSN* distribution is that when the skewness parameter $\boldsymbol{\gamma} = \mathbf{0}$, then (2) reduces to (3). Otherwise, the parameter vector $\boldsymbol{\gamma}$ regulates the skewness. Authors such as Arnold et al. (2002) and Azzalini and Capitanio (1999) have chosen $\delta_0 = 0$ so that (2) simplifies to

$$q(\mathbf{x}) = 2\phi_p(\mathbf{x};\ \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \Phi\left(\frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\sqrt{1 - \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}}} \right).$$

The *MSN* family has properties similar to those of the normal family with two exceptions: it lacks closure for the joint distribution of independent members of the *MSN* family and closure under linear combinations other than those given by nonsingular matrices.

2.2. The singular multivariate normal distribution

A well-known fact is that the *PDF* of a normally distributed random vector with singular covariance matrix does not exist with respect to the Lebesgue measure on $\mathbb{R}^p$. However, Khatri (1968) has shown that a multivariate normal *PDF* exists on a subspace of $\mathbb{R}^p$. The author (van Perlo-ten Kleij, 2004) has defined the singular multivariate normal distribution as follows.

Definition 2. Let $\mathbf{x}$ be a random vector with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathbb{R}_p^{\geq}$ such that $\operatorname{rank}(\boldsymbol{\Sigma}) = k < p$. Then, $\mathbf{x}$ is said to have a *singular multivariate normal* distribution, written $\mathbf{x} \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if the *PDF* of $\mathbf{x}$ is

$$\phi_p^S(\mathbf{x};\ \boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv (2\pi)^{-p/2} [\det_k \boldsymbol{\Sigma}]^{1/2} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{+} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \tag{4}$$