An EM algorithm for fitting matrix-variate skew-normal distributions on interval-censored and missing data

Victor H. Lachos- University of Connecticut, USA
(Joint work with Carlos Diniz, Mauricio Castro and Atila Prates)

January 09, 2026

Skew 2026
Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Padova, Italy

Skew 2026

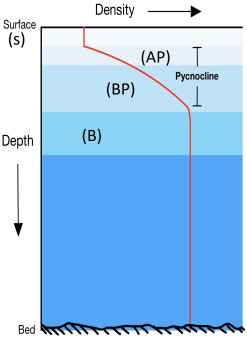
Outline

- Motivating example
- The matrix-variate normal (MVN) distribution
- The matrix-variate skew-normal (MVSN) and extended skew-normal (MVESN) distributions
- MVSN distribution for censoring and missing data
- The EM algorithm for ML estimation
- Simulation results
- Application
- Concluding remarks

Skew 2026

Motivating data - Water Quality

Water Quality: dataCensored dataset contained in R package **baytrends** (Murphy et al. 2023*)



- **8 Stations:** CB3.3C, CB4.1C, EE2.1, EE3.0, LE2.2, TF2.2, CB5.4, TF5.5
- **4 layers**: S, AP, BP, B
- **12 water quality variables**: Salinity (sal), Total Nitrogen (tn),…, Ammonium(nh4), Nitrite + Nitrate (no23)
- Measurement taken **every month** (1985-2016), contain **missing and interval-censored** data
- For each station, we have **n** matrices (12x4) of the form:

occasion 1

	S	AP	BP	B
sal	0.002	0.003	0.009	0.014
tn	[0, 0.05]	NA	0.112	[0.01, 0.017]
.	.	.	.	.
.	.	.	.	.
nh4	0.018	NA	0.058	0.088
no23	[0, 0.05]	0.020	0.058	NA

………

occasion n

	S	AP	BP	B
sal	0.005	NA	0.015	0.018
tn	NA	[0, 0.09]	0.122	[0.01, 0.23]
.	.	.	.	.
.	.	.	.	.
nh4	0.011	NA	NA	0.088
no23	[0, 0.03]	0.020	0.023	NA

*Murphy, R., E. Perry, J. Keisman, J. Harcum, and E. W. Leppo (2023). baytrends: Long Term Water Quality Trend Analysis. R package version 2.0.9.

Matrix Variate Normal (MVN) Distribution

MVN Distribution

An $p \times q$ random matrix $\boldsymbol{\mathcal{X}}$ follows an MVN distribution with mean parameter M ($p \times q$) and covariance matrices $\boldsymbol{\Sigma}$ ($p \times p$) and $\boldsymbol{\Psi}$ ($q \times q$), respectively, if its pdf is

$$\phi_{p\times q}(\mathcal{X}|\mathsf{M},\boldsymbol{\Sigma},\boldsymbol{\Psi}) = \frac{1}{(2\pi)^{pq/2}|\boldsymbol{\Psi}|^{p/2}|\boldsymbol{\Sigma}|^{q/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}[\boldsymbol{\Psi}^{-1}(\mathcal{X}-\mathsf{M})^\top\boldsymbol{\Sigma}^{-1}(\mathcal{X}-\mathsf{M})]\right\}.$$

Notation: $\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{p\times q}(\mathsf{M},\boldsymbol{\Sigma},\boldsymbol{\Psi})$.

Equivalent Definition:

$$\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{p\times q}(\mathsf{M},\boldsymbol{\Sigma},\boldsymbol{\Psi}) \iff \mathrm{vec}(\boldsymbol{\mathcal{X}}) = \mathsf{X} \sim \mathcal{N}_{pq}(\boldsymbol{\mu} = \mathrm{vec}(\mathsf{M}), \boldsymbol{\Lambda} = \boldsymbol{\Psi}\otimes\boldsymbol{\Sigma}).$$

where $\mathcal{N}_{pq}(\boldsymbol{\mu},\boldsymbol{\Lambda})$ denotes the MN distribution, $\mathrm{vec}(\cdot)$ is the vectorization operator and $\otimes$ denotes the Kronecker product.

It follows from this equivalence that

$$\mathbb{E}(\boldsymbol{\mathcal{X}}) = \mathsf{M} \text{ and } \mathrm{Var}(\mathrm{vec}(\boldsymbol{\mathcal{X}})) = \boldsymbol{\Psi}\otimes\boldsymbol{\Sigma}.$$

Skew 2026

MVN Distribution

- From Gupta and Nagar (1999, Theorem 2.3.5), we have that

$$C_{\text{row}} = \mathbb{E}[(\boldsymbol{\mathcal{X}} - \mathbb{E}(\boldsymbol{\mathcal{X}}))(\boldsymbol{\mathcal{X}} - \mathbb{E}(\boldsymbol{\mathcal{X}}))^\top] = \text{tr}(\boldsymbol{\Psi})\boldsymbol{\Sigma},$$
$$C_{\text{col}} = \mathbb{E}[(\boldsymbol{\mathcal{X}} - \mathbb{E}(\boldsymbol{\mathcal{X}}))^\top(\boldsymbol{\mathcal{X}} - \mathbb{E}(\boldsymbol{\mathcal{X}}))] = \text{tr}(\boldsymbol{\Sigma})\boldsymbol{\Psi},$$

  where the $p \times p$ matrix C_{row} represents the total variability among rows, and the $q \times q$ matrix C_{col} represents the total variability among columns.
- If $\{\boldsymbol{\mathcal{X}}_i,\ i = 1, \ldots, n\}$, are i.i.d random matrices from $N_{p \times q}(\mathsf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, then M, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Psi}$ have ML estimators given by

$$\widehat{\mathsf{M}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\mathcal{X}}_i,\ \ \widehat{\boldsymbol{\Psi}} = \frac{1}{np}\sum_{i=1}^{n}(\boldsymbol{\mathcal{X}}_i - \widehat{\mathsf{M}})^\top \widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\mathcal{X}}_i - \widehat{\mathsf{M}}) \text{ and } \widehat{\boldsymbol{\Sigma}} = \frac{1}{nq}\sum_{i=1}^{n}(\boldsymbol{\mathcal{X}}_i - \widehat{\mathsf{M}})\widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{\mathcal{X}}_i - \widehat{\mathsf{M}})^\top.$$

Must be computed iteratively until convergence!

MVN Distribution

Handling non-identifiability

If $\boldsymbol{\Sigma}^* = a\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}^* = a^{-1}\boldsymbol{\Psi}$, then $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma} = \boldsymbol{\Psi}^* \otimes \boldsymbol{\Sigma}^*$, so both estimates yield the same overall covariance matrix.

1. $(\boldsymbol{\Psi})_{11} = (\boldsymbol{\Sigma})_{11} = 1$ and add an scale parameter σ^2. (Glanz and Carvalho (2018))
2. Imposing the constraint $|\boldsymbol{\Psi}| = 1$. (Tomarchio et al. (2021))

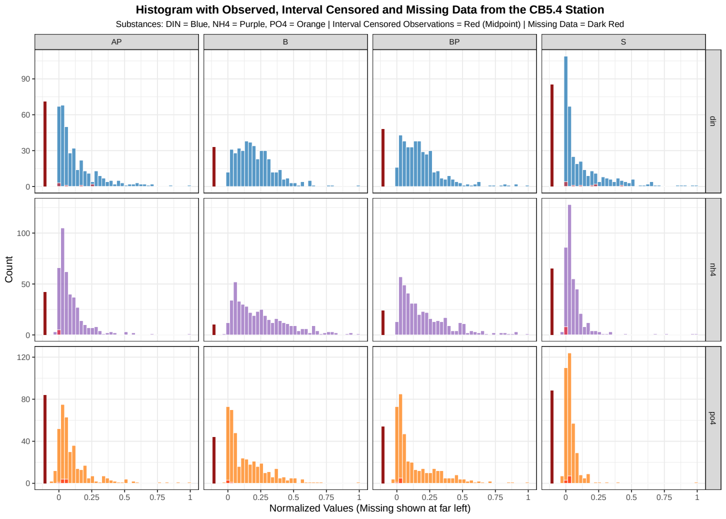
Some R packages: matrixNormal, MixMatrix, MatrixMixtures.

Benefits over Vectorization

A common strategy: Vectorization of the matrix-variate data (then use standard techniques). Here is a summary of the issues induced by such an approach.

1. **Interpretability**: the identification of the two sources of variability, governed by $\mathbf{\Sigma}$ and $\mathbf{\Psi}$, would not be possible if they were combined into a unique $\mathbf{\Lambda} = \mathbf{\Psi} \otimes \mathbf{\Sigma}$ matrix. This leads to a loss of interpretation of the data.
2. **Parsimony**: the number of free covariance parameters in the matrix-variate setting is $p(p+1)/2 + q(q+1)/2 - 1$. When vectorizing ($\mathbf{\Lambda}$), this number rises to $pq(pq+1)/2$, potentially resulting in an overparameterization of the multivariate models.
3. **Model selection**: strictly connected to the previous problem, the increase in the number of parameters in the multivariate setting can introduce challenges in model selection. This is a consequence of the increased weight assigned to the penalty term of widely used information criteria.

Motivating data - Water Quality



Dissolved Inorganic Nitrogen (DIN), Ammonium ($\mathrm{NH_4}$) and orthophosphorus ($\mathrm{PO_4}$).

Additional complications

- MVN distribution has been extensively investigated, but in practice, the distribution of the response variable is not normal.
- Another complication with these kinds of data arises when measures depart significantly from normality, for instance, high kurtosis (due to heavy tails or outliers) or skewness.
- In the multivariate context, the multivariate skew-normal (SN) distribution (Azallini and Dalla Valle, 1996; Arellano-Valle et al. 2005; Lin and Lee, 2008) is an appealing alternative.

Multivariate skew normal (SN) distribution

Let $W \sim \text{HN}(0,1)$, $\mathsf{X} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Delta})$, $W \perp \mathsf{X}$. Define

$$\mathsf{Y} = \boldsymbol{\mu} + W\mathsf{b} + \mathsf{X},$$

where $\boldsymbol{\mu}, \mathsf{b} \in \mathbb{R}^p$, and $\boldsymbol{\Delta} \in \mathbb{R}^{p \times p}$. Then the pdf of $\mathsf{Y} \in \mathbb{R}^p$ is given by

$$\begin{aligned} f_{\text{SN}}(\mathsf{y}) &= 2\phi_p(\mathsf{y} \mid \boldsymbol{\mu},\ \boldsymbol{\Delta} + \mathsf{b}\mathsf{b}^\top)\Phi\left(\frac{(\mathsf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Delta}^{-1}\mathsf{b}}{\sqrt{1+\mathsf{b}^\top\boldsymbol{\Delta}^{-1}\mathsf{b}}}\right), \\ &= 2\phi_p(\mathsf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Delta})\frac{1}{\sqrt{1+\mathsf{b}^\top\boldsymbol{\Delta}^{-1}\mathsf{b}}} \exp\left(\frac{1}{2}\frac{\left[(\mathsf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Delta}^{-1}\mathsf{b}\right]^2}{1+\mathsf{b}^\top\boldsymbol{\Delta}^{-1}\mathsf{b}}\right)\Phi\left(\frac{(\mathsf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Delta}^{-1}\mathsf{b}}{\sqrt{1+\mathsf{b}^\top\boldsymbol{\Delta}^{-1}\mathsf{b}}}\right), \end{aligned}$$

Notation: $\mathsf{Y} \sim \mathcal{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Delta}, \mathsf{b})$. See Bolfarine, et al. (2007).

Widely used reparameterization (Arellano-Valle el al. (2005))

$$\mathsf{Y} \sim \mathcal{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) \iff \boldsymbol{\lambda} = \frac{(\boldsymbol{\Delta} + \mathsf{b}\mathsf{b}^\top)^{-1/2}\mathsf{b}}{[1-\mathsf{b}^\top(\boldsymbol{\Delta}+\mathsf{b}\mathsf{b}^\top)^{-1}\mathsf{b}]^{1/2}}, \boldsymbol{\Sigma} = \boldsymbol{\Delta} + \mathsf{b}\mathsf{b}^\top.$$

with PDF: $\phi_{SN}(\mathsf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_p(\mathsf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathsf{y}-\boldsymbol{\mu}))$.

Multivariate extended SN (ESN) distribution

Let $W \sim \text{TN}_1(0,1;[-\widetilde{\kappa},\infty))$, $\mathbf{X} \sim \mathcal{N}_p(\mathbf{0},\boldsymbol{\Delta})$, $W \perp \mathbf{X}$. Define

$$\mathbf{Y} = \boldsymbol{\mu} + W\mathbf{b} + \mathbf{X},$$

where $\widetilde{\kappa} = \kappa/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, $\boldsymbol{\Omega} = \boldsymbol{\Delta} + \mathbf{b}\mathbf{b}^\top$, and $\mathbf{b} = \boldsymbol{\Omega}^{1/2}\boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$..
Then the pdf of $\mathbf{Y} \in \mathbb{R}^p$ is given by

$$f_{\text{ESN}}(\mathbf{y}) = \frac{1}{\Phi\left(\frac{\kappa}{\sqrt{1+\delta^2}}\right)}\frac{1}{\sqrt{1+\delta^2}}\,\phi_p(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Delta})\exp\left(\frac{1}{2}\frac{\left(\mathbf{b}^\top\boldsymbol{\Delta}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)^2}{1+\delta^2}\right)\Phi\left(\kappa+\frac{\mathbf{b}^\top\boldsymbol{\Delta}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{\sqrt{1+\delta^2}}\right) \tag{1}$$

$\delta^2 = \mathbf{b}^\top\boldsymbol{\Delta}^{-1}\mathbf{b}$. Notation: $\mathbf{Y} \sim \mathcal{ESN}_p(\boldsymbol{\mu},\boldsymbol{\Delta},\mathbf{b},\kappa)$.

Equivalent Definition: See Galarza et al. (2022).

$$\mathbf{Y} \sim \mathcal{ESN}_p(\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{\lambda},\kappa) \iff \boldsymbol{\lambda} = \frac{(\boldsymbol{\Delta}+\mathbf{b}\mathbf{b}^\top)^{-1/2}\mathbf{b}}{[1-\mathbf{b}^\top(\boldsymbol{\Delta}+\mathbf{b}\mathbf{b}^\top)^{-1}\mathbf{b}]^{1/2}}, \boldsymbol{\Omega} = \boldsymbol{\Delta}+\mathbf{b}\mathbf{b}^\top.$$

$$f_{\text{ESN}}(\mathbf{y}) = \frac{1}{\Phi\left(\kappa/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\right)}\phi_p(\mathbf{y}\mid\boldsymbol{\mu},\boldsymbol{\Omega})\,\Phi\left(\kappa+\boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right),$$

Papers on ESN distributition

METRON (2022) 80:231–253
https://doi.org/10.1007/s40300-021-00227-4

Check for updates

An EM algorithm for estimating the parameters of the multivariate skew-normal distribution with censored responses

Christian E. Galarza[1] · Larissa A. Matos[2] · Victor H. Lachos[3]

Received: 5 February 2020 / Accepted: 17 December 2021 / Published online: 28 January 2022

Abstract

Limited or censored data are collected in many studies. This occurs for many reasons in several practical situations, such as limitations in measuring equipment or from an experimental design. Consequently, the true value is recorded only if it falls within an interval range so that the responses can be either left, interval, or right-censored. Missing values can be seen just as a particular case. Linear and nonlinear regression models are routinely used to analyze these types of data. Most of these models are based on the normality assumption for the error term. However, such analyses might not provide robust inference when the normality assumption (or symmetry) is questionable. The need for asymmetric distributions for the random errors motivates us to develop a likelihood-based inference for linear models with censored responses based on the multivariate skew-normal distribution, where the missing/censoring mechanism is assumed to be "missing at random" (MAR). The proposed EM-type algorithm for maximum likelihood estimation uses closed-form expressions at the E-step based on formulas for the mean and variance of a truncated multivariate skew-normal distribution, available in the R package *MomTrunc*. Three datasets with censored and/or missing observations are analyzed and discussed.

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2021, VOL. 00, NO. 0, 1–11
https://doi.org/10.1080/10618600.2021.2000869

Taylor & Francis
Taylor & Francis Group

Check for updates

On Moments of Folded and Doubly Truncated Multivariate Extended Skew-Normal Distributions

Christian E. Galarza Morales[a], Larissa A. Matos[b], Dipak K. Dey[c], and Victor H. Lachos[c]

[a]ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas, FCNM, Guayaquil, Ecuador; [b]Department of Statistics, Campinas State University, Campinas, Brazil; [c]Department of Statistics, University of Connecticut, Mansfield, CT

ABSTRACT

This article develops recurrence relations for integrals that relate the density of multivariate extended skew-normal (ESN) distribution, including the well-known skew-normal (SN) distribution introduced by Azzalini and Dalla-Valle and the popular multivariate normal distribution. These recursions offer a fast computation of arbitrary order product moments of the multivariate truncated extended skew-normal and multivariate folded extended skew-normal distributions with the product moments as a byproduct. In addition to the recurrence approach, we realized that any arbitrary moment of the truncated multivariate extended skew-normal distribution can be computed using a corresponding moment of a truncated multivariate normal distribution, pointing the way to a faster algorithm since a less number of integrals is required for its computation which results much simpler to evaluate. Since there are several methods available to calculate the first two moments of a multivariate truncated normal distribution, we propose an optimized method that offers a better performance in terms of time and accuracy, in addition to consider extreme cases in which other methods fail. Finally, we present an application in finance where multivariate tail conditional expectation (MTCE) for SN distributed data is calculated using analytical expressions involving normal left-truncated moments. The R MomTrunc package provides these new efficient methods for practitioners. Supplementary files for this article are available online.

ARTICLE HISTORY
Received October 2019
Revised August 2021

KEYWORDS
Extended skew-normal distribution; Folded normal distribution; Product moments; Truncated distributions

R package available CRAN: MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions.

The MVSN distribution

A random matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{p\times q}$ follows a *MVSN* distribution with parameters $\mathsf{M}, \mathsf{A} \in \mathbb{R}^{p\times q}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and $\boldsymbol{\Psi} \in \mathbb{R}^{q\times q}$, if it admits the following stochastic representation:

$$\boldsymbol{\mathcal{Y}} = \mathsf{M} + W\mathsf{A} + \boldsymbol{\mathcal{X}},$$

where $\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{p\times q}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, $W \sim \mathrm{HN}(0,1) \perp \boldsymbol{\mathcal{X}}$; M is a location matrix; A is a skewness matrix; and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are row and column scale matrices, respectively, both symmetric and positive definite. We denote this as:

$$\boldsymbol{\mathcal{Y}} \sim \mathrm{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}).$$

PDF: $f_{\mathrm{MVSN}}(\mathcal{Y}) = \dfrac{2}{\tau}\exp\left(\dfrac{d_{\mathrm{A}}^2(\mathcal{Y})}{2\tau^2}\right)\phi_{p\times q}(\mathcal{Y} \mid \mathsf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})\Phi\left(\dfrac{d_{\mathrm{A}}(\mathcal{Y})}{\tau}\right)$,
where $\phi_{p\times q}(\cdot \mid \mathsf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ is the pdf of the matrix-variate normal distribution, $\Phi(\cdot)$ is the cdf of the standard normal distribution, $\tau = \sqrt{1 + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathsf{A}\boldsymbol{\Psi}^{-1}\mathsf{A}^\top)}$, and $d_{\mathrm{A}}(\mathcal{Y}) = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathcal{Y} - \mathsf{M})\boldsymbol{\Psi}^{-1}\mathsf{A}^\top\right)$.

- Equivalent Definition:

$$\mathcal{Y} \sim \mathrm{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) \quad \Longleftrightarrow \quad \mathrm{vec}(\mathcal{Y}) \sim \mathrm{SN}_{pq}(\mathrm{vec}(\mathsf{M}), \mathrm{vec}(\mathsf{A}), \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}),$$

where SN_{pq} denotes the multivariate skew-normal distribution.

- If $\mathcal{Y} \sim \mathrm{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, then its first two moments are

$$\mathbb{E}(\mathcal{Y}) = \mathsf{M} + \sqrt{\frac{2}{\pi}}\mathsf{A}, \;\; \mathrm{Var}(\mathcal{Y}) = \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi} + \sigma_W^2 \mathsf{A}\mathsf{A}^\top. \tag{2}$$

- If $\mathcal{Y} \sim \mathrm{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$.

$$C_{\mathrm{row}} = \mathbb{E}((\mathcal{Y} - \mathbb{E}(\mathcal{Y}))(\mathcal{Y} - \mathbb{E}(\mathcal{Y}))^\top) = \sigma_W^2 \mathsf{A}\mathsf{A}^\top + \mathrm{tr}(\boldsymbol{\Psi})\boldsymbol{\Sigma},$$
$$C_{\mathrm{col}} = \mathbb{E}((\mathcal{Y} - \mathbb{E}(\mathcal{Y}))^\top(\mathcal{Y} - \mathbb{E}(\mathcal{Y}))) = \sigma_W^2 \mathsf{A}^\top\mathsf{A} + \mathrm{tr}(\boldsymbol{\Sigma})\boldsymbol{\Psi},$$

where $\sigma_W^2 = 1 - 2/\pi$.

Useful properties cont...

- If $\mathcal{Y} \sim \text{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$. Then, for any $\mathsf{T} \in \mathbb{R}^{p\times q}$, the moment generating function of $\mathcal{Y}$ is given by

$$M_{\mathcal{Y}}(\mathsf{T}) = 2\,\text{etr}\left(\mathsf{T}\mathsf{M}^\top + \frac{1}{2}\boldsymbol{\Sigma}\mathsf{T}\boldsymbol{\Psi}\mathsf{T}^\top + \frac{1}{2}\mathsf{T}\mathsf{A}^\top\,\text{tr}(\mathsf{T}\mathsf{A}^\top)\right)\Phi\left(\text{tr}(\mathsf{T}\mathsf{A}^\top)\right).$$

- If $\mathcal{Y} \sim \mathrm{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, $\mathcal{Y} \in \mathbb{R}^{p\times q}$. Partition $\mathcal{Y} = \begin{bmatrix}\mathcal{Y}_1\\ \mathcal{Y}_2\end{bmatrix}$, with corresponding partitions of M, A, and $\boldsymbol{\Sigma}$. Let $\mathsf{T}_1 = \begin{bmatrix}\mathsf{I}_{p_1} & \mathbf{0}\end{bmatrix} \in \mathbb{R}^{p_1\times q}$ be the selection matrix extracting the first p_1 rows of $\mathcal{Y}$. Then

$$\mathcal{U} = \mathsf{T}_1\mathcal{Y} = \mathcal{Y}_1 \sim \text{MVSN}_{p_1\times q}(\mathsf{M}_1, \mathsf{A}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Psi}).$$

The MVESN distribution

A random matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{p\times q}$ follows a *MVESN* distribution with parameters $\mathsf{M}, \mathsf{A} \in \mathbb{R}^{p\times q}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and $\boldsymbol{\Psi} \in \mathbb{R}^{q\times q}$, if it admits the following stochastic representation:

$$\boldsymbol{\mathcal{Y}} = \mathsf{M} + W\mathsf{A} + \boldsymbol{\mathcal{X}},$$

where $\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{p\times q}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, $W \sim \mathrm{TN}(0, 1; (-\tilde{\kappa}, \infty)) \perp \boldsymbol{\mathcal{X}}$; $\widetilde{\kappa} = \kappa/\tau$, with $\tau^2 = 1 + \mathrm{vec}(\mathsf{A})^\top \boldsymbol{\Delta}^{-1} \mathrm{vec}(\mathsf{A})$ and $\boldsymbol{\Delta} = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ We denote this as:

$$\boldsymbol{\mathcal{Y}} \sim \mathrm{MVESN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \kappa).$$

PDF:

$$f_{\mathrm{MVESN}}(\mathcal{Y}) = \frac{1}{\Phi\left(\kappa/\tau\right)\tau}\phi_{p\times q}(\mathcal{Y} \mid \mathsf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) \exp\left(\frac{1}{2}\frac{d_{\mathsf{A}}(\mathcal{Y})^2}{\tau^2}\right)\Phi\left(\kappa + \frac{d_{\mathsf{A}}(\mathcal{Y})}{\tau}\right),$$

where $\phi_{p\times q}(\cdot \mid \mathsf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ is the pdf of the matrix-variate normal distribution, $\Phi(\cdot)$ is the cdf of the standard normal distribution, and $d_{\mathrm{A}}(\mathcal{Y}) = \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathcal{Y} - \mathsf{M})\boldsymbol{\Psi}^{-1}\mathsf{A}^\top\right)$.

Useful properties

- If $\mathcal{Y} \sim \mathrm{MVESN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \kappa)$, then

$$W \mid \mathcal{Y} \sim \mathrm{TN}\left(\frac{d_{\mathrm{A}}(\mathcal{Y})}{\tau^2},\ \frac{1}{\tau^2}; (-\tilde{\kappa}, \infty)\right),$$

- Equivalent Definition:

$$\mathcal{Y} \sim \mathrm{MVESN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \kappa) \iff \mathrm{vec}(\mathcal{Y}) \sim \mathrm{ESN}_{pq}(\mathrm{vec}(\mathsf{M}), \mathrm{vec}(\mathsf{A}), \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}, \kappa),$$

Recent papers on asymmetric matrix variate distributions

Computational Statistics and Data Analysis 199 (2024) 108016

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

ELSEVIER

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Check for updates

Three-way data clustering based on the mean-mixture of matrix-variate normal distributions

Mehdrad Naderi [a], Mostafa Tamandi [b], Elham Mirfarah [c], Wan-Lun Wang [d], Tsung-I Lin [e,f,*]

[a] Department of Mathematics, Physics and Electrical Engineering, Northumbria University, UK
[b] Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran
[c] School of Mathematics and Statistics, University of St Andrews, St Andrews, UK
[d] Department of Statistics and Institute of Data Science, National Cheng Kung University, Tainan 701, Taiwan
[e] Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan
[f] Department of Public Health, China Medical University, Taichung 404, Taiwan

A R T I C L E   I N F O

Dataset link: https://github.com/MerNaderi/FM-3MxN.git

Keywords:
ECME algorithm
Mean-mixture of matrix-variate normal distribution
Model-based clustering
Skew and heavy-tailed data
Three-way data

A B S T R A C T

With the steady growth of computer technologies, the application of statistical techniques t analyze extensive datasets has garnered substantial attention. The analysis of three-way (matrix variate) data has emerged as a burgeoning field that has inspired statisticians in recent year to develop novel analytical methods. This paper introduces a unified finite mixture mode that relies on the mean-mixture of matrix-variate normal distributions. The strength of ou proposed model lies in its capability to capture and cluster a wide range of three-way data tha exhibit heterogeneous, asymmetric and leptokurtic features. A computationally feasible ECMI algorithm is developed to compute the maximum likelihood (ML) estimates. Numerous simulatio studies are conducted to investigate the asymptotic properties of the ML estimators, validate the effectiveness of the Bayesian information criterion in selecting the appropriate model, an assess the classification ability in presence of contaminated noise. The utility of the propose methodology is demonstrated by analyzing a real-life data example.

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

ELSEVIER

Journal of Multivariate Analysis JMVA

Check for updates

Scale and shape mixtures of matrix variate extended skew normal distributions

Amir Rezaei [a], Fatemeh Yousefzadeh [a,*], Reinaldo B. Arellano-Valle [b]

[a] Department of Statistics, Faculty of Mathematics and Statistics, University of Birjand, Birjand, Iran
[b] Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile

A R T I C L E   I N F O

Article history:
Received 2 October 2019
Received in revised form 4 June 2020
Accepted 4 June 2020
Available online 23 June 2020

AMS 2010 subject classifications:
primary 62H05
secondary 62H12

Keywords:
EM algorithm
Matrix variate distributions
Matrix variate tail conditional expectation
Maximum likelihood estimator
Scale and shape mixtures
Skew normal distribution

A B S T R A C T

In this paper, we propose a matrix extension of the scale and shape mixtures of multivariate skew normal distributions and present some particular cases of this new class. We also present several formal properties of this class, such as the marginal distributions, the moment generating function, the distribution of linear and quadratic forms, and the selection and stochastic representations. In addition, we introduce the matrix variate tail conditional expectation measure and derive this risk measure for the scale and shape mixtures of matrix variate extended skew normal distributions. We present an efficient EM-type algorithm for the computation of maximum likelihood estimates of parameters in some special cases of the proposed class. Finally, we conduct a small simulation study and fit various special cases of the new class to a real dataset.

MVSN distribution for censoring and missing data

MVSN for censored and missing data

- Let's consider a set of i.i.d. random matrices

$$\mathcal{Y}_1, \ldots, \mathcal{Y}_n \overset{\text{iid}}{\sim} \mathsf{MVSN}_{p\times q}(\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}), \tag{3}$$

- The observed data for the ith subject are given by the matrices $(\mathcal{V}_i, \mathcal{C}_i)$, where $\mathcal{V}_i$ represents the matrix of uncensored/censored reading and $\mathcal{C}_i$ is the matrix of censoring indicators, satisfying

$$C_{ikj} = \begin{cases} 1 & \text{if } V_{1ijk} \leq Y_{ijk} \leq V_{2ijk} \\ 0 & \text{if } Y_{ijk} = V_{0ijk} \end{cases}, \tag{4}$$

  for all $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, p\}$ and $k \in \{1, \ldots, q\}$.
- Missing (at random) observations can be conveniently handled by considering $V_{1ijk} = -\infty$ and $V_{2ijk} = +\infty$.
- Observed data: $\widetilde{\mathcal{V}} = \{\mathcal{V}_1, \ldots, \mathcal{V}_i, \ldots, \mathcal{V}_n\}$ and $\widetilde{\mathcal{C}} = \{\mathcal{C}_1, \ldots, \mathcal{C}_i, \ldots, \mathcal{C}_n\}$.
- (3) and (4) define the MVSN interval-censored model (MVSNC model).

The likelihood function

For ease of likelihood computation, we use the vector representation. Specifically, let us consider $\text{vec}(\widetilde{\mathcal{Y}}) = \{\mathbf{y}_1, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_n\}$, $\text{vec}(\widetilde{\mathcal{V}}) = \{\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n\}$, and $\text{vec}(\widetilde{\mathcal{C}}) = \{\mathbf{c}_1, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_n\}$.

The first step is to partition each subject as $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where the superscripts o and c refer to the observed and censored parts, respectively. Accordingly and after reordering, we have $\mathbf{c}_i = (\mathbf{c}_i^{o\top}, \mathbf{c}_i^{c\top})^\top$, $\mathbf{v}_i = (\mathbf{v}_i^{o\top}, \mathbf{v}_i^{c\top})^\top$, with $\mathbf{v}_i^c = (\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c)$,

$$\mathbf{Y}_i^o \sim \mathcal{SN}_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Lambda}_i^{oo}, c_i^{oc}\boldsymbol{\Lambda}_i^{oo\,1/2}\tilde{\boldsymbol{\varphi}}_i^o) \quad (5)$$

$$\mathbf{Y}_i^c \mid \mathbf{Y}_i^o = \mathbf{y}_i^o \sim \mathcal{ESN}_{p_i^c}(\boldsymbol{\mu}_i^{co}, \boldsymbol{\Lambda}_i^{cc.o}, \boldsymbol{\Lambda}_i^{cc.o\,1/2}\boldsymbol{\varphi}_i^c, \tau_i^{co}) \quad (6)$$

Skew 2026

The likelihood function

The likelihood function of $\boldsymbol{\theta} = (\mathsf{M}, \mathsf{A}.\boldsymbol{\Sigma}, \boldsymbol{\Psi})$, for subject i, is given by

$$L_i(\boldsymbol{\theta}) = f(\mathsf{v}_{1i}^c \leq \mathsf{y}_i^c \leq \mathsf{v}_{2i}^c \mid \mathsf{y}_i^o, \boldsymbol{\theta}) f(\mathsf{y}_i^o \mid \boldsymbol{\theta})$$
$$= P_{p_i^c}((\mathsf{v}_{1i}^c, \mathsf{v}_{2i}^c) \mid \boldsymbol{\mu}_i^{co}, \boldsymbol{\Sigma}_i^{cc.o}, \boldsymbol{\Sigma}_i^{cc.o\,1/2} \boldsymbol{\varphi}_i^c, \tau_i^{co}) \mathrm{SN}_{p_i^o}(\mathsf{y}_i^o \mid \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, c_i^{oc} \boldsymbol{\Sigma}_i^{oo\,1/2} \tilde{\boldsymbol{\varphi}}_i^o) \equiv L_i.$$

where

- $\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Lambda}_i^{co} \boldsymbol{\Lambda}_i^{oo-1} (\mathsf{y}_i^o - \boldsymbol{\mu}_i^o), \quad \boldsymbol{\Lambda}_i^{cc.o} = \boldsymbol{\Lambda}_i^{cc} - \boldsymbol{\Lambda}_i^{co} (\boldsymbol{\Lambda}_i^{oo})^{-1} \boldsymbol{\Lambda}_i^{oc}$
- $\tilde{\boldsymbol{\varphi}}_i^o = \boldsymbol{\varphi}_i^o + \boldsymbol{\Lambda}_i^{oo\,-1} \boldsymbol{\Lambda}_i^{oc} \boldsymbol{\varphi}_i^c, \quad c_i^{oc} = (1 + \boldsymbol{\varphi}_i^{c\top} \boldsymbol{\Lambda}_i^{cc.o} \boldsymbol{\varphi}_i^c)^{-1/2}$ and $\tau_i^{co} = \tilde{\boldsymbol{\varphi}}_i^{o\top} (\mathsf{y}_i^o - \boldsymbol{\mu}_i^o)$.

$$\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^{o\top}, \boldsymbol{\mu}_i^{c\top})^\top, \quad \boldsymbol{\Lambda}_i = \begin{pmatrix} \boldsymbol{\Lambda}_i^{oo} & \boldsymbol{\Lambda}_i^{oc} \\ \boldsymbol{\Lambda}_i^{co} & \boldsymbol{\Lambda}_i^{cc} \end{pmatrix},$$
$$\boldsymbol{\lambda}_i = (\boldsymbol{\lambda}_i^{o\top}, \boldsymbol{\lambda}_i^{c\top})^\top \quad \text{and} \quad \boldsymbol{\varphi}_i = (\boldsymbol{\varphi}_i^{o\top}, \boldsymbol{\varphi}_i^{c\top})^\top.$$

The likelihood function is given by $L(\boldsymbol{\theta}) = \prod_{i=1}^n L_i(\boldsymbol{\theta})$. Easily computed via the R package MomTrunc!

Skew 2026

The EM Algorithm for ML estimation of the MVSNC

The EM Algorithm

Let $\boldsymbol{\theta}$ be the parameter vector and $\boldsymbol{y}_c = (\mathbf{y}^\top, \boldsymbol{q}^\top)$ be the vector of complete data, i.e., the observed data $\mathbf{y}^\top$ and the missing/censored data (or the latent variables, depending on the situation) $\boldsymbol{q}^\top$. The EM algorithm consists basically of two steps: the expectation (E-step) and the maximization (M-step).

- **E-Step:** Calculate the conditional expectation

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = E_{\boldsymbol{q}}\left[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right],$$

  where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the k-th iteration.

- **M-Step:** Update $\boldsymbol{\theta}^{(k)}$ according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = argmax_{\theta} Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}).$$

Skew 2026

Maximum likelihood (ML) estimation - EM algorithm

- $\boldsymbol{\theta} = (\mathsf{M}, \mathsf{A}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$;
- $\boldsymbol{\mathcal{Y}}_i \mid W_i = w_i \sim \mathcal{N}_{p\times q}(\mathsf{M} + w_i\mathsf{A},\ \boldsymbol{\Sigma},\ \boldsymbol{\Psi}),\ W_i \sim \mathrm{TN}(0, 1, [0, +\infty))$,
- Augmented data: $\mathsf{y}_c = \{\widetilde{\mathcal{Y}}, \mathsf{W}, \widetilde{\mathcal{V}}, \widetilde{\mathcal{C}}\}$;
- $\{\widetilde{\mathcal{Y}}, \mathsf{W}, \widetilde{\mathcal{V}}, \widetilde{\mathcal{C}}\} \Rightarrow [\widetilde{\mathcal{Y}}, \mathsf{W}\,]$;
- The complete log-likelihood function $\ell_c(\boldsymbol{\theta}|\mathsf{y}_c) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta})$, with

$$
\begin{aligned}
\ell_{ic}(\boldsymbol{\theta}) = & -\frac{q}{2}\log|\boldsymbol{\Sigma}| - \frac{p}{2}\log|\boldsymbol{\Psi}| \\
& - \frac{1}{2}\mathrm{tr}((\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}(\mathrm{vec}(\mathcal{Z}_i)\,\mathrm{vec}(\mathcal{Z}_i)^\top - w_i(\mathrm{vec}(\mathcal{Z}_i)\,\mathrm{vec}(\mathsf{A})^\top + \mathrm{vec}(\mathsf{A})\,\mathrm{vec}(\mathcal{Z}_i)^\top)) \\
& - \frac{1}{2}\mathrm{tr}((\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}(w_i^2\,\mathrm{vec}(\mathsf{A})\,\mathrm{vec}(\mathsf{A})^\top))
\end{aligned}
$$

with $\mathcal{Z}_i = \mathcal{Y}_i - \mathsf{M}_i,\ \ i = 1, \ldots, n$

Let $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)}$ be the current estimate at the kth step of the algorithm, the E-step provides the conditional expectation of the complete data log-likelihood function

$$
Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = E\left\{\ell_c(\boldsymbol{\theta}) \mid \widetilde{\mathcal{V}}, \widetilde{\mathcal{C}}, \widehat{\boldsymbol{\theta}}^{(k)}\right\} = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),
$$

The EM Algorithm

where

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{q}{2}\log|\boldsymbol{\Sigma}| - \frac{p}{2}\log|\boldsymbol{\Psi}| - \frac{1}{2}\mathrm{tr}\{(\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}\widehat{\boldsymbol{\Delta}}_i^{(k)}\},$$

and the expression of $\widehat{\boldsymbol{\Delta}}_i^{(k)}$ is given by:

$$\widehat{\boldsymbol{\Delta}}_i^{(k)} = \mathrm{vec}(\widehat{\mathcal{Z}_i^2})^{(k)} - \mathrm{vec}(\widehat{w\mathcal{Z}_i})^{(k)}\mathrm{vec}^\top(\mathbf{A}) - \mathrm{vec}(\mathbf{A})\,\mathrm{vec}^\top(\widehat{w\mathcal{Z}_i})^{(k)} + \widehat{w_i^2}^{(k)}\mathrm{vec}(\mathbf{A})\mathrm{vec}^\top(\mathbf{A}).$$

with $\mathcal{Z}_i = \mathcal{Y}_i - \mathbf{M}_i,\ i = 1,\ldots,n$ and

$$\begin{aligned}
\widehat{w}_i^{(k+1)} &= \mathbb{E}_{W_i,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)}(W_i \mid \mathrm{vec}(\widetilde{\mathcal{V}}_i), \mathrm{vec}(\widetilde{\mathcal{C}}_i), \widehat{\boldsymbol{\theta}}^{(k)}),\\
\widehat{w_i^2}^{(k+1)} &= \mathbb{E}_{W_i,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)}(W_i^2 \mid \mathrm{vec}(\widetilde{\mathcal{V}}_i), \mathrm{vec}(\widetilde{\mathcal{C}}_i), \widehat{\boldsymbol{\theta}}^{(k)}),\\
\mathrm{vec}(\widehat{w\mathcal{Y}_i})^{(k+1)} &= \mathbb{E}_{W_i,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)}(W_i\,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i) \mid \mathrm{vec}(\widetilde{\mathcal{V}}_i), \mathrm{vec}(\widetilde{\mathcal{C}}_i), \widehat{\boldsymbol{\theta}}^{(k)}),\\
\mathrm{vec}(\widehat{\mathcal{Y}}_i^{(k+1)}) &= \mathbb{E}_{W_i,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)}(\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i) \mid \mathrm{vec}(\widetilde{\mathcal{V}}_i), \mathrm{vec}(\widetilde{\mathcal{C}}_i), \widehat{\boldsymbol{\theta}}^{(k)}),\\
\mathrm{vec}(\widehat{\mathcal{Y}^2}_i)^{(k+1)} &= \mathbb{E}_{W_i,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)}(\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)\,\mathrm{vec}(\boldsymbol{\mathcal{Y}}_i)^\top \mid \mathrm{vec}(\widetilde{\mathcal{V}}_i), \mathrm{vec}(\widetilde{\mathcal{C}}_i), \widehat{\boldsymbol{\theta}}^{(k)}),
\end{aligned}$$

which can be derived following the results presented in Galarza et al. (2022) along with the Momtrunc R package.

M step. How to obtain closed-form expressions in the M-Step?

Let $\widehat{\mathsf{B}}_{ij}^{(k)}$ be a $p \times q$ matrix such that $\text{vec}(\widehat{\mathsf{B}}_{ij}^{(k)}) = \widehat{\mathsf{L}}_{ij}^{(k)}$. Here $\widehat{\mathsf{L}}_{ij}^{(k)}$ is jth column of the $pq \times pq$ lower triangular matrix $\widehat{\mathsf{L}}_i^{(k)}$, obtained from the Cholesky decomposition of the matrix

$$\widehat{\boldsymbol{\Delta}}_i^{(k)} = \text{vec}(\widehat{\mathcal{Z}_i^2})^{(k)} - \text{vec}(\widehat{w\mathcal{Z}_i})^{(k)}\text{vec}^\top(\mathsf{A}) - \text{vec}(\mathsf{A})\,\text{vec}^\top(\widehat{w\mathcal{Z}_i})^{(k)} + \widehat{w_i^2}^{(k)}\text{vec}(\mathsf{A})\text{vec}^\top(\mathsf{A}).$$

with $\mathcal{Z}_i = \mathcal{Y}_i - \mathsf{M}_i, \;\; i = 1, \ldots, n.$

$$\begin{aligned}
Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) &\propto -\frac{np}{2}\log(|\boldsymbol{\Psi}|) - \frac{nq}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left[(\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}\widehat{\mathsf{L}}_i^{(k)}\widehat{\mathsf{L}}_i^{(k)\top}\right] \\
&= -\frac{np}{2}\log(|\boldsymbol{\Psi}|) - \frac{nq}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left[\widehat{\mathsf{L}}_i^{(k)\top}(\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}\widehat{\mathsf{L}}_i^{(k)}\right] \\
&= -\frac{np}{2}\log(|\boldsymbol{\Psi}|) - \frac{nq}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{pq}\left[\widehat{\mathsf{L}}_{ij}^{(k)\top}(\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}\widehat{\mathsf{L}}_{ij}^{(k)}\right] \\
&= -\frac{np}{2}\log(|\boldsymbol{\Psi}|) - \frac{nq}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{pq}\left[\text{vec}(\widehat{\mathsf{B}}_{ij}^{(k)})^\top(\boldsymbol{\Psi}\otimes\boldsymbol{\Sigma})^{-1}\text{vec}(\widehat{\mathsf{B}}_{ij}^{(k)})\right] \\
&= -\frac{np}{2}\log(|\boldsymbol{\Psi}|) - \frac{nq}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{pq}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\widehat{\mathsf{B}}_{ij}^{(k)}\boldsymbol{\Psi}^{-1}\widehat{\mathsf{B}}_{ij}^{\top(k)}\right],
\end{aligned}$$

The EM Algorithm - M step

In the M step, $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ is conditionally maximized with respect to $\boldsymbol{\theta}$ and a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$ is obtained. Specifically, we have that

$$\widehat{\mathsf{M}}^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \left(\mathcal{Y}_i - \widehat{w}_i^{(k)} \widehat{\mathsf{A}}^{(k)} \right), \tag{7}$$

$$\widehat{\mathsf{A}}^{(k+1)} = \left(\sum_{i=1}^{n} \widehat{w_i^2}^{(k)} \right)^{-1} \sum_{i=1}^{n} \left\{ (\widehat{w\mathcal{Y}_i})^{(k)} - \widehat{w}_i^{(k+1)} \mathsf{M}^{(k)} \right\}. \tag{8}$$

$$\widehat{\boldsymbol{\Sigma}}^{(k+1)} = \frac{1}{nq} \sum_{i=1}^{n} \sum_{j=1}^{pq} \widehat{\mathsf{B}}_{ij}^{(k)} \widehat{\boldsymbol{\Psi}}^{(k)} \widehat{\mathsf{B}}_{ij}^{(k)\top}, \tag{9}$$

$$\widehat{\boldsymbol{\Psi}}^{(k+1)} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{pq} \widehat{\mathsf{B}}_{ij}^{(k)\top} \widehat{\boldsymbol{\Sigma}}^{(k+1)} \widehat{\mathsf{B}}_{ij}^{(k)}. \tag{10}$$

Closed-form expressions!

Skew 2026

The EM Algorithm - M step

Note that, to satisfy the identifiability constraint $|\Psi| = 1$ in the estimation process, the estimator of $\boldsymbol{\Psi}$ in (10) is then replaced by:

$$\widehat{\boldsymbol{\Psi}}^{(k+1)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{pq} \widehat{\mathsf{B}}_{ij}^{(k)\top}\widehat{\boldsymbol{\Sigma}}^{(k+1)}\widehat{\mathsf{B}}_{ij}^{(k)}}{|\sum_{i=1}^{n}\sum_{j=1}^{pq} \widehat{\mathsf{B}}_{ij}^{(k)\top}\widehat{\boldsymbol{\Sigma}}^{(k+1)}\widehat{\mathsf{B}}_{ij}^{(k)}|^{1/q}}. \tag{11}$$

We stopped the algorithm when $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}|\mathsf{V},\mathsf{C})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}|\mathsf{V},\mathsf{C}) - 1| < \epsilon$, with $\epsilon = 10^{-6}$, i.e., the algorithm stops when the relative distance between two successive evaluations of the log-likelihood is less than the tolerance.

Simulation studies

Simulation study

- We simulate data from an MVSN distribution having $p = 3$, $q = 4$, and the following parameters

$$\mathsf{M} = \begin{pmatrix} 1.00 & 2.00 & 1.00 & 2.00 \\ 2.00 & 2.00 & 1.00 & 1.00 \\ 3.00 & 3.00 & 2.00 & 3.00 \end{pmatrix}, \quad \mathsf{A} = \begin{pmatrix} -1.00 & -2.00 & -1.00 & -2.00 \\ 2.00 & -2.00 & 1.00 & -1.00 \\ 3.00 & -3.00 & 2.00 & -3.00 \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.50 & 0.60 & 0.24 \\ 0.60 & 1.50 & 0.60 \\ 0.24 & 0.60 & 1.50 \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} 2.15 & 1.72 & 1.38 & 1.10 \\ 1.72 & 2.15 & 1.72 & 1.38 \\ 1.38 & 1.72 & 2.15 & 1.72 \\ 1.10 & 1.38 & 1.72 & 2.15 \end{pmatrix}.$$

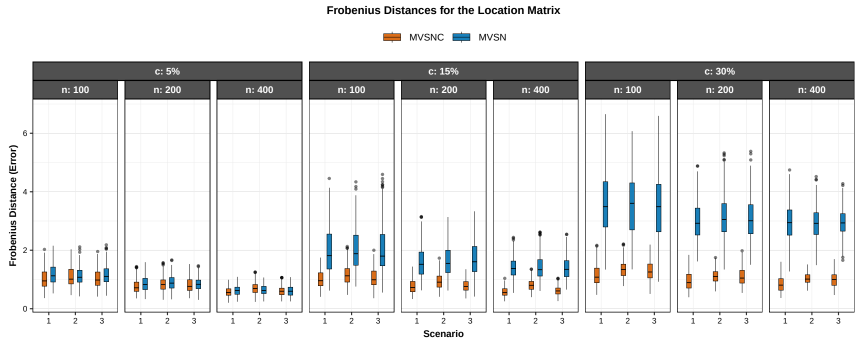
- We generate $s = 200$ independent replications.
- Three levels for the sample size, i.e. $n \in \{100, 200, 400\}$.
- Three levels for censoring $c \in \{5\%, 15\%, 30\%\}$.

Simulation study - Continued

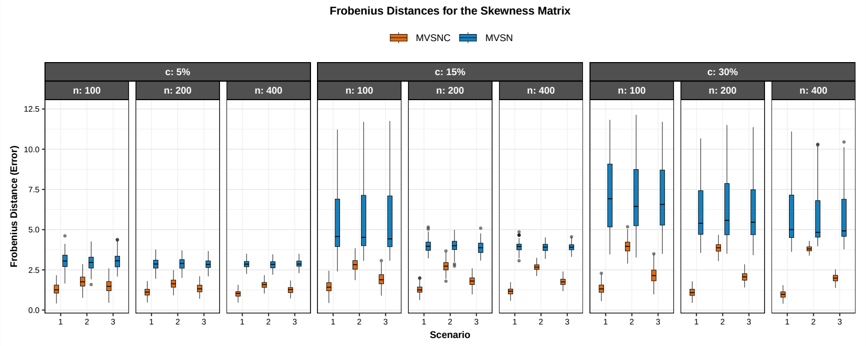
1. *Only interval-censored values*: For each selected value, two numbers a and b were randomly generated from a uniform distribution such that $c \in (a, b)$, which define the lower and upper bounds of the censoring interval.
2. *Only missing values*: All selected entries were substituted with $\pm\infty$.
3. *Mixed (50% missing)*: Half of the selected entries were replaced using item 2, and the remaining half followed item 1.

1. Therefore, by combining the experimental factors (c, s, and n), we obtain $3 \times 3 \times 3 = 27$ data configurations. Note that, for each data configuration, we generate 200 datasets, resulting in a total of 5400 simulated datasets.
2. The estimation error for each parameter matrix, M, A, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Psi}$ was measured using the Frobenius norm, defined as $\|\mathsf{H}\|_{\mathsf{F}} = \left(\mathrm{tr}(\mathsf{H}\mathsf{H}^\top)\right)^{1/2}$, where H denotes the corresponding error matrix in each case.
3. *Comparison*: After cleaning each dataset by removing the censoring and missing observations, we fit the complete-case MVSN model.

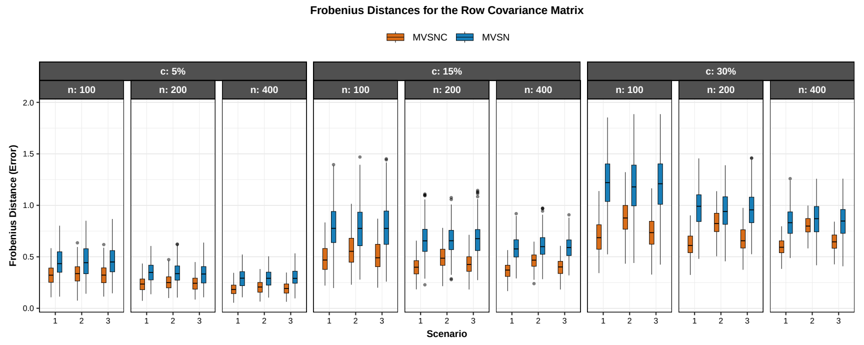
Simulation Study: Parameters recovery



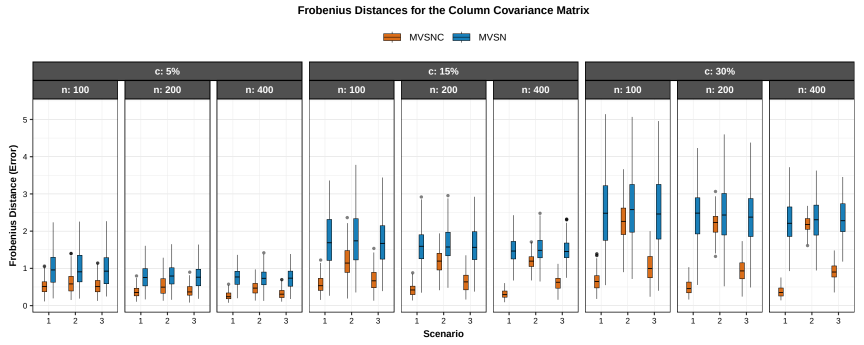
Simulation Study: Parameters recovery



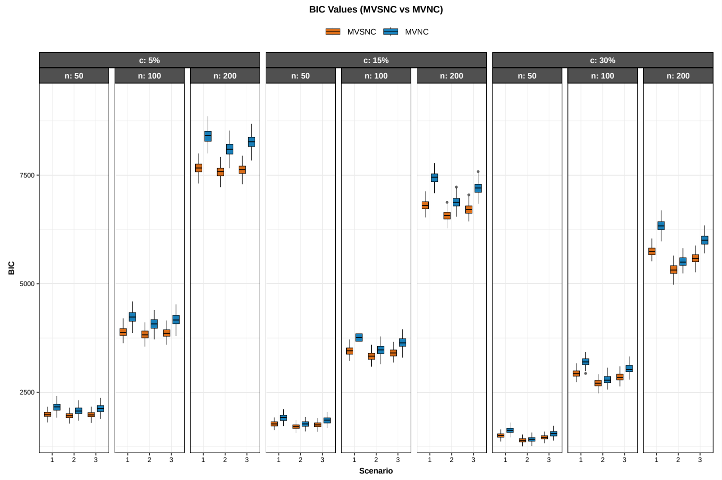
Simulation Study: Parameters recovery



Simulation Study: Parameters recovery



Simulation Study 2: Model Selection criteria (BIC)



In this case, data were generated from a matrix-variate normal inverse Gaussian (MVNIG) - Gallaugher and McNicholas (2019).

Application

Application: dataCensored

- We consider the *dataCensored* dataset contained in R package **baytrends** (Murphy et al., 2023).
- Attention in the station labeled as CB5.4, which exhibits interval-censored and missing data.
- For illustrative purposes, we consider the following p = 3 variables: orthophosphorus (po4), dissolved inorganic nitrogen (din), and ammonium (nh4). These variables present the highest levels of interval-censored and missing values in the dataset
- Specifically, ***po4*** has 14.93% of missing values and 1.60% of interval-censored values, totaling 16.56%. For ***din***, the proportions are 13.11% missing and 0.94% interval-censored, resulting in a total of 14.05%. For ***nh4*** shows proportions of 7.80% missing and 0.72% interval-censored values, totaling 8.52%.
- The resulting 3×4 matrices are then evaluated on $n = 452$ occasions.

Application: dataCensored

The estimated mean matrix is:

$$\widehat{\mathtt{LM}}$$

	S	AP	BP	B
$\mathrm{PO_4}$	0.002	0.003	0.008	0.013
DIN	0.122	0.116	0.125	0.129
$\mathrm{NH_4}$	0.017	0.025	0.054	0.082

- We observe a rise in the average values of both *po4* and *nh4* as they are assessed from the surface to the bottom.
- This trend could indicate that these substances are either being produced or accumulating at lower depths (Biological interpretation in organic decomposition).
- For *din*, the trend is similar, with the exception that its mean values decrease when moving from the surface to above the pycnocline ($S \Rightarrow AP$). The pycnocline is a layer of rapidly changing density with depth in the ocean.

Application: dataCensored

The correlation matrix $\mathsf{R}(\cdot)$, related to the total variance by rows $\boldsymbol{\Sigma}$.

$\widehat{\mathsf{R}}(\widehat{\mathsf{C}}^{\text{row}})$

	PO_4	*DIN*	NH_4
PO_4	1.000	0.394	0.577
DIN	0.394	1.000	0.522
NH_4	0.577	0.522	1.000

- There is a positive correlation between all the variables.
- *po4* and *nh4* are commonly associated with the decomposition of organic matter and wastewater discharge, and can explain the positive correlation.
- For the correlation between *din* and *nh4*, the positive value implies that elevated ammonium levels may lead to increased microbial activity, consequently giving rise to the production of various forms of dissolved inorganic nitrogen.

Application: dataCensored

The correlation matrix $\mathrm{R}(\cdot)$, related to the total variance by columns $\boldsymbol{\Psi}$.

$\widehat{\mathrm{R}}(\widehat{\mathrm{C}}^{\mathrm{column}})$

	S	AP	BP	B
S	1.000	0.854	0.474	0.357
AP	0.854	1.000	0.571	0.442
BP	0.474	0.571	1.000	0.824
B	0.357	0.442	0.824	1.000

- We note that the correlations suggest varying degrees of influence and connection between measurements at different depths.
- Surface conditions (S) seem to strongly influence conditions just above the pycnocline (AP), and to a lesser extent, conditions below the pycnocline (BP) and at the bottom (B).

Application: dataCensored

Skewness/Shape matrix.

$\widehat{\mathrm{A}}$

	S	AP	BP	B
PO_4	0.0002	0.0002	−0.0001	0.0008
DIN	0.0832	0.0731	0.0617	0.0554
NH_4	0.0042	0.0080	0.0148	0.0275

- The skewness values for PO_4 are close to zero and include a small negative entry at the BP layer (−0.0001), indicating that phosphate concentrations are nearly symmetric and stable across depths, without pronounced deviations from normality.

Concluding remarks

Conclusions

- The current literature on matrix-variate Models lack statistical approaches that simultaneously address skewness, missingness, and censoring.
- We introduce a novel approach for modeling data with interval-censored and missing values within a convenient skew-normal matrix-variate framework.
- A novel EM algorithm for obtaining ML estimates is developed by exploiting the statistical properties of matrix-variate skew-normal distributions.
- The experimental results and analysis of a real dataset support the usefulness and effectiveness of our proposal.
- The approach presented in this paper has the potential for extension in various directions. For instance, regression settings, model-based clustering, and skew-heavy-tailed matrix-variate distributions, such as the matrix-variate skew-t distribution.

Main references

Bolfarine, H., Montenegro, L. C., and Lachos, V. H. (2007). Influence diagnostics for skew-normal linear mixed models. *Sankhya: The Indian Journal of Statistics*, 69:648-670.

Galarza, C. E., Matos, L. A., Dey, D. K., and Lachos, V. H. (2022a). On moments of folded and doubly truncated multivariate extended skew-normal distributions. *Journal of Computational Graphical Statistics,*, 31:455–465.

Gallaugher, M. P. and McNicholas, P. D. (2019). Three skewed matrix variate distributions. *Statistics Probability Letters*, 145:103-109..

Gupta, A. K., and Nagar, D. K. (1999), Matrix Variate Distributions (Vol. 104), Boca Raton, FL: CRC Press.

Lachos, V. H., Tomarchio, S. D., Punzo, A., and Ingrassia, S. (2025). An EM algorithm for fitting matrix-variate normal distributions on interval-censored and missing data. *Statistics and Computing*, 35(2):1-11.

Valeriano, K. A. L., Galarza, C. E., Matos, L. A., and Lachos, V. H. Likelihood-based inference for the multivariate skew-t regression with censored or missing responses. *Journal of Multivariate Analysis*, 196:105174.

UCONN
UNIVERSITY OF CONNECTICUT

Victor Hugo Lachos Davila
Professor - Statistics Department
hlachos@uconn.edu

Skew 2026