Revista Colombiana de Estadística - Theoretical Statistics
January 2026, volume 49, issue 1, pp. 161–183
http://doi.org/10.15446/rce.v49n1.119621

Inference for Multivariate Interval Data: Bridging Frequentist and Bayesian Paradigms

Inferencia para datos intervălicos multivariados: un puente entre los paradigmas frecuentista y bayesiano

Ali Sadeghkhani[1,a], Abdolnasser Sadeghkhani[2,b], Abdulkadir Hussein[1,c]

[1]Department of Mathematics and Statistics, Faculty of Science, University of Windsor, Windsor, Ontario, Canada

[2]Department of Mathematics and Statistics, College of Science and Technology, North Carolina Agricultural and Technical State University, Greensboro, North Carolina, USA

[3]Department of Mathematics and Statistics, Faculty of Science, University of Windsor, Windsor, Ontario, Canada

Abstract

In recent years, the challenges posed by massive datasets have led researchers to explore aggregated representations, particularly interval-valued data, within the framework of symbolic data analysis. Although most recent research—apart from Samadi et al. (2024), who focused on the bivariate case—has primarily addressed parameter estimation in univariate settings, this paper extends these investigations to the general multivariate case for the first time. We derive maximum likelihood (ML) estimators for the parameters and establish their asymptotic distributions. Additionally, we develop a theoretical Bayesian framework, previously confined to the univariate setting, and extend it to multivariate interval-valued data. We provide a detailed exposition of the proposed estimators and conduct comparative performance analyses. Finally, we validate the effectiveness of our estimators through simulations and real-world data analysis.

***Keywords*:** Bayesian estimation; Entropy loss; Interval-valued data; L_2 loss; Maximum likelihood estimation.

[a]Ph.D. E-mail: sadeghk@uwindsor.ca
[b]Ph.D. E-mail: asadeghkhani@ncat.edu
[c]Ph.D. E-mail: ahussein@uwindsor.ca

Resumen

En los últimos años, los desafíos que plantean los conjuntos de datos masivos han llevado a los investigadores a explorar representaciones agregadas, en particular datos interválicos, en el marco del análisis de datos simbólicos. Aunque la investigación más reciente —salvo Samadi et al. (2024), quienes se centraron en el caso bivariado— ha abordado principalmente la estimación de parámetros en contextos univariados, este trabajo extiende por primera vez dichas investigaciones al caso multivariado general. Derivamos estimadores de máxima verosimilitud (MV) para los parámetros y establecemos sus distribuciones asintóticas. Además, desarrollamos un marco bayesiano teórico, previamente restringido al entorno univariado, y lo extendemos a datos interválicos multivariados. Presentamos una exposición detallada de los estimadores propuestos y realizamos análisis comparativos de desempeño. Finalmente, validamos la efectividad de nuestros estimadores mediante simulaciones y análisis de datos reales.

***Palabras clave*:** Datos interválicos; Estimación bayesiana; Estimación por máxima verosimilitud; Pérdida L_2; Pérdida por entropía.

1. Introduction

Symbolic data analysis (Diday, 1988) is a prominent field within statistical data analysis that focuses on understanding and modeling data represented in distributional form, known as symbols. These symbols can encompass various formats, including intervals, histograms, and other distributional representations. The foundational concept of symbolic data analysis is rooted in considering the symbol as the primary statistical unit of interest, necessitating inference at this level (Billard & Diday, 2012). Unlike a classical random variable taking values in $\mathbb{R}^p$, an interval-valued symbolic random variable is represented by an axis-aligned hyper-rectangle in $\mathbb{R}^p$ (a hypercube in the special case where all side lengths are equal). Interval-valued data, as a special case of symbolic data, provide a structured representation for information that inherently exists within intervals rather than precise point values.

Examples abound across various fields, illustrating the versatility and applicability of interval-valued data. At the same time, these data present non-standard modeling challenges, since each symbol jointly encodes location and variability, and standard multivariate tools designed for point-valued observations are not directly applicable. In finance, for instance, stock prices are often depicted as intervals to accommodate market fluctuations and volatility, providing analysts with a range rather than a single price point. Similarly, environmental monitoring utilizes intervals to report measurements like temperature or pollution levels, acknowledging variations and errors inherent in the data collection process. In medical diagnostics, interval-valued data emerge prominently, especially in scenarios where test results or patient parameters exhibit uncertainty and variability. Blood pressure readings or cholesterol levels, for instance, may be communicated as intervals rather than precise values, acknowledging the inherent uncertainty in medical measurements. See e.g., Billard & Diday (2003), Billard & Diday (2012) and Billard (2011) for an extensive, detailed overview and examples of symbolic data and their analysis, including interval-valued data. More recently, Zhu & Billard (2025) developed principal–component–based divisive clustering algorithms for interval-valued data, further illustrating the growing methodological interest in multivariate interval symbols beyond purely descriptive or aggregation–based analyses.

Beyond these classical contributions, symbolic and distribution-valued data analysis has undergone substantial development in recent decades. Foundational work on principal component methods and clustering for interval-valued data and histogram-valued data can be found in Lauro & Palumbo (2000), Irpino & Verde (2006), and Arroyo & Maté (2009), while Bock & Diday (2012) and Billard & Diday (2012) provide comprehensive treatments of symbolic data structures and associated multivariate methods. More recently, Beranger et al. (2023) proposed new model-based approaches for symbolic data, emphasizing likelihood-based and distributional modeling for complex symbolic objects.

In parallel, there has been growing interest in specific symbolic structures beyond simple intervals. Recent work by Sadeghkhani (2025) develops likelihood-based and Bayesian inference for multivariate triangular-valued data, and Sadeghkhani & Sadeghkhani (2025) study boxplot symbolic data with applications in climatology. These contributions illustrate how symbolic representations such as triangular distributions and five-number summaries can be modeled directly at the symbol level, further motivating flexible probabilistic frameworks for interval-valued and related symbolic data. The present paper fits into this emerging line of research by focusing on multivariate interval-valued observations and by providing analytically tractable likelihood and Bayesian procedures.

Thanks to advances in computational statistics, Bayesian methods provide a natural way to combine prior information with interval-valued observations. However, fully Bayesian treatments of interval-valued data remain relatively scarce. An important recent contribution is Xu & Qin (2024), who employ Jeffreys' prior and a Gibbs sampler to obtain posterior inference for interval-valued regression models. Their approach relies on Markov chain Monte Carlo to approximate the posterior distribution. In contrast, we propose a class of conjugate priors for the parameters of a multivariate interval-valued model that lead to closed-form posterior distributions. This avoids the need for Gibbs sampling, which is particularly attractive in higher dimensions (large p), and yields Bayes estimators that can be directly compared with the corresponding maximum likelihood (ML) estimators. Our development also complements and extends the frequentist results of Samadi et al. (2024) from the bivariate case ($p = 2$) to general p-variate interval-valued data.

This paper is organized as follows: Section 2 introduces key definitions, formulates the likelihood for p-variate interval-valued data, and derives the maximum likelihood (ML) estimators together with related asymptotic properties. Section 3 develops a Bayesian framework by specifying priors, obtaining closed-form posterior distributions and Bayes estimators, and showing that the resulting Bayesian estimators dominate their ML counterparts under L_2 loss for location and entropy-type loss for scale and dependence parameters. Section 4 reports simulation studies assessing finite-sample performance across several dimensions. Section 5 presents real-data illustrations based on interval-valued datasets. Finally, Section 6 concludes with a summary and directions for future work.

2. Multivariate Interval-Valued Likelihood Function

We begin this section with two definitions that will be used throughout the paper.

Definition 1 (Wishart Distribution)**.** We say that $\boldsymbol{A}$ is distributed as a Wishart distribution $\mathcal{W}_p(m, \boldsymbol{V})$, where $\boldsymbol{A}$ is a $p \times p$ positive definite symmetric matrix, m is the degrees of freedom, and $\boldsymbol{V}$ is the scale matrix. The corresponding probability density function (PDF) is given by

$$\mathcal{W}_p(\boldsymbol{A} \mid m, \boldsymbol{V}) = \frac{|\boldsymbol{A}|^{\frac{m-p-1}{2}}}{2^{\frac{mp}{2}} |\boldsymbol{V}|^{\frac{m}{2}} \Gamma_p\left(\frac{m}{2}\right)} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{A})\right), \tag{1}$$

where $|\cdot|$ denotes the determinant, $\mathrm{tr}(\cdot)$ denotes the trace, and $\Gamma_p(\cdot)$ is the multivariate generalization of the gamma function, given as

$$\Gamma_p(z) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma(z + \frac{1-i}{2}) = \int_{\boldsymbol{A}>0} \exp\left(-\mathrm{tr}(\boldsymbol{A})\right) |\boldsymbol{A}|^{a-\frac{p+1}{2}} \, d\boldsymbol{A}, \Re(a) > (p-1)/2.$$

Note that in Equation (1), we must have $m \geq p$ to ensure that the symmetric matrix $\boldsymbol{A}$ is invertible. Furthermore, $\mathbb{E}[\boldsymbol{A}] = m\boldsymbol{V}$.

Definition 2 (Inverse Wishart Distribution)**.** If $\boldsymbol{B} = \boldsymbol{A}^{-1}$, then $\boldsymbol{B}$ follows the inverse Wishart distribution $\mathcal{IW}_p(m, U)$, where the scale matrix is denoted as $U = \boldsymbol{V}^{-1}$, and its PDF is given by

$$\mathcal{IW}_p(\boldsymbol{B} \mid m, \boldsymbol{U}) = \frac{|\boldsymbol{U}|^{\frac{m}{2}} |\boldsymbol{B}|^{-\frac{m+p+1}{2}}}{2^{\frac{mp}{2}} \Gamma_p\left(\frac{m}{2}\right)} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{U}\boldsymbol{B}^{-1})\right), \tag{2}$$

with $\mathbb{E}[\boldsymbol{B}] = \boldsymbol{U}/(m-p-1)$ for $m > p+1$.

The first step in studying interval-valued data is through descriptive statistics. Bertrand & Goupil (2000) examined the univariate random interval, considering $X_{1i} = [a_{1i}, b_{1i}]$, where $a_{1i} < b_{1i}$ for $i = 1, \ldots, n$, under the assumption that points are uniformly spread across the intervals. They derived the sample mean and variance as $\bar{X}_1 = (2n)^{-1} \sum_{i=1}^{n} (a_{1i} + bi)$, and $S^2_{X_1} = (3n)^{-1} \sum_{i=1}^{n} (a^2_{1i} + a_{1i}b_{1i} + b^2_{1i}) - n^{-1}\bar{X}_1$. Billard (2008) examined the sample covariance function by considering a second random variable $X_{2i} = [a_{2i}, b_{2i}]$, where $a_{2i} < b_{2i}$, resulting in

$$\begin{aligned} S_{X_1X_2} =& (6n)^{-1}\sum_{i=1}^{n}\Big(2(a_{1i}-\bar{X}_1)(a_{2i}-\bar{X}_2) \\ &+ (a_{1i}-\bar{X}_1)(b_{2i}-\bar{X}_2) + (b_{1i}-\bar{X}_1)(a_{2i}-\bar{X}_2) \\ &+ 2(a_{1i}-\bar{X}_1)(b_{2i}-\bar{X}_2)\Big)\,. \end{aligned}$$

If $a_{1i} = b_{1i}$ for each $i = 1, \dots, n$, it means that each interval X_{1i} collapses into a single point rather than representing a range. This is essentially equivalent to dealing with point-valued data rather than interval-valued data. Therefore, in this case, the data can be treated as point-valued rather than interval-valued. Billard (2008) and Samadi et al. (2024) expanded upon the uniform distribution assumptions proposed by Bertrand & Goupil (2000), extending the results to include triangular and Pert distributions (Clark, 1962).

In order to construct the multivariate likelihood function of interval-valued data based on the uniformly spread assumption, we consider $\boldsymbol{X} = (X_1, \dots, X_p)$ representing a p-variate random variable with interval-valued realizations $\boldsymbol{X}i = (X1i, \dots, X_{pi})$, where $X_{ji} = [a_{ji}, b_{ji}]$ and $a_{ji} \leq b_{ji}$ (intervals can be open or closed at either end), for $i = 1, \dots, n$ and $j = 1, \dots, p$, representing hyper-rectangles in $\mathbb{R}^p$.

Given that each variable has aggregated observed values over an interval, it is necessary to consider the internal distribution of those values within the interval. Adapting from Le-Rademacher & Billard (2011), there exists a one-to-one correspondence between $\boldsymbol{X}_i = (X_{1i}, \dots, X_{pi})$ and $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)^\top$, where $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ represent the mean and the variance-covariance matrix of the internal distribution and can be obtained by

$$\boldsymbol{\Theta}_{i1} = \frac{1}{2}\left(a_{1i}+b_{1i}, \dots, a_{pi}+b_{pi}\right)^\top\,, \tag{3}$$

$$\boldsymbol{\Theta}_{i2} = \frac{1}{12}\mathrm{diag}((b_{1i}-a_{1i})^2, \dots, (b_{pi}-a_{1i})^2) + \frac{1}{12}\sum_{j\neq k}(b_{ji}-a_{ji})(b_{ki}-a_{ki})\,. \tag{4}$$

It is worth mentioning that Equation (4) represents a matrix with diagonal elements $\Theta_{i2}^{x_1}, \Theta_{i2}^{x_2}, \dots, \Theta_{i2}^{x_p}$ and off-diagonal elements $\Theta_{i2}^{x_jx_k}$ for $j \neq k$.

Since $\boldsymbol{X}_i$ is a random variable, the corresponding parameter $\boldsymbol{\Theta}$ varies and takes different values. Suppose the PDF of $\boldsymbol{X}_i$, denoted by $f_i^{\boldsymbol{X}_i}(\boldsymbol{x}_i; \boldsymbol{\Theta})$ and consequently can be expressed as a joint density of $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)^\top$ given by

$$\boldsymbol{\Theta}_{i1} = (\Theta_{i1}^{x_1}, \dots, \Theta_{i1}^{x_p})^\top \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})\,, \tag{5}$$

$$\boldsymbol{\Theta}_{i2} = \begin{pmatrix} \Theta_{i2}^{x_1} & \Theta_{i2}^{x_1x_2} & \cdots & \Theta_{i2}^{x_1x_{p-1}} & \Theta_{i2}^{x_1x_p} \\ \Theta_{i2}^{x_2x_1} & \Theta_{i2}^{x_2} & \cdots & \Theta_{i2}^{x_2x_{p-1}} & \Theta_{i2}^{x_2x_p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Theta_{i2}^{x_{p-1}x_1} & \Theta_{i2}^{x_{p-1}x_2} & \cdots & \Theta_{i2}^{x_{p-1}} & \Theta_{i2}^{x_{p-1}x_p} \\ \Theta_{i2}^{x_px_1} & \Theta_{i2}^{x_px_2} & \cdots & \Theta_{i2}^{x_px_{p-1}} & \Theta_{i2}^{x_p} \end{pmatrix} \sim \mathcal{W}_p(m, \boldsymbol{\Lambda})\,. \tag{6}$$

Next, we establish the likelihood function based on intervals.

2.1. ML Estimators and Related Properties

Let $\boldsymbol{S} = \sum_{i=1}^{n}(\boldsymbol{\Theta}_{i1} - \bar{\boldsymbol{\Theta}}_1)(\boldsymbol{\Theta}_{i1} - \bar{\boldsymbol{\Theta}}_1)^\top$, where $\bar{\boldsymbol{\Theta}}$ is the mean vector of parameters $\boldsymbol{\Theta}_{i1}$ for $i = 1, \dots, n$, given by $\bar{\boldsymbol{\Theta}}_1 = (\bar{\Theta}_1, \dots, \bar{\Theta}_n)^\top$.

Provided that $\{\boldsymbol{\Theta}_{i1}\}_{i=1}^{n}$ are independent and identically distributed (iid) and independent of iid $\{\boldsymbol{\Theta}_{i2}\}_{i=1}^{n}$, then the likelihood functions from equations (5) and (6) are given respectively by

$$L_1 = L_1(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{\Theta}_{11}, \dots, \boldsymbol{\Theta}_{n1}) = |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\theta} i1 - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} i1 - \boldsymbol{\mu})\right)$$

$$= |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) - \frac{n}{2}(\bar{\boldsymbol{\theta}}_1 - \boldsymbol{\mu})^\top(\bar{\boldsymbol{\theta}}_1 - \boldsymbol{\mu})\right), \tag{7}$$

$$L_2 = L_2(\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{12}, \dots, \boldsymbol{\Theta}_{n2}) = |\boldsymbol{\Lambda}|^{-\frac{nm}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\sum_{i=1}^{n}\boldsymbol{\theta}_{i2}\boldsymbol{\Delta}^{-1})\right). \tag{8}$$

The ML estimators of unknown vector $\boldsymbol{\mu}$, and matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are presented in the following theorem.

Theorem 1. *The ML estimators of parameters $\boldsymbol{\mu}$, and matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are given by*

$$\hat{\boldsymbol{\mu}}^{ML} = \bar{\boldsymbol{\Theta}}_1, \tag{9}$$

$$\hat{\boldsymbol{\Sigma}}^{ML} = \frac{\boldsymbol{S}}{n}, \tag{10}$$

$$\hat{\boldsymbol{\Lambda}}^{ML} = \frac{\sum_{i=1}^{n}\boldsymbol{\Theta}_{i2}}{nm}. \tag{11}$$

Proof. Considering the likelihood function L_1 in (7), and taking the derivative of L_1 with respect to $\boldsymbol{\mu}$ and setting it to zero gives

$$\begin{aligned}\frac{\partial L_1}{\partial \boldsymbol{\mu}} &= \frac{\partial}{\partial \boldsymbol{\mu}}\left(-\frac{n}{2}(\bar{\boldsymbol{\theta}}_1 - \boldsymbol{\mu})^\top(\bar{\boldsymbol{\theta}}_1 - \boldsymbol{\mu})\right) \\ &= -\frac{n}{2}\frac{\partial}{\partial \boldsymbol{\mu}}\left(\bar{\boldsymbol{\theta}}_1^\top\bar{\boldsymbol{\theta}}_1 - 2\bar{\boldsymbol{\theta}}_1^\top\boldsymbol{\mu} + \boldsymbol{\mu}^\top\boldsymbol{\mu}\right) \\ &= n(\bar{\boldsymbol{\theta}}_1 - \boldsymbol{\mu}) = 0,\end{aligned}$$

and solving for $\boldsymbol{\mu}$ gives Equation (9).

Analogously, by taking the derivative of L_1 with respect to $\boldsymbol{\Sigma}$ and setting it to zero, we have

$$\begin{aligned}\frac{\partial L_1}{\partial \boldsymbol{\Sigma}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}}\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\right) \\ &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}}\left(\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})\right) = 0,\end{aligned}$$

solving for $\boldsymbol{\Sigma}$ gives Equation (10).

In order to find the ML estimator $\mathbf{\Lambda}$, having the likelihood function L_2 in (8) and taking derivative of it with respect to $\mathbf{\Lambda}$ and setting it to zero gives

$$\frac{\partial L_2}{\partial \mathbf{\Lambda}} = -\frac{nm}{2}|\mathbf{\Lambda}|^{-\frac{nm}{2}-1} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\sum_{i=1}^{n} \boldsymbol{\theta}_{i2}\mathbf{\Delta}^{-1}\right)\right) = 0.$$

Solving for $\mathbf{\Lambda}$ yields Equation (11). This completes the proof.

□

2.2. Asymptotic Properties of ML Estimators

Theorem 2. *Consider Sym(p), the set of $p \times p$ real symmetric matrices, and let $\mathbb{P}(p) \subseteq Sym(p)$ represent the subset consisting of symmetric positive-definite matrices that forms a convex regular cone. Setting $\boldsymbol{\omega} = (\boldsymbol{\mu}, \mathbf{\Sigma}) \in \Omega = \mathbb{R}^p \times \mathbb{P}(p)$, then*

$$\sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}_m, I^{-1}(\boldsymbol{\omega}))\,,$$

with

$$I_{ij}(\boldsymbol{\omega}) \;=\; \left[\frac{\partial \boldsymbol{\mu}}{\partial \omega_i}\right]^{\top} \mathbf{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \omega_j} + \frac{1}{2}\,\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \omega_i}\mathbf{\Sigma}^{-1}\frac{\partial \mu}{\partial \omega_j}\right)\,,$$

and $m = dim(\Omega) = p(p+3)/2$.

***Proof*.** The symmetric semi-positive fisher information matrix (eg., Amari (2016)) is given by $I(\boldsymbol{\omega}) = \mathbb{V}[\nabla \log \mathcal{N}_m(\Theta_1 \mid \boldsymbol{\mu}, \mathbf{\Sigma})]$, where is a PDF of p-varaite normal with mean vector $\mathbb{E}[\Theta] = \boldsymbol{\mu}$, $\mathbb{V}[\Theta] = \mathbf{\Sigma}$, and $\mathbb{V}(\cdot)$ is the variance-covariance matrix. As discussed in Nielsen (2023), the fisher information matrix can be written as follows

$$\begin{aligned}
I(\boldsymbol{\omega}) &= \mathrm{Cov}\left[\nabla \log \mathcal{N}_m(\Theta_1 \mid \boldsymbol{\mu}, \mathbf{\Sigma}), \mathbf{\Sigma})\right] \\
&= \mathbb{E}\left[h\nabla \log \mathcal{N}_m(\Theta_1 \mid \boldsymbol{\mu}, \mathbf{\Sigma}))\nabla \log \mathcal{N}_m(\Theta_1 \mid \boldsymbol{\mu}, \mathbf{\Sigma}))^{\top}\right] \\
&= -\mathbb{E}\left[\nabla^2 \log \mathcal{N}_m(\Theta_1 \mid \boldsymbol{\mu}, \mathbf{\Sigma}))\right]
\end{aligned}$$

For multivariate distributions parameterized by an m-dimensional vector $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_p, \psi_{p+1}, \ldots, \psi_m) \in \mathbb{R}^m$, with $\boldsymbol{\mu} = (\psi_1, \ldots, \psi_p)$ and $\mathbf{\Sigma}(\boldsymbol{\psi}) = \mathrm{vech}(\psi_{p+1}, \ldots, \psi_m)$, where $\mathrm{vech}(\cdot)$ refers to the vech operator. Then we have $I(\boldsymbol{\omega}) = [I_{ij}(\boldsymbol{\omega})]$, with

$$I_{ij}(\boldsymbol{\omega}) \;=\; \left[\frac{\partial \boldsymbol{\mu}}{\partial \omega_i}\right]^{\top} \mathbf{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \omega_j} + \frac{1}{2}\,\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \omega_i}\mathbf{\Sigma}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \omega_j}\right).$$

see Skovgaard (1984), and Barachant et al. (2013) for more information. □

Proposition 1. *It can be checked that for $p = 1$ and $p = 2$, the Fisher information matrices are simplified to the following.*

$$I(\boldsymbol{\omega}) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/\sigma^4 \end{bmatrix}, \quad \textit{with} \quad \boldsymbol{\omega} = (\mu, \sigma^2),$$

$$I(\boldsymbol{\omega}) = \begin{bmatrix} \boldsymbol{A}_{2\times 2} & \mathbf{0}_{2\times 3} \\ \mathbf{0}_{3\times 2} & \boldsymbol{B}_{2\times 2} \end{bmatrix}, \quad \textit{with} \quad \boldsymbol{\omega} = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho),$$

and symmetric matrices are given by

$$\boldsymbol{A} = \begin{bmatrix} \frac{-1}{(1-\rho^2)\sigma_1^2} & \frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} \\ \frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{(\rho^2-1)\sigma_2^2} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \frac{-\rho^2+1}{(\rho^2-1)^2} & \frac{\rho}{(1-\rho^2)\sigma_1} & \frac{\rho}{(1-\rho^2)\sigma_2} \\ \frac{\rho}{(1-\rho^2)\sigma_1} & \frac{2-\rho^2}{(\rho^2-1)\sigma_1^2} & \frac{\rho^2}{(\rho^2-1)\sigma_1\sigma_2} \\ \frac{\rho}{(1-\rho^2)\sigma_2} & \frac{\rho^2}{(1-\rho^2)\sigma_1\sigma_2} & \frac{2-\rho^2}{(\rho^2-1)\sigma_2^2} \end{bmatrix}.$$

3. Bayesian Set Up

We begin by proposing prior distributions on the parameters and subsequently derive the posterior distribution based on the given likelihood functions (7) and (8) corresponding to models (5) and (6), respectively.

Consider the following priors on parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ in (5) and $\boldsymbol{\Lambda}$ in (6) as below

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+2}{2}}, \tag{12}$$

$$\pi(\boldsymbol{\Lambda}) \propto |\boldsymbol{\Lambda}|^{-\frac{p+1}{2}}. \tag{13}$$

Let $\boldsymbol{\Theta} = \left\{(\boldsymbol{\Theta}_{i1}, \boldsymbol{\Theta}_{i2})^\top\right\}_{i=1}^n$, and suppose that priors (12) and (13) are independent. Thus, we have

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) \propto |\boldsymbol{\Sigma}|^{-\frac{p+2}{2}} |\boldsymbol{\Lambda}|^{-\frac{p+1}{2}}. \tag{14}$$

The posterior distribution of $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}$ can then be written as

$$\begin{aligned} \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda} \mid \boldsymbol{\Theta}) \propto & L_1(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{\Theta}_{11}, \ldots, \boldsymbol{\Theta}_{1n})\, L_2(\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_{12}, \ldots, \boldsymbol{\Theta}_{n2})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) \\ \propto & |\boldsymbol{\Sigma}|^{-\frac{n+p+2}{2}} |\boldsymbol{\Lambda}|^{-\frac{nm+p+2}{2}} \\ & \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) - \frac{n}{2}\sum_{i=1}^n (\boldsymbol{\theta}_{i1} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta}_{i1} - \boldsymbol{\mu})\right) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}(\sum_{i=1}^n \boldsymbol{\theta}_{i2} \boldsymbol{\Lambda}^{-1})\right). \end{aligned} \tag{15}$$

The following lemma provides the full conditional posterior distributions associated with the posterior distribution in (15).

Lemma 1. *Full conditional distributions associated with the posterior distribution in (15) are given by*

$$\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \boldsymbol{\Theta}_1 \sim \mathcal{N}_p(\bar{\boldsymbol{\Theta}}_1, \boldsymbol{\Sigma}/n)\,, \tag{16}$$

$$\boldsymbol{\Sigma} \mid \boldsymbol{\mu}, \boldsymbol{\Theta}_1 \sim \mathcal{IW}_p(n+1, (\bar{\boldsymbol{\Theta}}_1 - \boldsymbol{\mu})(\bar{\boldsymbol{\Theta}}_1 - \boldsymbol{\mu})^\top/n)\,, \tag{17}$$

$$\boldsymbol{\Lambda} \mid \boldsymbol{\Theta}_2 \sim \mathcal{IW}_p(nm, \sum_{i=1}^{n} \boldsymbol{\Theta}_{i2})\,. \tag{18}$$

***Proof*.** The proof is straightforward and hence is omitted. □

3.1. Loss Functions

The most common loss function for estimating vector $\boldsymbol{\mu}$ using $\hat{\boldsymbol{\mu}}$ is L_2 loss, $||\boldsymbol{\mu} - \hat{\mu}||^2$, while the common loss function in the matrix form is the entropy loss (Stein, 1956)

$$\mathcal{L}(\boldsymbol{B}, \hat{\boldsymbol{B}}) = \operatorname{tr}\left(\hat{\boldsymbol{B}}\boldsymbol{B}^{-1}\right) - \log\left|\hat{\boldsymbol{B}}\boldsymbol{B}^{-1}\right| - p\,, \tag{19}$$

where B is a $p \times p$ symmetric matrix. The Bayesian estimator for the matrix estimator is the posterior mean.

The corresponding risk function to loss function (19) is given by

$$\mathcal{R}(\hat{\boldsymbol{B}}, \boldsymbol{B}) = \mathbb{E}\left[\mathcal{L}(\hat{\boldsymbol{B}}, \boldsymbol{B})\right]\,. \tag{20}$$

Theorem 3. *Consider model 5, 6, and the prior (14). The The Bayes estimators of parameters $\boldsymbol{\mu}$ (with respect to L_2 loss), $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ (with respect to entropy loss function in 19) are given by*

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{\Theta}}_1\,, \tag{21}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{\boldsymbol{S}}{n-p}\,, \tag{22}$$

$$\hat{\boldsymbol{\Lambda}} = \frac{\sum_{i=1}^{n} \boldsymbol{\Theta}_{i2}}{nm - p - 1}\,. \tag{23}$$

***Proof*.** Given that Bayes estimators are the expectations of corresponding marginal distributions, and with the posterior distributions available in (15), we integrate over $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$, and eventually over $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$, yielding

$$\begin{aligned}
\boldsymbol{\mu} \mid \boldsymbol{\Theta} &\sim \mathcal{T}_p(\bar{\boldsymbol{\Theta}}_1, \frac{\boldsymbol{S}}{n+p+1}, n+1-p)\,, \qquad (24)\\
\boldsymbol{\Sigma} \mid \boldsymbol{\Theta} &\sim \mathcal{IW}_p(n+1, \boldsymbol{S})\,,\\
\boldsymbol{\Lambda} \mid \boldsymbol{\Theta} &\sim \mathcal{IW}_p(nm, \sum_{i=1}^{n} \boldsymbol{\Theta}_{i2})\,,
\end{aligned}$$

where $\mathcal{T}_p(\boldsymbol{m}, \boldsymbol{A}, \nu)$ in (24) represents a multivariate Student's t-distribution with mean vector $\boldsymbol{m}$, variance matrix $\boldsymbol{A}$, and ν degrees of freedom with an expectation of $\boldsymbol{m}$. Completing the proof involves using the expectation of the inverse Wishart distribution, as given in Definition 2. □

Theorem 4. *Under the assumptions of Theorem 3, the Bayes estimators of the parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ obtained in equations (22) and (23) dominate the ML estimators (10) and (11) obtained in Theorem 1 under entropy loss function (19).*

Proof. Let Δ_Σ and Δ_Λ denote the difference in risk functions of the Bayes and ML estimators for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$, respectively. It can be easily seen that

$$\Delta_\Sigma = \mathbb{E}\left[\mathcal{L}(\hat{\boldsymbol{\Sigma}}^{ML}, \boldsymbol{\Sigma}) - \mathcal{L}(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})\right] = \log \frac{n}{n-1}\,. \tag{25}$$

Similarly, one can show that

$$\Delta_\Lambda = \frac{nm}{nm - p - 1}\,. \tag{26}$$

Both equations (25) and (26) confirm that the difference in risk functions is positive. This completes the proof. □

As with any Bayesian procedure, it is important to acknowledge that posterior inferences may exhibit some sensitivity to the choice of prior distributions; in this work we focus on the objective, conjugate priors in (12)–(13) that yield closed-form posteriors, and a more systematic prior-sensitivity study is left for future research.

3.2. Special Cases

3.2.1. Univariate Case

When $p = 1$, Equations (5) and (6) imply that $\{\Theta_{i1}\}_{i=1}^n$ are IID from $\mathcal{N}(\mu, \sigma^2)$ and are independent of $\{\Theta_{i2}\}_{i=1}^n$, which are IID from an exponential distribution $\mathcal{E}(\lambda)$ (equivalently $\mathcal{W}_1(2, 2\lambda)$). Furthermore, as shown by Le-Rademacher & Billard (2011), the ML estimators of parameters μ, σ^2, and λ are given by

$$\hat{\mu}^{ML} = \bar{\Theta}_1\,, \qquad \hat{\sigma}^{2ML} = \sum_{i=1}^n (\Theta_{i1} - \bar{\Theta}_1)^2/n \quad \hat{\lambda}^{ML} = \frac{\sum_{i=1}^n \Theta_{i2}}{n}.$$

The corresponding posterior distribution (15) in this case is $\pi(\mu, \sigma, \lambda) \propto \sigma^{-3}\lambda^{-1}$, which is also Jeffrey's prior studied by Xu & Qin (2024). Therefore, the conditional posterior distributions from equations (16), (17), and (18) are reduced respectively to the following

$$\mu \mid \Theta, \sigma^2, \gamma \sim \mathcal{N}(\bar{\Theta}_1, \sigma^2/n)\,,$$
$$\sigma^2 \mid \Theta, \mu, \lambda \sim \mathcal{IG}(\frac{n+1}{2}, \sum_{i=1}^{n}(\Theta_{i2}-\mu)^2/2)\,,$$
$$\lambda \mid \Theta, \mu, \sigma^2 \sim \mathcal{IG}(n, \sum_{i=2}^{n}\Theta_{i2})\,.$$

Moreover, the Bayes estimators can also be retrieved from Theorem (3), with $p = 1$ as below

$$\hat{\mu} = \bar{\Theta}_1\,, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\Theta_{i1}-\bar{\Theta}_1)^2}{n-1}\,, \quad \hat{\lambda} = \frac{\sum_{i=1}^{n}\Theta_{i2}}{n}\,.$$

Unlike Xu & Qin (2024), we have proposed closed-form Bayesian estimators for the parameters, making the Gibbs sampler method they employed unnecessary.

3.2.2. Bivariate Case

In this case $\{\boldsymbol{\Theta}_{i1}^{x_1,x_2}\}_{i=1}^{n}$ are iid from $\mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ (which is corresponding $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, and $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{smallmatrix}\right)$ in 5) is independent of $\{\boldsymbol{\Theta}_{i2}^{x_1,x_2}\}_{i=1}^{n}$ are iid from $\mathcal{W}_2(m, \boldsymbol{\Lambda} = \left(\begin{smallmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{smallmatrix}\right))$.

Therefore the conditional posterior distributions are obtained using (16), (17) and (18) with $p = 2$. In order to obtain the ML estimators of the parameters, lets rewrite the likelihood functions (7) and (8) in this case as follows.

$$L_1 = (2\pi\sigma_1\sigma_2)^{-n} \exp\left(\sum_{i=1^n}(\theta_{i1}^{x_1}-\mu_1)^2/\sigma_1^2 + \theta_{i1}^{x_2}-\mu_2)^2/\sigma_2^2\right) - 2\frac{\rho}{\sigma_1\sigma_2}\theta_{i1}^{x_1}-\mu_1)\theta_{i1}^{x_2}-\mu_2)/\sigma_2^2\,, \tag{27}$$

$$L_2 = \frac{(\lambda_{11}\lambda_{22}-\lambda_{12})^{2n}2^{-mn}\pi^{-n/2}}{(\Gamma(m/2)\Gamma((m-1)/2))^n}\prod_{i=1}^{n}(\theta_{i2}^{x_1}\theta_{i2}^{x_2}-\theta_{i2}^{x_1x2})^{m/2-1} \exp\left(-\frac{\lambda_{11}\lambda_{22}}{(\lambda_{11}\lambda_{22}-\lambda_{12}^2)}\left(\frac{\sum_{i=1}^{n}\theta_{i2}^{x_1}}{\lambda_{11}} + \frac{\sum_{i=1}^{n}\theta_{i2}^{x_2}}{\lambda_{22}} - \frac{2\lambda_{12}}{\lambda_{11}\lambda_{22}}\sum_{i=1}^{n}\theta_{i2}^{x_1x2}\right)\right). \tag{28}$$

Taking derivatives with respect to parameters of μ_i, σ_i for $i = 1, 2$, and ρ from log of likelihood functions (27) and (28) (see Samadi et al. (2024) for details) results to following ML estimators.

$$\hat{\mu}_i^{ML} = \bar{\Theta}_i^{x_i}, \quad \hat{\sigma}_i^{2ML} = \sum_{i=1}^{n}(\Theta_{i1}^{x_1} - \mu_i)^2/n, \text{ for } \quad i = 1, 2,$$

$$\hat{\rho}^{ML} = \frac{\sum_{i=1}^{n}(\Theta_{i1}^{x_i} - \mu_1)(\Theta_{i1}^{x_2} - \mu_2)}{\sqrt{\sum_{i=1}^{n}(\Theta_{i1}^{x_1} - \mu_1)^2 \sum_{i=1}^{n}(\Theta_{i1}^{x_1} - \mu_2)^2}},$$

$$\hat{\lambda}_{ii}^{ML} = \frac{\sum_{i=1}^{n}\Theta_{i2}^{x_i}}{nm}, \quad \text{for} \quad i = 1, 2, \hat{\lambda}_{12}^{ML} = \frac{\sum_{i=1}^{n}\Theta_{i2}^{x_1 x_2}}{nm},$$

which corresponds to Theorem 3 with $p = 2$.

4. Simulation Results

In the simulation, three scenarios are considered to generate samples of n from the random variables $\mathbf{\Theta}_{i1}$ and $\mathbf{\Theta}_{i2}$, as described in equations (5) and (6). In the first scenario (I), samples are generated from univariate distributions, where each random variable $\mathbf{\Theta}_{i1}$ and $\mathbf{\Theta}_{i2}$ was independently sampled from $\mathcal{N}(\mu = 2, \sigma^2 = 5)$, and $\mathcal{E}(\lambda = 2)$, respectively.

In the second scenario (II), bivariate distributions were employed, generating samples where the random variables $\mathbf{\Theta}_{i1}$ and $\mathbf{\Theta}_{i2}$ are independently sampled from $\mathcal{N}_2(\left(\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 & 3 \\ 3 & 9 \end{smallmatrix}\right))$, and $\mathcal{W}_2(m = 3, \mathbf{\Lambda} = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 5 \end{smallmatrix}\right))$. Finally, in the third scenario (III), trivariate distributions were utilized, resulting in samples where the random variables $\mathbf{\Theta}_{i1}$ and $\mathbf{\Theta}_{i2}$ are sampled from $\mathcal{N}_3(\boldsymbol{\mu}, \mathbf{\Sigma})$, and $\mathcal{W}_3(m = 3, \mathbf{\Lambda})$, with

$$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} 1 & 1.4 & 0.6 \\ 1.4 & 4 & 1.5 \\ 0.6 & 1.5 & 9 \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 5 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

These scenarios allow us to compare behavior and performance of the proposed Bayesian and ML parameter estimators for different dimensions.

For each simulation within each scenario, sample sizes of $n = 25, 50, 200$, and 500, with 10 000 iterations, are conducted. The estimated parameters using Theorems 1 and 3, associated with the ML and Bayes estimators, are tabulated in Tables 1, 2, and 3, corresponding to dimensions $p = 1$, $p = 2$, and $p = 3$ (scenarios I, II, and III).

According to Table 1 both methods yield similar estimates for μ across sample sizes, while Bayesian estimation tends to produce slightly higher estimates for σ^2 and λ compared to MLE, with standard deviations also presented.

In Table 2, representing a simulation with $p = 2$, both ML and Bayesian estimations exhibit consistency across various parameters and sample sizes. Analogous to Table 1, both methods yield similar estimates for μ_1 and μ_2, irrespective of sample size, with consistent standard deviations. However, for σ_1^2, σ_2^2, σ_{12}, λ_{11}, λ_{22}, and λ_{12}, Bayesian estimation tends to produce slightly higher estimates compared to ML, accompanied by corresponding standard deviations. This trend persists across different sample sizes, highlighting the robustness of Bayesian estimation in this scenario.

On the other hand, in Table 3, reflecting scenario III, we observe similar trends to those seen in simulation II (Table 2). Bayesian estimation consistently yields slightly higher parameter estimates compared to ML across various parameters and sample sizes.

Table 1: Simulation scenario I (univariate interval-valued model with $\Theta_{i1} \sim \mathcal{N}(\mu = 2, \sigma^2 = 5)$ and $\Theta_{i2} \sim \mathcal{E}(\lambda = 2)$): Monte Carlo means (with standard deviations in parentheses) of ML and Bayes estimators for sample sizes $n \in \{25, 50, 200, 500\}$.

Parameter	**MLE**	**Bayesian**
	$n = 25$	
μ	2.006555 (SD: 0.4395082)	2.006555 (SD: 0.4395082)
σ^2	4.824011 (SD: 1.396871)	5.025011 (SD: 1.455074)
λ	2.006947 (SD: 0.3253547)	2.061932 (SD: 0.3342685)
	$n = 50$	
μ	2.000126 (SD: 0.3111599)	2.000126 (SD: 0.3111599)
σ^2	4.912915 (SD: 0.9940378)	5.013179 (SD: 1.014324)
λ	2.001979 (SD: 0.2297495)	2.029032 (SD: 0.2328543)
	$n = 200$	
μ	1.997678 (SD: 0.1580845)	1.997678 (SD: 0.1580845)
σ^2	4.973424 (SD: 0.4938952)	4.998416 (SD: 0.4963771)
λ	1.999907 (SD: 0.1159792)	2.006596 (SD: 0.116367)
	$n = 500$	
μ	1.998748 (SD: 0.09948847)	1.998748 (SD: 0.09948847)
σ^2	4.990833 (SD: 0.3125381)	.000835 (SD: 0.3131644)
λ	2.000496 (SD: 0.07324693)	2.003167 (SD: 0.07334472)

Table 2: Simulation scenario II (bivariate interval-valued model with $\boldsymbol{\Theta}_{i1} \sim \mathcal{N}_2((2,4)^\top, \boldsymbol{\Sigma})$ and $\boldsymbol{\Theta}_{i2} \sim \mathcal{W}_2(m=3, \boldsymbol{\Lambda})$ as specified in Section 4): Monte Carlo means (with standard deviations in parentheses) of ML and Bayes estimators for sample sizes $n \in \{25, 50, 200, 500\}$.

Parameter	MLE	Bayesian
	$n = 25$	
μ_1	1.998343 (SD: 0.3971383)	1.998343 (SD: 0.3971383)
μ_2	3.998895 (SD: 0.5914762)	3.998895 (SD: 0.5914762)
σ_1^2	3.854855 (SD: 1.116438)	4.19006 (SD: 1.213519)
σ_2^2	8.67281 (SD: 2.542278)	9.426968 (SD: 2.763346)
σ_{12}	2.898315 (SD: 1.320275)	3.150342 (SD: 1.435081)
λ_{11}	2.001881 (SD: 0.324576)	2.085292 (SD: 0.3381)
λ_{22}	4.989752 (SD: 0.811227)	5.197658 (SD: 0.8450281)
λ_{12}	1.003482 (SD: 0.3770854)	1.045294 (SD: 0.3927973)
	$n = 50$	
μ_1	1.997207 (SD: 0.2815495)	1.997207 (SD: 0.2815495)
μ_2	3.995872 (SD: 0.4210236)	3.995872 (SD: 0.4210236)
σ_1^2	3.923179 (SD: 0.7913848)	4.086645 (SD: 0.8243592)
σ_2^2	8.821342 (SD: 1.775198)	9.188898 (SD: 1.849164)
σ_{12}	2.945679 (SD: 0.9464634)	3.068416 (SD: 0.9858993)
λ_{11}	1.999832 (SD: 0.2306271)	2.040645 (SD: 0.2353338)
λ_{22}	4.998042 (SD: 0.5803669)	5.100043 (SD: 0.5922111)
λ_{12}	1.002067 (SD: 0.2672165)	1.022517 (SD: 0.2726699)
	$n = 200$	
μ_1	2.001589 (SD: 0.1427427)	2.001589 (SD: 0.1427427)
μ_2	3.999514 (SD: 0.2126366)	3.999514 (SD: 0.2126366)
σ_1^2	3.979203 (SD: 0.3944584)	4.019397 (SD: 0.3984429)
σ_2^2	8.944955 (SD: 0.8901909)	9.035308 (SD: 0.8991827)
σ_{12}	2.975505 (SD: 0.4695272)	3.005561 (SD: 0.4742699)
λ_{11}	1.998426 (SD: 0.1153766)	2.008469 (SD: 0.1159564)
λ_{22}	4.996518 (SD: 0.2876331)	5.021626 (SD: 0.2890785)
λ_{12}	0.9984405 (SD: 0.1352488)	1.003458 (SD: 0.1359284)
	$n = 500$	
μ_1	2.000842 (SD: 0.09020568)	2.000842 (SD: 0.09020568)
μ_2	4.00109 (SD: 0.1340408)	4.00109 (SD: 0.1340408)
σ_1^2	3.993712 (SD: 0.2534548)	4.009751 (SD: 0.2544727)
σ_2^2	8.982588 (SD: 0.5689367)	9.018663 (SD: 0.5712216)
σ_{12}	2.996848 (SD: 0.3020145)	3.008884 (SD: 0.3032274)
λ_{11}	1.998547 (SD: 0.07260435)	2.002552 (SD: 0.07274985)
λ_{22}	5.000131 (SD: 0.1804525)	5.010151 (SD: 0.1808142)
λ_{12}	0.9986841 (SD: 0.08520831)	1.000685 (SD: 0.08537907)

TABLE 3: Simulation scenario III (trivariate interval-valued model with $\mathbf{\Theta}_{i1} \sim \mathcal{N}_3(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $\mathbf{\Theta}_{i2} \sim \mathcal{W}_3(m = 3, \mathbf{\Lambda})$ as specified in Section 4): Monte Carlo means (with standard deviations in parentheses) of ML and Bayes estimators for sample sizes $n \in \{25, 50, 200, 500\}$.

Parameter	**MLE**	**Bayesian**
	$n = 25$	
μ_1	2.001857 (SD: 0.1996053)	2.001857 (SD: 0.1996053)
μ_2	4.000511 (SD: 0.3989019)	4.000511 (SD: 0.3989019)
μ_3	6.00609 (SD: 0.5932688)	6.00609 (SD: 0.5932688)
σ_1^2	0.9618643 (SD: 0.2764356)	1.093028 (SD: 0.3141314)
σ_2^2	3.851198 (SD: 1.113989)	4.376362 (SD: 1.265897)
σ_3^2	8.655308 (SD: 2.536946)	9.835577 (SD: 2.882893)
σ_{12}	1.348785 (SD: 0.4805106)	1.53271 (SD: 0.5460348)
σ_{13}	0.5855684 (SD: 0.5996525)	0.6654186 (SD: 0.6814233)
σ_{23}	1.449337 (SD: 1.20976)	1.646974 (SD: 1.374727)
λ_{11}	2.004908 (SD: 0.3292506)	2.11786 (SD: 0.3478)
λ_{12}	1.002803 (SD: 0.3870634)	1.059299 (SD: 0.4088698)
λ_{13}	1.002417 (SD: 0.3094444)	1.058891 (SD: 0.3268779)
λ_{22}	5.010742 (SD: 0.8130529)	5.293038 (SD: 0.8588587)
λ_{23}	2.004597 (SD: 0.5016901)	2.117532 (SD: 0.5299543)
λ_{33}	2.999371 (SD: 0.4971395)	3.16835 (SD: 0.5251473)

TABLE 4: Simulation scenario III (continued): same trivariate interval-valued model as in Table 3; Monte Carlo means (with standard deviations in parentheses) of ML and Bayes estimators for $n = 50$.

Parameter	**MLE**	**Bayesian**
	$n = 50$	
μ_1	2.001289 (SD: 0.1414633)	2.001289 (SD: 0.1414633)
μ_2	4.006495 (SD: 0.2821739)	4.006495 (SD: 0.2821739)
μ_3	6.009829 (SD: 0.4182362)	6.009829 (SD: 0.4182362)
σ_1^2	0.9788372 (SD: 0.1987836)	1.041316 (SD: 0.2114719)
σ_2^2	3.91668 (SD: 0.7861622)	4.16668 (SD: 0.8363427)
σ_3^2	8.81281 (SD: 1.784628)	9.37533 (SD: 1.89854)
σ_{12}	1.371165 (SD: 0.3410975)	1.458686 (SD: 0.3628697)
σ_{13}	0.5826238 (SD: 0.4231293)	0.6198126 (SD: 0.4501376)
σ_{23}	1.456099 (SD: 0.8677798)	1.549041 (SD: 0.9231701)
λ_{11}	1.999239 (SD: 0.2297266)	2.054012 (SD: 0.2360205)
λ_{12}	0.9989928 (SD: 0.2741133)	1.026362 (SD: 0.2816232)
λ_{13}	0.9996421 (SD: 0.2188311)	1.02703 (SD: 0.2248265)
λ_{22}	5.000366 (SD: 0.5758727)	5.137362 (SD: 0.59165)
λ_{23}	1.999313 (SD: 0.354906)	2.054089 (SD: 0.3646294)
λ_{33}	2.996561 (SD: 0.3473413)	3.078658 (SD: 0.3568575)

Table 5: Simulation scenario III (continued): same trivariate interval-valued model as in Table 3; Monte Carlo means (with standard deviations in parentheses) of ML and Bayes estimators for $n = 200$.

Parameter	**MLE**	**Bayesian**
	$n = 200$	
μ_1	1.999994 (SD: 0.07056982)	1.999994 (SD: 0.07056982)
μ_2	4.000806 (SD: 0.142892)	4.000806 (SD: 0.142892)
μ_3	6.003733 (SD: 0.2112991)	6.003733 (SD: 0.2112991)
σ_1^2	0.9941807 (SD: 0.09982312)	1.009321 (SD: 0.1013433)
σ_2^2	3.981968 (SD: 0.3995839)	4.042607 (SD: 0.4056689)
σ_3^2	8.967562 (SD: 0.8961266)	9.104123 (SD: 0.9097732)
σ_{12}	1.392808 (SD: 0.1736081)	1.414019 (SD: 0.1762519)
σ_{13}	0.5962916 (SD: 0.2150788)	0.6053722 (SD: 0.2183541)
σ_{23}	1.490207 (SD: 0.436213)	1.5129 (SD: 0.4428558)
λ_{11}	1.999746 (SD: 0.1159753)	2.013167 (SD: 0.1167536)
λ_{12}	0.9999034 (SD: 0.1359524)	1.006614 (SD: 0.1368648)
λ_{13}	0.9997189 (SD: 0.1086336)	1.006428 (SD: 0.1093627)
λ_{22}	4.997789 (SD: 0.2901505)	5.031331 (SD: 0.2920978)
λ_{23}	1.998484 (SD: 0.1780972)	2.011896 (SD: 0.1792925)
λ_{33}	2.999204 (SD: 0.1738039)	3.019332 (SD: 0.1749704)

Table 6: Simulation scenario III (continued): same trivariate interval-valued model as in Table 3; Monte Carlo means (with standard deviations in parentheses) of ML and Bayes estimators for $n = 500$.

Parameter	**MLE**	**Bayesian**
	$n = 500$	
μ_1	1.999996 (SD: 0.04464038)	1.999996 (SD: 0.04464038)
μ_2	3.999947 (SD: 0.08934335)	3.999947 (SD: 0.08934335)
μ_3	5.999125 (SD: 0.1331876)	5.999125 (SD: 0.1331876)
σ_1^2	0.9984463 (SD: 0.0627888)	1.004473 (SD: 0.06316781)
σ_2^2	3.992415 (SD: 0.2512253)	4.016514 (SD: 0.2527417)
σ_3^2	8.990107 (SD: 0.5671697)	9.044374 (SD: 0.5705933)
σ_{12}	1.397712 (SD: 0.1085339)	1.406149 (SD: 0.109189)
σ_{13}	0.5989173 (SD: 0.136427)	0.6025325 (SD: 0.1372505)
σ_{23}	1.499177 (SD: 0.2751862)	1.508226 (SD: 0.2768472)
λ_{11}	1.999446 (SD: 0.0730402)	2.004792 (SD: 0.07323549)
λ_{12}	0.9997517 (SD: 0.08515412)	1.002425 (SD: 0.08538181)
λ_{13}	0.999939 (SD: 0.06844347)	1.002613 (SD: 0.06862647)
λ_{22}	4.999755 (SD: 0.1830031)	5.013123 (SD: 0.1834924)
λ_{23}	1.999826 (SD: 0.112514)	2.005173 (SD: 0.1128148)
λ_{33}	3.000691 (SD: 0.1095989)	3.008714 (SD: 0.109892)

5. Real Data

Example 1 (Medical dataset)**.** The data presented in Table 7 represent the range of pulse rate over a day (X_1), the range of systolic blood pressure over the same day (X_2), and the range of diastolic blood pressure over the same day (X_3). These observations were gathered from a sample of 59 patients, each suffering from various illnesses, out of a yearly hospitalized population of 3000. This data was used by Gil et al. (2007).

TABLE 7: Data on the ranges of pulse rate (X_1), systolic (X_2), and diastolic (X_3) blood pressure.

X_1	X_2	X_3	X_1	X_2	X_3
58–90	118–173	63–102	52–78	119–212	47–93
47–68	104–161	71–118	55–84	122–178	73–105
32–114	131–186	58–113	61–101	127–189	74–125
61–110	105–157	62–118	65–92	113–213	52–112
62–89	120–179	59–94	38–66	141–205	69–133
63–119	101–194	48–116	48–73	99–169	53–109
51–95	109–174	60–119	59–98	126–191	60–98
49–78	128–210	76–125	59–87	99–201	55–121
43–67	94–145	47–104	49–82	88–221	37–94
55–102	148–201	88–130	48–77	113–183	55–85
64–107	111–192	52–96	56–133	94–176	56–121
54–84	116–201	74–133	37–75	102–156	50–94
47–95	102–167	39–84	61–94	103–159	52–95
56–90	104–161	55–98	44–110	102–185	63–118
44–108	106–167	45–95	46–83	111–199	57–113
63–109	112–162	62–116	52–98	130–180	64–121
62–95	136–201	67–122	56–84	103–161	55–97
48–107	90–177	52–104	54–92	125–192	59–101
26–109	116–168	58–109	53–120	97–182	54–104
61–108	98–157	50–111	49–88	124–226	57–101
54–78	98–160	47–108	75–124	120–180	59–90
53–103	97–154	60–107	58–99	100–161	54–104
47–86	87–150	47–86	59–78	159–214	99–127
70–132	141–256	77–158	55–89	138–221	70–118
63–115	108–147	62–107	55–80	87–152	50–95
47–83	115–196	65–117	70–105	120–188	53–105
56–103	99–172	42–86	40–80	95–166	54–100
71–121	113–176	57–95	56–97	92–173	45–107
68–91	114–186	46–103	37–86	83–140	45–91
62–100	145–210	100–136			

Despite $\Theta_{i1}^{x_3}$ being slightly deviant from the normality assumption, overall, having p-values 0.14, and 0.08 using Mardia's test or the Shapiro-Wilk test (see Korkmaz et al. (2014)) indicates that there is no significant evidence against multivariate normality of $\boldsymbol{\Theta}_{i1} = (\Theta_{i1}^{x_1}, \Theta_{i1}^{x_2}, \Theta_{i1}^{x_3})^\top$. As a means to verify whether

$$\boldsymbol{\Theta}_{i2} = \begin{bmatrix} \Theta_{i2}^{x_1} & \Theta_{i2}^{x_1 x_2} & \Theta_{i2}^{x_1 x_3} \\ \Theta_{i2}^{x_1 x_2} & \Theta_{i2}^{x_2} & \Theta_{i2}^{x_2 X_3} \\ \Theta_{i2}^{x_1 x_3} & \Theta_{i2}^{x_2 X_3} & \Theta_{i2}^{X_3} \end{bmatrix}$$

is consistent with being drawn from a Wishart distribution, we follow Algorithm 1 with $n = 59$, $p = 3$, degrees of freedom, $df = n - p + 1 = 57$, and ML estimator of Λ from Table 8.

Algorithm 1 Goodness-of-Fit Test for Wishart Distribution

Input: Observed matrices $\{O_1, O_2, ..., O_n\}$, Wishart parameters df and $\hat{\Lambda}$
Output: Test statistic and p-value
Simulate n Wishart-distributed matrices $\{S_1, S_2, ..., S_n\}$ with parameters df and $\hat{\Lambda}$
for $i = 1$ to n **do**
    Compute the sample covariance matrix C_i of O_i
    Reshape C_i into a vector v_i
end for
Reshape each S_i into a vector u_i
Perform a statistical test to compare the distributions of $\{v_1, v_2, ..., v_n\}$ and $\{u_1, u_2, ..., u_n\}$ (Chi-squared test)
Compute the test statistic and p-value
return p-value

Having returned p-value=0.2 does not reject the assumption that $\mathbf{\Theta}_{i2}$ is from a Wishart distribution.

The ML and Bayesian estimators for mean vector $\boldsymbol{\mu}$, variance covariance matrix $\mathbf{\Sigma}$, and scale matrix $\mathbf{\Lambda}$ are presented in Table 8.

TABLE 8: ML and Bayesian estimates of $\boldsymbol{\mu}$, $\mathbf{\Sigma}$, and $\mathbf{\Lambda}$ for the trivariate medical interval-valued dataset in Table 7 (ranges of pulse rate, systolic and diastolic blood pressure; $n = 59$ patients).

Parameter	MLE	Bayesian
$\boldsymbol{\mu}$	$\begin{bmatrix} 74.5169 \\ 146.7034 \\ 83.4491 \end{bmatrix}$	$\begin{bmatrix} 74.5169 \\ 146.7034 \\ 83.4491 \end{bmatrix}$
$\mathbf{\Sigma}$	$\begin{bmatrix} 116.08446 & 27.03893 & 18.16188 \\ 27.03893 & 329.96711 & 149.77729 \\ 18.16188 & 149.77729 & 157.47199 \end{bmatrix}$	$\begin{bmatrix} 122.30327 & 28.48744 & 19.13484 \\ 28.48744 & 347.64392 & 157.80107 \\ 19.13484 & 157.80107 & 165.90799 \end{bmatrix}$
$\mathbf{\Lambda}$	$\begin{bmatrix} 2.7849 & 4.1862 & 3.033 \\ 4.1862 & 7.5626 & 5.1744 \\ 3.0329 & 5.1744 & 3.7496 \end{bmatrix}$	$\begin{bmatrix} 2.7883 & 4.1912 & 3.036 \\ 4.1912 & 7.5716 & 5.1806 \\ 3.0365 & 5.1806 & 3.754 \end{bmatrix}$

Example 2. (Car dataset)

The data in Table 9 provides measurements for 8 different car models. These measurements include four variables: X_1 represents the price of the car in thousands of euros, X_2 denotes the maximum velocity, X_3 indicates the acceleration time required to reach a given speed, and X_4 represents the cylinder capacity of the car. These variables are utilized as per Billard & Diday (2012).

TABLE 9: Cars Data.

Car Model	X_1	X_2	X_3	X_4
Aston Martin	$[260.5, 460.0]$	$[298, 306]$	$[4.7, 5.0]$	$[5935, 5935]$
Audi A6	$[68.2, 140.3]$	$[216, 250]$	$[6.7, 9.7]$	$[1781, 4172]$
Audi A8	$[123.8, 171.4]$	$[232, 250]$	$[5.4, 10.1]$	$[2771, 4172]$
BMW 7	$[104.9, 276.8]$	$[228, 240]$	$[7.0, 8.6]$	$[2793, 5397]$
Ferrari	$[240.3, 391.7]$	$[295, 298]$	$[4.5, 5.2]$	$[3586, 5474]$
Honda NSR	$[205.2, 215.2]$	$[260, 270]$	$[5.7, 6.5]$	$[2977, 3179]$
Mercedes C	$[55.9, 115.2]$	$[210, 250]$	$[5.2, 11.0]$	$[1998, 3199]$
Porsche	$[147.7, 246.4]$	$[280, 305]$	$[4.2, 5.2]$	$[3387, 3600]$

One can easily verify that there is no evidence to reject $\mathbf{\Theta}_{i1} = (\Theta_{i1}^{x_1}, \Theta_{i1}^{x_2}, \Theta_{i1}^{x_3}, \Theta_{i1}^{x_4})^\top$ as a multivariate normal. Due to the small sample size, checking whether

$$\mathbf{\Theta}_{i2} = \begin{bmatrix} \Theta_{i2}^{x_1} & \Theta_{i2}^{x_1x_2} & \Theta_{i2}^{x_1x_3} & \Theta_{i2}^{x_1x_4} \\ \Theta_{i2}^{x_1x_2} & \Theta_{i2}^{x_2} & \Theta_{i2}^{x_2X_3} & \Theta_{i2}^{x_2x_4} \\ \Theta_{i2}^{x_1x_3} & \Theta_{i2}^{x_2X_3} & \Theta_{i2}^{X_3} & \Theta_{i2}^{X_3X_4} \\ \Theta_{i2}^{X_1X_4} & \Theta_{i2}^{X_2X_4} & \Theta_{i2}^{X_3X_4} & \Theta_{i2}^{X_4} \end{bmatrix}$$

follows a Wishart distribution is quite challenging. However, we attempt to modify Algorithm 1 to incorporate bootstrapping by resampling from the observed matrices with replacement to generate additional samples. This can be seen in Algorithm 2, with $n = 8$, $B = 100$, $p = 4$, and $df = 5$.

The results (p-value $=$ 0.33) confirms that we can Wishart distribution assumption is not violated. In Table 9 one can find the ML and Bayesian estimation of the parameters in this example.

TABLE 10: ML and Bayesian estimates of $\boldsymbol{\mu}$, $\mathbf{\Sigma}$, and $\mathbf{\Lambda}$ for the four-dimensional car interval-valued dataset in Table 9 ($n = 8$ car models; price, maximum velocity, acceleration time, and cylinder capacity).

Parameter	MLE	Bayesian
$\boldsymbol{\mu}$	$\begin{bmatrix} 201.4687 \\ 261.75 \\ 6.5437 \\ 3772.25 \end{bmatrix}$	$\begin{bmatrix} 201.4687 \\ 261.75 \\ 6.5437 \\ 3772.25 \end{bmatrix}$
$\mathbf{\Sigma}$	$\begin{bmatrix} 8040.986 & 2268.914 & -109.7983 & 81444.44 \\ 2268.914 & 852.37 & -42.7609 & 19997.9 \\ -109.798 & -42.7609 & 2.191 & -903.314 \\ 81444.44 & 19997.906 & -903.314 & 1002110.312 \end{bmatrix}$	$\begin{bmatrix} 16081.973 & 4537.8281 & -219.596 & 162888.88 \\ 4537.828 & 1704.75 & -85.521 & 39995.81 \\ -219.59 & -85.521 & 4.3830 & -1806.628 \\ 162888.884 & 39995.8125 & -1806.62812 & 2004220.625 \end{bmatrix}$
$\mathbf{\Lambda}$	$\begin{bmatrix} 235.613 & 25.751 & 2.774 & 2222.525 \\ 25.751 & 8.38 & 0.99 & 414.185 \\ 2.774 & 0.99 & 0.145 & 55.38 \\ 2222.525 & 414.185 & 55.38 & 40736.7 \end{bmatrix}$	$\begin{bmatrix} 269.271 & 29.43 & 3.17 & 2540.02 \\ 29.43 & 9.576 & 1.131 & 473.35 \\ 3.17 & 1.131 & 0.165 & 63.3 \\ 2540.03 & 473.354 & 63.3 & 46556.228 \end{bmatrix}$

***Observation* 1.** As can be seen from Example 2, when dealing with a scenario where the sample size is small ($n = 8$) and the dimensionality of the data is relatively large compared to the sample size ($p = 4$), it is common to observe significant divergence between the ML and Bayesian estimators. This discrepancy arises due to several factors. Firstly, in general, the limited amount of data can lead to sparse representations in the high-dimensional space, making it challenging for the ML method to accurately estimate parameters. On the other hand, Bayesian estimation offers advantages in this context. By incorporating prior information about

Algorithm 2 Goodness-of-Fit Test for Wishart Distribution with Bootstrapping

Input: Observed matrices $\{O_1, O_2, ..., O_n\}$, Wishart parameters df and $\hat{\Lambda}$, Bootstrap iterations B
Output: Test statistic and p-value
for $b = 1$ to B **do**
  Randomly sample n matrices with replacement from $\{O_1, O_2, ..., O_n\}$ to obtain $\{O_{b1}, O_{b2}, ..., O_{bn}\}$
  Simulate n Wishart-distributed matrices $\{S_{b1}, S_{b2}, ..., S_{bn}\}$ with parameters df and $\hat{\Lambda}$
  for $i = 1$ to n **do**
    Compute the sample covariance matrix C_{bi} of O_{bi}
    Reshape C_{bi} into a vector v_{bi}
  end for
  Reshape each S_{bi} into a vector u_{bi}
  Perform a statistical test to compare the distributions of $\{v_{b1}, v_{b2}, ..., v_{bn}\}$ and $\{u_{b1}, u_{b2}, ..., u_{bn}\}$ (Chi-squared test)
  Compute the test statistic and p-value for the bth bootstrap iteration
end for
Compute the test statistic and p-value based on the distribution of the bootstrap test statistics
return p-value

the parameters, Bayesian estimation can provide regularization and help stabilize parameter estimates, particularly when the sample size is small. Moreover, Bayesian methods allow for the integration of domain knowledge and uncertainty quantification, which can improve the robustness of the estimates. Overall, this divergence underscores the superiority of Bayesian estimation over MLE in scenarios with small sample sizes and high dimensionality, highlighting the importance of adopting Bayesian approaches when dealing with such data constraints, while ML relies on large sample theory and can have problems in smaller samples.

6. Concluding remarks

So far, we have assumed that values within intervals are uniformly distributed. Hence, the ML and Bayesian estimators of the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ from Theorems 1 and 3, can be expressed in terms of a_{ji} and b_{ji} for $i = 1, \ldots, n$ and $j = 1, \ldots, p$, using equations (3) and (4).

Additionally, there are other cases where we might assume that the internal distribution follows other distributions such as the triangular or Pert distribution. Interested readers are referred to Samadi et al. (2024) for more information.

Note that in all aforementioned cases, we maintain the assumption that $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, the mean and the variance-covariance matrix are multivariate normal and Wishart distributed.

Alternatively, some researchers have recently considered (e.g., Brito & Duarte Silva (2012); Lin et al. (2022); and Beranger et al. (2023)) the skew-normal distribution as a generalization when the normality assumption is not met. In the univariate case, $p = 1$, a Box-Cox transformation

$$\Theta_{i1}^{*x_j} = \begin{cases} \frac{(\Theta_{i1}^{x_1})^{\kappa_1} - 1}{\kappa_1} & \text{if } \kappa_1 \neq 0 \\ \log(\Theta_{i1}^{x_1}) & \text{if } \kappa_1 = 0 \end{cases},$$

or a power transformation

$$\Theta_{i2}^{*x_1} = \begin{cases} (\Theta_{i2}^{x_1})^{\kappa_2} & \text{if } \kappa_2 \neq 0 \\ \log(\Theta_{i2}^{x_1}) & \text{if } \kappa_2 = 0 \end{cases},$$

were suggested by Xu & Qin (2024) to convert $\Theta_{ik}^{x_1}$, for $k = 1, 2$, and $i = 1, \ldots, n$, into $\Theta_{ik}^{*x_1}$ (which is normal for $k = 1$, and exponential for $k = 2$). Parameters κ_k for $k = 1, 2$ vary between $(-3, 3)$, and their optimal values can be found numerically.

In conclusion, this paper comprehensively explored both Bayesian and frequentist approaches for handling multivariate interval-valued data. From a theoretical perspective, we derived ML and Bayesian estimators for the underlying mean vector and covariance structures, established some asymptotic properties of the ML estimators, and compared the procedures under L_2 loss for location parameters and entropy-type losses for scale and dependence parameters. The study involved investigating some asymptotic properties of the ML estimators and comparing ML methods with Bayesian methods in terms of the risk function associated with L_2 loss and entropy losses. From an applied perspective, we assessed the finite-sample performance of the proposed estimators and examined how the Bayesian and ML approaches behave in realistic data-analytic settings. Through extensive simulation studies, we demonstrated the performance of the proposed estimators and illustrated their applicability to real interval-valued datasets. Overall, our findings underscore the significance of leveraging both Bayesian and frequentist methodologies in addressing the complexities of multivariate interval-valued data, providing valuable insights for future research and practical applications in various domains, particularly in areas where interval-valued and symbolic data naturally arise as aggregated summaries of underlying measurements.

Acknowledgments

The authors thank the editor and the anonymous reviewers for their careful reading of the manuscript and their constructive comments, which helped improve the clarity and presentation of this work.

[Received: April 2025 — Accepted: December 2025]

References

Amari, S.-i. (2016), *Information geometry and its applications*, Vol. 194, Springer.

Arroyo, J. & Maté, C. (2009), 'Forecasting histogram time series with k-nearest neighbours methods', *International Journal of Forecasting* **25**(1), 192–207.

Barachant, A., Bonnet, S., Congedo, M. & Jutten, C. (2013), 'Classification of covariance matrices using a Riemannian-based kernel for BCI applications', *Neurocomputing* **112**, 172–178.

Beranger, B., Lin, H. & Sisson, S. (2023), 'New models for symbolic data analysis', *Advances in Data Analysis and Classification* **17**(3), 659–699.

Bertrand, P. & Goupil, F. (2000), Descriptive statistics for symbolic data, *in* 'Analysis of symbolic data: exploratory methods for extracting statistical information from complex data', Springer, pp. 106–124.

Billard, L. (2008), Sample covariance functions for complex quantitative data, *in* 'Proceedings of World IASC Conference, Yokohama, Japan', pp. 157–163.

Billard, L. (2011), 'Brief overview of symbolic data and analytic issues', *Statistical Analysis and Data Mining: The ASA Data Science Journal* **4**(2), 149–156.

Billard, L. & Diday, E. (2003), 'From the statistics of data to the statistics of knowledge: symbolic data analysis', *Journal of the American Statistical Association* **98**(462), 470–487.

Billard, L. & Diday, E. (2012), *Symbolic data analysis: Conceptual statistics and data mining*, John Wiley & Sons.

Bock, H.-H. & Diday, E. (2012), *Analysis of symbolic data: exploratory methods for extracting statistical information from complex data*, Springer Science & Business Media.

Brito, P. & Duarte Silva, A. P. (2012), 'Modelling interval data with normal and skew-normal distributions', *Journal of Applied Statistics* **39**(1), 3–20.

Clark, C. E. (1962), 'The pert model for the distribution of an activity time.', *Operations Research* **10**(3).

Diday, E. (1988), The symbolic approach in clustering and related methods of data analysis: the basic choices, *in* 'Classification and Related Methods of Data Analysis, Proceedings of the First Conference of the International Federation of Classification Societies (IFCS-87: Technical University of Aachen', North Holland, pp. 673–684.

Gil, M. Á., González-Rodríguez, G., Colubi, A. & Montenegro, M. (2007), 'Testing linear independence in linear models with interval-valued data', *Computational Statistics & Data Analysis* **51**(6), 3002–3015.

Irpino, A. & Verde, R. (2006), A new Wasserstein based distance for the hierarchical clustering of histogram symbolic data, *in* 'Data science and classification', Springer, pp. 185–192.

Korkmaz, S., Goksuluk, D. & Zararsiz, G. (2014), 'MVN: an R package for assessing multivariate normality'.

Lauro, C. N. & Palumbo, F. (2000), 'Principal component analysis of interval data: a symbolic data analysis approach', *Computational statistics* **15**(1), 73–87.

Le-Rademacher, J. & Billard, L. (2011), 'Likelihood functions and some maximum likelihood estimators for symbolic data', *Journal of Statistical Planning and Inference* **141**(4), 1593–1602.

Lin, H., Caley, M. J. & Sisson, S. A. (2022), 'Estimating global species richness using symbolic data meta-analysis', *Ecography* **2022**(3), e05617.

Nielsen, F. (2023), 'A simple approximation method for the Fisher–Rao distance between multivariate normal distributions', *Entropy* **25**(4), 654.

Sadeghkhani, A. (2025), 'On multivariate triangular-valued data', *Discover Data* **3**(1), 53.

Sadeghkhani, A. & Sadeghkhani, A. (2025), 'On inference of boxplot symbolic data: applications in climatology', *Advances in Statistical Climatology, Meteorology and Oceanography* **11**(1), 73–87.

Samadi, S. Y., Billard, L., Guo, J.-H. & Xu, W. (2024), 'MLE for the parameters of bivariate interval-valued model', *Advances in Data Analysis and Classification* **18**(4), 827–850.

Skovgaard, L. T. (1984), 'A Riemannian geometry of the multivariate normal model', *Scandinavian journal of statistics* pp. 211–223.

Stein, C. (1956), Some problems in multivariate analysis, part i, Technical report, Department of Statistics, Stanford University.

Xu, M. & Qin, Z. (2024), 'Bayesian framework for interval-valued data using Jeffreys' prior and posterior predictive checking methods', *Communications in Statistics-Simulation and Computation* **53**(5), 2425–2443.

Zhu, J. & Billard, L. (2025), 'Clustering interval-valued data using principal components', *Journal of Statistical Theory and Practice* **19**(4), 78.