STAPRO: 8316 Model 3G pp. 1–7 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Three skewed matrix variate distributions

Michael P.B. Gallaugher, Paul D. McNicholas *

Department of Mathematics & Statistics, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 3 August 2018
Received in revised form 13 August 2018
Accepted 19 August 2018
Available online xxxx

Keywords:
Matrix variate
Skewed distributions

ABSTRACT

Three-way data can be conveniently modelled by using matrix variate distributions. Although there has been a lot of work for the matrix variate normal distribution, there is little work in the area of matrix skew distributions. Three matrix variate distributions that incorporate skewness, as well as other flexible properties such as concentration, are discussed. Equivalences to multivariate analogues are presented, and moment generating functions are derived. Maximum likelihood parameter estimation is discussed, and simulated data is used for illustration.

1. Introduction

Matrix variate distributions are useful in modelling three way data, e.g., multivariate longitudinal data. Although the matrix normal distribution is widely used, there is relative paucity in the area of matrix skewed distributions. Herein, we discuss matrix variate extensions of three already well established multivariate distributions using matrix normal variance–mean mixtures. Specifically, we consider a matrix variate generalized hyperbolic distribution, a matrix variate variance–gamma distribution, and a matrix variate normal inverse Gaussian (NIG) distribution. Along with the matrix variate skew-t distribution, mixtures of these respective distributions have been used for clustering (Gallaugher and McNicholas, 2018); however, unlike the matrix variate skew-t distribution (Gallaugher and McNicholas, 2017), their properties have yet to be discussed and this letter aims to fill that gap.

2. Background

2.1. The matrix variate normal and related distributions

One of the most mathematically tractable examples of a matrix variate distribution is the matrix variate normal distribution. An $n \times p$ random matrix $\mathscr{X}$ follows a matrix variate normal distribution if its probability density function can be written as

$$f(\mathbf{X}|\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{p}{2}}|\boldsymbol{\Psi}|^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\boldsymbol{\Psi}^{-1}(\mathbf{X}-\mathbf{M})'\right)\right\},$$

where $\mathbf{M}$ is an $n \times p$ location matrix, $\boldsymbol{\Sigma}$ is an $n \times n$ scale matrix for the rows of $\mathscr{X}$ and $\boldsymbol{\Psi}$ is a $p \times p$ scale matrix for the columns of $\mathscr{X}$. We denote this distribution by $\mathcal{N}_{n\times p}(\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ and, for notational clarity, we will denote the random matrix by $\mathscr{X}$ and its realization by $\mathbf{X}$. One useful property of the matrix variate normal distribution, as given in Harrar and Gupta (2008), is

$$\mathscr{X} \sim \mathcal{N}_{n\times p}(\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) \iff \mathrm{vec}(\mathscr{X}) \sim \mathcal{N}_{np}(\mathrm{vec}(\mathbf{M}), \boldsymbol{\Psi}\otimes\boldsymbol{\Sigma}), \tag{1}$$

* Corresponding author.
E-mail address: paul@math.mcmaster.ca (P.D. McNicholas).

https://doi.org/10.1016/j.spl.2018.08.012

Please cite this article in press as: Gallaugher M.P.B., McNicholas P.D., Three skewed matrix variate distributions. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.08.012.

where $\mathcal{N}_{np}(\cdot)$ denotes the multivariate normal distribution with dimension np, $\mathrm{vec}(\cdot)$ denotes the vectorization operator, and $\otimes$ denotes the Kronecker product.

Although the matrix variate normal is probably the most well known matrix variate distribution, there are other examples. For example, the Wishart distribution (Wishart, 1928) was shown to be the distribution of the sample covariance matrix for a random sample from a multivariate normal distribution. There are also a few examples of a matrix variate skew normal distribution such as Chen and Gupta (2005), Domínguez-Molina et al. (2007) and Harrar and Gupta (2008). Most recently, Gallaugher and McNicholas (2017), considered a matrix variate skew-t distribution using a matrix normal variance–mean mixture.

There are also a few examples of mixtures of matrix variate distributions. Anderlucci and Viroli (2015) considered a mixture of matrix variate normal distributions for clustering multivariate longitudinal data and Doğru et al. (2016) considered a mixture of matrix variate t distributions.

2.2. The inverse and generalized inverse Gaussian distributions

The derivation of the matrix distributions and parameter estimation discussed in Section 3, will rely heavily on the generalized inverse Gaussian distribution, and to a lesser extent the inverse Gaussian distribution. A random variable Y follows an inverse Gaussian distribution if its probability density function is of the form

$$f(y|\delta,\gamma)=\frac{\delta}{\sqrt{2\pi}}\exp\{\delta\gamma\}y^{-\frac{3}{2}}\exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{y}+\gamma^2y\right)\right\}$$

for $y>0$, where $\delta,\gamma>0$. For notational purposes, we will denote this distribution by $\mathrm{IG}(\delta,\gamma)$.

The generalized inverse Gaussian distribution has two different parameterizations, both of which will be useful. A random variable Y has a generalized inverse Gaussian distribution parameterized by $a,b>0$ and $\lambda\in\mathbb{R}$, denoted by $\mathrm{GIG}(a,b,\lambda)$, if its probability density function can be written as

$$f(y|a,b,\lambda)=\frac{(a/b)^{\frac{\lambda}{2}}y^{\lambda-1}}{2K_\lambda(\sqrt{ab})}\exp\left\{-\frac{ay+b/y}{2}\right\}$$

for $y>0$, where

$$K_\lambda(u)=\frac{1}{2}\int_0^\infty z^{\lambda-1}\exp\left\{-\frac{u}{2}\left(z+\frac{1}{z}\right)\right\}dz$$

is the modified Bessel function of the third kind with index λ. Some expectations of functions of a GIG random variable with this parameterization have a mathematically tractable form, e.g.,

$$\mathbb{E}(Y)=\sqrt{\frac{b}{a}}\frac{K_{\lambda+1}(\sqrt{ab})}{K_\lambda(\sqrt{ab})},\qquad \mathbb{E}\left(1/Y\right)=\sqrt{\frac{a}{b}}\frac{K_{\lambda+1}(\sqrt{ab})}{K_\lambda(\sqrt{ab})}-\frac{2\lambda}{b},$$
$$\mathbb{E}(\log Y)=\log\left(\sqrt{\frac{b}{a}}\right)+\frac{1}{K_\lambda(\sqrt{ab})}\frac{\partial}{\partial\lambda}K_\lambda(\sqrt{ab}).$$

Although this parameterization of the GIG distribution will be useful for parameter estimation, for the purposes of deriving the density of the matrix variate generalized hyperbolic distribution, it is more useful to take the parameterization

$$g(y|\omega,\eta,\lambda)=\frac{(w/\eta)^{\lambda-1}}{2\eta K_\lambda(\omega)}\exp\left\{-\frac{\omega}{2}\left(\frac{w}{\eta}+\frac{\eta}{w}\right)\right\},\tag{2}$$

where $\omega=\sqrt{ab}$ and $\eta=\sqrt{a/b}$. For notational clarity, we will denote the parameterization given in (2) by $\mathrm{I}(\omega,\eta,\lambda)$.

2.3. Variance–mean mixtures

A p-variate random vector $\mathbf{X}$ defined in terms of a variance–mean mixture, has a probability density function of the form

$$f(\mathbf{x})=\int_0^\infty\phi_p(\mathbf{x}|\boldsymbol{\mu}+w\boldsymbol{\alpha},w\boldsymbol{\Sigma})h(w|\boldsymbol{\theta})dw,$$

where the random variable $W>0$ has density function $h(w|\boldsymbol{\theta})$, and $\phi_p(\cdot)$ represents the density function of the p-variate Gaussian distribution. This representation is equivalent to writing

$$\mathbf{X}=\boldsymbol{\mu}+W\boldsymbol{\alpha}+\sqrt{W}\mathbf{V},\tag{3}$$

where $\boldsymbol{\mu}$ is a location parameter, $\boldsymbol{\alpha}$ is the skewness, $\mathbf{V}\sim\mathcal{N}_p(\mathbf{0},\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ as the scale matrix, and W has density function $h(w|\boldsymbol{\theta})$. Note that W and $\mathbf{V}$ are independent. Many multivariate distributions can be obtained through a variance–mean mixture by changing the distribution of W. For example, the p-dimensional generalized hyperbolic distribution,