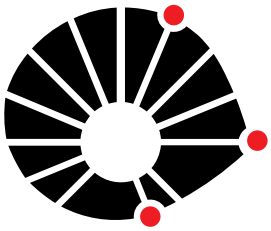


UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

CHRISTIAN EDUARDO GALARZA MORALES

On Moments of doubly truncated multivariate distributions

Momentos de distribuições multivariadas duplamente truncadas

Campinas

2020

Christian Eduardo Galarza Morales

On Moments of doubly truncated multivariate distributions

Momentos de distribuições multivariadas duplamente truncadas

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Estatística.

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Supervisor: Larissa Avila Matos

Co-supervisor: Victor Hugo Lachos Dávila

Este exemplar corresponde à versão final da Tese defendida pelo aluno Christian Eduardo Galarza Morales e orientada pela Profa. Dra. Larissa Avila Matos.

Campinas

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

G131o

Galarza Morales, Christian Eduardo, 1988-
On moments of doubly truncated multivariate distributions / Christian Eduardo Galarza Morales. – Campinas, SP : [s.n.], 2020.

Orientador: Larissa Avila Matos.
Coorientador: Víctor Hugo Lachos Dávila.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Análise multivariada. 2. Distribuição normal assimétrica. 3. Observações censuradas (Estatística). 4. Estimativa de parâmetro. I. Matos, Larissa Avila, 1987-. II. Lachos Dávila, Víctor Hugo, 1973-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Momentos de distribuições multivariadas duplamente truncadas
Palavras-chave em inglês:
Multivariate analysis
Skew-normal distributions
Censored observations (Statistics)
Parameter estimation
Área de concentração: Estatística
Titulação: Doutor em Estatística
Banca examinadora:
Larissa Avila Matos [Orientador]
Filidor Edilfonso Vilca Labra
Silvia Lopes de Paula Ferrari
Elcio Lebensztayn
Clécio da Silva Ferreira
Data de defesa: 23-03-2020
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-4818-6006
- Currículo Lattes do autor: http://lattes.cnpq.br/1340975521059252

Tese de Doutorado defendida em 23 de março de 2020 e aprovada

pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). LARISSA AVILA MATOS

Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA

Prof(a). Dr(a). SILVIA LOPES DE PAULA FERRARI

Prof(a). Dr(a). ELCIO LEBENSZTAYN

Prof(a). Dr(a). CLÉCIO DA SILVA FERREIRA

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

To my beloved wife, Silvia.

Acknowledgements

At first and foremost I would like to thank my invisible God for having been my strength and biggest support throughout the course of this doctorate.

I would like to express my gratitude to everyone who has supported to me during these long four years, in a very especial way to my lovely wife Silvia, who helped me not to lose my mind and to my parents, Víctor and Esther. This achievement is ours.

I am really grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP-Brazil (grants 2015/17110-9 and 2018/11580-1) for the economic support during this PhD, without what, I would not be able to develop this thesis successfully. It would have been impossible to reach my goal without its help.

I also like to express my warm thanks to Larissa Ávila Matos and Víctor Hugo Lachos, my thesis supervisor and co-supervisor respectively, for their aspiring guidance, invaluably constructive criticism and friendship. Thanks again to Mr. Gaudencio Zurita for teaching me that a talent should not be lost.

I shall never forget my entire close friends and teachers during my five academic years here at UNICAMP and one at University of Connecticut-UCONN; sincerely, they have made this journey easier. Finally I wanted to thank to Brazil, an open arms country whose wealth is not in their lands but in its people.

"He aquí mi secreto, que no puede ser más simple:
sólo con el corazón se puede ver bien;
lo esencial es invisible para los ojos."
Antoine de Saint-Exupéry

Resumo

Nesta tese, calculamos os momentos duplamente truncados, ou seja, em um hiper-retângulo, para uma classe geral de distribuições assimétricas denominada família de seleção elíptica multivariada. Essa grande família de distribuições inclui versões assimétricas multivariadas complexas de distribuições elípticas bem conhecidas como as distributições normal, t de Student, exponencial potência, hiperbólica, Slash, Pearson tipo II, normal contaminada, entre outras.

Em quatro capítulos baseados em artigos, apresentamos formulações recorrentes para os momentos de distribuições multivariadas duplamente truncadas e dobradas, expressões explícitas para casos particulares como momentos univariados de ordem inferior, condições suficientes e necessárias para a existência dos momentos truncados, comparação da eficiência computacional entre modelos, estudos de simulação, abordagens otimizadas e aproximações numéricas para casos especiais como casos limites, e momentos quando uma partição tem volume quase zero ou não é truncada. Métodos para realizar estimação em modelos de regressão multivariados assimétricos censurados são apresentados e mostrados através de três aplicações da vida real. Além disso, resultados gerais para distribuições da família mistura de escala normal são apresentados.

Os métodos propostos foram implementados no pacote `MomTrunc` do software `R`, um pacote altamente otimizado que inclui rotinas `C++` por meio do `Rcpp`, que fornece momentos teóricos truncados, momentos Monte Carlo e outras funções de interesse como funções de densidade de probabilidade, distribuições acumuladas e funções geradoras de variáveis aleatórias para várias distribuições multivariadas simétricas e assimétricas.

Palavras-chave: Distribuições elípticas, Distribuições dobradas, Distribuições de seleção, Distribuições truncadas, Momentos truncados

Abstract

In this thesis, we calculate doubly truncated moments, that is, in a hyper-rectangle, for a general class of asymmetric distributions called the selection elliptical family multivariate. This large family of distributions includes complex multivariate asymmetric versions of well-known elliptical distributions as the normal, Student's t, exponential power, hyperbolic, Slash, Pearson type II, contaminated normal, among others.

In four paper-based chapters, we present recurrent formulations for moments of doubly truncated and folded multivariate distributions, explicit expressions for particular cases as univariate lower order moments, sufficient and necessary conditions for the existence of truncated moments, comparison of computational efficiency between models, simulation studies, optimized approaches as well as numerical approximation for special cases such as limiting cases, and moments when a partition has almost zero volume or no truncation. Methods for performing estimation on censored skewed multivariate regression models are presented and showed through three real-life applications. Furthermore, general results for the scale-mixture of normal distributions are presented.

The methods proposed have been implemented in the `MomTrunc` package of the `R` software, a highly optimized package including `C++` routines through `Rcpp`, that offers theoretical, Monte Carlo truncated moments and other functions of interest as probability density, cumulative distribution, and random generator functions for various symmetric and asymmetric multivariate distributions.

Keywords: Elliptical distributions, Folded distributions, Selection distributions, Truncated distributions, Truncated moments

List of Figures

List of Tables

List of Algorithms

List of abbreviations and acronyms

CN	Contaminated normal
EM	Expectation-maximization
ESN	Extended skew-normal
EST	Extended skew-t
FESN	Folded extended skew-normal
FMD	Folded multivariate distribution
FMVN	Folded multivariate normal
FMVT	Folded multivariate t
MGF	Moment generating function
ML	Maximum likelihood
MVN	Multivariate normal
MVT	Multivariate (Student's) t
SE	Selection elliptical
SMN	Scale-mixture of normal
SMSUN	Scale-mixture of unified skew-normal
SN	Skew-normal
ST	Skew-t
SUE	Unified-skew elliptical
SUN	Unified-skew normal
SUT	Unified-skew t
TE	Truncated elliptical
TESN	Truncated extended skew-normal
TMD	Truncated multivariate distribution

TMVN	Truncated multivariate normal
TMVT	Truncated multivariate t
TSE	Truncated multivariate skew-elliptical
TSMN	Truncated scale mixutre of normal
TSUT	Truncated unified-skew t

Contents

Introduction

From a probabilistic point of view, doubly truncated expectations have been a problem of interest for a long time. From the first two one-sided truncated moments for the normal distribution, useful in Tobin's model (Tobin, 1958), its evolution led to its extension to the multivariate case (Tallis, 1961), double truncation (Manjunath & Wilhelm, 2009), heavy tails when considering the Student's t bivariate case in Nadarajah (2007), and finally the first two moments for the multivariate Student's t case in Ho *et al.* (2012).

Truncated expectations are usually used for estimation models in environmental areas, survival analysis, finance, among others. Doubly truncated moments are very important not only to model responses restricted in some interval (for example, ratios, grades, assets portfolio return, etc), but in the context of censored interval models.

Limited or censored data are collected in many studies. This occurs, in several practical situations, for many reasons such as limitations in measuring equipment or from an experimental design. In consequence, the extra true value is recorded only if it falls within an interval range, so, the responses can be either left, interval or right censored. Missing values can be seen just as a particular case.

In addition to censored models, from a frequentist framework, moments are required in step E of the Expectation-Maximization (EM) Dempster *et al.* (1977) algorithm, when we consider the response $\mathbf{Y}_i$, $i = 1, \ldots, n$, being an i.i.d. sample from a given distribution of interest. Knowing these expectations leads to closed EM algorithms, circumventing Monte Carlo methods for estimating the E-step of the algorithm and consequently making possible to fit complex models in a fraction of a time.

We center our attention to the selection elliptical (SE) family of distributions (Arellano-Valle *et al.*, 2006a), a wide class of multivariate asymmetric elliptical distributions. Given the flexibility of this family, we can model features such as asymmetry, heavy tails and multimodality, while interval censoring allows to consider additionally to interval-censoring, missing values (at random) and left censoring for strictly positive responses when considering intervals of the form $(-\infty, \infty)$ and $(0, c]$ respectively. The general results for the SE family, involve all the moments previously used in the literature of censored models in a frequentist point of view. Evidence of applicability and importance of these models, are the three articles already submitted and one more currently in progress. It is worth noting that the `CensMFM` package from `R` (De Alencar *et al.*, 2019b) uses our moments to model finite mixtures of censored or missing multivariate data. We know of at least two jobs that are currently being developed using our moments.

A brief walk-through

This work has been organized in six chapters, where chapters from 2 to 5 are papers (see technical production subsection 6.1 for more details). These last are in chronological order with the purpose of passing the idea of the evolution of the research to the reader.

A recursive approach: The main idea is to compute an arbitrary doubly truncated product moment of a variable $\mathbf{X}$, that is,

$$\mathbb{E}[\mathbf{X}^{\mathbf{k}} \mid \mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b}] = \mathbb{E}[X_1^{k_1}, \ldots, X_p^{k_p} \mid a_1 \leqslant X_1 \leqslant b_1, \ldots, a_p \leqslant X_p \leqslant b_p],$$

using a recursive approach departing from the probability $\mathbb{P}(\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b}) = \mathbb{P}(a_1 \leqslant X_1 \leqslant b_1, \ldots, a_p \leqslant X_p \leqslant b_p)$ as initial condition. Depending of the distribution of $\mathbf{X}$, the probability above can be hard to compute. For the normal and Student's t distribution, there exists efficient methods well implemented and available in most statistical softwares; for instance, in `R` language, the `mvtnorm` (Genz & Bretz, 2009; Genz *et al.*, 2020) and `tlrmvnmvt` (Cao *et al.*, 2019a,b) packages, this last being released lately in November 2019.

Recursion is developed by establishing a differential equation that involves the density of $\mathbf{X}$ in question (see Kan & Robotti (2017)). This recursion allows calculating the moments for truncated distributions which may have a complex moment generating function (MGF) and cannot be treated by differentiation. Based on this recursive approach were written two articles for the doubly truncated Student's t and the extended skew-normal (ESN) distribution, which are resumed in chapters 2 and 3 respectively. Also, interesting general results for the scale mixture of normal distribution family using this recursion can be found in Chapter 1. Notice that the ESN distribution includes the well-known skew-normal (SN) distribution as particular case. Both articles aforementioned also considered the moments for the positive multivariate variable $|\mathbf{X}|$, that is, its folded version.

In Chapter 3, we additionally established a 1-1 relationship between the moments of a truncated ESN distribution and the moments of a truncated normal distribution. This led to a more efficient (and faster) algorithm since the number of required integrals is smaller. Chapter 4 proposed estimation on interval-censored models for skew-normal responses based on this last approach. Naturally, next step was to move forward to the extended skew-t (EST) distribution (Arellano-Valle & Genton, 2010) in order to incorporate heavy tails, however, since ESN distribution results to be a member of the SLCT-EC family, it was possible to write a general result for this class, which contains the EST distribution itself.

A 1-1 relation: Assume that $\mathbf{Y}$ follows a distribution belonging to the SE class. Then, we are able to compute any arbitrary moment of $\mathbf{Y} \mid (\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b})$, that is, a doubly

truncated selection elliptical distribution just using an unique corresponding moment of a doubly truncated elliptical distribution $\mathbf{X} \mid (\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta})$, its symmetric version.

For instance, consider $\mathbf{Y}$ following a unified skew-t (SUT) distribution, a complex multivariate asymmetric heavy-tailed distribution which includes the EST and the skew-t (ST) distribution (Azzalini & Capitanio, 2003). Then, the first and second truncated moment of $\mathbf{Y}$ can be calculated only using the first and second truncated moment of a symmetric Student's t distribution $\mathbf{X}$, say, which moments were already proposed in Chapter 2. It is worth mentioning that for any of the t distributions above their normal analogous versions (the unified skew-normal (SUN), ESN and the SN distribution) are retrieved when $\nu \uparrow \infty$.

This 1-1 relation is highly convenient since doubly truncated moments for some members of the elliptical family of distributions are already available in the literature and statistical softwares. For this reason, this thesis focuses mainly in complex asymmetric versions of the normal and Student's t distributions which probabilities of the form $\mathbb{P}(\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b})$ are available. Chapter 5 summarized this general results given some emphasis to the doubly truncated SUT distribution, embedding all theoretical results in chapters before.

Implementation: All methods described above have been coded in the `R` package `MomTrunc` (Galarza *et al.*, 2018). The package is able to calculate $\mathbb{E}[\mathbf{X}^{\mathbf{k}} \mid \mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b}]$ even for extreme cases as when the probability $\mathbb{P}(\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b}) \approx 0$ due to extreme parameter settings, integration limits, or even the numerical precision of the machine. For example, the only other package that calculates truncated moments for the normal case, the `tmvtnorm` package (Wilhelm & Manjunath, 2015), under the extreme conditions mentioned above, it returns values `NaNs` and even negative variances. In particular, in contrast with the `TTmoment` package (Ho *et al.*, 2015) for Student's t case, the package is capable of calculating the moments for degrees of freedom $\nu < 5$ and even decimals, for example $\nu = 2.17$. It worth mentioning that moments for a double truncated variable always exist (for Student t case, for $\nu > 0$), since it is limited. In addition to the moments, our package provides $\mathbb{P}(\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b})$ probabilities for different members of the multivariate SE family, including the option of returning the logarithm in base 2, useful when true probability is much less than the precision of the machine.

Algorithms for the normal case have been coded in `R` from their original versions in `Matlab` available in Kan & Robotti (2017). To the best of our knowledge, until the beginning of 2018, there were only one available package in `R` offering doubly truncated moments for the normal distribution (`tmvtnorm`) and only one for the Student's t case (`TTmoment`). Since its release in February 2018, the `MomTrunc` package has been downloaded almost 9000 times, a significant number considering that this is a specialized package.

Structure of the thesis

The organization of the thesis is as follows:

Chapter 1: We provide some background material. We review some definitions, methodologies and we describe some datasets used throughout the thesis.

Chapter 2: This Chapter develops recurrence relations for integrals that involve the density of multivariate Student's t distributions. The proposed techniques allow for fast computation of arbitrary-order product moments of folded and truncated multivariate Student's t distributions and offer explicit expressions of their low-order moments. We propose an optimized algorithm that outperforms other methods in the literature, which can deal with missing data in the response at not computational cost. The usefulness and effectiveness of the proposed techniques are demonstrated through both simulated and real data, where we show its usefulness on censored regression models with missing data.

Chapter 3: We extend the recurrence approach to integrals related to asymmetric multivariate densities. Specifically, we compute recurrence relations involving the density of the ESN distribution, including the well-known SN distribution introduced by Azzalini & Dalla-Valle (1996) and the popular multivariate normal distribution. These recursions offer a fast computation of arbitrary order product moments of the multivariate truncated ESN and multivariate folded ESN (FESN) 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 result 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.

Chapter 4: The need for asymmetric distributions for the random errors on linear censored models, motivate us to develop a likelihood-based inference for linear models with censored responses based on the multivariate SN distribution. Most linear and nonlinear regression models used to analyze censored data 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 proposed EM algorithm for maximum likelihood estimation uses closed-form expressions at the E-step, that are based on formulas for the mean and variance of a truncated multivariate skew-normal distribution, computed in the Chapter before. Three datasets with censored and/or missing observations are analyzed and discussed.

Chapter 5: We generalize all results before to the class of asymmetric distributions called the selection elliptical (SE) family of distributions, a family including complex multivariate asymmetric versions of well-known elliptical distributions as the normal, Student's t, among others. We address the moments for doubly truncated members of this family, establishing neat formulation for high order moments as well as for its first two moments. We establish sufficient and necessary conditions for the existence of these truncated moments. Also, we propose optimized methods able to deal with extreme setting of the parameters, partitions with almost zero volume or no truncation. A brief numerical study is presented in order to validate the methodology. A direct application of ST truncated moments is developed in the context of risk measurement in Finance. Useful expressions in censored modeling are presented, which have been particularized to the SUT distribution, a complex multivariate asymmetric heavy-tailed distribution which includes the EST (ESN) and ST (SN) distribution as particular cases. Finally, we conclude the chapter proposing estimation on interval-censored models for skew-t responses based on this last expressions.

Chapter 6: We present some final remarks, technical production and further researches related to this thesis.

1 Preliminaries

We begin our exposition by defining the notation and presenting some basic concepts and some useful results which are used throughout the development of this thesis. As is usual in probability theory and its applications, we denote a random variable by an upper-case letter and its realization by the correspondent lower case and use boldface letters for vectors and matrices. Let $\mathbf{I}_p$ represent a $p \times p$ identity matrix, $\mathbf{A}^\top$ be the transpose of $\mathbf{A}$, and $|\mathbf{X}| = (|X_1|. \ldots, |X_p|)^\top$ denote the absolute value of each component of the vector $\mathbf{X}$. For multiple integrals, we use the shorthand notation

$$\int_{\mathbf{a}}^{\mathbf{b}} f(\mathbf{x}) \mathrm{d}\mathbf{x} = \int_{a_1}^{b_1} \ldots \int_{a_p}^{b_p} f(x_1, \ldots, x_p) \mathrm{d}x_p \ldots \mathrm{d}x_1.$$

where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$. For two p-dimensional random vectors $\mathbf{x} = (x_1, \ldots, x_p)^\top$ and $\boldsymbol{\kappa} = (k_1, \ldots, k_p)^\top$, let $\mathbf{x}^{\boldsymbol{\kappa}}$ stand for $(x_1^{\kappa_1}, x_2^{\kappa_2}, \ldots, x_p^{\kappa_p})$. General results to compute the probability of a random vector lying in a hyper-rectangle are summarized in the following results.

Lemma 1.1. *Let $\mathbf{X}$ be a p-variate random vector with joint probability density function (pdf) $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ and cumulative density function (cdf) $F_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$. Let $\mathbb{A}$ be a Borel set in $\mathbb{R}^p$ of the form*

$$\mathbb{A} = \{(x_1, \ldots, x_p) \in \mathbb{R}^p : \ a_1 \leqslant x_1 \leqslant b_1, \ldots, a_p \leqslant x_p \leqslant b_p\} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{a} \leqslant \mathbf{x} \leqslant \mathbf{b}\}. \quad (1.1)$$

Then $\mathbb{P}(\mathbf{X} \in \mathbb{A}) = \sum_{\mathbf{s} \in S(\mathbf{a}, \mathbf{b})} (-1)^{n_s} F_{\mathbf{X}}(\mathbf{s}; \boldsymbol{\theta})$, where $S(\mathbf{a}, \mathbf{b}) = \{\mathbf{s} : \mathbf{s} = (s_1, \ldots, s_p)$ with $s_i = \{a_i, b_i\}$, $i = 1, \ldots, p\}$ and $n_s = \sum_{i=1}^{p} \mathbb{1}(s_i = a_i)$ with $\mathbb{1}(\cdot)$ being the indicator function.

Proof. Based on the inclusion-exclusion principle, the probability $\mathbb{P}(\mathbf{X} \in \mathbb{A}) = \mathbb{P}(\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b})$ can be computed by summing the 2^p terms corresponding to the $\mathbf{s}$ elements in the solution space of $S(\mathbf{a}, \mathbf{b})$, where the term signs depend on the number of a's elements in the vector $\mathbf{s}$, i.e., n_s. □

Theorem 1.1. *Let $\mathbf{X}$ be a p-variate random vector with joint pdf $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ and joint cdf $F_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$. If $\mathbf{Y} = |\mathbf{X}|$, then the joint pdf and cdf of $\mathbf{Y}$ that follows a folded distribution are given, respectively, by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} f_{\mathbf{X}}(\boldsymbol{\Lambda}_s \mathbf{y}; \boldsymbol{\theta}), \quad \text{for } \mathbf{y} \geqslant \mathbf{0},$$

and $F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} \pi_s F_{\mathbf{X}}(\boldsymbol{\Lambda}_s \mathbf{y}; \boldsymbol{\theta})$, where $S(p) = \{\mathbf{s} : \mathbf{s} = (s_1, \ldots, s_p)$, with $s_i = \pm 1, i = 1, \ldots, p\}$, $\boldsymbol{\Lambda}_s = Diag(\mathbf{s})$ and $\pi_s = \prod_{i=1}^{p} s_i$.

Proof. The distribution function $F_{\mathbf{Y}}(\mathbf{y})$ can be calculated as a particular case of Lemma 1.1, when $\mathbf{a} = -\mathbf{y}$ and $\mathbf{b} = \mathbf{y}$. It follows that

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{y}) &= \mathbb{P}(-\mathbf{y} \leqslant \mathbf{X} \leqslant \mathbf{y}) \\ &= \mathbb{P}(-y_1 \leqslant X_1 \leqslant y_1, -y_2 \leqslant X_2 \leqslant y_2, \ldots, -y_p \leqslant X_p \leqslant y_p) \\ &= F_{\mathbf{X}}(\mathbf{y}) - \sum_i F_{\mathbf{X}}(\mathbf{y}_{\text{-}(i)}) + \sum_{i<j} F_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j)}) - \sum_{i<j<k} F_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j,k)}) + \ldots + (-1)^p F_{\mathbf{X}}(-\mathbf{y}), \end{aligned} \tag{1.2}$$

where $\mathbf{y}_{-(i)}$ denotes the $\mathbf{y}$ vector with its ith elements multiplied by -1. For instance, we have that $\mathbf{y}_{\text{-}(i)} = (y_1, y_2, \ldots, y_{i-1}, -y_i, y_{i+1}, \ldots, y_p)$. It is easy to see that $F_{\mathbf{Y}}(\mathbf{y})$ can be written as $F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} \pi_s F_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y}; \boldsymbol{\theta})$, with the constant $\pi_s = \prod_{i=1}^{p} s_i$ providing the signs $(-1, 1)$ correctly for each summand in (1.2). As a result, we have the joint pdf of $\mathbf{Y} = |\mathbf{X}|$ given by

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{\partial^p}{\partial y_1 \partial y_2 \ldots \partial y_p} F_{\mathbf{Y}}(\mathbf{y}) \\ &= f_{\mathbf{X}}(\mathbf{y}) - (-1) \sum_i f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i)}) + (-1)^2 \sum_{i<j} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j)}) - (-1)^3 \sum_{i<j<k} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j,k)}) \\ &\quad + \cdots + (-1)^{2p} f_{\mathbf{X}}(-\mathbf{y}) \\ &= f_{\mathbf{X}}(\mathbf{y}) + \sum_i f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i)}) + \sum_{i<j} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j)}) + \sum_{i<j<k} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j,k)}) + \ldots + f_{\mathbf{X}}(-\mathbf{y}) \\ &= \sum_{\mathbf{s} \in S(p)} f_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y}; \boldsymbol{\theta}). \end{aligned}$$

Note that we have conveniently used $f_{\mathbf{X}}(\mathbf{x})$ instead of $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ for simplicity. □

Corollary 1.1. *If $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\xi}, \mathbf{\Psi})$ belongs to the location-scale family of distributions with location and scale parameters $\boldsymbol{\xi}$ and $\mathbf{\Psi}$ respectively, then the joint pdf and cdf of $\mathbf{Y} = |\mathbf{X}|$ are given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} f_{\mathbf{X}}(\mathbf{y}; \mathbf{\Lambda}_s \boldsymbol{\xi}, \mathbf{\Lambda}_s \mathbf{\Psi} \mathbf{\Lambda}_s), \quad \textit{for } \mathbf{y} \geqslant \mathbf{0},$$

and

$$F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} \pi_s F_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y}; \boldsymbol{\xi}, \mathbf{\Psi}).$$

Proof. By using the change-of-variable method for $\mathbf{Z}_s = \mathbf{\Lambda}_s \mathbf{X}$, then $f_{\mathbf{Z}_s}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y})$ since $\mathbf{\Lambda}_s^{-1} = \mathbf{\Lambda}_s$, $\mathbf{J} = \mathbf{\Lambda}_s$ and $|\det(\mathbf{J})| = 1$, where $\mathbf{J}$ is the Jacobian matrix of the transformation. Additionally, if $\mathbf{X} \sim f_{\mathbf{X}}(\cdot; \boldsymbol{\xi}, \mathbf{\Psi})$ belongs to the location-scale family of distributions with location and scale parameters $\boldsymbol{\xi}$ and $\mathbf{\Psi}$, respectively, then $\mathbf{Z}_s = \mathbf{\Lambda}_s \mathbf{X} \sim f_{\mathbf{X}}(\mathbf{z}; \mathbf{\Lambda}_s \boldsymbol{\xi}, \mathbf{\Lambda}_s \mathbf{\Psi} \mathbf{\Lambda}_s)$. By Theorem 1.1, we obtain $f_{\mathbf{Y}}(\mathbf{y})$ and $F_{\mathbf{Y}}(\mathbf{y})$ accordingly. □

Corollary 1.1 generalizes the results of Chakraborty & Chatterjee (2013) for the folded multivariate normal (FMVN) case to all distributions belong to the multivariate location-scale family.

Corollary 1.2. *Under the same conditions of Corollary 1.1, we have that*

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^{+\boldsymbol{\kappa}}], \quad \textit{where } \mathbf{X}^{+} = \mathbf{X}\cdot\mathbb{1}(\mathbf{X} > \mathbf{0}).$$

Proof. By the simple property of probability theory, we can deduce that

$$\begin{aligned}\int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y} &= \sum_{\mathbf{s}\in S(p)} \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{X}}(\mathbf{y}; \boldsymbol{\Lambda}_s\boldsymbol{\xi}, \boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s)\mathrm{d}\mathbf{y} \\ &= \sum_{\mathbf{s}\in S(p)} \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{Z}_s}(\mathbf{y})\mathrm{d}\mathbf{y} \\ &= \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^{+\boldsymbol{\kappa}}].\end{aligned}$$

□

1.1 Moments of doubly truncated scale mixture of normal distributions

1.1.1 Scale mixture of normal distributions (SMN)

An element of the symmetrical class of scale mixture of multivariate normal distributions (Andrews & Mallows, 1974; Lange & Sinsheimer, 1993) is defined as the distribution of the p-variate random vector

$$\mathbf{y} = \boldsymbol{\mu} + \zeta(U)^{1/2}\mathbf{Z}, \tag{1.3}$$

where $\boldsymbol{\mu}$ is a location vector, $\mathbf{Z}$ is a normal random vector with mean vector $\mathbf{0}$, variance–covariance matrix $\boldsymbol{\Sigma}$, U is a positive random variable with cumulative distribution function (cdf) $\mathrm{H}(u;\boldsymbol{\nu})$ and probability density function (pdf) $h(u;\boldsymbol{\nu})$, independent of $\mathbf{Z}$, where $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of U and $\zeta(\cdot)$ is the weight function. Note that given $U = u$, $\mathbf{y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\zeta(u)\boldsymbol{\Sigma}$. Hence, the pdf of $\mathbf{y}$ is given by

$$SMN_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu}) = \int_0^{\infty} \phi_p(\mathbf{y};\boldsymbol{\mu},\zeta(u)\boldsymbol{\Sigma})d\mathrm{H}(u;\boldsymbol{\nu}), \tag{1.4}$$

where $\phi_p(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ stands for the pdf of the p–variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$. We use the notation $\mathrm{SMN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma};\mathrm{H})$ when $\mathbf{Y}$ has distribution in the SMN class.

Three scale mixture of normal distributions are commonly used for robust estimation which share same weight function $\zeta(u) = 1/u$:

- *The multivariate Student–t distribution*, $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where ν is called the degrees of freedom, can be derived from the mixture model (1.3), where U is distributed as $Gamma(\frac{\nu}{2}, \frac{\nu}{2})$, with $\nu > 0$. The pdf of $\mathbf{y}$ takes the following hierarchical form

$$t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) h_U(u) \mathrm{d}u. \tag{1.5}$$

  been equivalent to (2.1).

- *The multivariate slash distribution*, $\mathrm{SL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, arises when the distribution of U is $Beta(\nu, 1)$, with $u \in (0, 1)$ and $\nu > 0$. Its pdf is given by

$$SL_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) du, \quad \mathbf{y} \in \mathbb{R}^p,$$

  and can be evaluated through numerical method, for example, using the `R` function *integrate*.

- *The multivariate contaminated normal distribution*, $CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \rho)$, where $\nu, \rho \in (0, 1)$. Here, U is a discrete random variable taking one of two states and with probability function given by

$$h(u; \boldsymbol{\nu}) = \nu \mathbb{1}\{u = \rho\} + (1 - \nu)\mathbb{1}\{u = 1\},$$

  where $\boldsymbol{\nu} = (\nu, \rho)$. The associated density is

$$CN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \nu \phi_p(\mathbf{y}; \boldsymbol{\mu}, \rho^{-1}\boldsymbol{\Sigma}) + (1 - \nu)\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

  The parameter ν can be interpreted as the proportion of outliers while ρ may be interpreted as a scale factor.

Now, let $\mathrm{TSMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{H}; \mathbb{A})$ represent a p-variate truncated SMN (TSMN) distribution for $\mathrm{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathrm{H})$ lying in the hyper-rectangle $\mathbb{A}$ as defined in (1.1). We may also use the notation $\mathrm{TSMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ for simplicity. Specifically, we say that the p-dimensional vector $\mathbf{X} \sim \mathrm{TSMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{H}; \mathbb{A})$, if its density is given by:

$$TSMN_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}; \mathbb{A}) = \frac{SMN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{y}}, \ \mathbf{a} \leqslant \mathbf{x} \leqslant \mathbf{b}. \tag{1.6}$$

1.1.2 A recursive approach for TSMN moments

Let suppose that $\mathbf{Y} \sim \mathrm{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathrm{H})$. From (1.4) we have that the density function of $\mathbf{Y}$, $f_{\mathbf{Y}}(\mathbf{y}) \triangleq SMN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \mathbb{E}_U \left[\phi_p(\mathbf{y}; \boldsymbol{\mu}, \zeta(U)\boldsymbol{\Sigma})\right]. \tag{1.7}$$

Kan & Robotti (2017) proposed a recurrence relation for the moments of a multivariate normal (MVN) distribution based on a differential equation of its pdf $\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})$, that is

$$-\frac{\partial}{\partial \mathbf{y}}\boldsymbol{\phi}_p\left(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}\right) = \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}\right), \tag{1.8}$$

which is obtained by multiplying both sides by $\mathbf{y}^{\boldsymbol{\kappa}}$ and then integrating both sides from $\mathbf{a}$ to $\mathbf{b}$ with respect to $\mathbf{y}$. Following the same exercise for the pdf of $\mathbf{Y}$, say $f_{\mathbf{Y}}(\mathbf{y})$, it follows from the equation above that

$$\begin{aligned} -\frac{\partial}{\partial \mathbf{y}} f_{\mathbf{Y}}(\mathbf{y}) &= \mathbb{E}_U\Big[\tfrac{\partial}{\partial \mathbf{y}}\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},\zeta(U)\boldsymbol{\Sigma}\right)\Big] \\ &= \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\,\mathbb{E}_U\big[\zeta(U)^{-1}\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},\zeta(U)\boldsymbol{\Sigma}\right)\big] \\ &= \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\,\mathbb{E}_U\big[\zeta(U)^{-1} f_{\mathbf{Y}|U}(\mathbf{y})\big], \end{aligned} \tag{1.9}$$

with $\mathbf{Y}|U=u \sim \mathrm{N}_p(\boldsymbol{\mu},\zeta(u)\boldsymbol{\Sigma})$. Note that, the derivative and the expectation in the first line of (1.9) can be interchanged due to the Leibniz rule. As seen, the right side depends on the mixture variable U. Next, we compute $\mathbb{E}_U\big[\zeta(U)^{-1} f_{\mathbf{Y}|U}(\mathbf{y})\big]$ for particular cases of interest.

1.1.2.1 Particular Cases

1. *Multivariate normal distribution:* This is trivial , since U is a degenerated random variable in 1, that is, $\mathbb{P}(U=1)=1$. Then $\zeta(u)^{-1} f_{\mathbf{Y}|U}(\mathbf{y}) = f_{\mathbf{Y}}(\mathbf{y})$ and consequently we recover expression (1.8).

2. *Multivariate Student-t distribution:* For $U \sim \text{Gamma}(\frac{\nu}{2},\frac{\nu}{2})$ and a weight function $\zeta(u)=u^{-1}$, it follows from equation (1.9) that

$$\begin{aligned} -\frac{\partial}{\partial \mathbf{y}} t_p\left(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu\right) &= \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\,\mathbb{E}_U[U\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},U^{-1}\boldsymbol{\Sigma}\right)] \\ &= \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\, t_p\big(\mathbf{y};\boldsymbol{\mu},\tfrac{\nu}{\nu+2}\boldsymbol{\Sigma},\nu+2\big). \end{aligned} \tag{1.10}$$

Proof. Setting $\zeta(u)=u^{-1}$, with $U \sim Gamma(\frac{\nu}{2},\frac{\nu}{2})$, it follows that,

$$\begin{aligned} \mathbb{E}_U\left[U\phi_p\left(\mathbf{y};\boldsymbol{\mu},U^{-1}\boldsymbol{\Sigma}\right)\right] &= \int_0^\infty \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} u^{\frac{\nu}{2}} e^{-\frac{\nu}{2}u}\,\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma}\right)\mathrm{d}u \\ &= \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}}\Gamma\left(\tfrac{p+\nu+2}{2},\tfrac{\delta(\mathbf{y})+\nu}{2}\right), \\ &= \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}\left(\frac{\nu+2}{\nu}\right)^{\frac{p+\nu+2}{2}}}{\Gamma\left(\frac{\nu}{2}\right)|2\pi\boldsymbol{\Sigma}|^{\frac{1}{2}}}\Gamma\left(\tfrac{p+\nu+2}{2},\tfrac{\nu+2}{\nu}\left(\tfrac{\delta(\mathbf{y})+\nu}{2}\right)\right) \end{aligned}$$

where $\delta(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ is the Mahalanobis distance and $\Gamma(\alpha,\lambda)$ represents the two parameter Gamma function,

$$\Gamma(\alpha,\lambda) = \int_0^\infty u^{\alpha-1}\exp(-\lambda u)\mathrm{d}u.$$

After some algebra we obtain,

$$\begin{aligned}
\mathbb{E}_U\left[U\phi_p\left(\mathbf{y};\boldsymbol{\mu},U^{-1}\boldsymbol{\Sigma}\right)\right] &= \int_0^\infty \frac{1}{(2\pi)^{\frac{p}{2}}\left|(\frac{\nu}{\nu+2})u^{-1}\boldsymbol{\Sigma}\right|^{\frac{1}{2}}}\exp\left\{-\tfrac{u}{2}\left(\tfrac{\nu+2}{\nu}\right)\delta(\mathbf{y})\right\} \\
&\quad \times\frac{\left(\frac{\nu+2}{2}\right)^{\frac{\nu+2}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}u^{\frac{\nu+2}{2}-1}\exp\left\{-\left(\tfrac{\nu+2}{2}\right)u\right\}\mathrm{d}u, \\
&= \int_0^\infty \boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},u^{-1}\tfrac{\nu}{\nu+2}\boldsymbol{\Sigma}\right)h_V(u)\mathrm{d}u, \\
&= t_p\big(\mathbf{y};\boldsymbol{\mu},\tfrac{\nu}{\nu+2}\boldsymbol{\Sigma},\nu+2\big),
\end{aligned}$$

where $V \sim Gamma\left(\frac{\nu+2}{2},\frac{\nu+2}{2}\right)$. This completes the proof. □

3. *Multivariate Slash distribution:* For this case, we have that $U \sim \text{Beta}(\nu,1)$ and same weight function $\zeta(u) = u^{-1}$. Then,

$$\begin{aligned}
\mathbb{E}_U[U\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},U^{-1}\boldsymbol{\Sigma}\right)] &= \int u\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma}\right)h(u)\mathrm{d}u \\
&= \int_0^1 \nu u^{(\nu+1)-1}\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma}\right)\mathrm{d}u \\
&= \frac{\nu}{\nu+1}SL(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu). \qquad (1.11)
\end{aligned}$$

Hence,

$$-\frac{\partial}{\partial\mathbf{y}}SL_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{\nu}{\nu+1}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})SL_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu+1). \qquad (1.12)$$

4. *Multivariate Contaminated normal (CN) distribution:* Since the pdf of $\mathbf{Y}$ is a finite mixture of two normal densities, it follows directly that

$$\begin{aligned}
-\frac{\partial}{\partial\mathbf{y}}CN_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu}) &= -\frac{\partial}{\partial\mathbf{y}}\left[\nu\phi_p(\mathbf{y};\boldsymbol{\mu},\rho^{-1}\boldsymbol{\Sigma})+(1-\nu)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\right] \\
&= \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\left[\nu\rho\phi_p(\mathbf{y};\boldsymbol{\mu},\rho^{-1}\boldsymbol{\Sigma})+(1-\nu)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\right].
\end{aligned}$$

Note that $\mathbb{E}_U[U\boldsymbol{\phi}_p\left(\mathbf{y};\boldsymbol{\mu},U^{-1}\boldsymbol{\Sigma}\right)] = \nu\rho\phi_p(\mathbf{y};\boldsymbol{\mu},\rho^{-1}\boldsymbol{\Sigma})+(1-\nu)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})$, where this last is not proportional to a CN density. This breaks the recurrence relation of CN moments; however, it is easy to realize that any CN moment is a finite mixture of normal moments as well.

1.1.3 Mean and Covariance Matrix of Truncated Multivariate SMN distributions

Let $\mathbf{Z} \sim \mathrm{TN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbb{A})$ be a doubly truncated multivariate normal over the truncation region $\mathbb{A}$, that is with truncation limits $\mathbf{a}$ and $\mathbf{b}$. Kan & Robotti (2017) showed that

$$\begin{aligned}\mathbb{E}[Z_i] = \mu_i + \frac{1}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})} \sum_{j=1}^{p} \sigma_{ij} [\phi_1(a_j; \mu_j, \sigma_j^2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}, \tilde{\boldsymbol{\mu}}_j^a, \tilde{\boldsymbol{\Sigma}}_j) \\ - \phi_1(b_j; \mu_j, \sigma_j^2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}, \tilde{\boldsymbol{\mu}}_j^b, \tilde{\boldsymbol{\Sigma}}_j)], \quad i = 1, \ldots, p, \end{aligned} \tag{1.13}$$

where

$$\tilde{\boldsymbol{\mu}}_j^a = \boldsymbol{\mu}_{(j)} + \boldsymbol{\Sigma}_{(j),j} \frac{a_j - \mu_j}{\sigma_j^2}, \tag{1.14}$$

$$\tilde{\boldsymbol{\mu}}_j^b = \boldsymbol{\mu}_{(j)} + \boldsymbol{\Sigma}_{(j),j} \frac{b_j - \mu_j}{\sigma_j^2}, \tag{1.15}$$

$$\tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}_{(j),(j)} - \frac{1}{\sigma_j^2} \boldsymbol{\Sigma}_{(j),j} \boldsymbol{\Sigma}_{j,(j)}. \tag{1.16}$$

and $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathbf{a}}^{\mathbf{b}} \phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{x}$.

Let $\mathbf{W} \sim \mathrm{TSMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{H}; \mathbb{A})$ be a truncated multivariate SMN distribution with density function as in 1.6. Then, its mean is given by

$$\begin{aligned}\mathbb{E}[\mathbf{W}] &= \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{x}} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{w}\, SMN_p(\mathbf{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{w} \\ &= \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{x}} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{w} \int_0^{\infty} \phi_p(\mathbf{w}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) h_U(u) \mathrm{d}u \mathrm{d}\mathbf{w} \\ &= \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{x}} \int_0^{\infty} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{w} \phi_p(\mathbf{w}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \mathrm{d}\mathbf{w} h_U(u) \mathrm{d}u, \end{aligned}$$

where we have used expression (1.4) and Fubini's rule.

Noting that, $\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x} \phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\mathbf{x} = L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathbb{E}[\mathbf{Z}]$, it follows from (1.13) that

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{w} \phi_p(\mathbf{w}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \mathrm{d}\mathbf{w} = \boldsymbol{\mu} L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) + u^{-1}\boldsymbol{\Sigma}\mathbf{d},$$

where the j-th element of $\mathbf{d}$ is given by

$$\begin{aligned} d_j = \phi_1(a_j; \mu_j, u^{-1}\sigma_j^2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^a, u^{-1}\tilde{\boldsymbol{\Sigma}}_j) \\ - \phi_1(b_j; \mu_j, u^{-1}\sigma_j^2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^b, u^{-1}\tilde{\boldsymbol{\Sigma}}_j). \end{aligned}$$

It follows that

$$\mathbb{E}[\mathbf{W}] = \boldsymbol{\mu} + \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{x}} \boldsymbol{\Sigma}(\mathbf{q}_a^H - \mathbf{q}_b^H), \tag{1.17}$$

where the j-th element of $\mathbf{q}_a^H$ and $\mathbf{q}_b^H$ are

$$q_{a,j}^H = \mathbb{E}_U[U^{-1}\phi_1(a_j;\mu_j,U^{-1}\sigma_j^2)L_{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^a,U^{-1}\tilde{\boldsymbol{\Sigma}}_j)], \tag{1.18}$$

$$q_{b,j}^H = \mathbb{E}_U[U^{-1}\phi_1(b_j;\mu_j,U^{-1}\sigma_j^2)L_{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^b,U^{-1}\tilde{\boldsymbol{\Sigma}}_j)]. \tag{1.19}$$

Furthermore, let $\mathbf{e}_i = [\mathbf{0}_{i-1}, 1, \mathbf{0}_{p-i}]$, that is, a vector of zeros with a 1 in the ith position. Hence,

$$\begin{aligned}
\mathbb{E}[\mathbf{W}^{\boldsymbol{\kappa}+\mathbf{e}_i}] &= \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x}}\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{w}^{\boldsymbol{\kappa}+\mathbf{e}_i}SMN_p(\mathbf{w};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{w}\\
&= \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x}}\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{w}^{\boldsymbol{\kappa}+\mathbf{e}_i}\int_0^{\infty}\phi_p(\mathbf{w};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})h_U(u)\mathrm{d}u\mathrm{d}\mathbf{w}\\
&= \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x}}\int_0^{\infty}\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{w}^{\boldsymbol{\kappa}+\mathbf{e}_i}\phi_p(\mathbf{w};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})\mathrm{d}\mathbf{w}h_U(u)\mathrm{d}u.
\end{aligned}$$

From Kan & Robotti (2017) (Theorem 1), we have

$$\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{w}^{\boldsymbol{\kappa}+\mathbf{e}_i}\phi_p(\mathbf{w};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})\mathrm{d}\mathbf{w} = \mu_i\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{w}^{\boldsymbol{\kappa}}\phi_p(\mathbf{w};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})\mathrm{d}\mathbf{w} + u^{-1}\mathbf{e}_i^{\top}\boldsymbol{\Sigma}\mathbf{c}_{\boldsymbol{\kappa}},$$

where $\mathbf{c}_{\boldsymbol{\kappa}}$ is an p-vector with j-th element

$$\begin{aligned}
c_{\boldsymbol{\kappa},j} &= -\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{w}^{\boldsymbol{\kappa}}\frac{\partial\phi_p(\mathbf{w};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma})}{\partial w_j}\mathrm{d}\mathbf{w}\\
&= k_j F_{\boldsymbol{\kappa}-\mathbf{e}_j}^p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},u^{-1}\boldsymbol{\Sigma}) + a_j^{k_j}\phi_1(a_j;\mu_j,u^{-1}\sigma_j^2)F_{\boldsymbol{\kappa}_{(j)}}^{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^a,u^{-1}\tilde{\boldsymbol{\Sigma}}_j)\\
&\quad - b_j^{k_j}\phi_1(b_j;\mu_j,u^{-1}\sigma_j^2)F_{\boldsymbol{\kappa}_{(j)}}^{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^b,u^{-1}\tilde{\boldsymbol{\Sigma}}_j),
\end{aligned}$$

Using the above equation, we obtain the following recurrence relation

$$\mathbb{E}[\mathbf{W}^{\boldsymbol{\kappa}+\mathbf{e}_i}] = \mu_i\mathbb{E}[\mathbf{W}^{\boldsymbol{\kappa}}] + \frac{\mathbf{e}_i^{\top}\boldsymbol{\Sigma}\mathbf{d}_{\boldsymbol{\kappa}}^H}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x}}, \tag{1.20}$$

where $\mathbf{d}_{\kappa}^H = \mathbb{E}_U[U^{-1}\mathbf{c}_{\boldsymbol{\kappa}}]$. Using these results and let $\mathbf{D}^H = [\mathbf{d}_{\mathbf{e}_1}^H,\ldots,\mathbf{d}_{\mathbf{e}_p}^H]$, we can write

$$\mathbb{E}[\mathbf{W}\mathbf{W}^{\top}] = \boldsymbol{\mu}\mathbb{E}[\mathbf{W}]^{\top} + \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x}}\boldsymbol{\Sigma}\mathbf{D}^H, \tag{1.21}$$

$$\mathrm{cov}[\mathbf{W}] = \frac{1}{\int_{\mathbf{a}}^{\mathbf{b}} SMN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu})\mathrm{d}\mathbf{x}}\boldsymbol{\Sigma}(\mathbf{D}^H - (\mathbf{q}_a^H - \mathbf{q}_b^H)\mathbb{E}[\mathbf{W}]^{\top}). \tag{1.22}$$

Using the recurrence formula in (1.20), we are able to compute any product moment of $\mathbf{W}$ with the vector $\mathbf{d}_{\boldsymbol{\kappa}}^H$ depending on the mixture distribution $\mathrm{H}(u;\boldsymbol{\nu})$. Particular expressions for expectations in terms $\mathbf{q}_a^H$, $\mathbf{q}_b^H$ and $\mathbf{D}^H$, involved in the first two moments of a TSMN distribution are presented next.

1.1.3.1 Student-t case

Lemma 1.2. *Suppose* $U \sim G(\frac{\nu}{2}, \frac{\nu}{2})$. *For* $\nu > 2$, *we have that*

$$\mathbb{E}_U[U^{-1}\phi_p(\mathbf{w}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})] = \frac{\nu}{\nu-2} t_p\big(\mathbf{w}; \boldsymbol{\mu}, \tfrac{\nu}{\nu-2}\boldsymbol{\Sigma}, \nu-2\big)\,, \tag{1.23}$$

and hence

$$\begin{aligned}\mathbb{E}_U[U^{-1}\phi_p(\mathbf{w}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})|W_j = a_j] &= \mathbb{E}_U[U^{-1}\phi_1(a_j; \mu_j, u^{-1}\sigma_j^2)\phi_{p-1}(\mathbf{w}_{(j)}, \tilde{\boldsymbol{\mu}}_j^a, u^{-1}\tilde{\boldsymbol{\Sigma}}_j)] \\ &= \frac{\nu}{\nu-2} t_1(a_j; \mu_j, \sigma_j^{*2}, \nu-2) t_{p-1}(\mathbf{w}_{(j)}; \tilde{\boldsymbol{\mu}}_j^a, \tilde{\boldsymbol{\Sigma}}_j^a, \nu-1).\end{aligned}$$

Note that last equation holds since the Student's t distribution is closed under conditioning. Proof for lemma 1.2 is similar to proof of (1.10). Integrating both sides of (1.23) from $\mathbf{a}$ to $\mathbf{b}$, it is easy to see that

$$\mathbb{E}_U[U^{-1}L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})] = \frac{\nu}{\nu-2} L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \tfrac{\nu}{\nu-2}\boldsymbol{\Sigma}, \nu-2),$$

where $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} t_p(\mathbf{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})\mathrm{d}\mathbf{w}$.

1.1.3.2 Slash case

For the truncated multivariate Slash distribution, we can propose the following lemma.

Lemma 1.3. *Suppose* $U \sim Beta(\nu, 1)$. *We have*

$$\begin{aligned}\mathbb{E}_U[U^{-1}\phi_p(\mathbf{w}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})] &= \int_0^1 \nu u^{(\nu-1)-1}\phi_p\left(\mathbf{y}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right)\mathrm{d}u \\ &= \frac{\nu}{\nu-1} SL_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu-1), \qquad \textit{for } \nu > 1.\end{aligned}$$

Unfortunately, we can not derive analogous closed form expressions for the Slash case as in lemma 1.2 since the lack of closure over conditioning property of the Slash distribution. Hence, only the first summand of $d_{\boldsymbol{\kappa},j}$ can be simplified and consequently we should appeal to numerical methods for the other two terms. This would lead to an inefficient recurrence scheme so it will not be part of this work.

1.1.3.3 Contaminated normal case

Since the multivariate contaminated normal distribution is a finite mixture of two MVN distributions, any arbitrary moment for its truncated version can be computed as a mixture of TMVN moments as well. That is,

$$\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} CN_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})\mathrm{d}\mathbf{x} = \nu \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} \phi_p(\mathbf{x}; \boldsymbol{\mu}, \rho^{-1}\boldsymbol{\Sigma})\mathrm{d}\mathbf{x} + (1-\nu)\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} \phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{d}\mathbf{x},$$

$$= \nu F_{\boldsymbol{\kappa}}^{p}(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \rho^{-1}\boldsymbol{\Sigma}) + (1-\nu) F_{\boldsymbol{\kappa}}^{p}(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{1.24}$$

For $\mathbf{X}_0 \sim \mathrm{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu})$, it follows that

$$\mathbb{E}[\mathbf{X}_0^{\boldsymbol{\kappa}} | \mathbf{a} \leqslant \mathbf{X}_0 \leqslant \mathbf{b}] = \frac{\nu\pi_1 \mathbb{E}[\mathbf{X}_1^{\boldsymbol{\kappa}} | \mathbf{a} \leqslant \mathbf{X}_1 \leqslant \mathbf{b}] + (1-\nu)\pi_2 \mathbb{E}[\mathbf{X}_2^{\boldsymbol{\kappa}} | \mathbf{a} \leqslant \mathbf{X}_1 \leqslant \mathbf{b}]}{\pi_0}, \tag{1.25}$$

where $\pi_0 = \mathbb{P}(\mathbf{a} \leqslant \mathbf{X}_0 \leqslant \mathbf{b})$, $\pi_1 = \mathbb{P}(\mathbf{a} \leqslant \mathbf{X}_1 \leqslant \mathbf{b})$ and $\pi_2 = \mathbb{P}(\mathbf{a} \leqslant \mathbf{X}_2 \leqslant \mathbf{b})$, with $\mathbf{X}_1 \sim N_p(\boldsymbol{\mu}, \rho^{-1}\boldsymbol{\Sigma})$ and $\mathbf{X}_2 \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1.2 Case studies

In this section we present the motivating datasets, which will be analyzed in this thesis.

1.2.1 Concentration levels data

This dataset consists of concentration levels of certain dissolved trace metals in freshwater streams across the Commonwealth of Virginia. The data were provided by the Virginia Department of Environment Quality (VDEQ). It is very important to determine the quality of Virginia's water resources across the state to guide their safe use. The methodology adopted must neither underestimate nor overestimate the levels of contamination, as otherwise the results can compromise public health, environmental safety or can unfairly restrict local industry.

The data consist of the concentration levels of the dissolved trace metals copper (Cu), lead (Pb), zinc (Zn), calcium (Ca) and magnesium (Mg) from 184 independent randomly selected sites in freshwater streams across Virginia. The Cu, Pb, and Zn concentrations are reported in μg/L of water, whereas Ca and Mg concentrations are suitably reported in mg/L of water. Since the measurements are taken at different times, the presence of multiple limit of detection values is possible for each trace metal (VDEQ, 2003). The limit of detection is 0.1μg/L for Cu and Pb, 1.0mg/L for Zn, 0.5mg/L for Ca and 1.0mg/L for Mg. The percentages of left-censored values are 2.7% for Ca, 4.9% for Cu, 9.8% for Mg, which are small in comparison to 78.3% for Pb and 38.6% for Zn. Also note that 17.9% of the streams had 0 non-detected trace metals, 39.1% had 1, 37.0% had 2, 3.8% had 3, 1.1% had 4, and 1.1% had 5. Figure 1 shows the histograms for the concentration levels study.

1.2.2 Apple data

Apple data (Little & Rubin, 1987) is a small dataset frequently used in missing data literature which contains partially observed measurements of hundreds of fruits y_{i1} and 100 times the percentage of wormy fruits y_{i2} on 18 apple trees. In this dataset, the

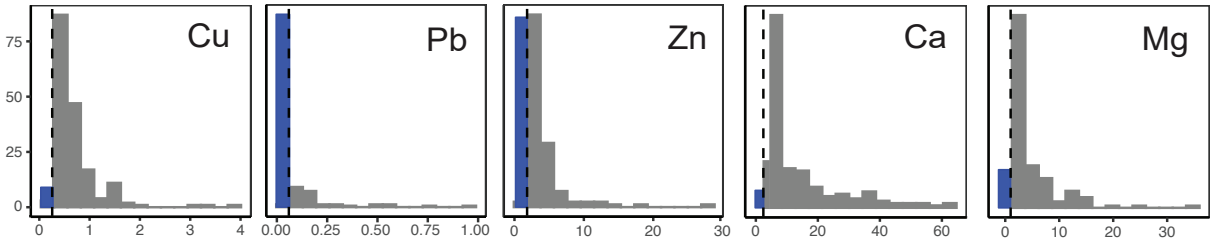


Figure 1 – VDEQ data. Histograms for the concentration levels study. Complete observed points are represented in gray bins while censored observations are represented by blue bins. Limits of detection are represented in dashed lines.

observations y_{1i} and y_{2i}, for $i = 1, \ldots, 12$, are fully observed, while y_{2i}, for $i = 13, \ldots, 18$, are missing (see Table 1).

Table 1 – Apple data.

y_{1k}	8	6	11	22	14	17	18	24	19	23	26	40	4	4	5	6	8	10
y_{2k}	59	58	56	53	60	45	43	42	39	38	30	27	-	-	-	-	-	-

1.2.3 Wine data

Wine data represent 27 chemical measurements on each of 178 wine specimens belonging to three types of wine produced in the Piedmont region of Italy and is available in the `SN` library at CRAN repository (Azzalini, 2020). The data have been presented and examined by Forina et al. (1986) and were freely accessible until a few years ago.

The 28 variables are: `wine`, wine name (categorical variable, i.e. factor, with levels Barbera, Barolo, Grignolino); `alcohol`, alcohol percentage (numeric); `sugar`, sugar-free extract (numeric); `acidity`, fixed acidity (numeric); `tartaric`, tartaric acid (numeric); `malic`, malic acid (numeric); `uronic`, uronic acids (numeric); `pH`, pH (numeric); `ash`, ash (numeric); `alcal_ash`, alcalinity of ash (numeric); `potassium`, potassium (numeric); `calcium`, calcium (numeric); `magnesium`, magnesium (numeric); `phosphate`, phosphate (numeric); `cloride`, chloride (numeric); `phenols`, total phenols (numeric); `flavanoids`, flavanoids (numeric); `nonflavanoids`, nonflavanoid phenols (numeric); `proanthocyanins`, proanthocyanins (numeric); `colour`, colour intensity (numeric); `hue`, hue (numeric), `OD_dw` OD_{280}/OD_{315} of diluted wines (numeric); `OD_fl`, OD_{280}/OD_{315} of flavanoids (numeric); `glycerol`, glycerol (numeric); `butanediol`, 2,3-butanediol (numeric); `nitrogen`, total nitrogen (numeric), `proline`, proline (numeric); and `methanol`, methanol (numeric).

This dataset avaiable in `SN` package does not contain the“sulphate” variable.

2 On moments of folded and truncated multivariate Student's t distributions via recurrence relations

2.1 Introduction

The multivariate Student's t (MVT) distribution has played over the past decades a crucial role in statistical analysis because it offers a more viable alternative with respect to real-world data, in particular due to its properties of having harmonizing parameter (called the degrees of freedom) to control the thickness of tails and including the multivariate normal (MVN) distribution as a limiting case. Both the MVT and the MVN are members of the general family of elliptically symmetric distributions whose properties have been widely studied (Fang *et al.*, 1990). Some recent applications in the areas such as spatial models (De Bastiani *et al.*, 2015), linear mixed effects models (Pinheiro *et al.*, 2001; Savalli *et al.*, 2006), multivariate linear mixed effects models (Wang & Fan, 2011; Wang & Lin, 2014), mixture modelling (Peel & McLachlan, 2000), missing data imputation (Wang *et al.*, 2017) and Bayesian statistical modeling (Fonseca *et al.*, 2008; Wang & Lin, 2015), have been broadly studied.

On the other hand, for many applications on simulations or experimental studies, the researches often generate a large number of datasets with values restricted to fixed intervals. For example, variables such as pH, grades, viral load in HIV studies and humidity in environmental studies, have upper and lower bounds due to detection limits, and the support of their densities is restricted to some given intervals. Thus, the necessity of studying the truncated distributions along with their properties arises naturally. In this context, there has been a growing interest in evaluating the moments of truncated distributions. For instance, Tallis (1961) provided the formulae for the first two moments of truncated multivariate normal (TN) distributions. Lien (1985) gave the expressions for the moments of truncated bivariate log-normal distributions with applications to testing the Houthakker effect in future markets. Jawitz (2004) derived the truncated moments of several continuous univariate distributions commonly applied to hydrologic problems. Kim (2008) provided analytical formulae for moments of the truncated univariate Student's t distribution in a recursive form. Flecher *et al.* (2010) obtained expressions for the moments of truncated skew-normal distributions (Azzalini, 1985) and applied the results to model the relative humidity data. Genç (2013) studied the moments of a doubly truncated member of the symmetrical class of normal/independent distributions and their applications to

the actuarial data. Ho *et al.* (2012) presented a general formula to compute the first two moments of the truncated multivariate Student's t (TMVT) distribution based on the moment generating function (MGF) of the TMVN by expressing a TMVT random variable as a TMVN scale mixture variable. Arismendi (2013) provided explicit expressions for computing arbitrary-order product moments of the TMVN distribution by using the MGF. However, the calculation of this approach relies on differentiation of the MGF and can be prohibitively time consuming.

Instead of differentiating the MGF of the TN distribution, Kan & Robotti (2017) recently presented recurrence relations for integrals that involve directly the density of the MVN distribution for computing arbitrary order product moments of the TMVN distribution. These recursions offer fast computation of the moments of folded (FMVN) and TMVN distributions, which require evaluating p-dimensional integrals that involve the MVN density. Explicit expressions for some low order moments of FMVN and TMVN distributions are presented. Although some proposals to calculate the moments of the truncated Student's t distribution (Kim, 2008; Ho *et al.*, 2012) have been recently published so far, to the best of our knowledge, there is no attempt on studying the product moments of folded (FMVT) and TMVT distributions. In this paper, we develop recurrence relations for integrals involving the density of MVT distributions based on the idea of Kan & Robotti (2017). The proposed recursions allow fast computation of the product moments of the FMVT and TMVT distributions. The proposed new methodology has been implemented in the R package `MomTrunc` (Galarza *et al.*, 2018) available on CRAN repository.

The rest of this paper is organized as follows. In Section 2.2, we define the notation and briefly discuss some preliminary results related to the MVT, TMVT and FMVT distributions. Section 2.3 presents a recurrence formula of an integral for evaluating product moments of the FMVT and TMVT distributions. Explicit expressions for the first two moments of the FMVT and TMVT distributions are also presented. Section 2.4 presents maximum likelihood (ML) estimation for the MVT distribution with the presence interval censored responses. The proposed method is illustrated in Section 2.5 through a simulation study and a real-data example concerning the concentration levels data. Some concluding remarks and implications for future research are given in Section 2.6. Technical details and additional information are sketched in the Appendix A.

2.2 Preliminaries

2.2.1 The MVT and FMVT distributions and main properties

A random variable $\mathbf{X}$ having a p-variate t distribution with location vector $\boldsymbol{\mu}$, positive-definite scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν, denoted by

$\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, has the pdf:

$$t_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma\left(\frac{p+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{\delta(\mathbf{x})}{\nu}\right)^{-(p+\nu)/2}, \tag{2.1}$$

where $\Gamma(\cdot)$ is the standard gamma function and $\delta(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ is the Mahalanobis distance. Let $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ be $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} t_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\nu}) \mathrm{d}\mathbf{x}$, where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$. The cdf of $\mathbf{X}$ is denoted as

$$T_p(\mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{-\infty}^{\mathbf{b}} t_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{x} = L_p(-\infty, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu).$$

In light of Theorem 1.1, we have $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \sum_{\mathbf{s} \in S(\mathbf{a}, \mathbf{b})} (-1)^{n_s} T_p(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, where $S(\mathbf{a}, \mathbf{b})$ and n_s are defined as in Theorem 1.1.

It is known that as $\nu \to \infty$, $\mathbf{X}$ converges in distribution to a multivariate normal with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, denoted by $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. An important property of the random vector $\mathbf{X}$ is that it can be written as a scale mixture of the MVN random vector coupled with a independent positive random variable $U \sim \text{Gamma}(\nu/2, \nu/2)$, where its pdf can be obtained as in (1.5).

The following properties of the MVT distribution are useful for our theoretical developments. We start with the marginal-conditional decomposition of a MVT random vector. The proof of the following propositions can be found in Arellano-Valle & Bolfarine (1995).

Proposition 2.1. *Let* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *partitioned as* $\mathbf{X}^\top = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ *with* $dim(\mathbf{X}_1) = p_1$, $dim(\mathbf{Y}_2) = p_2$, *where* $p_1 + p_2 = p$. *Let* $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$ *and* $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$ *be the corresponding partitions of* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$. *Then, we have*

(*i*) $\mathbf{X}_1 \sim t_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$; *and*

(*ii*) *The conditional distribution of* $\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}_1)$ *is given by*

$$\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}_1) \sim t_{p_2}\left(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1\right),$$

where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ *and* $\widetilde{\boldsymbol{\Sigma}}_{22.1} = \left(\frac{\nu + \delta_1}{\nu + p_1}\right) \boldsymbol{\Sigma}_{22.1}$ *with* $\delta_1 = (\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ *and* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Proposition 2.2. *Let* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. *Then for any fixed vector* $\mathbf{b} \in \mathbb{R}^m$ *and matrix* $\mathbf{A} \in \mathbb{R}^{m \times p}$ *of full rank we get*

$$\mathbf{V} = \mathbf{b} + \mathbf{A}\mathbf{X} \sim t_m(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \nu).$$

We are interested in computing $E\left[|X_1|^{k_1}\dots|X_p|^{k_p}\right]$ and $\mathbb{E}[X_1^{k_1}\dots X_p^{k_p}|a_i < X_i < b_i, i = 1,\dots,p]$ for any non-negative integer values $k_i = 0, 1, 2, \dots$, where the former is the moment of a FMVT distribution $|\mathbf{X}| = (|X_1|,\dots,|X_p|)^\top$, and the later is the moment of a TMVT distribution, with X_i truncated at the lower limit a_i and upper limit b_i, $i = 1,\dots,p$. Remark that some of the $a_i's$ can be $-\infty$ and some of the $b_i's$ can be $+\infty$ in the second expression. When all the b_i's are ∞, the distribution is called the lower TMVT, and when all the a_i's are $-\infty$, the distribution is called the upper TMVT.

2.2.2 The TMVT distribution and main properties

A p-dimensional random vector $\mathbf{Y}$ is said to follow a doubly truncated Student's t distribution with location vector $\boldsymbol{\mu}$, scale-covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν over the truncation region $\mathbb{A}$ defined in (1.1), denoted by $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A})$, if it has the pdf:

$$Tt_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}, \ \mathbf{a} \leqslant \mathbf{y} \leqslant \mathbf{b}. \tag{2.2}$$

Note that equation above is a special case of equation (1.6). Besides, the cdf of $\mathbf{Y}$ evaluated at the region $\mathbf{a} \leqslant \mathbf{y} \leqslant \mathbf{b}$ is

$$TT_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{1}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} \int_{\mathbf{a}}^{\mathbf{y}} t_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) d\mathbf{x} = \frac{L_p(\mathbf{a}, \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}.$$

The following propositions are related to the marginal and conditional moments of the first two moments of the TMVT distributions under a double truncation. The proof is similar to those given in Matos *et al.* (2013). In what follows, we shall use the notation $\mathbf{Y}^{(0)} = 1$, $\mathbf{Y}^{(1)} = \mathbf{Y}$, $\mathbf{Y}^{(2)} = \mathbf{Y}\mathbf{Y}^\top$, and $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ stands for a p-variate doubly truncated Student's t distribution on $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$.

Proposition 2.3. *If* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ *then it holds that*

$$\mathbb{E}\left[\left(\frac{\nu + p}{\nu + \delta(\mathbf{Y})}\right)^r \mathbf{Y}^{(k)}\right] = c_p(\nu, r) \frac{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r)}{L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)} \mathbb{E}[\mathbf{W}^{(k)}],$$

where

$$c_p(\nu, r) = \left(\frac{\nu + p}{\nu}\right)^r \frac{\Gamma\left(\frac{p+\nu}{2}\right)\Gamma\left(\frac{\nu+2r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p+\nu+2r}{2}\right)},$$

$\boldsymbol{\Sigma}^* = \nu\boldsymbol{\Sigma}/(\nu + 2r)$ *and* $\nu + 2r > 0$, *with* $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu + 2r; (\mathbf{a}, \mathbf{b}))$.

Notice that Proposition 2.3 depends on formulas for $\mathbb{E}[\mathbf{W}]$ and $\mathbb{E}[\mathbf{W}\mathbf{W}^\top]$, where $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$. Having established the formula on the k-order moment of $\mathbf{Y}$, we provide an explicit formula for the conditional moments with respect to a two-component partition of $\mathbf{Y}$.

Proposition 2.4. *Let* $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$. *Consider the partition* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)$ *with* $\dim(\mathbf{Y}_1) = p_1$, $\dim(\mathbf{Y}_2) = p_2$, $p_1 + p_2 = p$, *and the corresponding partitions of* $\mathbf{a}$, $\mathbf{b}$, $\boldsymbol{\mu}$, *and* $\boldsymbol{\Sigma}$. *Then,*

$$\mathbb{E}\left[\left(\frac{\nu+p}{\nu+\delta(\mathbf{Y})}\right)^r \mathbf{Y}_2^{(k)} \mid \mathbf{Y}_1\right] = \frac{d_p(p_1,\nu,r)}{(\nu+\delta(\mathbf{Y}_1))^r}\frac{L_{p_2}(\mathbf{a}_2,\mathbf{b}_2;\boldsymbol{\mu}_{2.1},\widetilde{\boldsymbol{\Sigma}}^*_{22.1},\nu+p_1+2r)}{L_{p_2}(\mathbf{a}_2,\mathbf{b}_2;\boldsymbol{\mu}_{2.1},\widetilde{\boldsymbol{\Sigma}}_{22.1},\nu+p_1)}\mathbb{E}[\mathbf{W}_2^{(k)}],$$

for $\nu + p_1 + 2r > 0$, *with* $\delta(\mathbf{Y}_1) = \delta(\mathbf{Y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$,

$$\widetilde{\boldsymbol{\Sigma}}^*_{22.1} = \left(\frac{\nu+\delta_1}{\nu+2r+p_1}\right)\boldsymbol{\Sigma}_{22.1}, \quad \textit{and} \quad d_p(p_1,\nu,r) = (\nu+p)^r\,\frac{\Gamma\left(\frac{p+\nu}{2}\right)\Gamma\left(\frac{p_1+\nu+2r}{2}\right)}{\Gamma\left(\frac{p_1+\nu}{2}\right)\Gamma\left(\frac{p+\nu+2r}{2}\right)},$$

where $\boldsymbol{\Sigma}_{22.1}$ *is defined as in proposition 2.1. Moreover,* $\mathbf{W}_2 \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \widetilde{\boldsymbol{\Sigma}}^*_{22.1}, \nu + p_1 + 2r; (\mathbf{a}_2, \mathbf{b}_2))$.

2.3 The recurrence relation for the multivariate Student's t integral

Let $\mathbf{a}_{(i)}$ be a vector $\mathbf{a}$ with its ith element being removed. For a matrix $\boldsymbol{\Delta}$, we let $\boldsymbol{\Delta}_{i(j)}$ stand for the ith row of $\boldsymbol{\Delta}$ with its jth element being removed. Similarly, $\boldsymbol{\Delta}_{(i,j)}$ stands for the matrix $\boldsymbol{\Delta}$ with its ith row and jth columns being removed. Besides, let $\mathbf{e}_i$ denote a $p \times 1$ vector with its ith element equaling one and zero otherwise.

The integral that we are interested in evaluating is

$$\mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)\mathrm{d}\mathbf{x}.$$

The initial condition is obviously $\mathrm{F}^p_0(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$. The recurrence relation for the normal case has been recently presented by Kan & Robotti (2017). When $p = 1$, the use of integration by parts straightforwardly leads to

$$\begin{aligned}
\mathrm{F}^1_0(a,b;\mu,\sigma^2,\nu) &= T_1(b;\mu,\sigma^2,\nu) - T_1(a;\mu,\sigma^2,\nu),\\
\mathrm{F}^1_{k+1}(a,b;\mu,\sigma^2,\nu) &= \mu\mathrm{F}^1_k(a,b;\mu,\sigma^2,\nu) + \tfrac{k\nu\sigma^2}{(\nu-2)}\mathrm{F}^1_{k-1}(a,b;\mu,\tfrac{\nu}{\nu-2}\sigma^2,\nu-2)\\
&\quad + \tfrac{\nu\sigma^2}{(\nu-2)}[a^k t_1(a;\mu,\tfrac{\nu}{\nu-2}\sigma^2,\nu-2) - b^k t_1(b;\mu,\tfrac{\nu}{\nu-2}\sigma^2,\nu-2)],\ (k \geqslant 0).
\end{aligned} \tag{2.3}$$

When $p > 1$, we need a similar recurrence relation in order to compute $\mathrm{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ which we propose in the following Theorem:

Theorem 2.1. *For* $p \geqslant 1$ *and* $i = 1, \ldots, p$,

$$F^p_{\boldsymbol{\kappa}+\mathbf{e}_i}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \mu_i F^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) + \tfrac{\nu}{\nu-2}\mathbf{e}_i^\top\boldsymbol{\Sigma}\mathbf{c}_{\boldsymbol{\kappa}}, \tag{2.4}$$

where $\mathbf{c}_{\boldsymbol{\kappa}}$ *is a* $p \times 1$ *vector with the* j*th element being*

$$c_{\boldsymbol{\kappa},j} = k_j F^p_{\boldsymbol{\kappa}-\mathbf{e}_j}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)$$

$$+a_j^{k_j} t_1(a_j;\mu_j,\sigma_j^{*2},\nu-2) F^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)$$
$$-b_j^{k_j} t_1(b_j;\mu_j,\sigma_j^{*2},\nu-2) F^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1), \tag{2.5}$$

and

$$\tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}^*_{(j)(j)} - \frac{1}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j}\boldsymbol{\Sigma}^*_{j(j)},\ \tilde{\delta}_j^{\mathbf{a}} = \frac{\nu-2+\frac{(a_j-\mu_j)^2}{\sigma_j^{*2}}}{\nu-1},\ \tilde{\delta}_j^{\mathbf{b}} = \frac{\nu-2+\frac{(b_j-\mu_j)^2}{\sigma_j^{*2}}}{\nu-1},$$
$$\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}} = \boldsymbol{\mu}_{(j)} + \tfrac{(a_j-\mu_j)}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j},\quad \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}} = \boldsymbol{\mu}_{(j)} + \tfrac{(b_j-\mu_j)}{\sigma_j^{*2}}\boldsymbol{\Sigma}^*_{(j)j},\ \boldsymbol{\Sigma}^* = \tfrac{\nu}{\nu-2}\boldsymbol{\Sigma},\ \sigma_j^{*2} = \frac{\nu}{\nu-2}\sigma_j^2.$$

When $k_j = 0$, the first term in (2.5) vanishes. When $a_j = -\infty$ and $k_j \leqslant \nu-2$, the second term vanishes, and when $b_j = +\infty$ and $k_j \leqslant \nu-2$, the third term vanishes.

Proof. In light of equation (1.10), we have that

$$-\frac{\partial t_p(\mathbf{x};\boldsymbol{\mu},\frac{\nu}{\nu-2}\boldsymbol{\Sigma},\nu-2)}{\partial \mathbf{x}} = \frac{\nu-2}{\nu} t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}). \tag{2.6}$$

Multiplying each element on both sides by $\mathbf{x}^{\boldsymbol{\kappa}}$ and integrating $\mathbf{x}$ from $\mathbf{a}$ to $\mathbf{b}$, we have

$$\mathbf{c}_{\boldsymbol{\kappa}} = \tfrac{\nu-2}{\nu}\boldsymbol{\Sigma}^{-1}\begin{bmatrix} \mathrm{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_1} - \mu_1\mathrm{F}^p_{\boldsymbol{\kappa}} \\ \mathrm{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_2} - \mu_2\mathrm{F}^p_{\boldsymbol{\kappa}} \\ \vdots \\ \mathrm{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_p} - \mu_p\mathrm{F}^p_{\boldsymbol{\kappa}}. \end{bmatrix},$$

Using integration by parts, the jth element of the left-hand side is

$$c_{\boldsymbol{\kappa},j} = -\int_{\mathbf{a}_{(j)}}^{\mathbf{b}_{(j)}} \mathbf{x}^{\boldsymbol{\kappa}} t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)|_{x_j=a_j}^{b_j}\mathrm{d}\mathbf{x}_{(j)} + \int_{\mathbf{a}}^{\mathbf{b}} k_j\mathbf{x}^{\boldsymbol{\kappa}-\mathbf{e}_j} t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)\mathrm{d}\mathbf{x}. \tag{2.7}$$

Using the fact that

$$t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)|_{x_j=a_j} = t_1(a_j;\mu_j,\sigma_j^{*2},\nu-2)t_{p-1}(\mathbf{x}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1) \quad \text{and}$$
$$t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2)|_{x_j=b_j} = t_1(b_j;\mu_j,\sigma_j^{*2},\nu-2)t_{p-1}(\mathbf{x}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1),$$

we get

$$c_{\boldsymbol{\kappa},j} = k_j\mathrm{F}^p_{\boldsymbol{\kappa}-\mathbf{e}_j}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2) + a_j^{k_j}t_1(a_j;\mu_j,\sigma_j^{*2},\nu-2)\mathrm{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)$$
$$-b_j^{k_j}t_1(b_j;\mu_j,\sigma_j^{*2},\nu-2)\mathrm{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1).$$

When $k_j = 0$, the last integral in (2.7) is equal to zero, and the first term of $c_{\boldsymbol{\kappa},j}$ vanishes. When $a_j \to -\infty$ and $k_j \leqslant \nu-2$, $a_j^{k_j}t_1(a_j;\mu_j,\sigma_j^{*2},\nu-2) \to 0$, so the second term of $c_{\boldsymbol{\kappa},j}$ vanishes. Similarly when $b_j \to \infty$ the third term of $c_{\boldsymbol{\kappa},j}$ vanishes. Finally, the desired result is obtained by multiplying $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$ on both sides of (2.4). □

As a consequence, $\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}}]$ always exists for $\sum_{j=1}^{p}\kappa_j < \nu$. When all the $a_i's$ are $-\infty$ or all the $b_i's$ are $+\infty$, the length of the recurrence relation is reduced to $2p+1$ rather than the original $3p+1$. When all the $a_i's$ are $-\infty$ and all the $b_i's$ are $+\infty$, we have

$$\mathrm{F}_{\boldsymbol{\kappa}}^{p}(-\infty,+\infty;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}}],\quad \mathbf{X}\sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$$

and the recursive relation of length $(p+1)$ is

$$\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}+\mathbf{e}_i}]=\mu_i E[\mathbf{X}^{\boldsymbol{\kappa}}]+\sum_{j=1}^{p}\sigma_{ij}^{*}k_j\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}-\mathbf{e}_i}],\ \mathbf{Y}\sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma}^{*},\nu-2),\ i=1,\ldots,p.$$

Another special case of interest occurs when $a_i=0$ and $b_i=+\infty$, $i=1,\ldots,p$. In this scenario, we have $I_{\boldsymbol{\kappa}}^{p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mathrm{F}_{\boldsymbol{\kappa}}^{p}(\mathbf{0},+\infty;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$. The recurrence relation for $I_{\boldsymbol{\kappa}}^{p}$ can be written as

$$I_{\boldsymbol{\kappa}+\mathbf{e}_i}^{p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)=\mu_i I_{\boldsymbol{\kappa}}^{p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)+\sum_{j=1}^{p}\sigma_{ij}^{*}d_{\kappa,j},\ i=1,\ldots,p,$$

where

$$d_{\kappa,j}=\begin{cases} k_j I_{\boldsymbol{\kappa}-\mathbf{e}_i}^{p}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) & \text{for } k_j>0,\\ t_1(0|\mu_j,\sigma_j^{*2},\nu-2)I_{\boldsymbol{\kappa}_{(j)}}^{p-1}(\tilde{\boldsymbol{\mu}}_j,\tilde{\delta}_j\tilde{\boldsymbol{\Sigma}}_j,\nu-1) & \text{for } k_j=0,\end{cases}$$

with

$$\tilde{\boldsymbol{\mu}}_j=\boldsymbol{\mu}_{(j)}-\frac{\mu_j}{\sigma_j^{*2}}\boldsymbol{\Sigma}_{(j)j}^{*},\ \tilde{\boldsymbol{\Sigma}}_j=\boldsymbol{\Sigma}_{(j)(j)}^{*}-\frac{1}{\sigma_j^{*2}}\boldsymbol{\Sigma}_{(j)j}^{*}\boldsymbol{\Sigma}_{j(j)}^{*},\ \text{and}\ \tilde{\delta}_j=\frac{\nu-2+\frac{\mu_j^2}{\sigma_j^{*2}}}{\nu-1}.$$

2.3.1 The first two moments of the doubly TMVT distribution

Let $\mathbf{X}\sim t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ and $\mathbf{Z}=\mathbf{X}\mid(\mathbf{a}\leqslant\mathbf{X}\leqslant\mathbf{b})\sim Tt_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;(\mathbf{a},\mathbf{b}))$. It follows that

$$\mathbb{E}[\mathbf{Z}^{\boldsymbol{\kappa}}]=\frac{1}{L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}\int_{\mathbf{a}}^{\mathbf{b}}\mathbf{x}^{\boldsymbol{\kappa}}t_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)d\mathbf{x}=\frac{\mathrm{F}_{\boldsymbol{\kappa}}^{p}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}{L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}.$$

Furthermore, let denote $\mathrm{F}_{\boldsymbol{\kappa}}^{p}\equiv\mathrm{F}_{\boldsymbol{\kappa}}^{p}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ and $L\equiv L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ for simplicity. In light of Theorem 2.1, it is straightforward that to see that

$$\mathbb{E}[Z_i]=\frac{F_{\mathbf{e}_i}^{p}}{L}=\mu_i+\frac{1}{L}\mathbf{e}_i^{\top}\boldsymbol{\Sigma}^{*}\mathbf{c_0}\ \text{ and }\ \mathbb{E}[Z_iZ_j]=\frac{F_{\mathbf{e}_i+\mathbf{e}_j}^{p}}{L}=\mu_j\mathbb{E}[Z_i]+\frac{1}{L}\mathbf{e}_j^{\top}\boldsymbol{\Sigma}^{*}\mathbf{c}_{\mathbf{e}_i},\quad(2.8)$$

where $\mathbf{c_0}=\mathbf{c_a}-\mathbf{c_b}$, with

$$c_{\mathbf{a}}=\left[t_1(a_j;\mu_j,\sigma_j^{*2},\nu-2)\,L_{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)\right]_{j=1}^{p},\quad(2.9)$$

$$c_{\mathbf{b}}=\left[t_1(b_j;\mu_j,\sigma_j^{*2},\nu-2)\,L_{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)\right]_{j=1}^{p},\quad(2.10)$$

and

$$c_{\mathbf{e}_i} = \Big[\mathbf{e}_{ij} L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2) + a_j t_1(a_j;\mu_j,\sigma_j^{*2},\nu-2)\mathrm{F}^{p-1}_{\mathbf{e}_{i(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1) - b_j t_1(b_j;\mu_j,\sigma_j^{*2},\nu-2)\mathrm{F}^{p-1}_{\mathbf{e}_{i(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)\Big]_{j=1}^{p}. \tag{2.11}$$

where

$$\begin{aligned} c_{\mathbf{e}_{ii}} &= L_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma}^*,\nu-2) + a_j c_{\mathbf{a}i} - b_j c_{\mathbf{b}i}, \\ c_{\mathbf{e}_{ij}} &\overset{i\neq j}{=} a_j c_{\mathbf{a}i}\mathbb{E}[(\mathbf{X}_{(j)} \mid X_j = a_j) \mid \mathbf{a}_{(j)} \leqslant \mathbf{X}_{(j)} \leqslant \mathbf{b}_{(j)}] \\ &\quad -b_j c_{\mathbf{b}i}\mathbb{E}[(\mathbf{X}_{(j)} \mid X_j = b_j) \mid \mathbf{a}_{(j)} \leqslant \mathbf{X}_{(j)} \leqslant \mathbf{b}_{(j)}]. \end{aligned}$$

This last equality is obtained by noting that

$$\mathbb{P}(\mathbf{a}_{(j)} \leqslant \mathbf{X}_{(j)} \leqslant \mathbf{b}_{(j)} | X_j = a_j) = L_{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)$$

and $\mathbb{P}(\mathbf{a}_{(j)} \leqslant \mathbf{X}_{(j)} \leqslant \mathbf{b}_{(j)} | X_j = b_j) = L_{p-1}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j,\nu-1)$.

Let $\mathbf{C} = (\mathbf{c}_{\mathbf{e}_1}, \mathbf{c}_{\mathbf{e}_2}, \ldots, \mathbf{c}_{\mathbf{e}_p})$. From expressions in (2.8), we can note that for $\mathbb{E}[Z_i]$, $\mathbf{c_0}$ does not depend on i and, for $\mathbb{E}[Z_i Z_j]$, $\mathbf{c}_{\mathbf{e}_i}$ does not depend on j. Then, it is easy to establish that the mean vector $\boldsymbol{\xi} = \mathbb{E}[\mathbf{Z}]$ and variance-covariance matrix $\boldsymbol{\Psi} = \mathrm{cov}[\mathbf{Z}]$ are given by

$$\boldsymbol{\xi} = \boldsymbol{\mu} + \frac{1}{L}\boldsymbol{\Sigma}^* \boldsymbol{c_0}, \tag{2.12}$$

$$\boldsymbol{\Psi} = \frac{1}{L}\boldsymbol{\Sigma}^*(\mathbf{C} - \boldsymbol{c_0}\boldsymbol{\xi}^\top), \tag{2.13}$$

where $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top] = \boldsymbol{\mu}\boldsymbol{\xi}^\top + \frac{1}{L}\mathbf{C}\boldsymbol{\Sigma}^*$.

Methods for computing the mean and variance-covariance matrix of $\mathbf{Z}$ are summarized in algorithms 1 and 2. Note that, to calculate the variance-covariance matrix

Algorithm 1 – Mean vector for $\mathbf{Z} \sim Tt_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu;(\mathbf{a},\mathbf{b}))$

```
mean(a, b, θ)
L ← L_p(a, b; μ, Σ, ν); c_a ← 0;  c_b ← 0;
for j = 1 : p do
    θ_j^a ← (μ̃_j^a, δ̃_j^a Σ̃_j, ν − 1);  θ_j^b ← (μ̃_j^b, δ̃_j^b Σ̃_j, ν − 1);
    if a_j ≠ ∞ then
        c_a(j) ← t_1(a_j; μ_j, σ*²_j, ν − 2) L_{p−1}(a_(j), b_(j); μ̃_j^a, δ̃_j^a Σ̃_j, ν − 1);
    end
    if b_j ≠ ∞ then
        c_b(j) ← t_1(b_j; μ_j, σ*²_j, ν − 2) L_{p−1}(a_(j), b_(j); μ̃_j^b, δ̃_j^b Σ̃_j, ν − 1);
    end
end
ξ ← μ + ν/(ν − 2) Σ(c_a − c_b)/L;
return ξ;
```

Algorithm 2 – Mean vector and variance-covariance matrix for $\mathbf{Z} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$

meanvar$(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta})$
$L \leftarrow L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$; $L^* \leftarrow L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \nu - 2)$;
$\mathbf{W_a} \leftarrow \mathbf{0}_{p\times p}$; $\mathbf{W_b} \leftarrow \mathbf{0}_{p\times p}$;
for $j = 1 : p$ **do**
  $\boldsymbol{\theta}_j^{\mathbf{a}} \leftarrow (\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$; $\boldsymbol{\theta}_j^{\mathbf{b}} \leftarrow (\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
  if $a_j \neq \infty$ **then**
    $\boldsymbol{c}_{\mathbf{a}}(j) \leftarrow t_1(a_j; \mu_j, \sigma_j^{*2}, \nu - 2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\delta}_j^{\mathbf{a}}\tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
    $\mathbf{W_a}(-j, j) \leftarrow \text{mean}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}, \boldsymbol{\theta}_{(j)}^{\mathbf{a}})$;
    $\mathbf{W_a}(j, j) \leftarrow \mathbf{a}(j)$;
  end
  if $b_j \neq \infty$ **then**
    $\boldsymbol{c}_{\mathbf{b}}(j) \leftarrow t_1(b_j; \mu_j, \sigma_j^{*2}, \nu - 2) L_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\delta}_j^{\mathbf{b}}\tilde{\boldsymbol{\Sigma}}_j, \nu - 1)$;
    $\mathbf{W_b}(-j, j) \leftarrow \text{mean}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}, \boldsymbol{\theta}_{(j)}^{\mathbf{b}})$;
    $\mathbf{W_b}(j, j) \leftarrow \mathbf{b}(j)$;
  end
end
$\boldsymbol{\xi} \leftarrow \boldsymbol{\mu} + \boldsymbol{\Sigma}^*(\boldsymbol{c}_{\mathbf{a}} - \boldsymbol{c}_{\mathbf{b}})/L$;
$\boldsymbol{\Psi} \leftarrow (L^*\text{diag}(p) + \mathbf{W_a}\text{diag}(\mathbf{c_a}) - \mathbf{W_b}\text{diag}(\mathbf{c_b}))\boldsymbol{\Sigma}^*/L$;
return $\boldsymbol{\xi}$, $\boldsymbol{\Psi}$;

$\boldsymbol{\Psi}$ in Algorithm 2, it is necessary to compute $2p$ $(p-1)$-variate mean vectors (lines 8 and 13) through Algorithm 1. This schema leads to only $1 + 2p$ necessary integrals to compute the mean and additional $1 + 2p + 4p(p-1)$ integrals for the variance-covariance matrix. It is noteworthy to mention that *i)* probabilities between lines 7 and 12 in Algorithm 2, can be recycled from the mean$(\mathbf{a}, \mathbf{b}, \boldsymbol{\theta})$ function, and *ii)* $\mathbf{C}$ is not symmetric, however both of its (i, j)-th and (j, i)-th elements $c_{\mathbf{e}_j i}$ and $c_{\mathbf{e}_i j}$ depends on probabilities of the form $\mathbb{P}\left(\mathbf{a}_{(i,j)} \leqslant \mathbf{X}_{(i,j)} \leqslant \mathbf{b}_{(i,j)} \mid (X_i, X_j) = (x_i, x_j)\right)$, with $(x_i, x_j) \in \{a_i, b_i\} \times \{a_j, b_j\}$. This leads to an optimal schema with a maximum total of $2(1 + p^2)$ integrals to compute the mean and the variance-covariance matrix in the case that the distribution is doubly truncated. Lastly, we remark that this recurrence is limited to work for real degrees of freedom greater than 3 due to the computation of $\boldsymbol{\Sigma}^{**} = \nu^*\boldsymbol{\Sigma}/(\nu^* - 2)$ when $\nu^* = \nu - 1$. For $\nu = 3$, we approximate its value taking its right-hand limit, which showed a good performance in terms of precision and stability. Finally, expressions for the mean vector and variance-covariance matrix derived in subsection 1.1.3 are equivalent but less efficient.

2.3.2 The first two moments of the TMVT distribution when a non-truncated partition exists

We describe a trick for fast computation of the first two moments of the TMVT distribution when there are double infinite limits. Consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$

such that $dim(\mathbf{X}_1) = p_1$, $dim(\mathbf{X}_2) = p_2$, where $p_1 + p_2 = p$. Using the law of total expectations, we have

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix} \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2] \\ \mathbf{X}_2 \end{bmatrix}$$

and

$$\text{cov}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[\text{cov}[\mathbf{X}_1|\mathbf{X}_2]] + \text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] & \text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2], \mathbf{X}_2] \\ \text{cov}[\mathbf{X}_2, \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] & \text{cov}[\mathbf{X}_2] \end{bmatrix}.$$

Let p_1 be the number of pairs in $[\mathbf{a}, \mathbf{b}]$ that are both infinite. We consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ in which the upper and lower truncation limits associated with $\mathbf{X}_1$ are both infinite, but at least one of the truncation limits associated with $\mathbf{X}_2$ is finite. Let

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \mathbf{a} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top)^\top \text{ and } \mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top,$$

be the corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a}$ and $\mathbf{b}$. Since $\mathbf{a}_1 = -\infty$ and $\mathbf{b}_1 = \infty$, it follows that $\mathbf{X}_2 \sim \text{Tt}_{p_2}\big(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \nu; [\mathbf{a}_2, \mathbf{b}_2]\big)$ and $\mathbf{X}_1|\mathbf{X}_2 \sim t_{p_1}\big(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})(\nu + \delta(\mathbf{x}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}))/(\nu + p_2), \nu + p_2\big)$. This leads to

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\xi}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\xi}_2 \end{bmatrix}. \tag{2.14}$$

On the other hand, we have that $\text{cov}[\mathbf{X}_2, \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \text{cov}[\mathbf{X}_2, \mathbf{X}_2\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}] = \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, $\text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ and $\mathbb{E}[\text{cov}[\mathbf{X}_1|\mathbf{X}_2]] = \omega_{1.2}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$ with $\boldsymbol{\xi}_2 = \mathbb{E}[\mathbf{X}_2]$ and $\boldsymbol{\Psi}_{22} = \text{cov}[\mathbf{X}_2]$, and

$$\omega_{1.2} = \mathbb{E}\left(\frac{\nu + \delta(\mathbf{X}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})}{\nu + p_2 - 2}\right) = \left(\frac{\nu}{\nu - 2}\right)\frac{L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}^*, \nu - 2)}{L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \nu)}, \tag{2.15}$$

with $\boldsymbol{\Sigma}_{22}^* = \nu\boldsymbol{\Sigma}_{22}/(\nu - 2)$. This last expression follows from Proposition 2.3. Finally,

$$\text{cov}[\mathbf{X}] = \begin{bmatrix} \omega_{1.2}\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\big(\omega_{1.2}\mathbf{I}_{p_2} - \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\big)\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Psi}_{22} \\ \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{\Psi}_{22} \end{bmatrix}, \tag{2.16}$$

where $\boldsymbol{\xi}_2$ and $\boldsymbol{\Psi}_{22}$ are the mean vector and variance-covariance matrix of the TMVT distribution, which can be computed by using (2.12) and (2.13), respectively.

Note that $\mathbf{X}_1$ does not follow a non-truncated t distribution, that is, $\mathbf{X}_1 \not\sim t_{p_1}\big(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu\big)$, even though $\mathbf{a}_1 = -\infty$ and $\mathbf{b}_1 = \infty$. In general, the marginal distributions of a TMt distribution are not TMt, however this holds for $\mathbf{X}_2$ due to the particular case $\mathbf{a}_1 = -\infty$ and $\mathbf{b}_1 = \infty$. Also note that obtaining (2.15) does not require the computation of additional integrals given that the probabilities $L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}^*, \nu - 2)$ and $L_p(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, \nu)$ are involved in the calculation of $\boldsymbol{\xi}_2$ and $\boldsymbol{\Psi}_{22}$ (see Algorithm 2, Line 2).

It is important to emphasize that $\mathbb{E}[\mathbf{X}]$ and $\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ exist if and only if $\nu + p_2 > 1$ and $\nu + p_2 > 2$ respectively. This is equivalent to say that, (2.14) exists if at least one dimension containing a finite limit exists. Besides, (2.16) exists if at least two dimensions containing a finite limit exist.

As can be seen, we can use equations (2.14) and (2.16) to deal with double infinite limits, where the truncated moments are computed only over a p_2-variate partition, avoiding some unnecessary integrals and saving a significant computational cost.

2.3.3 Folded Multivariate Student's t distribution

Let $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, we now turn our attention to discuss the computation of any arbitrary order moment of $|\mathbf{X}|$. First, we established the following corollary.

Corollary 2.1. *If* $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *then* $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X} \sim t_p(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)$ *and consequently the joint* pdf, cdf *and the* $\boldsymbol{\kappa}$*th raw moment of* $\mathbf{Y} = |\mathbf{X}|$ *are, respectively, given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} t_p(\mathbf{y}; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu), \quad F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} \pi_s T_p(\mathbf{y}_s; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu),$$

and

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \sum_{\mathbf{s}\in S(p)} I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu),$$

where $\mathbf{y}_s = \boldsymbol{\Lambda}_s\mathbf{y}$, $\boldsymbol{\mu}_s = \boldsymbol{\Lambda}_s\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_s\boldsymbol{\Sigma}\boldsymbol{\Lambda}_s$ *and* $I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu) = \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} t_p(\mathbf{y}; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)\, d\mathbf{y}$.

Proof. The proof follows straightforwardly from the definition of probability theory and basic matrix algebra and thus is omitted. □

Thus the product moments of $\mathbf{Y}$ can be calculated easily using $I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)$ terms as stated above. In particular, in light of Corollary 1.2, we have that the mean vector $\boldsymbol{\xi}$ and variance-covariance matrix $\boldsymbol{\Psi}$ of $\mathbf{Y}$ can be calculated as

$$\boldsymbol{\xi} = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^+], \tag{2.17}$$

$$\boldsymbol{\Psi} = \sum_{\mathbf{s}\in S(p)} \mathbb{E}\left[\mathbf{Z}_s^+ \mathbf{Z}_s^{+\top}\right] - \boldsymbol{\xi}\boldsymbol{\xi}^\top, \tag{2.18}$$

where $\mathbf{Z}_s^+$ is the positive component of $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X} \sim t_p(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \nu)$ from Corollary 1.1. Note that there are 2^p times more integrals to be calculated as compared to the non-folded case. Specifically, $(1 + p)2^p$ integrals are required for the mean vector, and additional $(1 + 2p + 2p^2)2^p$ integrals for the variance-covariance matrix. For the univariate case, the

explicit expressions for the first four raw moments of $Y = |X|$, where $X \sim t(\mu, \sigma^2, \nu)$, based on (2.3) and (2.5) can be obtained as

$$\begin{aligned}
\mathbb{E}[Y] &= \mu[1 - 2T_1(0; \mu, \sigma^2, \nu)] + 2\sigma^{*2} t_1(0; \mu, \sigma^{*2}, \nu - 2), \\
\mathbb{E}[Y^2] &= \mu^2 + \sigma^{*2}, \\
\mathbb{E}[Y^3] &= \mu^2 \mathbb{E}[Y] + 3\mu\sigma^{*2}[1 - 2T_1(0; \mu, \sigma^{*2}, \nu - 2)] + \tfrac{4(\nu-2)\sigma^{*4}}{\nu-4} t_1(0; \mu, \tfrac{\nu}{\nu-4}\sigma^2, \nu - 4), \\
\mathbb{E}[Y^4] &= \mu^4 + 6\mu^2\sigma^{*2} + \tfrac{3(\nu-2)}{\nu-4}\sigma^{*4}.
\end{aligned}$$

Illustrative results via the implementation of `R` package `MomTrunc` are presented in the Appendix B.

2.4 Inference for MVT with Interval Censored Responses

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a $p \times 1$ response vector for the ith sample unit, for $i \in \{1, \ldots, n\}$, and consider the set of independent and identically distributed samples:

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu), \tag{2.19}$$

where the location vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$ and the dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ depend on an unknown and reduced parameter vector $\boldsymbol{\alpha}$. However, the response vector $\mathbf{Y}_i$ may not be fully observed due to censoring, so we define $(\mathbf{V}_i, \mathbf{C}_i)$ the observed data for the ith sample, where $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$ represents either an uncensored observation ($V_{ik} = V_{0i}$) or the interval censoring level ($V_{ik} \in [V_{1ik}, V_{2ik}]$), and $\mathbf{C}_i = (C_{i1}, \ldots, C_{ip})^\top$ is the vector of censoring indicators, satisfying

$$C_{ik} = \begin{cases} 1 & \text{if } V_{1ik} \leqslant Y_{ik} \leqslant V_{2ik}\,, \\ 0 & \text{if } Y_{ik} = V_{0i}\,. \end{cases} \tag{2.20}$$

for all $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, p\}$, i.e., $C_{ik} = 1$ if Y_{ik} is located within a specific interval. In this case, (2.19) along with (2.20) defines the multivariate Student's t interval censored model (hereafter, the MVT-IC model). Notice that a left censoring structure causes truncation from the lower limit of the support of the distribution, since we only know that the true observation Y_{ik} is less than or equal to the observed quantity V_{2ik}. This situation has been studied by Lachos *et al.* (2017). Missing observations can be handled by considering $V_{1ik} = -\infty$ and $V_{2ik} = +\infty$.

2.4.1 The likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$ is a realization of $\mathbf{Y}_i \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. To obtain the likelihood function of the MVT-IC model, we firstly treat the

observed and censored components of $\mathbf{y}_i$, separately, i.e., $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $C_{ik} = 0$ for all elements in the p_i^o-dimensional vector $\mathbf{y}_i^o$, and $C_{ik} = 1$ for all elements in the p_i^c-dimensional vector $\mathbf{y}_i^c$. Accordingly, we write $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $\mathbf{V}_i^c = (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c)$ with

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

Then, using Proposition 2.1, we have that $\mathbf{Y}_i^o \sim t_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu)$ and $\mathbf{Y}_i^c \mid \mathbf{Y}_i^o = \mathbf{y}_i^o \sim t_{p_i^c}(\boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o)$, where

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co} \boldsymbol{\Sigma}_i^{oo-1} (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \quad \mathbf{S}_i^{co} = \left\{ \frac{\nu + \delta(\mathbf{y}_i^o)}{\nu + p_i^o} \right\} \boldsymbol{\Sigma}_i^{cc.o}, \tag{2.21}$$

$$\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co} (\boldsymbol{\Sigma}_i^{oo})^{-1} \boldsymbol{\Sigma}_i^{oc} \quad \text{and} \quad \delta(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)^\top (\boldsymbol{\Sigma}_i^{oo})^{-1} (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o). \tag{2.22}$$

Let $\mathbf{V} = \text{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ denote the observed data. Therefore, the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}^\top, \nu)^\top$, where $\boldsymbol{\alpha} = \text{vech}(\boldsymbol{\Sigma})$, for the observed data $(\mathbf{V}, \mathbf{C})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln L_i, \tag{2.23}$$

where L_i represents the likelihood function of $\boldsymbol{\theta}$ for the ith sample, given by

$$\begin{aligned} L_i \equiv & L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i) = f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{V}_{1i}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{V}_{2i}^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ & = L_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \mathbf{S}_i^{co}, \nu + p_i^o) t_{p_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \nu). \end{aligned}$$

2.4.2 Parameter estimation via the EM algorithm

We describe how to carry out ML estimation for the MVT-IC model. The EM algorithm, originally proposed by Dempster *et al.* (1977), is a very popular iterative optimization strategy and commonly used to obtain ML estimates for incomplete-data problems. This algorithm has many attractive features such as the numerical stability, the simplicity of implementation and quite reasonable memory requirements (McLachlan & Krishnan, 2008).

By the essential property of MVT distribution, we can write $\mathbf{Y}_i | (U_i = u_i) \sim N_p(\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma})$ and $u_i \sim \text{Gamma}(\nu/2, \nu/2)$. The complete-data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ is given by $\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta})$, where the individual complete-data log-likelihood is

$$\ell_{ic}(\boldsymbol{\theta}) = -\frac{1}{2}\big\{\ln|\boldsymbol{\Sigma}| + u_i(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\big\} + \ln h(u_i; \nu) + c,$$

where c is a constant irrelevant of $\boldsymbol{\theta}$ and $h(u_i; \nu)$ is the pdf of Gamma$(\nu/2, \nu/2)$. In summary, the EM algorithm for the MVT-IC model can be adopted as follows:

E-step: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \hat{\nu}^{(k)})$ at the kth step, the E-step provides the conditional expectation of the complete data log-likelihood function

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \mathbb{E}\Big[\ell_c(\boldsymbol{\theta}) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\Big] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\left[\left\{\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\boldsymbol{\mu}^\top - \boldsymbol{\mu}(\widehat{u\mathbf{y}}_i^{(k)})^\top + \hat{u}_i^{(k)}\boldsymbol{\mu}\boldsymbol{\mu}^\top\right\}\boldsymbol{\Sigma}^{-1}\right]$$

with $\widehat{u\mathbf{y}}_i^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{u\mathbf{y}_i^2}^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\hat{u}_i^{(k)} = \mathbb{E}[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ which are collected in Appendix A. Note that, since ν is fixed, the calculation of $\mathbb{E}[\ln h(U_i;\nu) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$ is unnecessary.

M-step: Conditionally maximizing $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ with respect to each entry of $\boldsymbol{\theta}$, we update the estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu^{(k)})$ by

$$\begin{aligned}
\hat{\boldsymbol{\mu}}^{(k+1)} &= \left(\sum_{i=1}^{n}\hat{u}_i^{(k)}\right)^{-1}\sum_{i=1}^{n}\widehat{u\mathbf{y}}_i^{(k)}, \\
\widehat{\boldsymbol{\Sigma}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}}_i^{(k)}\hat{\boldsymbol{\mu}}^{(k+1)\top} - \hat{\boldsymbol{\mu}}^{(k+1)}(\widehat{u\mathbf{y}}_i^{(k)})^\top + \hat{u}_i^{(k)}\hat{\boldsymbol{\mu}}^{(k+1)}\hat{\boldsymbol{\mu}}^{(k+1)\top}\right\}.
\end{aligned}$$

Then we update the parameter ν by maximizing the marginal log-likelihood function for $\mathbf{y}$, that is, $\hat{\nu}^{(k+1)} = \arg\max_{\nu}\sum_{i=1}^{n}\log f(\mathbf{V}_i \mid \mathbf{C}_i, \hat{\boldsymbol{\mu}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}^{(k+1)}; \nu^{(k)})$.

The algorithm is iterated until a suitable convergence rule is satisfied. In the later analysis, the algorithm is terminated when the difference between two successive evaluations of the log-likelihood defined in (2.23) is less than a tolerance, i.e., $\ell(\widehat{\boldsymbol{\theta}}^{(k+1)} \mid \mathbf{V}, \mathbf{C}) - \ell(\widehat{\boldsymbol{\theta}}^{(k)} \mid \mathbf{V}, \mathbf{C}) < \epsilon$, for example, $\epsilon = 10^{-6}$.

2.5 Numerical Illustrations

2.5.1 Simulation study

We conduct a simple simulation study to show how Monte Carlo (MC) estimates for the mean vector and variance-covariance matrix elements converge to the real values computed by our method. We consider a 5-variate t distribution $\mathbf{X} \sim t_5(\mathbf{0}, \boldsymbol{\Sigma}, 4)$, where $\boldsymbol{\Sigma}$ is a positive-definite matrix with unit diagonal elements and off-diagonal elements $\sigma_{ij} = \sigma_i\sigma_j$ for $i \neq j = 1, \ldots, 5$ where $\sigma_1 = -0.4$, $\sigma_2 = -0.7$, $\sigma_3 = 1$, $\sigma_4 = 0.7$, and $\sigma_5 = 0.4$. Let $\mathbf{Y} \overset{d}{=} \mathbf{X} \mid (\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b})$ be a TMVT random variable with lower and upper truncation limits $\mathbf{a} = (-\infty, -\infty, -\infty, -3, -3)^\top$ and $\mathbf{b} = (\infty, \infty, 1, 1, \infty)^\top$, respectively. Note that the

first two dimensions are not truncated, while the other three are upper, interval and lower truncated, respectively. Hence, we can write $\mathbf{a} = (-\infty_2, \mathbf{a}_2)$ and $\mathbf{b} = (\infty_2, \mathbf{b}_2)$, with $\mathbf{a}_2 = (-\infty, -3, -3)$ and $\mathbf{b}_2 = (1, 1, \infty)$. Consider the partitions $\mathbf{X}_1 = (X_1, X_2)^\top$ and $\mathbf{X}_2 = (X_3, X_4, X_5)^\top$. Since a non-truncated partition $\mathbf{X}_1$ exists, we use relations (2.14) and (2.16) given in Subsection 3.2 to compute $\boldsymbol{\xi} = \mathbb{E}[\mathbf{Y}]$ and $\boldsymbol{\Omega} = \mathrm{cov}[\mathbf{Y}]$ and obtain the true values:

$$\boldsymbol{\xi} = \left(\begin{array}{c} 0.167 \\ 0.292 \\ \hline \mathbf{-0.417} \\ \mathbf{-0.397} \\ \mathbf{-0.110} \end{array}\right) \quad \text{and} \quad \boldsymbol{\Omega} = \left(\begin{array}{ccc|ccc} 1.355 & & & & & \\ 0.224 & 1.137 & & & & \\ \hline -0.321 & -0.561 & \mathbf{0.802} & & & \\ -0.166 & -0.290 & \mathbf{0.414} & \mathbf{0.698} & & \\ -0.101 & -0.177 & \mathbf{0.253} & \mathbf{0.131} & \mathbf{1.165} \end{array}\right).$$

In this scenario, lower partitions of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ (values in bold) correspond to $\boldsymbol{\xi}_2 = \mathbb{E}[\mathbf{X}_2 \mid \mathbf{a}_2 \leqslant \mathbf{X}_2 \leqslant \mathbf{b}_2]$ and $\boldsymbol{\Omega}_{22} = \mathrm{cov}[\mathbf{X}_2 \mid \mathbf{a}_2 \leqslant \mathbf{X}_2 \leqslant \mathbf{b}_2]$ due to $\mathbb{P}(\mathbf{a} \leqslant \mathbf{X} \leqslant \mathbf{b}) = \mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{X}_2 \leqslant \mathbf{b}_2)$, which are computed using our recurrence-based formulae (2.12) and (2.13), while the reminder are computed using basic algebra where no integrals are needed.

A total of 10,000 realizations of $\mathbf{Y}$ were generated, and then the sample mean and the sample variance-covariance matrix are computed. Figures 2 and 3 shows the evolution trace of the MC estimates for the distinct elements of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$, denoted by $\hat{\xi}_i$ and $\hat{\omega}_{ij}$ for $i, j = 1, \ldots, 5$ with $i \neq j$, along with true values depicted as blue dashed lines. Note that even with 10000 MC simulations there exist slight variation in the chains for some elements as depicted in Figure 3.

Remark that the computation of the first two moments of $\mathbf{Y}$ based on $\boldsymbol{\xi}_2$ and $\boldsymbol{\Omega}_{22}$ using our method discussed in Subsection 3.2 results in 1.2 times faster than that considering the full vector $\mathbf{Y}$ with the non-truncated partition. Even though the evaluation of integrals involving infinite values are faster, the number of integrals required increases exponentially as the dimension p increases. For instance, considering a vector of dimension $p = 20$, where 15 (75%) of its dimensions are non-truncated, the computational time for evaluating the first two moments based on (2.14) and (2.16) is 10 times faster than that using the crude method. Our method indeed improves the computational efficiency.

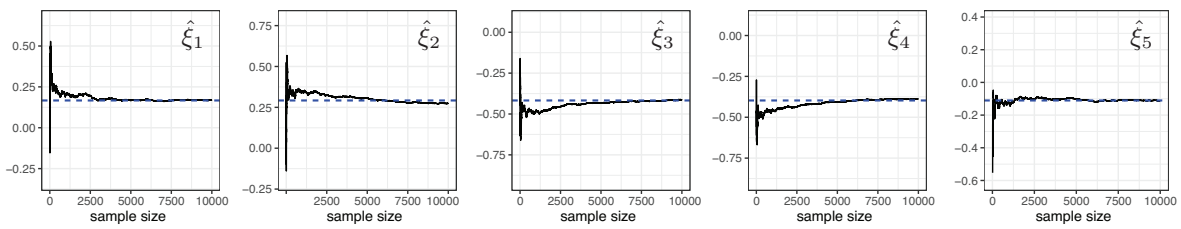


Figure 2 – MC estimates for the elements of $\boldsymbol{\xi} = \mathbb{E}[\mathbf{Y}]$.

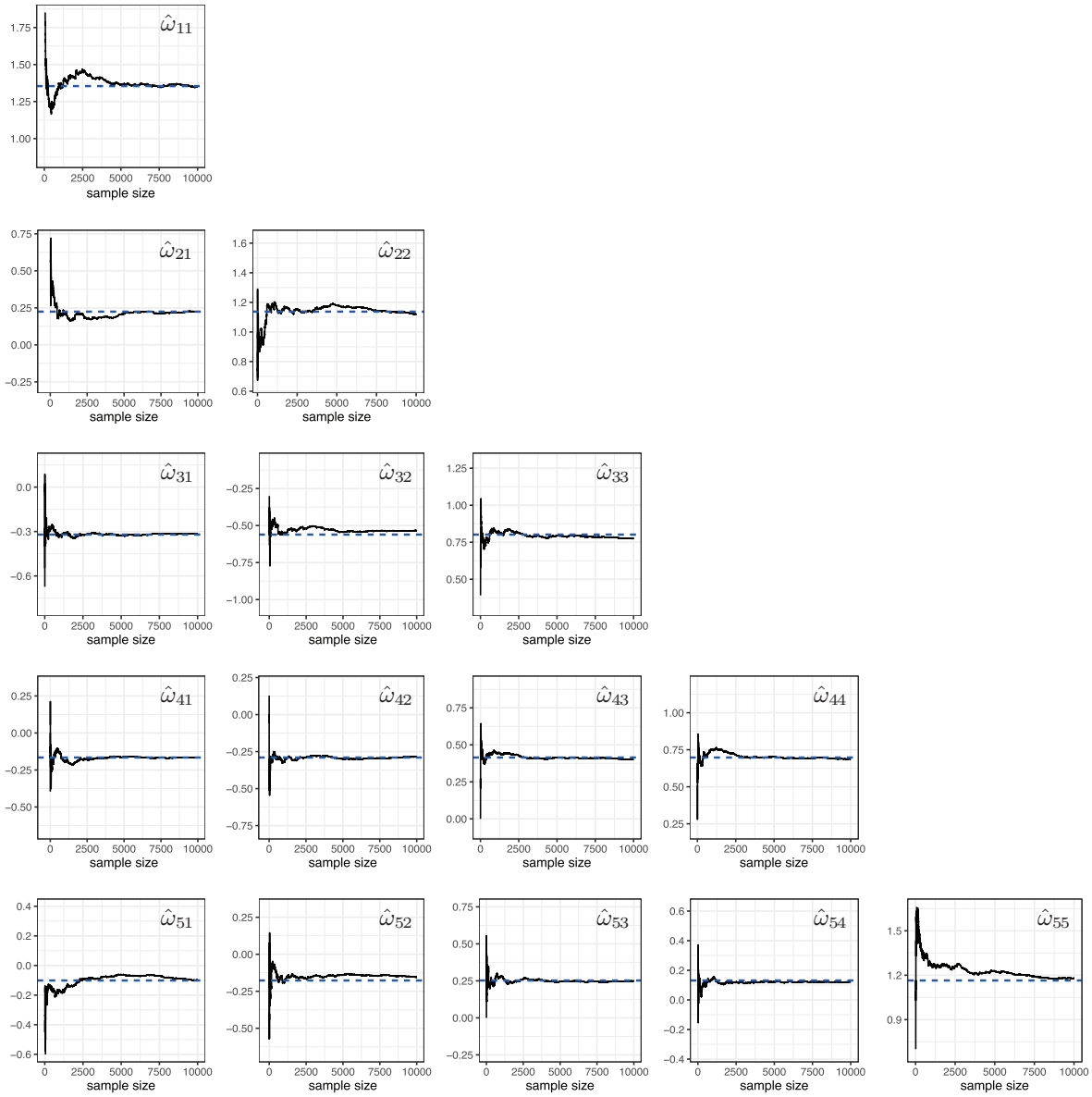


Figure 3 – MC estimates for the distinct elements of $\boldsymbol{\Omega} = \text{cov}[\mathbf{Y}]$.

2.5.2 Concentration levels study

In order to study the performance of our proposed model and algorithm, we analyze the concentrations level dataset introduced before in subsection 1.2.1. Thus, we fit the MVT-IC model to the data which contain $p = 5$ attributes, and thus we assume that $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i5}) \sim t_5(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$. For the sake of comparison, we also fit the MVN-IC model which can be considered as a limiting case when $\nu \to \infty$. As the concentration levels are strictly positive measures, to guarantee this, we consider an interval-censoring analysis by setting all lower limit of detection to equal 0 for all trace metals. Due to the different scales for each trace metal, we standardize the dataset to have zero mean and unit variance as in Wang *et al.* (2019), which considered this study as a left-censoring problem without taking in account the possibility of predicting negative concentration levels for the trace metals. For instance, we can see from Figure 5 that Pb censored concentrations

take values on the small interval $[0, 0.1]$.

Table 2 – VDEQ data. Estimated mean and ML estimate for ν and model criteria.

Model	Cu	Pb	Zn	Ca	Mg	ν	$\ell(\hat{\boldsymbol{\theta}}\|\mathbf{Y})$	AIC
Normal	0.556	0.099	2.314	12.083	3.814	-	−1351.60	2743.19
Student's t	0.557	0.102	2.329	12.084	3.817	3	−1040.21	2122.42

The ML estimates of model parameters are obtained using the EM algorithm described in Subsection 4.2. The estimated mean of the trace metals, degrees of freedom ν as well as the maximized log-likelihood and Akaike information criterion (AIC; Akaike, 1974) are shown in Table 2. Here, we can see that the estimated mean values are quite similar for both models. The estimated value of ν is fairly small, taking the minimum possible value that our algorithm supports (see subsection 2.3.1). This indicates a lack of adequacy of the normal assumption for the VDEQ data. This finding can be also confirmed from Figure 4 where the profile log-likelihood values are depicted for a grid of values of ν. As expected, the AIC value for our MVT-IC model is lower than that for MVN-IC model.

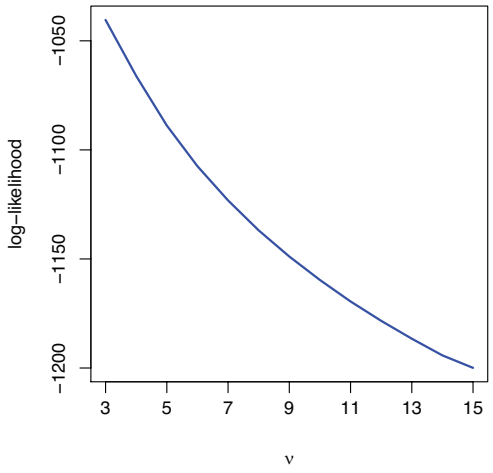


Figure 4 – VDEQ data. Plot of the profile log-likelihood of the degrees of freedom ν.

Figure 5 shows the histograms on diagonal and pair-wise scatter plots for the concentration levels study. From the histograms in diagonal of the matrix plot, we observe how censored observations (taking values over the dashed lines) are distributed to the left (blue bins) after fitting our proposed model, while gray bins represent complete observed points. On the other hand, the scatter plots in off-diagonal of the matrix plot show complete observed (black) points and the predicted observations using the multivariate SN-C model (blue triangles).

With the aim of validating the proposed censored model approach, we compare the correlation matrices of the data by considering 5 strategies: (a) *Original:* original data, (b) *Omitting:* zeros are not considered, (c) *Manipulating:* multiplying the limit of detection by the factor 0.75, (d) MVN-IC model, and (e) MVT-IC model.

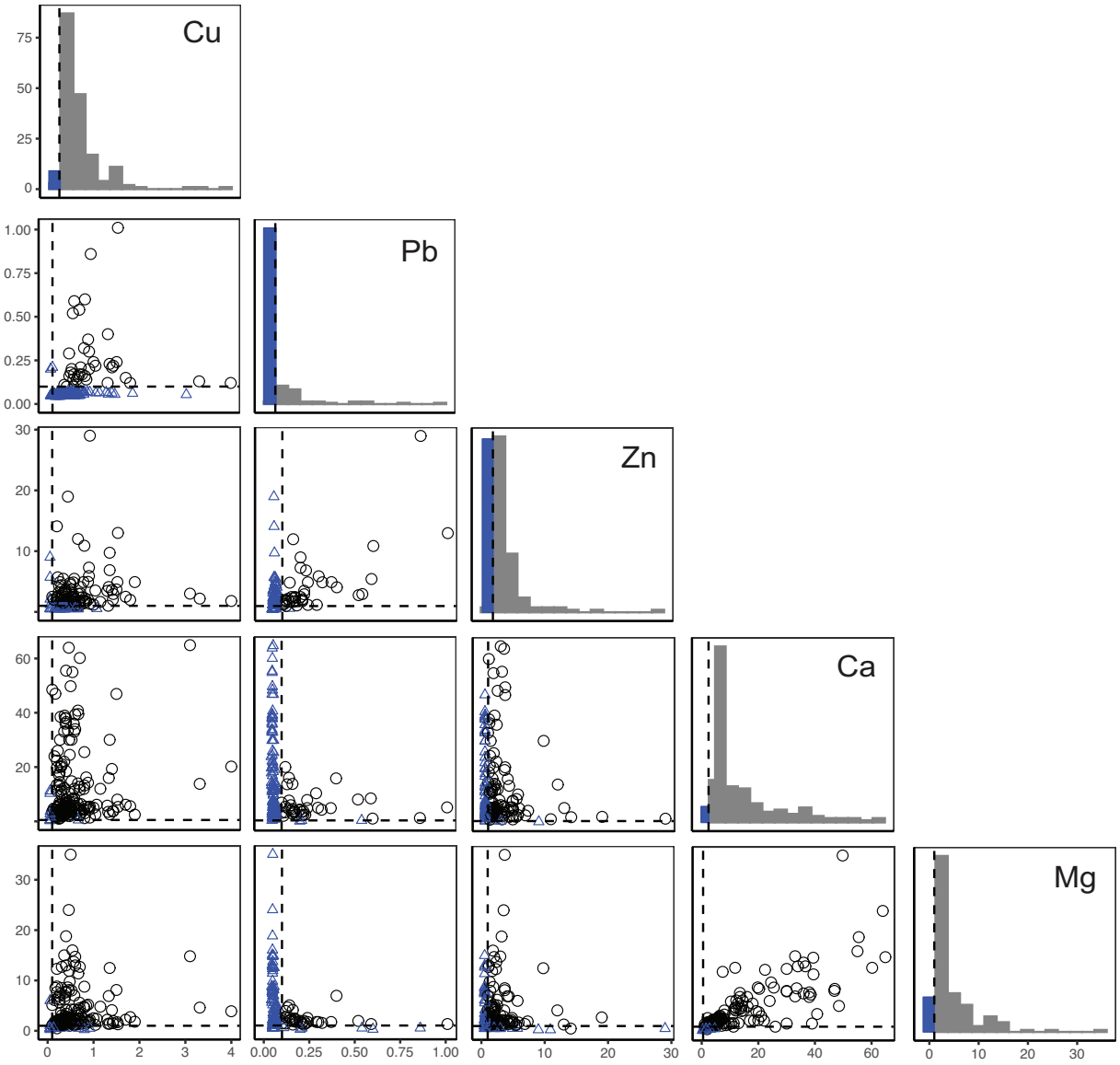


Figure 5 – VDEQ data. Histograms (diagonal) and pair-wise scatter plots (lower-triangle) for the concentration levels study. Complete observed points are represented in black points (gray bins) and MVT predicted observations in blue triangles (bins). Limit of detection are represented in dashed lines.

From the results depicted in Figure 6, we can find that the correlation matrices for the MVN-IC and MVT-IC models are similar. Based on the AIC, we consider the second one as a reference. We can get very decent results for this study by using the original data (a) or even manipulating the data (c), with both tending to underestimate the correlations. Omitting (b) is by far the worst strategy. For example, the correlation between Pb and Cu is poorly estimated to the point that they have the sign changed. Similar problems arise for the correlations between Zn and other three elements. Given the large number of censored observations, the *omitting* method leads to loss of information (say the case where the correlation between Ca and Pb and that between Ca and Mg are estimated to be zero).

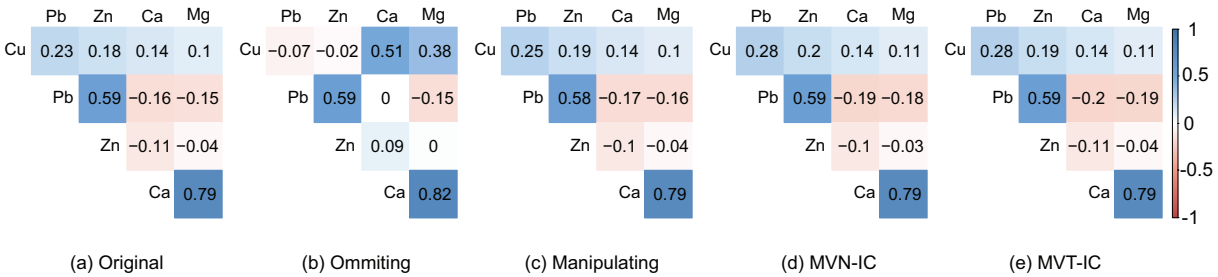


Figure 6 – VDEQ data. Correlation matrices of the concentration levels for 5 different strategies.

2.6 Conclusion

We have developed recurrence relations for integrals involving the density of MVT distribution and provided explicit expressions for the first two moments of the TMVT and FMVT distributions. These recursions allow fast computation of arbitrary-order product moments of TMVT and FMVT distributions. As an illustration, we have shown the practicability of our methods through a real-data example that contains positive censored observations. Our methods can also be applied in the context of missing observations (Lin *et al.*, 2009). The proposed methodology has been implemented in the `R MomTrunc` package, which is available on CRAN repository.

3 Efficient computation of moments of folded and doubly truncated multivariate extended skew-normal distributions

3.1 Introduction

Many applications on simulations or experimental studies, the researches often generate a large number of datasets with restricted values to fixed intervals. For example, variables such as pH, grades, viral load in HIV studies and humidity in environmental studies, have upper and lower bounds due to detection limits, and the support of their densities is restricted to some given intervals. Thus, the need to study truncated distributions along with their properties naturally arises. In this context, there has been a growing interest in evaluating the moments of truncated distributions. These variables are also often skewed, departing from the traditional assumption of using symmetric distributions. For instance, Tallis (1961) provided the formulae for the first two moments of truncated multivariate normal (TMVN) distributions. Lien (1985) gave the expressions for the moments of truncated bivariate log-normal distributions with applications to test the Houthakker effect (Houthakker, 1959) in future markets. Jawitz (2004) derived the truncated moments of several continuous univariate distributions commonly applied to hydrologic problems. Kim (2008) provided analytical formulae for moments of the truncated univariate Student-t distribution in a recursive form. Flecher *et al.* (2010) obtained expressions for the moments of truncated univariate skew-normal distributions (Azzalini, 1985) and applied the results to model the relative humidity data. Genç (2013) studied the moments of a doubly truncated member of the symmetrical class of univariate normal/independent distributions and their applications to the actuarial data. Ho *et al.* (2012) presented a general formula based on the slice sampling algorithm to approximate the first two moments of the truncated multivariate Student-t (TMVT) distribution under the double truncation. Arismendi (2013) provided explicit expressions for computing arbitrary order product moments of the TMVN multivariate distribution by using the moment generating function (MGF). However, the calculation of this approach relies on differentiation of the MGF and can be somewhat time consuming.

Instead of differentiating the MGF of the TMVN distribution, Kan & Robotti (2017) recently presented recurrence relations for integrals that are directly related to the density of the multivariate normal distribution for computing arbitrary order product moments of the TMVN distribution. These recursions offer a fast computation of the

moments of folded (FN) and TMVN distributions, which require evaluating p-dimensional integrals that involve the Normal (N) density. Explicit expressions for some low order moments of FN and TMVN distributions are presented in a clever way, although some proposals to calculate the moments of the univariate truncated skew-normal distribution and truncated univariate skew-normal/independent distribution (Flecher *et al.*, 2010) has recently been published. So far, to the best of our knowledge, there has not been attempt on studying neither moments nor product moments of the folded multivariate extended skew-normal (FESN) and truncated multivariate extended skew-normal (TESN) distributions. Moreover, our proposed methods allow to compute, as a by-product, the product moments of folded and truncated distributions, of the N (Kan & Robotti, 2017), SN (Azzalini & Dalla-Valle, 1996), and their respective univariate versions. The proposed algorithm and methods are implemented in the new `R` package `MomTrunc`.

The rest of this paper is organized as follows. In Section 3.2 we briefly discuss some preliminary results related to the multivariate SN, ESN and TESN distributions and some of its key properties. The section 3.3 presents a recurrence formula of an integral to be applied in the essential evaluation of moments of the TESN distribution as well as explicit expressions for the first two moments of the TESN and TMVN distributions. A direct relation between the moments of the TESN and TMVN distribution is also presented which is used to improved the proposed methods. In section 3.4, by means of approximations, we propose strategies to circumvent some numerical problems that arise on limiting distributions and extreme cases. We compare our proposal with others popular methods of the literature in Section 3.5. Finally, Section 3.6 is devoted to the moments of the FESN distribution, several related results are discussed. Explicit expressions are presented for high order moments for the univariate case and the mean vector and variance-covariance matrix of the multivariate FESN distribution. Finally, some concluding remarks are presented in Section 3.7.

3.2 Preliminaries

3.2.1 The multivariate skew-normal distribution

In this subsection we present the skew-normal distribution and some of its properties. We say that a $p \times 1$ random vector $\mathbf{Y}$ follows a multivariate SN distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$ and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, and we write $\mathbf{Y} \sim \mathrm{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its joint probability density function (pdf) is given by

$$SN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{3.1}$$

where $\Phi_1(\cdot)$ represents the cumulative distribution function (cdf) of the standard univariate normal distribution. If $\boldsymbol{\lambda} = \mathbf{0}$ then (3.1) reduces to the symmetric $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pdf. Except

by a straightforward difference in the parametrization considered in (3.1), this model corresponds to the one introduced by Azzalini & Dalla-Valle (1996), whose properties were extensively studied in Azzalini & Capitanio (1999). See also Arellano-Valle & Genton (2005).

Proposition 3.1 (***cdf* of the SN**). *If* $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, *then for any* $\mathbf{y} \in \mathbb{R}^p$

$$F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}(\mathbf{Y} \leqslant \mathbf{y}) = 2\Phi_{p+1}\big((\mathbf{y}^\top, 0)^\top; \boldsymbol{\mu}^*, \boldsymbol{\Omega}\big), \tag{3.2}$$

where $\boldsymbol{\mu}^* = (\boldsymbol{\mu}^\top, 0)^\top$ *and* $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & -\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}^\top & 1 \end{pmatrix}$, *with* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

Proof of proposition 3.1 can be found in Azzalini & Dalla-Valle (1996). It is worth mentioning that the multivariate skew-normal distribution is closed over marginalization but not conditioning. Next we present its extended version which holds both properties, called, the multivariate ESN distribution.

3.2.2 The extended multivariate skew-normal distribution

We say that a $p \times 1$ random vector $\mathbf{Y}$ follows a ESN distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$, a $p \times 1$ skewness parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, and a shift or extension parameter $\tau \in \mathbb{R}$, denoted by $\mathbf{Y} \sim \text{ESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, if its pdf is given by

$$ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \xi^{-1}\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{3.3}$$

with $\xi = \Phi_1(\tau/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})$. Note that when $\tau = 0$, we retrieve the skew-normal distribution defined in (3.1), that is, $ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0) = SN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Here, we used a slightly different parametrization of the ESN distribution than the one given in Arellano-Valle & Azzalini (2006a) and Arellano-Valle & Genton (2010). Futhermore, Arellano-Valle & Genton (2010) deals with the multivariate extended skew-t (EST) distribution, in which the ESN is a particular case when the degrees of freedom ν goes to infinity. From this last work, it is straightforward to see that

$$ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) \longrightarrow \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ as } \tau \to +\infty.$$

Also, letting $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})$, it follows that $\mathbf{Z} \sim \text{ESN}_p(\mathbf{0}, \mathbf{I}, \boldsymbol{\lambda}, \tau)$, with mean vector and variance-covariance matrix

$$\mathbb{E}[\mathbf{Z}] = \eta\boldsymbol{\lambda} \qquad \text{and} \qquad \text{cov}[\mathbf{Z}] = \mathbf{I}_p - \mathbb{E}[\mathbf{Z}]\left(\mathbb{E}[\mathbf{Z}] - \frac{\tau}{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\boldsymbol{\lambda}\right)^\top,$$

with $\eta = \phi_1(\tau; 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})/\xi$. Then, the mean vector and variance-covariance matrix of $\mathbf{Y}$ can be easily computed as $\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbb{E}[\mathbf{Z}]$ and $\text{cov}[\mathbf{Y}] = \boldsymbol{\Sigma}^{1/2}\text{cov}[\mathbf{Z}]\boldsymbol{\Sigma}^{1/2}$.

Next, we present some propositions that are crucial to develop our methods. First, we propose the marginal and conditional distribution of the ESN with pdf as in (3.3) (proof can be found in the Appendix B), while the second and third proposition comes from Arellano-Valle & Azzalini (2006a) and Arellano-Valle & Genton (2010).

Proposition 3.2 (*Marginal and conditional distribution of the ESN*). *Let* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ *and* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 $(p_1 + p_2 = p)$, *respectively. Let*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top \quad \text{and} \quad \boldsymbol{\varphi} = (\boldsymbol{\varphi}_1^\top, \boldsymbol{\varphi}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ *and* $\boldsymbol{\varphi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. *Then,*

$$\mathbf{Y}_1 \sim ESN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1, c_{12}\tau), \quad \mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1 \sim ESN_{p_2}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2, \tau_{2.1})$$

where $c_{12} = (1 + \boldsymbol{\varphi}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ *and* $\tau_{2.1} = \tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)$.

Proposition 3.3 (*Stochastic representation of the ESN*). *Let* $\mathbf{X} \sim N_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Omega})$ *with* $\mathbf{X}$ *part as* $\mathbf{X} = (\mathbf{X}_1{}^\top, X_2)^\top$. *If*

$$\mathbf{Y} \overset{d}{=} (\mathbf{X}_1 | X_2 < \tilde{\tau}), \tag{3.4}$$

it follows that $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, *with* $\boldsymbol{\mu}^*$ *and* $\boldsymbol{\Omega}$ *as defined in Proposition 3.1, and* $\tau = \tilde{\tau}(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

Proposition 3.4 (cdf *of the ESN*). *If* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, *then for any* $\mathbf{y} \in \mathbb{R}^p$

$$F_{\mathbf{Y}}(\mathbf{y}) = \mathbb{P}(\mathbf{Y} \leqslant \mathbf{y}) = \frac{\Phi_{p+1}\big((\mathbf{y}^\top, \tilde{\tau})^\top; \boldsymbol{\mu}^*, \boldsymbol{\Omega}\big)}{\Phi(\tilde{\tau})}. \tag{3.5}$$

Proof is direct from proposition 3.3. Hereinafter, for $\mathbf{Y} \sim \text{ESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, we will denote to its *cdf* as $F_{\mathbf{Y}}(\mathbf{y}) \equiv \tilde{\Phi}_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ for simplicity.

Let $\mathbb{A}$ be a Borel set in $\mathbb{R}^p$. We say that the random vector $\mathbf{Y}$ has a truncated extended skew-normal distribution on $\mathbb{A}$ when $\mathbf{Y}$ has the same distribution as $\mathbf{Y}|(\mathbf{Y} \in \mathbb{A})$. In this case, the pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)}{\mathbb{P}(\mathbf{Y} \in \mathbb{A})}\mathbf{1}_{\mathbb{A}}(\mathbf{y}),$$

where $\mathbf{1}_{\mathbb{A}}$ is the indicator function of $\mathbb{A}$. We use the notation $\mathbf{Y} \sim \text{TESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau; \mathbb{A})$. If $\mathbb{A}$ has the form

$$\mathbb{A} = \{(x_1, \ldots, x_p) \in \mathbb{R}^p : \ a_1 \leqslant x_1 \leqslant b_1, \ldots, a_p \leqslant x_p \leqslant b_p\} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{a} \leqslant \mathbf{x} \leqslant \mathbf{b}\}, \tag{3.6}$$

then we use the notation $\{\mathbf{Y} \in \mathbb{A}\} = \{\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}\}$, where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$. Here, we say that the distribution of $\mathbf{Y}$ is doubly truncated. Analogously we define $\{\mathbf{Y} \geqslant \mathbf{a}\}$ and $\{\mathbf{Y} \leqslant \mathbf{b}\}$. Thus, we say that the distribution of $\mathbf{Y}$ is truncated from below and truncated from above, respectively. For convenience, we also use the notation $\mathbf{Y} \sim \text{TESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau; (\mathbf{a}, \mathbf{b}))$.

3.3 On moments of the doubly truncated multivariate ESN distribution

3.3.1 A recurrence relation

For two p-dimensional vectors $\mathbf{x} = (x_1, \ldots, x_p)^\top$ and $\boldsymbol{\kappa} = (k_1, \ldots, k_p)^\top$, let $\mathbf{x}^{\boldsymbol{\kappa}}$ stand for $(x_1^{\kappa_1}, \ldots, x_p^{\kappa_p})$, and let $\mathbf{a}_{(i)}$ be a vector $\mathbf{a}$ with its ith element being removed. For a matrix $\mathbf{A}$, we let $\mathbf{A}_{i(j)}$ stand for the ith row of $\mathbf{A}$ with its jth element being removed. Similarly, $\mathbf{A}_{(i,j)}$ stands for the matrix $\mathbf{A}$ with its ith row and jth columns being removed. Besides, let $\mathbf{e}_i$ denote a $p \times 1$ vector with its ith element equaling one and zero otherwise. Let

$$\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \int_{\mathbf{a}}^{\mathbf{b}} ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)\mathrm{d}\mathbf{x}.$$

We are interested in evaluating the integral

$$\mathcal{F}_{\boldsymbol{\kappa}}^p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)\mathrm{d}\mathbf{x}. \tag{3.7}$$

The initial condition is obviously $\mathcal{F}_0^p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$. When $\boldsymbol{\lambda} = \mathbf{0}$ and $\tau = 0$, we recover the multivariate normal case, and then

$$\mathcal{F}_{\boldsymbol{\kappa}}^p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0}, 0) \equiv F_{\boldsymbol{\kappa}}^p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x}^{\boldsymbol{\kappa}} \phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{d}\mathbf{x}, \tag{3.8}$$

with initial condition

$$\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0}, 0) \equiv L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathbf{a}}^{\mathbf{b}} \phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{d}\mathbf{x}. \tag{3.9}$$

Note that we use calligraphic style for the integrals of interest $\mathcal{F}_{\boldsymbol{\kappa}}^p$ and $\mathcal{L}_p$ when we work with the skewed version. In both expressions (3.8) and (3.9), for the normal case, we are using compatible notation with the one used by Kan & Robotti (2017).

3.3.1.1 Univariate case

Let $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda, \tau)^\top$ be the set of parameters. When $p = 1$, it is straightforward to use integration by parts to show that

$$\mathcal{F}_0^1(a, b; \boldsymbol{\theta}) \;=\; \xi^{-1}\left[\Phi_2\left((b-\mu, \tau)^\top; \mathbf{0}, \boldsymbol{\Omega}\right) - \Phi_2\left((a-\mu, \tau)^\top; \mathbf{0}, \boldsymbol{\Omega}\right)\right],$$

$$\begin{aligned}\mathcal{F}^1_{k+1}(a,b;\boldsymbol{\theta}) &= \mu\mathcal{F}^1_k(a,b;\boldsymbol{\theta}) + k\sigma^2\mathcal{F}^1_{k-1}(a,b;\boldsymbol{\theta}) + \lambda\sigma\eta F^1_k(a,b;\mu-\mu_b,\gamma^2)\\ &\quad +\sigma^2\big(a^k ESN_1(a;\boldsymbol{\theta}) - b^k ESN_1(b;\boldsymbol{\theta})\big);\quad \text{for } k\geqslant 0,\end{aligned}$$

where $\boldsymbol{\Omega} = \begin{pmatrix} \sigma^2 & -\sigma\psi \\ -\sigma\psi & 1 \end{pmatrix}$, $\psi = \lambda/\sqrt{1+\lambda^2}$, $\mu_b = \dfrac{1}{\sigma}\lambda\tau\gamma^2$ and $\gamma = \sigma/\sqrt{1+\lambda^2}$.

When $p > 1$, we need a similar recurrence relation in order to compute $\mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b},\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$ which we propose in the next theorem.

3.3.1.2 Multivariate case

Theorem 3.1. *For $p\geqslant 1$ and $i = 1,\ldots,p$,*

$$\mathcal{F}^p_{\boldsymbol{\kappa}+\mathbf{e}_i}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \mu_i\mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) + \delta_i F^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu}-\boldsymbol{\mu}_b,\boldsymbol{\Gamma}) + \mathbf{e}_i^\top\boldsymbol{\Sigma}\mathbf{d}_{\boldsymbol{\kappa}}, \tag{3.10}$$

where $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_p)^\top = \eta\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$, $\boldsymbol{\mu}_b = \tilde{\tau}\boldsymbol{\Delta}$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}-\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ and $\mathbf{d}_{\boldsymbol{\kappa}}$ is a p-vector with jth element

$$\begin{aligned} d_{\boldsymbol{\kappa},j} &= k_j\mathcal{F}^p_{\boldsymbol{\kappa}-\mathbf{e}_j}(\mathbf{a},\mathbf{b},\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\\ &\quad +a_j^{k_j} ESN_1(a_j;\mu_j,\sigma_j^2,c_j\sigma_j\tilde{\varphi}_j,c_j\tau)\mathcal{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\boldsymbol{\Sigma}}_j,\tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)},\tilde{\tau}_j^{\mathbf{a}})\\ &\quad -b_j^{k_j} ESN_1(b_j;\mu_j,\sigma_j^2,c_j\sigma_j\tilde{\varphi}_j,c_j\tau)\mathcal{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\boldsymbol{\Sigma}}_j,\tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)},\tilde{\tau}_j^{\mathbf{b}}),\end{aligned} \tag{3.11}$$

with $\tilde{\tau}_j^{\mathbf{a}} = \tau+\tilde{\varphi}_j(a_j-\mu_j)$ and $\tilde{\tau}_j^{\mathbf{b}} = \tau+\tilde{\varphi}_j(b_j-\mu_j)$, where

$$\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}} = \boldsymbol{\mu}_{(j)} + \boldsymbol{\Sigma}_{(j),j}\frac{a_j-\mu_j}{\sigma_j^2},\ \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}} = \boldsymbol{\mu}_{(j)} + \boldsymbol{\Sigma}_{(j),j}\frac{b_j-\mu_j}{\sigma_j^2},\ \tilde{\varphi}_j = \varphi_j + \frac{1}{\sigma_j^2}\boldsymbol{\Sigma}_{j(j)}\boldsymbol{\varphi}_{(j)},$$

$$c_j = \frac{1}{(1+\boldsymbol{\varphi}_{(j)}^\top\tilde{\boldsymbol{\Sigma}}_j\boldsymbol{\varphi}_{(j)})^{1/2}},\qquad \textit{and}\qquad \tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}_{(j),(j)} - \frac{1}{\sigma_j^2}\boldsymbol{\Sigma}_{(j),j}\boldsymbol{\Sigma}_{j,(j)}.$$

Proof. Taking the derivative of the ESN density, we have

$$\begin{aligned}\frac{\partial}{\partial\mathbf{x}}ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) &= \xi^{-1}\Big\{\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\frac{\partial}{\partial\mathbf{x}}\Phi_1\big(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\big) + \frac{\partial}{\partial\mathbf{x}}\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\\ &\qquad \Phi(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\Big\},\\ &= \xi^{-1}\Big\{\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}\phi_1(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) - \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\\ &\qquad \times\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi_1(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\Big\},\\ &= -\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) + \xi^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}\,\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\\ &\qquad \times\phi_1(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})).\end{aligned} \tag{3.12}$$

On the other hand we have that

$$\begin{aligned}\phi_1(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{|\boldsymbol{\Sigma}^{-1}|^{1/2}}{(2\pi)^{\frac{p+1}{2}}}\exp\left\{-\frac{1}{2}\left[\delta(\mathbf{x})+(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))^2\right]\right\},\\ &= |\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}|^{1/2}\exp\left\{\frac{\boldsymbol{\mu}_b^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}_b}{2}\right\}\times\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\tau^2}{2}\right\}\\ &\quad\times\frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Gamma}|^{1/2}}\exp\left\{-\frac{\delta(\mathbf{x};\boldsymbol{\mu}-\boldsymbol{\mu}_b,\boldsymbol{\Gamma})}{2}\right\},\\ &= |\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}|^{1/2}\exp\left\{\frac{\boldsymbol{\mu}_b^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}_b}{2}\right\}\phi_1(\tau)\phi_p(\mathbf{x};\boldsymbol{\mu}-\boldsymbol{\mu}_b,\boldsymbol{\Gamma}),\end{aligned} \tag{3.13}$$

where we use the fact that,

$$\begin{aligned}\delta(\mathbf{x})+(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))^2 &= (\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\mu})+2\boldsymbol{\mu}_b^\top\boldsymbol{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\mu})+\tau^2,\\ &= (\mathbf{x}-(\boldsymbol{\mu}-\boldsymbol{\mu}_b))^\top\boldsymbol{\Gamma}^{-1}(\mathbf{x}-(\boldsymbol{\mu}-\boldsymbol{\mu}_b))-\boldsymbol{\mu}_b^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}_b+\tau^2,\\ &= \delta(\mathbf{x};\boldsymbol{\mu}-\boldsymbol{\mu}_b,\boldsymbol{\Gamma})-\boldsymbol{\mu}_b^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}_b+\tau^2,\end{aligned}$$

and where $\delta(\mathbf{x})\equiv\delta(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})=(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is the Mahalanobis distance and with $\boldsymbol{\Gamma}=\boldsymbol{\Sigma}^{1/2}(\mathbf{I}_p+\boldsymbol{\lambda}\boldsymbol{\lambda}^\top)^{-1}\boldsymbol{\Sigma}^{1/2}=\boldsymbol{\Sigma}-\boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ (using the fact that, $(\mathbf{I}_p+\boldsymbol{\lambda}\boldsymbol{\lambda}^\top)^{-1}=\mathbf{I}_p-(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top$) and $\boldsymbol{\mu}_b=\tau\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}=\tilde{\tau}\boldsymbol{\Delta}$.

Plugging (3.13) in (3.12), we obtain

$$-\frac{\partial}{\partial\mathbf{x}}ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)=\boldsymbol{\Sigma}^{-1}\left[(\mathbf{x}-\boldsymbol{\mu})ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)-\boldsymbol{\delta}\phi_p(\mathbf{x};\boldsymbol{\mu}-\boldsymbol{\mu}_b,\boldsymbol{\Gamma})\right],$$

where $\boldsymbol{\delta}=\eta\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$ and

$$\begin{aligned}\eta &= |\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}|^{1/2}\times\frac{\phi_1(\tau)}{\xi}\exp\left\{\frac{\boldsymbol{\mu}_b^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\mu}_b}{2}\right\},\\ &= |\mathbf{I}_p+\boldsymbol{\lambda}\boldsymbol{\lambda}^\top|^{-1/2}\times\frac{\phi_1(\tau)}{\xi}\exp\left\{\frac{\tau^2\boldsymbol{\lambda}^\top(\mathbf{I}_p+\boldsymbol{\lambda}\boldsymbol{\lambda}^\top)^{-1}\boldsymbol{\lambda}}{2}\right\},\\ &= \frac{1}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}\times\frac{1}{\xi\sqrt{2\pi}}\exp\left\{-\frac{\tau^2}{2}\left[1-\frac{\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\right]\right\},\\ &= \frac{1}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}\times\frac{1}{\xi\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{\tau^2}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\right\},\\ &= \frac{\phi_1(\tau;0,1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}{\xi}\end{aligned} \tag{3.14}$$

with $det(\mathbf{I}_p+\boldsymbol{\lambda}\boldsymbol{\lambda}^\top)=1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}$ from the Sylvester's determinant identity (Harville, 1997).

Multiplying both sides by $\mathbf{x}^{\boldsymbol{\kappa}}$ and integrating from $\mathbf{a}$ to $\mathbf{b}$, we have (after suppressing the arguments of $\mathcal{F}_{\boldsymbol{\kappa}}^p$ and $F_{\boldsymbol{\kappa}}^p$) that

$$\mathbf{d}_{\boldsymbol{\kappa}}=\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathcal{F}_{\boldsymbol{\kappa}+\mathbf{e}_1}^p & - & \mu_1\mathcal{F}_{\boldsymbol{\kappa}}^p & - & \delta_1F_{\boldsymbol{\kappa}}^p\\ \mathcal{F}_{\boldsymbol{\kappa}+\mathbf{e}_2}^p & - & \mu_2\mathcal{F}_{\boldsymbol{\kappa}}^p & - & \delta_2F_{\boldsymbol{\kappa}}^p\\ & \vdots & & \vdots & \\ \mathcal{F}_{\boldsymbol{\kappa}+\mathbf{e}_p}^p & - & \mu_p\mathcal{F}_{\boldsymbol{\kappa}}^p & - & \delta_pF_{\boldsymbol{\kappa}}^p\end{bmatrix},$$

and the jth element of the left hand side is

$$d_{\boldsymbol{\kappa},j} = -\int_{\mathbf{a}(j)}^{\mathbf{b}(j)} \mathbf{x}^{\boldsymbol{\kappa}} ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\Big|_{x_j=a_j}^{b_j} \mathrm{d}\mathbf{x}(j) + \int_{\mathbf{a}}^{\mathbf{b}} k_j \mathbf{x}^{\boldsymbol{\kappa}-\mathbf{e}_j} ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\mathrm{d}\mathbf{x}$$

by using integration by parts. Using Proposition 3.2, we know that

$$ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\big|_{x_j=a_j} = ESN_1(a_j;\mu_j,\sigma_j^2,c_j\sigma_j\tilde{\varphi}_j,c_j\tau)ESN_{p-1}(\mathbf{x}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\boldsymbol{\Sigma}}_j,\tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)},\tilde{\tau}_j^{\mathbf{a}}),$$

$$ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\big|_{x_j=b_j} = ESN_1(b_j;\mu_j,\sigma_j^2,c_j\sigma_j\tilde{\varphi}_j,c_j\tau)ESN_{p-1}(\mathbf{x}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\boldsymbol{\Sigma}}_j,\tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)},\tilde{\tau}_j^{\mathbf{b}}),$$

and we obtain

$$\begin{aligned} d_{\boldsymbol{\kappa},j} \;=\; & k_j\mathcal{F}^p_{\boldsymbol{\kappa}-\mathbf{e}_j}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) \\ & + a_j^{k_j} ESN_1(a_j;\mu_j,\sigma_j^2,c_j\sigma_j\tilde{\varphi}_j,c_j\tau)\mathcal{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{a}},\tilde{\boldsymbol{\Sigma}}_j,\tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)},\tilde{\tau}_j^{\mathbf{a}}) \\ & - b_j^{k_j} ESN_1(b_j;\mu_j,\sigma_j^2,c_j\sigma_j\tilde{\varphi}_j,c_j\tau)\mathcal{F}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\mathbf{a}_{(j)},\mathbf{b}_{(j)};\tilde{\boldsymbol{\mu}}_j^{\mathbf{b}},\tilde{\boldsymbol{\Sigma}}_j,\tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)},\tilde{\tau}_j^{\mathbf{b}}). \end{aligned}$$

Finally, multiplying both sides by $\boldsymbol{\Sigma}$, we obtain (3.10). This completes the proof. □

This delivers a simple way to compute any arbitrary moments of multivariate TSN distribution $\mathcal{F}^p_{\boldsymbol{\kappa}}$ based on at most $3p+1$ lower order terms, with $p+1$ of them being p-dimensional integrals, the rest being $(p-1)$-dimensional integrals, and a normal integral $F^p_{\boldsymbol{\kappa}}$ that can be easily computed through our proposed `R` package `MomTrunc` available at CRAN. When $k_j = 0$, the first term in (3.11) vanishes. When $a_j = -\infty$, the second term vanishes, and when $b_j = +\infty$, the third term vanishes. When we have no truncation, that is, all the $a_i's$ are $-\infty$ and all the $b_i's$ are $+\infty$, for $\mathbf{Y} \sim \mathrm{ESN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$, we have that

$$\mathcal{F}^p_{\boldsymbol{\kappa}}(-\infty,+\infty;\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}],$$

and in this case the recursive relation is

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}+\mathbf{e}_i}] = \mu_i\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] + \delta_i\mathbb{E}[\mathbf{W}^{\boldsymbol{\kappa}}] + \sum_{j=1}^{p}\sigma_{ij}k_j\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}-\mathbf{e}_i}], \quad i = 1,\ldots,p,$$

with $\mathbf{W} \sim N_p(\boldsymbol{\mu}-\boldsymbol{\mu}_b,\boldsymbol{\Gamma})$.

It is worth to stress that any arbitrary truncated moment of $\mathbf{Y}$, that is,

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}|\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \frac{\mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)}{\mathcal{L}_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)}, \tag{3.15}$$

can be computed using the recurrence relation given in Theorem 3.1. In the next section, we proposed another approach to compute (3.15) using a unique corresponding arbitrary moment to a truncated normal vector.

3.3.2 Computing ESN moments based on normal moments

Next, we present a theorem establishing a 1-1 relation between the ESN integral $\mathcal{F}_{\boldsymbol{\kappa}}$ and the normal integral $F_{\boldsymbol{\kappa}}$.

Theorem 3.2. *It holds that*

$$\mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \xi^{-1} F^{p+1}_{\boldsymbol{\kappa}^*}(\mathbf{a}^*,\mathbf{b}^*;\boldsymbol{\mu}^*,\boldsymbol{\Omega}), \tag{3.16}$$

with $\boldsymbol{\mu}^*$ *and* $\boldsymbol{\Omega}$ *as defined in Proposition 3.1, and* $\boldsymbol{\kappa}^* = (\boldsymbol{\kappa}^\top, 0)^\top$, $\mathbf{a}^* = (\mathbf{a}^\top, -\infty)^\top$ *and* $\mathbf{b}^* = (\mathbf{b}^\top, \tilde{\tau})^\top$.

In particular, for $\boldsymbol{\kappa} = \mathbf{0}$, *then*

$$\mathcal{L}_p(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \xi^{-1} L_{p+1}(\mathbf{a}^*,\mathbf{b}^*;\boldsymbol{\mu}^*,\boldsymbol{\Omega}). \tag{3.17}$$

Proof. From the inclusion-exclusion principle, we have that

$$\mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} \int_{-\infty}^{\mathbf{s}} \mathbf{x}^{\boldsymbol{\kappa}} ESN_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\mathrm{d}\mathbf{x},$$

where $S(\mathbf{a},\mathbf{b}) = \mathbf{a}\times\mathbf{b}$ is a cartesian product with 2^p elements of the form $\mathbf{s} = (s_1,\ldots,s_p)^\top$, with $s_i \in \{a_i, b_i\}$ for $i = 1,\ldots,p$ and $n_s = \sum_{i=1}^{p} \mathbb{1}(s_i = a_i)$. For $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$ and $\mathbf{X} = (\mathbf{X}_1{}^\top, X_2)^\top \sim N_{p+1}(\boldsymbol{\mu}^*,\boldsymbol{\Omega})$, it follows from its stochastic representation (3.4) that

$$\begin{aligned}
\mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{a},\mathbf{b};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) &= \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} \int_{-\infty}^{\mathbf{s}} \mathbf{x}_1^{\boldsymbol{\kappa}} f_{\mathbf{X}_1}(\mathbf{x}_1|X_2<\tilde{\tau})\mathrm{d}\mathbf{x}_1,\\
&= \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} \int_{-\infty}^{\mathbf{s}} \int_{-\infty}^{\tilde{\tau}} \mathbf{x}_1^{\boldsymbol{\kappa}} \frac{f_{\mathbf{Y}}(\mathbf{x}_1,x_2)}{P(X_2<\tilde{\tau})}\mathrm{d}x_2\mathrm{d}\mathbf{x}_1,\\
&= \xi^{-1} \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} \int_{-\infty}^{\mathbf{s}} \int_{-\infty}^{\tilde{\tau}} \mathbf{x}^{\boldsymbol{\kappa}^*} f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x},\\
&= \xi^{-1} \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} \int_{-\infty}^{\mathbf{s}_1} \mathbf{x}^{\boldsymbol{\kappa}^*} \phi_{p+1}(\mathbf{x};\boldsymbol{\mu}^*,\boldsymbol{\Omega})\mathrm{d}\mathbf{x},\\
&= \xi^{-1} \sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} F^{p+1}_{\boldsymbol{\kappa}^*}(-\infty,\mathbf{s}_1;\boldsymbol{\mu}^*,\boldsymbol{\Omega})
\end{aligned}$$

with $\mathbf{s}_1 = (\mathbf{s}^\top, \tilde{\tau})^\top$ being a vector of dimension $p+1$. For convenience, we preserve the index $\mathbf{s}$ in the summation due to $\mathbf{s}_1$ is a one-to-one transformation. Similarly, we define the vector $\mathbf{s}_0 = (\mathbf{s}^\top, -\infty)^\top$.

Let U_0 and U_1 be the two sets $U_0 = \bigcup_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} \mathbf{s}_0$ and $U_1 = \bigcup_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} \mathbf{s}_1$, both with 2^p elements. Then, $U_0 \cup U_1$ contains the same 2^{p+1} elements $\mathbf{s}^*$ in $S(\mathbf{a}^*,\mathbf{b}^*)$ for $\mathbf{a}^* = (\mathbf{a}^\top, -\infty)^\top$

and $\mathbf{b}^* = (\mathbf{b}^\top, \tilde{\tau})^\top$. Since $\mathrm{F}^{p+1}_{\boldsymbol{\kappa}^*}(\mathbf{s}_0; \boldsymbol{\mu}^*, \boldsymbol{\Omega}) = 0$ for all $\mathbf{s}_0 \in U_0$ and $n_{s^*} = n_s + \mathbb{1}(s_{p+1} = -\infty)$, then

$$\begin{aligned}
\sum_{\mathbf{s}\in S(\mathbf{a},\mathbf{b})} (-1)^{n_s} F^{p+1}_{\boldsymbol{\kappa}^*}(-\infty, \mathbf{s}_1; \boldsymbol{\mu}^*, \boldsymbol{\Omega}) &= \sum_{\mathbf{s}_0\in U_0} (-1)^{n_s+1} F^{p+1}_{\boldsymbol{\kappa}^*}(-\infty, \mathbf{s}_0; \boldsymbol{\mu}^*, \boldsymbol{\Omega}) \\
&\quad + \sum_{\mathbf{s}_1\in U_1} (-1)^{n_s} F^{p+1}_{\boldsymbol{\kappa}^*}(-\infty, \mathbf{s}_1; \boldsymbol{\mu}^*, \boldsymbol{\Omega}), \\
&= \sum_{\mathbf{s}^*\in S(\mathbf{a}^*,\mathbf{b}^*)} (-1)^{n_{s^*}} F^{p+1}_{\boldsymbol{\kappa}^*}(-\infty, \mathbf{s}^*; \boldsymbol{\mu}^*, \boldsymbol{\Omega}), \\
&= F^{p+1}_{\boldsymbol{\kappa}^*}(\mathbf{a}^*, \mathbf{b}^*; \boldsymbol{\mu}^*, \boldsymbol{\Omega}).
\end{aligned}$$

□

Equation (3.17) offers us in a very convenient manner to compute the probability $\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, since efficient algorithms already exist to calculate $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ (e.g., see Genz (1992)), which avoids performing 2^p evaluations of *cdf* of the multivariate N distribution.

Corollary 3.1. *For* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ *and* $\mathbf{X} \sim N_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Omega})$*, it follows from Theorem 3.2 that*

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}} | \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}^*} | \mathbf{a}^* \leqslant \mathbf{X} \leqslant \mathbf{b}^*],$$

with $\mathbf{a}^*$*,* $\mathbf{b}^*$*,* $\boldsymbol{\kappa}^*$*,* $\boldsymbol{\mu}^*$ *and* $\boldsymbol{\Omega}$ *as defined in Theorem 3.2.*

3.3.3 Mean and covariance matrix of multivariate TESN distributions

Let us consider $\mathbf{Y} \sim \mathrm{TESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, (\mathbf{a}, \mathbf{b}))$. In light of Theorem 3.1, we have that

$$\begin{aligned}
\mathbb{E}[Y_i] = \frac{1}{\mathcal{L}} \Bigg[\delta_i L + \sum_{j=1}^{p} \sigma_{ij} \big[& ESN_1(a_j; \mu_j, \sigma_j^2, c_j\sigma_j\tilde{\varphi}_j, c_j\tau) \mathcal{L}_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\boldsymbol{\Sigma}}_j, \tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)}, \tilde{\tau}_j^{\mathbf{a}}) \\
& - ESN_1(b_j; \mu_j, \sigma_j^2, c_j\sigma_j\tilde{\varphi}_j, c_j\tau) \mathcal{L}_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\boldsymbol{\Sigma}}_j, \tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)}, \tilde{\tau}_j^{\mathbf{b}}) \big] \Bigg] + \mu_i,
\end{aligned}$$

for $i = 1, \ldots, p$, where $\mathcal{L} \equiv \mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ and $L \equiv L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})$.

It follows that

$$\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu} + \frac{1}{\mathcal{L}} [L\boldsymbol{\delta} + \boldsymbol{\Sigma}(\mathbf{q}_a - \mathbf{q}_b)], \tag{3.18}$$

where the j-th element of $\mathbf{q}_a$ and $\mathbf{q}_b$ are

$$q_{a,j} = ESN_1(a_j; \mu_j, \sigma_j^2, c_j\sigma_j\tilde{\varphi}_j, c_j\tau) \mathcal{L}_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}, \tilde{\boldsymbol{\Sigma}}_j, \tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)}, \tilde{\tau}_j^{\mathbf{a}}), \tag{3.19}$$

$$q_{b,j} = ESN_1(b_j; \mu_j, \sigma_j^2, c_j\sigma_j\tilde{\varphi}_j, c_j\tau) \mathcal{L}_{p-1}(\mathbf{a}_{(j)}, \mathbf{b}_{(j)}; \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}, \tilde{\boldsymbol{\Sigma}}_j, \tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)}, \tilde{\tau}_j^{\mathbf{b}}). \tag{3.20}$$

Denoting $\mathbf{D} = [\mathbf{d}_{\mathbf{e}_1}, \ldots, \mathbf{d}_{\mathbf{e}_p}]$, we can write

$$\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] = \boldsymbol{\mu}\mathbb{E}[\mathbf{Y}]^\top + \frac{1}{\mathcal{L}}[L\boldsymbol{\delta}\mathbb{E}[\mathbf{W}]^\top + \boldsymbol{\Sigma}\mathbf{D}], \tag{3.21}$$

$$\text{cov}[\mathbf{Y}] = \big[\boldsymbol{\mu} - \mathbb{E}[\mathbf{Y}]\big]\mathbb{E}[\mathbf{Y}]^\top + \frac{1}{\mathcal{L}}[L\boldsymbol{\delta}\mathbb{E}[\mathbf{W}]^\top + \boldsymbol{\Sigma}\mathbf{D}], \tag{3.22}$$

where $\mathbf{W} \sim \text{TN}_p(\boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}, (\mathbf{a}, \mathbf{b}))$, that is a p-variate truncated normal distribution on $(\mathbf{a}, \mathbf{b})$.

Besides, from Corollary 3.1, we have that the first two moments of $\mathbf{Y}$ can be also computed as

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{X}]_{(p+1)}, \tag{3.23}$$

$$\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]_{(p+1,p+1)}, \tag{3.24}$$

with $\mathbf{X} \sim \text{TN}_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Omega}; (\mathbf{a}^*, \mathbf{b}^*))$. Note that $\text{cov}[\mathbf{Y}] = \mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] - \mathbb{E}[\mathbf{Y}]\mathbb{E}[\mathbf{Y}^\top]$. Equations (3.23) and (3.24) are more convenient for computing $\mathbb{E}[\mathbf{Y}]$ and $\text{cov}[\mathbf{Y}]$ since all boils down to compute the mean and the variance-covariance matrix for a $p+1$-variate TMVN distribution which integrals are less complex than the ESN ones.

3.3.4 Mean and covariance matrix of TMVN distributions

Some approaches exists to compute the moments of a TMVN distribution. For instance, for doubly truncation, Manjunath & Wilhelm (2009) (method available through the `tmvtnorm R` package) computed the mean and variance of $\mathbf{X}$ directly deriving the MGF of the TMVN distribution. On the other hand, Kan & Robotti (2017) (method available through the `MomTrunc R` package) is able to compute arbitrary higher order TMVN moments using a recursive approach as a result of differentiating the multivariate normal density. For right truncation, Vaida & Liu (2009) (see Appendix B) proposed a method to compute the mean and variance of $\mathbf{X}$ also by differentiating the MGF, but where the off-diagonal elements of the Hessian matrix are recycled in order to compute its diagonal, leading to a faster algorithm. Next, we present an extension of Vaida & Liu (2009) algorithm to handle doubly truncation.

3.3.5 Deriving the first two moments of a double truncated MVN distribution through its MGF

Theorem 3.3. *Let* $\mathbf{X} \sim TN_p(\mathbf{0}, \mathbf{R}; (\mathbf{a}, \mathbf{b}))$*, with* $\mathbf{R}$ *being a correlation matrix of order* $p \times p$*. Then, the first two moments of* $\mathbf{X}$ *are given by*

$$\mathbb{E}[\mathbf{X}] = \left.\frac{\partial m(\mathbf{t})}{\partial \mathbf{t}}\right|^\top_{\mathbf{t}=\mathbf{0}} = -\frac{1}{L}\mathbf{R}\mathbf{q},$$

$$\mathbb{E}[\mathbf{X}\mathbf{X}^\top] = \left.\frac{\partial^2 m(\mathbf{t})}{\partial \mathbf{t}\partial \mathbf{t}^\top}\right|_{\mathbf{t}=\mathbf{0}} = \mathbf{R} + \frac{1}{L}\mathbf{R}\mathbf{H}\mathbf{R},$$

and consequently,

$$\operatorname{cov}[\mathbf{X}] = \mathbf{R} + \frac{1}{L^2}\mathbf{R}\big(L\mathbf{H} - \mathbf{q}\mathbf{q}^\top\big)\mathbf{R},$$

where $L \equiv L_p(\mathbf{a}, \mathbf{b}; \mathbf{0}, \mathbf{R})$, $\mathbf{q} = \mathbf{q}_a - \mathbf{q}_b$, *with the* i*-th element of* $\mathbf{q}_a$ *and* $\mathbf{q}_b$ *as*

$$q_{a,i} = \phi_1(a_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; a_i\mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) \qquad \text{and} \qquad q_{b,i} = \phi_1(b_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; b_i\mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i),$$

$\mathbf{H}$ *being a symmetric matrix of dimension* p*, with off-diagonal elements* h_{ij} *given by*

$$\begin{aligned} h_{ij} &= h_{ij}^{aa} - h_{ij}^{ba} - h_{ij}^{ab} + h_{ij}^{bb}, \\ &= \phi_2(a_i, a_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{aa}, \tilde{\mathbf{R}}_{ij}) - \phi_2(b_i, a_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{ba}, \tilde{\mathbf{R}}_{ij}) \\ &\quad - \phi_2(a_i, b_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{ab}, \tilde{\mathbf{R}}_{ij}) + \phi_2(b_i, b_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{bb}, \tilde{\mathbf{R}}_{ij}), \end{aligned} \tag{3.25}$$

and diagonal elements

$$h_{ii} = a_i q_{a_i} - b_i q_{b_i} - \mathbf{R}_{i,(i)}\mathbf{H}_{(i),i}, \tag{3.26}$$

with $\tilde{\mathbf{R}}_i = \mathbf{R}_{(i),(i)} - \mathbf{R}_{(i),i}\mathbf{R}_{i,(i)}$, $\boldsymbol{\mu}_{ij}^{\alpha\beta} = \mathbf{R}_{(ij),[i,j]}(\alpha_i, \beta_j)^\top$ *and* $\tilde{\mathbf{R}}_{ij} = \mathbf{R}_{(i,j),(i,j)} - \mathbf{R}_{(i,j),[i,j]}\mathbf{R}_{[i,j],(i,j)}$.

Proof. See Appendix B.

The main difference of our proposal in Theorem 3.3 and other approaches deriving the MGF relies on (3.26), where the diagonal elements are recycled using the off-diagonal elements h_{ij}, $1 \leqslant i \neq j \leqslant p$. Furthermore, for $\mathbf{W} \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; (\tilde{\mathbf{a}}, \tilde{\mathbf{b}}))$, we have that

$$\mathbb{E}[\mathbf{W}] = \boldsymbol{\mu} - \mathbf{S}\,\mathbb{E}[\mathbf{X}], \tag{3.27}$$

$$\operatorname{cov}[\mathbf{W}] = \mathbf{S}\operatorname{cov}[\mathbf{X}]\,\mathbf{S}, \tag{3.28}$$

where $\boldsymbol{\Sigma}$ being a positive-definite matrix, $\mathbf{S} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$, and truncation limits $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ such that $\mathbf{a} = \mathbf{S}^{-1}(\tilde{\mathbf{a}} - \boldsymbol{\mu})$ and $\mathbf{b} = \mathbf{S}^{-1}(\tilde{\mathbf{b}} - \boldsymbol{\mu})$.

3.4 Dealing with limiting and extreme cases

Let consider $\mathbf{Y} \sim \mathrm{ESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$. As $\tau \to \infty$, we have that $\xi = \Phi(\tilde{\tau}) \to 1$. Besides, as $\tau \to -\infty$, we have that $\xi \to 0$ and consequently $\mathcal{F}_{\boldsymbol{\kappa}}^p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \xi^{-1} F_{\boldsymbol{\kappa}^*}^{p+1}(\mathbf{a}^*, \mathbf{b}^*; \boldsymbol{\mu}^*, \boldsymbol{\Omega}) \to \infty$. Thus, for negative $\tilde{\tau}$ values small enough, we are not able to compute $\mathbb{E}[\mathbf{Y}^\kappa]$ due to computation precision. For instance, in `R` software, $\Phi(\tilde{\tau}) = 0$ for $\tilde{\tau} < -37$. The next proposition helps us to circumvent this problem.

Proposition 3.5. *(Limiting distribution for the ESN) As* $\tau \to -\infty$,

$$ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\tau}) \longrightarrow \phi_p(\mathbf{y}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}). \tag{3.29}$$

Proof. Let $X_2 \sim N(0,1)$. As $\tilde{\tau} \to -\infty$, we have that $P(X_2 \leqslant \tilde{\tau}) \to 0$, $\mathbb{E}[X_2|X_2 \leqslant \tilde{\tau}] \to \tilde{\tau}$ and $\text{var}[X_2|X_2 \leqslant \tilde{\tau}] \to 0$ (i.e., X_2 is (i.e., $\mathbf{X}_2$ is degenerated on $\tilde{\tau}$). In light of Proposition 3.3, $\mathbf{Y} \overset{d}{=} (\mathbf{X}_1|X_2 = \tilde{\tau})$, and by the conditional distribution of a multivariate normal, it is straightforward to show that $\mathbb{E}[\mathbf{X}_1|X_2 = \tilde{\tau}] = \boldsymbol{\mu} - \boldsymbol{\mu}_b$ and $\text{cov}[\mathbf{X}_1|X_2 = \tilde{\tau}] = \boldsymbol{\Gamma}$, which concludes the proof. □

3.4.1 Approximating the mean and variance-covariance of a TMVN distribution for extreme cases

While using the normal relation (3.23) and (3.24), we may also face numerical problems for extreme settings of $\boldsymbol{\lambda}$ and τ due to the scale matrix $\boldsymbol{\Omega}$ does depend on them. Most common problem is that the normalizing constant $L_p(\mathbf{a}^*, \mathbf{b}^*; \boldsymbol{\mu}^*, \boldsymbol{\Omega})$ is approximately zero, because the probability density has been shifted far from the integration region. It is worth mentioning that, for these cases, it is not even possible to estimate the moments generating Monte Carlo (MC) samples due to the high rejection ratio when subsetting to a small integration region.

For instance, consider a bivariate truncated normal vector $\mathbf{X} = (X_1, X_2)^\top$, with X_1 and X_2 having zero mean and unit variance, $\text{cov}(X_1, X_2) = -0.5$ and truncation limits $\mathbf{a} = (-20, -10)^\top$ and $\mathbf{b} = (-9, 10)^\top$. Then, we have that the limits of X_1 are far from the density mass since $P(-20 \leqslant X_1 \leqslant -9) \approx 0$. For this case, both the `mtmvnorm` function from the `tmvtnorm R` package and the `Matlab` codes provided in Kan & Robotti (2017) return wrong mean values outside the truncation interval $(\mathbf{a}, \mathbf{b})$ and negative variances. Values are quite high too, with mean values greater than 1×10^{10} and all the elements of the variance-covariance matrix greater than 1×10^{20}. When changing the first upper limit from -9 to -13, that is $\mathbf{b} = (-13, 10)^\top$, both routines return `Inf` and `NaN` values for all the elements.

Although the above scenarios seem unusual, extreme situations that require correction are more common than expected. For instance, this occurs when the elements of the scale matrix $\boldsymbol{\Sigma}$ are small, even if the location parameter $\boldsymbol{\mu}$ is close to the integration region. Furthermore, the development of this part was motivated as we identified this problem when we fit censored regression models, with high asymmetry and presence of outliers. Hence, we present correction method in order to approximate the mean and the variance-covariance of a multivariate TMVN distribution even when the numerical precision of the software is a limitation.

Dealing with out-of-bounds limits

Consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ such that $dim(\mathbf{X}_1) = p_1$, $dim(\mathbf{X}_2) = p_2$, where $p_1 + p_2 = p$. It is well known that

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix}\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2] \\ \mathbf{X}_2\end{bmatrix}$$

and

$$\text{cov}[\mathbf{X}] = \begin{bmatrix}\mathbb{E}[\text{cov}[\mathbf{X}_1|\mathbf{X}_2]] + \text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] & \text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2], \mathbf{X}_2] \\ \text{cov}[\mathbf{X}_2, \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] & \text{cov}[\mathbf{X}_2]\end{bmatrix}.$$

Now, consider $\mathbf{X} \sim \text{TN}_p\big(\boldsymbol{\mu}, \boldsymbol{\Sigma}, (\mathbf{a}, \mathbf{b})\big)$ to be partitioned as above. Also consider the corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{a} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top)^\top$ and $\mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$. We say that the limits $[\mathbf{a}_2, \mathbf{b}_2]$ of $\mathbf{X}_2$ are out-of-bounds if $P(\mathbf{a}_2 \leqslant \mathbf{X}_2 \leqslant \mathbf{b}_2) \approx 0$. Let us consider the case where we are not able to compute any moment of $\mathbf{X}$, because there exists a partition $\mathbf{X}_2$ of $\mathbf{X}$ of dimension p_2 that is out-of-bounds. Note this happens because $L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \leqslant P(\mathbf{a}_2 \leqslant \mathbf{X}_2 \leqslant \mathbf{b}_2) \approx 0$. Also, we consider the partition $\mathbf{X}_1$ such that $P(\mathbf{a}_1 \leqslant \mathbf{X}_1 \leqslant \mathbf{b}_1) > 0$. Since the limits of $\mathbf{X}_2$ are out-of-bounds (and $\mathbf{a}_2 < \mathbf{b}_2$), we have two possible cases: $\mathbf{b}_2 \to -\infty$ or $\mathbf{a}_2 \to \infty$. For convenience, let $\boldsymbol{\xi}_2 = \mathbb{E}[\mathbf{X}_2]$ and $\boldsymbol{\Psi}_{22} = \text{cov}[\mathbf{X}_2]$. For the first case, as $\mathbf{b}_2 \to -\infty$, we have that $\boldsymbol{\xi}_2 \to \mathbf{b}_2$ and $\boldsymbol{\Psi}_{22} \to \mathbf{0}_{p_2 \times p_2}$. Analogously, we have that $\boldsymbol{\xi}_2 \to \mathbf{a}_2$ and $\boldsymbol{\Psi}_{22} \to \mathbf{0}_{p_2 \times p_2}$ as $\mathbf{a}_2 \to \infty$.

Then $\mathbf{X}_1 \sim \text{TN}_{p_1}\big(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; [\mathbf{a}_1, \mathbf{b}_1]\big)$, $\mathbf{X}_2 \sim N_{p_2}\big(\boldsymbol{\xi}_2, \mathbf{0}\big)$ (i.e., $\mathbf{X}_2$ is degenerated on $\boldsymbol{\xi}_2$) and $\mathbf{X}_1|\mathbf{X}_2 \sim \text{TN}_{p_1}\big(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\xi}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}; [\mathbf{a}_1, \mathbf{b}_1]\big)$. Given that $\text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \mathbf{0}_{p_1 \times p_2}$ and $\text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2], \mathbf{X}_2] = \mathbf{0}_{p_2 \times p_2}$, it follows that

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix}\boldsymbol{\xi}_{1.2} \\ \boldsymbol{\xi}_2\end{bmatrix} \qquad \text{and} \qquad \text{cov}[\mathbf{X}] = \begin{bmatrix}\boldsymbol{\Psi}_{11.2} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{0}_{p_2 \times p_2}\end{bmatrix}, \tag{3.30}$$

with $\boldsymbol{\xi}_{1.2} = \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]$ and $\boldsymbol{\Psi}_{11.2} = \text{cov}[\mathbf{X}_1|\mathbf{X}_2]$ being the mean and variance-covariance matrix of a TMVN distribution, which can be computed using (3.27) and (3.28).

In the event that there are double infinite limits, we can partition the vector as well, in order to avoid unnecessary calculation of these integrals.

Dealing with double infinite limits

Let p_1 be the number of pairs in $(\mathbf{a}, \mathbf{b})$ that are both infinite. We consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$, such that the upper and lower truncation limits associated with $\mathbf{X}_1$ are both infinite, but at least one of the truncation limits associated with $\mathbf{X}_2$ is finite. Since $\mathbf{a}_1 = -\infty$ and $\mathbf{b}_1 = \infty$, it follows that $\mathbf{X}_2 \sim \text{TN}_{p_2}\big(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, [\mathbf{a}_2, \mathbf{b}_2]\big)$ and

$\mathbf{X}_1|\mathbf{X}_2 \sim N_{p_1}\big(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big)$. This leads to

$$\mathbb{E}[\mathbf{X}] = \mathbb{E}\begin{bmatrix}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)\\ \mathbf{X}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{\xi}_2 - \boldsymbol{\mu}_2)\\ \boldsymbol{\xi}_2\end{bmatrix}, \tag{3.31}$$

and

$$\operatorname{cov}[\mathbf{X}] = \begin{bmatrix}\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\big(\mathbf{I}_{p_2} - \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\big)\boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Psi}_{22}\\ \boldsymbol{\Psi}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{\Psi}_{22}\end{bmatrix}, \tag{3.32}$$

with $\boldsymbol{\xi}_2$ and $\boldsymbol{\Psi}_{22}$ being the mean vector and variance-covariance matrix of a TMVN distribution, which can be computed using (3.27) and (3.28) as well.

As can be seen, we can use equations (3.31) and (3.32) to deal with double infinite limits, where the truncated moments are computed only over a p_2-variate partition, avoiding some unnecessary integrals and saving some computational effort. On the other hand, expression (3.30) let us to approximate the mean and the variance-covariance matrix for cases where the computational precision is a limitation.

3.5 Comparison of our proposal with existent methods

Now, we compare different approaches to compute the mean vector and variance-covariance matrix of a p-variate TESN distribution. We consider our proposal derived from Theorem 3.1, as well as the normal relation given in Theorem 3.2 using different (some existent) methods for computing the mean and variance-covariance of a TMVN distribution. The methods that we compare are the following:

Proposal 1: Theorem 3.1, i.e., equations (3.18), and (3.24),
Proposal 2: Normal relation (NR) in Theorem 3.2 using Theorem 3.3,
K&R: NR in Theorem 3.2 using the `Matlab` routine from Kan & Robotti (2017),
tmvtnorm: NR in Theorem 3.2 using the `tmvtnorm R` function from Manjunath & Wilhelm (2009).

Left panel of Figure 7 shows the number of integrals required to achieve this for different dimensions p. We compare the proposal 1 for a p-variate TESN distribution and the equivalent $p+1$-variate normal approaches K&R and proposal 2.

It is clear that the importance of the new proposed method since it reduces the number of integral involved almost to half, this compared to the TESN direct results from proposal 1, when we consider the double truncation. In particular, for left/right truncation, we have that the equivalent $p+1$-variate normal approach along with Vaida & Liu (2009) (now, a special case of proposal 2) requires up to 4 times less integrals than when we use

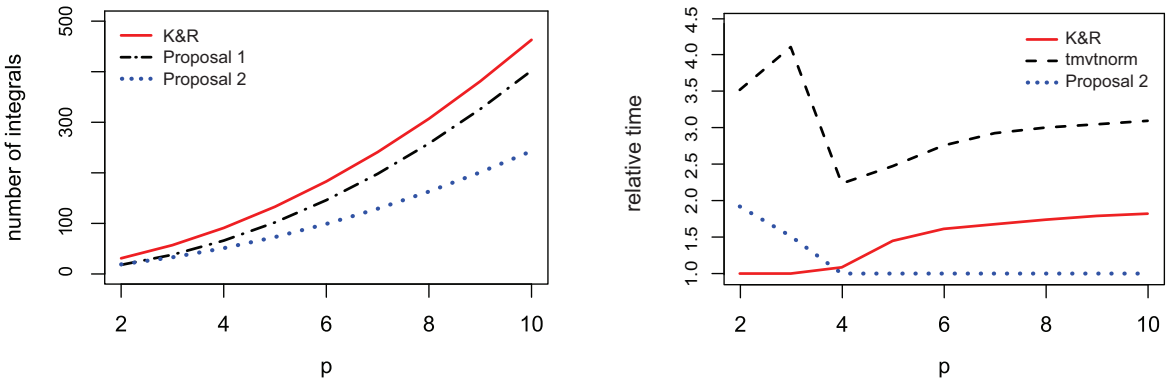


Figure 7 – Number of integrals required and relative processing time for computing the mean vector and variance-covariance matrix for a p-variate truncated ESN and MVN distribution respectively, for 3 different approaches under double truncation.

the method K&R. As seen before, the normal relation proposal 2 outperforms the proposal 1, that is, the equivalent normal approach always resulted faster even it considers one more dimension, that is a $p+1$-variate normal vector, due to its integrals are less complex than for the ESN case.

Processing time when using the equivalent normal approach are depicted in the right panel of Figure 7. Here, we compare the relative processing time of the mean and variance-covariance of a TMVN distribution under the methods `tmvtnorm`, K&R and our proposal 2, for different dimensions p. Note that a method with relative processing time equal to 1 means this is the fastest one. In general, our proposal is the fastest one, as expected. Method K&R resulted better only for $p \leqslant 3$, which confirms the necessity for a faster algorithm, in order to deal with high dimensional problems. Method `tmvtnorm` resulted to be the slowest one by far. Our `MonTrunc R` package computes the mean and the variance of a TMVN distribution in an optimal way, such that it uses the method proposed by K&R for $p < 4$ and otherwise proposal 2.

3.6 On moments of folded multivariate ESN distributions

First, we established some general results for the pdf, cdf and moments of a folded multivariate distributions (FMD). These extend the results found in Chakraborty & Chatterjee (2013) for a folded normal (FN) distribution to any multivariate distribution, as well as the multivariate location-scale family. The proofs are given in Appendix B.

Theorem 3.4 (***pdf* and *cdf* of a MFD**). *Let* $\mathbf{X} \in \mathbb{R}^p$ *be a* p*-variate random vector with pdf* $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ *and cdf* $F_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ *with* $\boldsymbol{\theta}$ *being a set of parameters characterizing such distribution. If* $\mathbf{Y} = |\mathbf{X}|$*, then the joint pdf and cdf of* $\mathbf{Y}$ *that follows a folded distribution*

of $\mathbf{X}$ *are given, respectively, by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} f_{\mathbf{X}}(\boldsymbol{\Lambda}_s\mathbf{y};\boldsymbol{\theta}) \qquad \textit{and} \qquad F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} \pi_s F_{\mathbf{X}}(\boldsymbol{\Lambda}_s\mathbf{y};\boldsymbol{\theta}), \quad \textit{for } \mathbf{y}\geqslant\mathbf{0},$$

where $S(p) = \{-1,1\}^p$ *is a cartesian product with* 2^p *elements, each of the form* $\mathbf{s} = (s_1,\ldots,s_p)$, $\boldsymbol{\Lambda}_s = Diag(\mathbf{s})$ *and* $\pi_s = \prod_{i=1}^{p} s_i$.

Corollary 3.2. *If* $\mathbf{X}\sim f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\xi},\boldsymbol{\Psi})$ *belongs to the location-scale family of distributions with location and scale parameters* $\boldsymbol{\xi}$ *and* $\boldsymbol{\Psi}$ *respectively, then* $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X}\sim f_{\mathbf{X}}(\mathbf{z};\boldsymbol{\Lambda}_s\boldsymbol{\xi},\boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s)$ *and consequently the joint pdf and cdf of* $\mathbf{Y} = |\mathbf{X}|$ *are given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} f_{\mathbf{X}}(\mathbf{y};\boldsymbol{\Lambda}_s\boldsymbol{\xi},\boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s) \quad \textit{and} \quad F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} \pi_s F_{\mathbf{X}}(\boldsymbol{\Lambda}_s\mathbf{y};\boldsymbol{\xi},\boldsymbol{\Psi}), \quad \textit{for } \mathbf{y}\geqslant\mathbf{0}.$$

Hence, the $\boldsymbol{\kappa}$*-th moment of* $\mathbf{Y}$ *follows as*

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[(\mathbf{Z}_s^{\boldsymbol{\kappa}})^+],$$

where $\mathbf{X}^+$ *denotes the positive component of the random vector* $\mathbf{X}$.

Let $\mathbf{X}\sim ESN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$, we now turn our attention to discuss the computation of any arbitrary order moment of $|\mathbf{X}|$, a multivariate folded ESN (FESN) distribution. Let define the $\mathcal{I}^p_{\boldsymbol{\kappa}} \equiv \mathcal{I}^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$ function as

$$\mathcal{I}^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} ESN_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)\mathrm{d}\mathbf{y}. \tag{3.33}$$

Note that $\mathcal{I}^p_{\boldsymbol{\kappa}}$ is a special case of $\mathcal{F}^p_{\boldsymbol{\kappa}}$ that occurs when $a_i = 0$ and $b_i = +\infty$, $i = 1,\ldots,p$. In this scenario we have

$$\mathcal{I}^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) = \mathcal{F}^p_{\boldsymbol{\kappa}}(\mathbf{0},+\infty;\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau).$$

When $\boldsymbol{\lambda} = \mathbf{0}$ and $\tau = 0$, that is, the normal case we write $\mathcal{I}^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\mathbf{0},0) = I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

Proposition 3.6. *If* $\mathbf{X}\sim ESN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$ *then* $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X}\sim ESN_p(\boldsymbol{\mu}_s,\boldsymbol{\Sigma}_s,\boldsymbol{\lambda}_s,\tau)$ *and consequently the joint* pdf, cdf *and the* $\boldsymbol{\kappa}$*th raw moment of* $\mathbf{Y} = |\mathbf{X}|$ *are, respectively, given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s}\in S(p)} ESN_p(\mathbf{y}_p;\boldsymbol{\mu}_s,\boldsymbol{\Sigma}_s,\boldsymbol{\lambda}_s,\tau),$$

$$F_{\mathbf{Y}}(\mathbf{y}) = \mathcal{L}_p(-\mathbf{y},\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau),$$

and

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \sum_{\mathbf{s}\in S(p)} \mathcal{I}^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}_s,\boldsymbol{\Sigma}_s,\boldsymbol{\lambda}_s,\tau),$$

where $\mathbf{y}_s = \boldsymbol{\Lambda}_s\mathbf{y}$, $\boldsymbol{\mu}_s = \boldsymbol{\Lambda}_s\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_s\boldsymbol{\Sigma}\boldsymbol{\Lambda}_s$ *and* $\boldsymbol{\lambda}_s = \boldsymbol{\Lambda}_s\boldsymbol{\lambda}$.

Proof. Note that is suffices to show that *if* $\mathbf{X} \sim \text{ESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, *then* $\mathbf{Z}_s = \boldsymbol{\Lambda}_s \mathbf{X} \sim \text{ESN}_p(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s, \tau)$ since the rest of the corollary is straightforward. We have that

$$
\begin{aligned}
&= ESN_p(\mathbf{x}; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s, \tau) \\
&= \xi^{-1} \phi_p(\mathbf{x}; \boldsymbol{\Lambda}_s \boldsymbol{\mu}, \boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s) \times \Phi_1\big(\tau + (\boldsymbol{\Lambda}_s \boldsymbol{\lambda})^\top (\boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s)^{-1/2} (\mathbf{x} - \boldsymbol{\Lambda}_s \boldsymbol{\mu})\big) \\
&= \xi^{-1} |\boldsymbol{\Lambda}_s \boldsymbol{\Lambda}_s|^{1/2} \phi_p(\boldsymbol{\Lambda}_s^{-1} \mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \Phi_1\big(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_s (\boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s)^{-1/2} \boldsymbol{\Lambda}_s (\boldsymbol{\Lambda}_s^{-1} \mathbf{x} - \boldsymbol{\mu})\big) \\
&= \xi^{-1} \phi_p(\boldsymbol{\Lambda}_s \mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \Phi_1\big(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Lambda}_s (\boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s)^{-1/2} \boldsymbol{\Lambda}_s (\boldsymbol{\Lambda}_s \mathbf{x} - \boldsymbol{\mu})\big) \qquad (3.34) \\
&\overset{?}{=} \xi^{-1} \phi_p(\boldsymbol{\Lambda}_s \mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \Phi_1\big(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2} (\boldsymbol{\Lambda}_s \mathbf{x} - \boldsymbol{\mu})\big) \qquad (3.35) \\
&= ESN_p(\boldsymbol{\Lambda}_s \mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau),
\end{aligned}
$$

where $\xi^{-1} = \Phi_1\big(\tau / \sqrt{1 + \boldsymbol{\lambda}_s^\top \boldsymbol{\lambda}_s}\big)$ due to $\boldsymbol{\lambda}_s^\top \boldsymbol{\lambda}_s = \boldsymbol{\lambda}^\top \boldsymbol{\lambda}$.

In order to equalize (3.34) and (3.35), we see that it suffices to show that $\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Lambda}_s (\boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s)^{-1/2} \boldsymbol{\Lambda}_s$. This is equivalent to show that $\mathbf{A} = \mathbf{B}$ for $\mathbf{A} = (\boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s)^{1/2}$ and $\mathbf{B} = \boldsymbol{\Lambda}_s \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Lambda}_s$. We have that both matrices $\mathbf{A}$ and $\mathbf{B}$ are positive-definite matrices since $(\boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s)^{1/2}$ and $\boldsymbol{\Sigma}^{1/2}$ are too, as a consequence that they are obtained using Singular Value Decomposition (SVD). Finally, given that $\mathbf{A}^2 = \mathbf{B}^2 = \boldsymbol{\Lambda}_s \boldsymbol{\Sigma} \boldsymbol{\Lambda}_s$ and any positive-definite matrix has an unique positive-definite square root, we conclude that $\mathbf{A} = \mathbf{B}$ by uniqueness, which concludes the proof. □

Observation 3.1. *As a consequence of Proposition 3.6, we also have the new vectors* $\boldsymbol{\delta}_s = \boldsymbol{\Lambda}_s \boldsymbol{\delta}$, $\boldsymbol{\mu}_{b_s} = \boldsymbol{\Lambda}_s \boldsymbol{\mu}_b$, $\boldsymbol{\varphi}_s = \boldsymbol{\Lambda}_s \boldsymbol{\varphi}$, $\tilde{\boldsymbol{\varphi}}_s = \boldsymbol{\Lambda}_s \tilde{\boldsymbol{\varphi}}$, $\tilde{\boldsymbol{\mu}}_{js}^{\mathbf{a}} = \boldsymbol{\Lambda}_{s(j)} \tilde{\boldsymbol{\mu}}_j^{\mathbf{a}}$ *and* $\tilde{\boldsymbol{\mu}}_{js}^{\mathbf{b}} = \boldsymbol{\Lambda}_{s(j)} \tilde{\boldsymbol{\mu}}_j^{\mathbf{b}}$, *and matrix* $\boldsymbol{\Gamma}_s = \boldsymbol{\Lambda}_s \boldsymbol{\Gamma} \boldsymbol{\Lambda}_s$, *while the constants* ξ, η, c_j ,$\tilde{\boldsymbol{\Sigma}}_j$, *and* $\tilde{\tau}_j$ *remain invariant with respect to* $\mathbf{s}$.

A bivariate case

For instance, let us consider a bivariate case. We denote that $\mathbf{X}$ follows a bivariate ESN distribution as $\mathbf{X} \sim \text{ESN}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_{12}, \sigma_2^2, \lambda_1, \lambda_2, \tau)$. For $\mathbf{Y} = |\mathbf{X}|$, it follows that

$$
\begin{aligned}
f_{\mathbf{Y}}(y_1, y_2) &= f_{\mathbf{X}}(y_1, y_2) + f_{\mathbf{X}}(-y_1, y_2) + f_{\mathbf{X}}(y_1, -y_2) + f_{\mathbf{X}}(-y_1, -y_2) \qquad (3.36) \\
&= f_{\mathbf{X}_1}(y_1, y_2) + f_{\mathbf{X}_2}(y_1, y_2) + f_{\mathbf{X}_3}(y_1, y_2) + f_{\mathbf{X}_4}(y_1, y_2) \qquad (3.37) \\
&= \sum_{i=1}^{4} f_{\mathbf{X}_i}(y_1, y_2),
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{X}_1 &\sim \text{ESN}_2(+\mu_1, +\mu_2, \sigma_1^2, +\sigma_{12}, \sigma_2^2, +\lambda_1, +\lambda_2, \tau), \\
\mathbf{X}_2 &\sim \text{ESN}_2(-\mu_1, +\mu_2, \sigma_1^2, -\sigma_{12}, \sigma_2^2, -\lambda_1, +\lambda_2, \tau), \\
\mathbf{X}_3 &\sim \text{ESN}_2(+\mu_1, -\mu_2, \sigma_1^2, -\sigma_{12}, \sigma_2^2, +\lambda_1, -\lambda_2, \tau),
\end{aligned}
$$

$$\mathbf{X}_4 \sim \mathrm{ESN}_2(-\mu_1, -\mu_2, \sigma_1^2, +\sigma_{12}, \sigma_2^2, -\lambda_1, -\lambda_2, \tau).$$

Equation (3.36) stands from Theorem 3.4. Its four summands are respectively equivalent to the four terms in (3.37). This can be seen through a briefly comparison of the density regions at the points on Figure 8 above and the four points in the respective densities in Figure 16 in the Appendix section B. As noted, to pass the signs to the parameters, let us to fix the point (y_1, y_2) and then integrating in both sides, we can obtain any arbitrary moment for $|\mathbf{X}|$ as a sum of other 2^p moments.

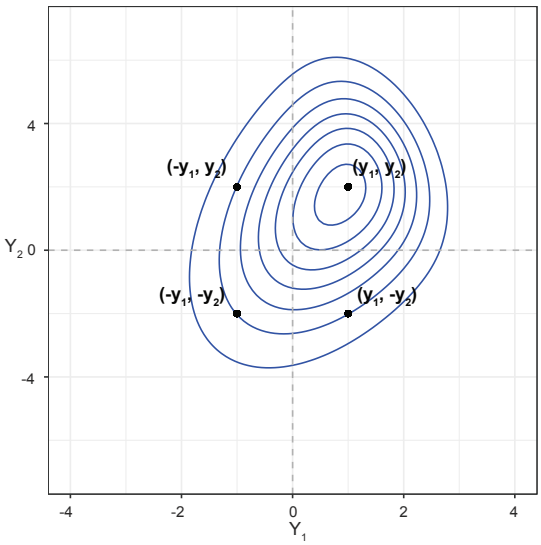


Figure 8 – Density of $\mathbf{X}$.

From Proposition 3.6, we can compute any arbitrary moment of a FESN distribution as a sum of $\mathcal{I}^p_{\boldsymbol{\kappa}}$ integrals. In light of Theorem 3.1, the recurrence relation for $\mathcal{I}^p_{\boldsymbol{\kappa}}$ can be written as

$$\mathcal{I}^p_{\boldsymbol{\kappa}+\mathbf{e}_i}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \mu_i \mathcal{I}^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) + \delta_i I^p_{\boldsymbol{\kappa}}(\boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}) + \sum_{j=1}^{p} \sigma_{ij} d_{\kappa,j}, \ i = 1, \ldots, p, \quad (3.38)$$

where

$$d_{\kappa,j} = \begin{cases} k_j \mathcal{I}^p_{\boldsymbol{\kappa}-\mathbf{e}_i}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) & ; \quad \text{for } k_j > 0 \\ ESN_1(0; \mu_j, \sigma_j^2, c_j\sigma_j\tilde{\varphi}_j, c_j\tau)\mathcal{I}^{p-1}_{\boldsymbol{\kappa}_{(j)}}(\tilde{\boldsymbol{\mu}}_j, \tilde{\boldsymbol{\Sigma}}_j, \tilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{\varphi}_{(j)}, \tilde{\tau}_j) & ; \quad \text{for } k_j = 0 \end{cases}$$

with $\tilde{\boldsymbol{\mu}}_j = \boldsymbol{\mu}_{(j)} - \dfrac{\mu_j}{\sigma_j^2}\boldsymbol{\Sigma}_{(j)j}$ and $\tilde{\tau}_j = \tau - \tilde{\varphi}_j\mu_j$.

It is also possible to use the normal relation in Theorem 3.2 to compute $\mathbb{E}[|\mathbf{X}|^{\boldsymbol{\kappa}}]$ in a simpler manner as in next proposition.

Proposition 3.7. *Let* $\mathbf{Y} = |\mathbf{X}|$*, with* $\mathbf{X} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$*. In light of Theorem 3.4, It follows that*

$$\mathbb{E}[\mathbf{Y}^{\boldsymbol{\kappa}}] = \xi^{-1} \sum_{\mathbf{s} \in S(p)} I^{p+1}_{\boldsymbol{\kappa}^*}(\boldsymbol{\mu}^*_s, \boldsymbol{\Omega}^-_s), \quad (3.39)$$

where $I_{\boldsymbol{\kappa}}^{p}(\boldsymbol{\mu},\boldsymbol{\Sigma}) \equiv F_{\boldsymbol{\kappa}}^{p}(\mathbf{0},\infty;\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{\mu}_s^* = (\boldsymbol{\mu}_s^\top, \tilde{\tau})^\top$ *and* $\boldsymbol{\Omega}_s = \begin{pmatrix} \boldsymbol{\Sigma}_s & -\boldsymbol{\Delta}_s \\ -\boldsymbol{\Delta}_s^\top & 1 \end{pmatrix}$, *with* $\boldsymbol{\mu}_s = \boldsymbol{\Lambda}_s\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_s = \boldsymbol{\Lambda}_s\boldsymbol{\Sigma}\boldsymbol{\Lambda}_s$, $\boldsymbol{\Delta}_s = \boldsymbol{\Lambda}_s\boldsymbol{\Delta}$ *and* $\boldsymbol{\Omega}_s^-$ *standing for the block matrix* $\boldsymbol{\Omega}_s$ *with all its off-diagonal block elements signs changed.*

Proof is direct from Theorem 3.2 as $\mathcal{I}_{\boldsymbol{\kappa}}^{p}$ is a special case of $\mathcal{F}_{\boldsymbol{\kappa}}^{p}$. From proposition 3.2, we have that the mean and variance-covariance matrix can be calculated as a sum of 2^p terms as well, that is

$$\mathbb{E}[\mathbf{Y}] = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[\mathbf{Z}_s^+], \tag{3.40}$$

$$\operatorname{cov}[\mathbf{Y}] = \sum_{\mathbf{s}\in S(p)} \mathbb{E}\big[\mathbf{Z}_s^+\mathbf{Z}_s^{+\top}\big] - \mathbb{E}[\mathbf{Y}]\mathbb{E}[\mathbf{Y}]^\top, \tag{3.41}$$

where $\mathbf{Z}_s^+$ is the positive component of $\mathbf{Z}_s = \boldsymbol{\Lambda}_s\mathbf{X} \sim \mathrm{ESN}_p(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s, \tau)$. Note that there are 2^p times more integrals to be calculated as compared to the non-folded case, representing a huge computational effort for high dimensional problems.

In order to circumvent this, we can use the fact that $\mathbb{E}[\mathbf{Y}] = (\mathbb{E}[Y_1], \ldots, \mathbb{E}[Y_p])^\top$ and the elements of $\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top]$ are given by the second moments $\mathbb{E}[Y_i^2]$ and $\mathbb{E}[Y_iY_j]$, $1 \leqslant i \neq j \leqslant p$. Thus, it is possible to calculate explicit expressions for the mean vector and variance-covariance matrix of the FESN only based on the marginal univariate means and variances of Y_i, as well as the covariance terms $\operatorname{cov}(Y_i, Y_j)$.

Univariate case

Using the recurrence relation on $\mathcal{I}_k$ in (3.38), and following the notation in Subsection 3.3.1.1, we can find explicit expressions for $\mathbb{E}[|X|^k]$ for its first four raw moments, as well as for others univariate folded distributions that are special cases of the ESN distribution. For instance, setting $\tau = 0$, we obtain the moments for the univariate folded skew-normal $Y = |X|$, with $X \sim SN_1(\mu, \sigma^2, \lambda)$. Additionally, if we set the skewness parameter as $\lambda = 0$, we obtain the moments for the folded normal distribution where $X \sim N(\mu, \sigma^2)$. Furthermore, moments for the well-known half normal distribution can be obtained when we set $\mu = \lambda = \tau = 0$. Explicit expressions for all these special cases can be found the Appendix B.

Next, we propose explicit expressions for the mean and the variance-covariance of the multivariate FESN distribution.

3.6.1 Explicit expressions for mean and covariance matrix of multivariate folded ESN distribution

Let $\mathbf{X} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$. To obtain the mean and covariance matrix of $|\mathbf{X}|$ boils down to compute $\mathbb{E}[|X_i|]$, $\mathbb{E}[|X_i^2|]$ and $\mathbb{E}[|X_iX_j|]$. Consider X_i to be the i-th marginal partition

of $\mathbf{X}$ distributed as $X_i \sim \text{ESN}(\mu_i, \sigma_i^2, \lambda_i, \tau_i)$. In light of proposition 3.6 it follows that

$$\mathbb{E}[|X_i|^k] = \mathcal{I}_k^1(\mu_i, \sigma_i^2, \lambda_i, \tau_i) + \mathcal{I}_k^1(-\mu_i, \sigma_i^2, -\lambda_i, \tau_i).$$

Thus, using the recurrence relation on $\mathcal{I}_k$ in (3.38), and following the notation in Subsection 3.3.1.1, we can write explicit expressions for $\mathbb{E}[|X_i|]$ and $\mathbb{E}[|X_i|^2]$. High order moments for the univariate FESN and others related distributions are detailed in Appendix B.

It remains to obtain $\mathbb{E}[|X_i X_j|]$ for $i \neq j$, which can be obtained as

$$\begin{aligned}\mathbb{E}[|X_i X_j|] =& \mathcal{I}_{1,1}^2(\mu_i, \mu_j, \sigma_i^2, \sigma_{ij}, \sigma_j^2, \lambda_i, \lambda_j, \tau) + \mathcal{I}_{1,1}^2(\mu_i, -\mu_j, \sigma_i^2, -\sigma_{ij}, \sigma_j^2, \lambda_i, -\lambda_j, \tau) \\ &+ \mathcal{I}_{1,1}^2(-\mu_i, \mu_j, \sigma_i^2, -\sigma_{ij}, \sigma_j^2, -\lambda_i, \lambda_j, \tau) + \mathcal{I}_{1,1}^2(-\mu_i, -\mu_j, \sigma_i^2, \sigma_{ij}, \sigma_j^2, -\lambda_i, -\lambda_j, \tau),\end{aligned} \tag{3.42}$$

as pointed in proposition 3.6, with (X_i, X_j) denoting an arbitrary bivariate partition of $\mathbf{X}$. Without loss of generality, let's consider the partition $(X_1, X_2) \sim \text{ESN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ and $(W_1, W_2) \sim N_2(\mathbf{m}, \boldsymbol{\Gamma})$ with $\mathbf{m} = \boldsymbol{\mu} - \boldsymbol{\mu}_b$. For simplicity, we denote $\mathcal{I}_{1,1}^2 \equiv \mathcal{I}_{1,1}^2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, and the normalizing constants $\mathcal{L}_2 \equiv \mathcal{L}_2(\mathbf{0}, \infty; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ and $L_2 \equiv L_2(\mathbf{0}, \infty; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})$.

Using the recurrence relation on $\mathcal{I}_{\boldsymbol{\kappa}+\mathbf{e}_i}^2$ in (3.38), we can obtain $\mathcal{I}_{1,1}^2$ for $\boldsymbol{\kappa} = (1,0)^\top$ and $\mathbf{e}_2 = (0,1)^\top$ as

$$\begin{aligned}\mathcal{I}_{1,1}^2 =& (\mu_1\mu_2 + \sigma_{12})\mathcal{L}_2 + (\delta_1\mu_2 + \delta_2(\mu_1 - \mu_{b1}))\, L_2 + (\mu_2\sigma_1^2 + \sigma_{12})\tilde{\phi}^{(1)}(1 - \tilde{\Phi}^{(2.1)}) \\ &+ \delta_2\left[\gamma_1^2\phi(\mu_1; \mu_{b1}, \gamma_1^2)(1 - \Phi(0; m_{2.1}, \gamma_{2.1}^2)) + \gamma_{12}\phi(\mu_2; \mu_{b2}, \gamma_2^2)(1 - \Phi(0; m_{1.2}, \gamma_{1.2}^2))\right] \\ &+ \mu_2\sigma_{12}\tilde{\phi}^{(2)}(1 - \tilde{\Phi}^{(1.2)}) + \sigma_2^2\tilde{\phi}^{(2)}\mathcal{I}_1^1(\mu_{1.2}, \sigma_{11.2}^2, \sigma_{11.2}\varphi_1, \tau_{1.2}),\end{aligned} \tag{3.43}$$

where $m_{2.1} = m_2 - \gamma_{12}m_1/\gamma_1^2$, $m_{1.2} = m_1 - \gamma_{12}m_2/\gamma_2^2$, $\gamma_{2.1}^2 = \gamma_2^2 - \gamma_{12}/\gamma_1^2$, $\gamma_{1.2}^2 = \gamma_1^2 - \gamma_{12}/\gamma_2^2$, and in light of Proposition 3.2 we have that $\tilde{\Phi}^{(2.1)} \equiv \tilde{\Phi}_1(0; \mu_{2.1}, \sigma_{2.1}^2, \sigma_{2.1}\varphi_2, \tau_{2.1})$, $\tilde{\Phi}^{(1.2)} \equiv \tilde{\Phi}_1(0; \mu_{1.2}, \sigma_{1.2}^2, \sigma_{1.2}\varphi_1, \tau_{1.2})$, and $\tilde{\phi}^{(\ell)} \equiv ESN_1(0; \mu_\ell, \sigma_\ell^2, c_\ell\sigma_\ell\tilde{\varphi}_\ell, c_\ell\tau)$ for $\ell = \{1, 2\}$.

Using Remark 1 along with (3.42), we finally obtain an explicit expression for $\mathbb{E}[|X_i X_j|]$ as

$$\begin{aligned}&= \mathbb{E}[|X_i X_j|] \\ &= (\mu_i\mu_j + \sigma_{ij})(1 - 2(\tilde{\Phi}^{(i)} + \tilde{\Phi}^{(j)})) + (\delta_i\mu_j + \delta_j(\mu_i - \mu_{bi}))\,(1 - 2(\Phi^{(i)} + \Phi^{(j)})) \\ &\quad + 2\mu_j\left[\sigma_i^2\tilde{\phi}^{(i)}(1 - 2\tilde{\Phi}^{(i)}) + \sigma_{ij}\tilde{\phi}^{(j)}(1 - 2\tilde{\Phi}^{(j)})\right] + 2\sigma_j^2\tilde{\phi}^{(j)}\mathbb{E}[|Y_{i.j}|] \\ &\quad + 2\delta_j\left[\gamma_i^2\phi(\mu_i; \mu_{bi}, \gamma_i^2)(1 - 2\Phi(0; m_{j.i}, \gamma_{j.i}^2)) + \gamma_{ij}\phi(\mu_j; \mu_{bj}, \gamma_j^2)(1 - 2\Phi(0; m_{i.j}, \gamma_{i.j}^2))\right]\end{aligned} \tag{3.44}$$

with $X_{i.j} \sim \text{ESN}_i(\mu_{i.j}, \sigma_{i.j}^2, \sigma_{i.j}\varphi_i, \tau_{i.j})$. Furthermore,

$$\tilde{\Phi}^{(1)} \equiv \tilde{\Phi}_2(\mathbf{0}; (-\mu_i, \mu_j)^\top, \boldsymbol{\Sigma}^-, (-\lambda_i, \lambda_j)^\top, \tau), \qquad \tilde{\Phi}^{(2)} \equiv \tilde{\Phi}_2(\mathbf{0}; (\mu_i, -\mu_j)^\top, \boldsymbol{\Sigma}^-, (\lambda_i, -\lambda_j)^\top, \tau),$$

$$\Phi^{(1)} \equiv \Phi_2(\mathbf{0}; (-m_i, m_j)^\top, \boldsymbol{\Gamma}^-) \qquad \text{and} \qquad \Phi^{(2)} \equiv \Phi_2(\mathbf{0}; (m_i, -m_j)^\top, \boldsymbol{\Gamma}^-),$$

with $\boldsymbol{\Sigma}^{-}$ ($\boldsymbol{\Gamma}^{-}$) denoting the $\boldsymbol{\Sigma} = [\sigma_{ij}]$ ($\boldsymbol{\Gamma} = [\gamma_{ij}]$) matrix with all its signs of covariances (off-diagonal elements) changed. Here, we have simplified the expressions above using that $\sum_{\mathbf{s}\in S(p)} L_p(\mathbf{0}, \infty, \mathbf{m}_s, \boldsymbol{\Gamma}_s) = 1$, along the equivalences

$$\begin{aligned}
\mathcal{L}_p(\mathbf{0}, \infty; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}_s, \tau) &= \tilde{\Phi}_p(\mathbf{0}; -\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, -\boldsymbol{\lambda}_s, \tau), && \text{for } \mathbf{s} \in S(p) \\
ESN_p(\mathbf{0}; \boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q, \boldsymbol{\lambda}_q, \tau) &= ESN_p(\mathbf{0}; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r, \boldsymbol{\lambda}_r, \tau), && \text{for } \mathbf{q}, \mathbf{r} \in S(p) \\
P(Y_1 Y_2 \cdots Y_p > 0) &= \sum_{\mathbf{s}\in S(p)} \pi_s \mathcal{L}_p(\mathbf{0}, \infty; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s, \tau),
\end{aligned}$$

with $\pi_s = \prod_{i=1}^{p} s_i$ as in Theorem 3.4 and $\sum_{\mathbf{s}\in S(p)} \mathcal{L}_p(\mathbf{0}, \infty; \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s, \tau) = 1$. It is worth mentioning that these expressions hold for the normal case, when $\boldsymbol{\lambda} = \mathbf{0}$ and $\tau = 0$.

As expected, this approach is much faster than the one using equations (3.40) and (3.41). For instance, when we consider a trivariate folded ESN distribution, we have that it is approximately 56x times faster than using MC methods and 10x times faster than using equations (3.40) and (3.41). Time comparison (summarized in Figure 15, right panel) as well as sample codes of our `MomTrunc R` package are provided in the Appendices B.3 and B.4, respectively. Contours of different FESN densities can be found in Figure 17 given in Appendix B.3 as well.

3.7 Conclusions

In this paper, we have developed a recurrence approach for computing order product moments of TESN and FESN distributions as well as explicit expressions for the first two moments as a byproduct, generalizing results obtained by Kan & Robotti (2017) for the normal case. The proposed methids also includes the moments of the well-known truncated multivariate SN distribution, introduced by Azzalini & Dalla-Valle (1996). For the TESN, we have proposed an optimized robust algorithm based only in normal integrals, which for the limiting normal case outperforms the existing popular method for computing the first two moments, even computing these two moments for extreme cases where all available algorithms fail. The proposed method (including its limiting and special cases) has been coded and implemented in the `R MomTrunc` package, which is available for the users on CRAN repository.

During the last decade or so, censored modeling approaches have been used in various ways to accommodate increasingly complicated applications. Many of these extensions involve using Normal (Vaida & Liu, 2009) and Student-t (Matos *et al.*, 2013; Lachos *et al.*, 2017), however statistical models based on distributions to accommodate censored and skewness, simultaneously, so far have remained relatively unexplored in

the statistical literature. We hope that by making the codes available to the community, we will encourage researchers of different fields to use our newly methods. For instance, now it is possible to derive analytical expressions on the E-step of the EM algorithm for multivariate SN responses with censored observation as in Matos *et al.* (2013).

4 Likelihood-based inference for multivariate skew-normal censored responses

4.1 Introduction

In many applications on simulations or on experimental studies, the researches often generate a large number of datasets with values restricted to fixed intervals. For example, variables such as pH, grades, viral load in HIV studies and humidity in environmental studies, have upper and lower bounds due to detection limits, being the support of their densities restricted to some given intervals. On the other hand, during the last decade or so, censored modeling approaches have been used in several ways to accommodate increasingly complicated applications. Many of these extensions involve using normal and its symmetrical extension. For instance, Massuia *et al.* (2015) proposed the Student-t censored regression model. Garay *et al.* (2017) (see also, Matos *et al.* (2013)) advocated the use of the multivariate Student-t distribution in the context of censored regression models, where a simple and efficient EM-type algorithm for iteratively computing ML estimates of the parameters was also presented. More recently, Wang *et al.* (2018) proposed a multivariate extension of the models of Garay *et al.* (2017) and Matos *et al.* (2013), for analyzing multi-outcome longitudinal data with censored observations, where they established a feasible EM algorithm that admits closed-form expressions at E-steps and tractable solutions at M-steps. They demonstrated its robustness against outliers through extensive simulations. A common drawback of these proposals is that they are not appropriate when the observed data exhibit skewness, which might lead to bias estimates (Azzalini & Capitanio, 1999).

In this paper, we propose to use the multivariate skew-normal distribution to analyze censored data, so that the SN-C model is defined and a fully likelihood-based approach is carried out, including the implementation of an exact EM-type algorithm for the ML estimation. Like Garay *et al.* (2017), we show that the E-step reduces to computing the first two moments of a truncated multivariate skew-normal distribution, which are implemented in the `R` package `MomTrunc` (Galarza *et al.*, 2018). The likelihood function is easily computed as a byproduct of the E-step and it is used for monitoring convergence and for model selection.

The rest of this paper is organized as follows. In Section 4.2 we briefly discuss some preliminary results related to the multivariate SN, extended SN (ESN) and some of its key properties. Moreover, the truncated extended skew-normal is presented along with some sketch of the computation of its first two moments. Section 4.3 presents the EM

algorithm for estimating the model parameters of multivariate SN responses as well as in a regression SN setting. Section 4.4 implements the proposed algorithm to real datasets and finally, some concluding remarks are presented in Section 4.5.

4.2 Preliminaries

In this section we present some properties of the multivariate skew-normal distribution and its extended version, the extended skew-normal distribution.

4.2.1 The multivariate skew-normal distribution

We say that a $p \times 1$ random vector $\mathbf{Y}$ follows a multivariate SN distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$ and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, and we write $\mathbf{Y} \sim \mathrm{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its joint probability density function (pdf) is given by

$$SN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{4.1}$$

where $\Phi_1(\cdot)$ represents the cumulative distribution function (cdf) of the standard univariate normal distribution. If $\boldsymbol{\lambda} = \mathbf{0}$, then (4.1) reduces to the symmetric $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pdf. Except by a straightforward difference in the parametrization considered in (4.1), this model corresponds to that introduced by Azzalini & Dalla-Valle (1996).

Proposition 4.1 (*cdf* of the SN)**.** *If* $\mathbf{Y} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$*, then for any* $\mathbf{y} \in \mathbb{R}^p$*,*

$$F_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{Y} \leqslant \mathbf{y}) = 2\Phi_{p+1}\big((\mathbf{y}^\top, 0)^\top; \boldsymbol{\mu}^*, \boldsymbol{\Omega}\big), \tag{4.2}$$

where $\boldsymbol{\mu}^* = (\boldsymbol{\mu}^\top, 0)^\top$ *and* $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & -\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}^\top & 1 \end{pmatrix}$*, with* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

It is worth mentioning that the multivariate skew-normal distribution is closed over marginalization but not conditioning. Next we present its extended version which holds both properties, called, the multivariate ESN distribution.

4.2.2 The extended multivariate skew-normal distribution

We say that a $p \times 1$ random vector $\mathbf{Y}$ follows a ESN distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$, a $p \times 1$ skewness parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, and a shift or extension parameter $\tau \in \mathbb{R}$, denoted by $\mathbf{Y} \sim \mathrm{ESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, if its pdf is given by

$$ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \xi^{-1}\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \tag{4.3}$$

with $\xi = \Phi_1(\tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})$. Note that, when $\tau = 0$, we retrieve the skew-normal distribution defined in (4.1), that is, $ESN_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},0) = SN_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$. In this work, we uses a slightly different parametrization of the ESN distribution found in Arellano-Valle & Azzalini (2006a) and Arellano-Valle & Genton (2010). Furthermore, Arellano-Valle & Genton (2010) deals with the multivariate extended skew-t (EST) distribution, in which the ESN is a particular case when the degrees of freedom ν go to infinity. From Arellano-Valle & Genton (2010), it is straightforward to see that

$$ESN_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau) \longrightarrow \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}), \text{ as } \tau \to +\infty.$$

The following propositions will allow us to develop our methods.

Proposition 4.2 (*Marginal and conditional distribution of the ESN*)**.** *Let* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$ *partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top,\mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 *($p_1+p_2=p$), respectively. Let*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top,\boldsymbol{\mu}_2^\top)^\top, \quad \boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top,\boldsymbol{\lambda}_2^\top)^\top \quad \text{and} \quad \boldsymbol{\varphi} = (\boldsymbol{\varphi}_1^\top,\boldsymbol{\varphi}_2^\top)^\top,$$

be the corresponding partitions of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ *and* $\boldsymbol{\varphi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. *Then,*

$$\mathbf{Y}_1 \sim ESN_{p_1}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1,c_{12}\tau), \quad \mathbf{Y}_2|\mathbf{Y}_1=\mathbf{y}_1 \sim ESN_{p_2}(\boldsymbol{\mu}_{2.1},\boldsymbol{\Sigma}_{22.1},\boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2,\tau_{2.1})$$

where $c_{12} = (1+\boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1-\boldsymbol{\mu}_1)$ *and* $\tau_{2.1} = \tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1-\boldsymbol{\mu}_1)$.

Proof. The proof can be found in Appendix section B. □

Proposition 4.3 (*Stochastic representation by convolution*)**.** *Assume that* $\mathbf{X}$ *and* T *are independent variables with distribution* $N_p(\mathbf{0},\mathbf{I}_p)$ *and* $TN_1(0,1;[-\tilde{\tau},\infty])$ *(which represents a truncated standard normal distribution on* $[-\tilde{\tau},\infty)$*) respectively. Henceforth,*

$$\mathbf{Y} \overset{d}{=} \boldsymbol{\mu} + T\boldsymbol{\Delta} + \boldsymbol{\Gamma}^{1/2}\mathbf{X}, \tag{4.4}$$

is distributed as $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$, *where* $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ *with* $\boldsymbol{\Delta}$ *as defined in Proposition 4.1.*

Proof. The proof can be found in Arellano-Valle & Azzalini (2006a). □

Proposition 4.4 (*Stochastic representation by conditioning*)**.** *Let* $\mathbf{X} = (\mathbf{X}_1{}^\top, X_2)^\top \sim N_{p+1}(\boldsymbol{\mu}^*,\boldsymbol{\Omega})$. *If*

$$\mathbf{Y} \overset{d}{=} (\mathbf{X}_1|X_2 < \tilde{\tau}), \tag{4.5}$$

it follows that $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau)$, *with* $\boldsymbol{\mu}^*$ *and* $\boldsymbol{\Omega}$ *as defined in Proposition 4.1, and* $\tilde{\tau} = \tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

The stochastic representation in Equation (4.5) can be derived from Proposition 1 in Arellano-Valle & Genton (2010).

Proposition 4.5 (cdf *of the ESN*). *If* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, *then for any* $\mathbf{y} \in \mathbb{R}^p$,

$$F_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{Y} \leqslant \mathbf{y}) = \frac{\Phi_{p+1}\big((\mathbf{y}^\top, \tilde{\tau})^\top; \boldsymbol{\mu}^*, \boldsymbol{\Omega}\big)}{\Phi(\tilde{\tau})}. \tag{4.6}$$

Proof. The proof comes from proposition 4.4. Hereinafter, for $\mathbf{Y} \sim \mathrm{ESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, we will denote its *cdf* as $F_{\mathbf{Y}}(\mathbf{y}) \equiv \tilde{\Phi}_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, for simplicity. □

4.2.3 The truncated extended skew-normal distribution

Let $\mathbb{A}$ be a Borel set in $\mathbb{R}^p$. We say that the random vector $\mathbf{Y}$ has a truncated extended skew-normal distribution on $\mathbb{A}$ when $\mathbf{Y}$ has the same distribution as $\mathbf{Y}|(\mathbf{Y} \in \mathbb{A})$. In this case, the pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; \mathbb{A}) = \frac{ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)}{P(\mathbf{Y} \in \mathbb{A})} \mathbf{1}_{\mathbb{A}}(\mathbf{y}),$$

where $\mathbf{1}_{\mathbb{A}}$ is the indicator function of $\mathbb{A}$. We use the notation $\mathbf{Y} \sim \mathrm{TESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau; \mathbb{A})$. If $\mathbb{A}$ has the form

$$\mathbb{A} = \{(x_1, \ldots, x_p) \in \mathbb{R}^p : \ a_1 \leqslant x_1 \leqslant b_1, \ldots, a_p \leqslant x_p \leqslant b_p\} = \{\mathbf{x} \in \mathbb{R}^p : \mathbf{a} \leqslant \mathbf{x} \leqslant \mathbf{b}\}, \tag{4.7}$$

we use the notation $\{\mathbf{Y} \in \mathbb{A}\} = \{\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}\}$, where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$. In this case, we say that the distribution of $\mathbf{Y}$ is double truncated. Analogously, we define $\{\mathbf{Y} \geqslant \mathbf{a}\}$ and $\{\mathbf{Y} \leqslant \mathbf{b}\}$, so, we say that the distribution of $\mathbf{Y}$ is truncated from below and truncated from above, respectively. For convenience, we also use the notation $\mathbf{Y} \sim \mathrm{TESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau; [\mathbf{a}, \mathbf{b}])$. In particular, for a truncated p-variate skew-normal and normal distribution on $[\mathbf{a}, \mathbf{b}]$, we use the notations $\mathbf{X} \sim \mathrm{TSN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; [\mathbf{a}, \mathbf{b}])$ and $\mathbf{W} \sim \mathrm{TN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; [\mathbf{a}, \mathbf{b}])$, respectively. We also define the normalizing constant $\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = P(\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b})$ as

$$\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = \int_{\mathbf{a}}^{\mathbf{b}} ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) \mathrm{d}\mathbf{y}.$$

If all $\boldsymbol{\lambda}$ and τ are equal to zero, we have a normal integral $\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0}, 0) = L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathbf{a}}^{\mathbf{b}} \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\mathbf{y}$.

Remark: Note that, we use calligraphic style $\mathcal{L}_p$ to denote the skewed extended normal integral and roman style L_p for the symmetric one.

Let $\mathbf{a}_{(i)}$ denote a vector $\mathbf{a}$ with its ith element being removed. For a matrix $\mathbf{A}$, we let $\mathbf{A}_{(i,j)}$ stands for its ith row and jth column being removed. Then, for $\mathbf{Y} \sim$

$\text{TESN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau; [\mathbf{a}, \mathbf{b}])$, it follows from corollary 3.1 that the first two moments of $\mathbf{Y}$ can be computed from its corresponding Normal random variable, as

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{W}]_{(p+1)}, \tag{4.8}$$

$$\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] = \mathbb{E}[\mathbf{W}\mathbf{W}^\top]_{(p+1,p+1)}, \tag{4.9}$$

where $\mathbf{W} \sim \text{TN}_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Omega}; [\mathbf{a}^*, \mathbf{b}^*])$, with $\boldsymbol{\mu}^*$ and $\boldsymbol{\Omega}$ as defined in Proposition 4.1, $\mathbf{a}^* = (\mathbf{a}^\top, -\infty)^\top$ and $\mathbf{b}^* = (\mathbf{b}^\top, \tilde{\tau})^\top$.

The first two moments of $\mathbf{Y}$ obtained from equations (4.8) and (4.9) are available through the `MomTrunc R` package (Galarza *et al.*, 2018), which so far, is the unique method to compute the moments of the TESN and TSN. In the following, we present some useful propositions and corollaries related to TESN random vectors.

Proposition 4.6. *Let* $\mathbf{Y} \sim TESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, [\mathbf{a}, \mathbf{b}])$. *For any measurable function* $g(\cdot)$, *we have that*

$$\mathbb{E}\left[g(\mathbf{Y})\frac{\phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\right] = \frac{\eta L}{\mathcal{L}}\mathbb{E}[g(\mathbf{W})], \tag{4.10}$$

with $\eta = \phi_1(\tau; 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})/\xi$, $\boldsymbol{\mu}_b = \tilde{\tau}\boldsymbol{\Delta}$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, $L \equiv L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})$, $\mathcal{L} \equiv \mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ *and* $\mathbf{W} \sim TN_p(\boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}, [\mathbf{a}, \mathbf{b}])$.

Proof. Let $\mathbf{X} = (\mathbf{X}_1{}^\top, X_2)^\top \sim N_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Omega})$ as in Proposition 4.4. From the conditional distribution of a multivariate normal distribution , it is straightforward to show that $\mathbf{X}_1|X_2 \sim N_p(\boldsymbol{\mu} - X_2\boldsymbol{\Delta}, \boldsymbol{\Gamma})$ and $X_2|\mathbf{X}_1 \sim N_1(-\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}), 1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta})$. Then, it holds that

$$\begin{aligned}
f_{X_2|\mathbf{x}_1}(\tilde{\tau}; \mathbf{x}) f_{\mathbf{x}_1}(\mathbf{x}) &= f_{X_2}(\tilde{\tau}) f_{\mathbf{x}_1|X_2}(\mathbf{x}; \tilde{\tau}) \\
\phi_1(\tilde{\tau}; -\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}), 1 - \boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta})\phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \phi_1(\tilde{\tau})\phi_p(\mathbf{x}; \boldsymbol{\mu} - \tilde{\tau}\boldsymbol{\Delta}, \boldsymbol{\Gamma}) \\
\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}} \times \phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \phi_1(\tilde{\tau})\phi_p(\mathbf{x}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}) \\
\phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \phi_1(\tau; 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})\phi_p(\mathbf{x}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}),
\end{aligned} \tag{4.11}$$

where we have used that $\boldsymbol{\Delta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta} = -\boldsymbol{\lambda}^\top\boldsymbol{\lambda}$. Thus,

$$\begin{aligned}
\mathbb{E}\left[g(\mathbf{Y})\frac{\phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\right] &= \int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{y})\frac{\phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))}{\Phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))}\frac{\text{ESN}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)}{\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)}\mathrm{d}\mathbf{y} \\
&= \int_{\mathbf{a}}^{\mathbf{b}} \frac{\phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\xi\mathcal{L}_p(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)} g(\mathbf{y})\mathrm{d}\mathbf{y} \\
&= \frac{\phi_1(\tau; 0, 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})}{\xi\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)} \int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{y})\phi_p(\mathbf{y}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})\mathrm{d}\mathbf{y} \\
&= \frac{\eta L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})}{\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)} \int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{w})\frac{\phi_p(\mathbf{w}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})}{L_p(\mathbf{a}, \mathbf{b}, \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})}\mathrm{d}\mathbf{w} \\
&= \frac{\eta L}{\mathcal{L}}\mathbb{E}[g(\mathbf{W})],
\end{aligned}$$

for $\mathbf{W} \sim TN_p(\boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}, [\mathbf{a}, \mathbf{b}])$. □

Corollary 4.1. *Let* $\mathbf{Y} \sim TESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, [\mathbf{a}, \mathbf{b}])$. *As* $\tau \longrightarrow -\infty$, *we have from Proposition 4.6 that*

$$\mathbb{E}\left[g(\mathbf{Y})\frac{\phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\right] = -\frac{\tau}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\mathbb{E}[g(\mathbf{W})], \tag{4.12}$$

where $\mathbf{W} \sim TN_p(\boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}, [\mathbf{a}, \mathbf{b}])$.

Proof. As $\tau \to -\infty$, we have

$$\frac{\Phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))}{\Phi_1(\tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})} \longrightarrow \frac{0}{0}.$$

Using L'Hospital,

$$\begin{aligned}
\lim_{\tau\to-\infty} ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) &= \lim_{\tau\to-\infty} \frac{\phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))\phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{-1/2} \times \phi_1(\tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})} \\
&\overset{(4.11)}{=} \lim_{\tau\to-\infty} \frac{\phi_1(\tau; 0; 1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})\,\phi_p(\mathbf{y}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})}{(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{-1/2} \times \phi_1(\tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})} \\
&= \lim_{\tau\to-\infty} \phi_p(\mathbf{y}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma}).
\end{aligned}$$

Therefore, $ESN_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\tau}) \longrightarrow \phi_p(\mathbf{y}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})$, $\mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) \longrightarrow L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu} - \boldsymbol{\mu}_b, \boldsymbol{\Gamma})$ and $\eta \longrightarrow -\tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})$, as $\tau \longrightarrow -\infty$. This last holds from the inverse Mill's ratio, since $\phi(x)/\Phi(x) \longrightarrow -x$ as $x \longrightarrow -\infty$. It is enough to replace the limiting terms in Proposition 4.6. □

Corollary 4.2. *Setting* $\tau = 0$ *in Corollary 1, it follows that* $\mathbf{Y} \sim TSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, [\mathbf{a}, \mathbf{b}])$ *and*

$$\mathbb{E}\left[g(\mathbf{Y})\frac{\phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\right] = \frac{L_0}{\sqrt{\frac{\pi}{2}(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}\,\mathcal{L}_0}\mathbb{E}[g(\mathbf{W}_0)], \tag{4.13}$$

with $L_0 = L_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Gamma})$, $\mathcal{L}_0 = \mathcal{L}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0)$ *and* $\mathbf{W}_0 \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Gamma}, [\mathbf{a}, \mathbf{b}])$.

Proof. The proof is straightforward. Setting $\tau = 0$, it suffices to show that $\boldsymbol{\mu}_b = \mathbf{0}$ and $\eta = \sqrt{2/\pi(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}$. □

Proposition 4.7. *Let* $\mathbf{Y} \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$. *Also, let* $\mathbf{Y}$ *be partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 $(p_1 + p_2 = p)$, *with corresponding partitions of* $\mathbf{a}$, $\mathbf{b}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ *and* $\boldsymbol{\varphi}$. *Then, for any measurable function* $g(\cdot)$, *we have that*

$$\mathbb{E}_{\mathbf{Y}_2}\left[g(\mathbf{Y}_2)\frac{\phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}{\Phi_1(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}))}\middle|\, \mathbf{Y}_1\right] = \frac{\eta_{2.1}L_{2.1}}{\mathcal{L}_{2.1}}\mathbb{E}[g(\mathbf{W}_2)], \tag{4.14}$$

where $L_{2.1} = L_{p_2}(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_{2.1} - \boldsymbol{\mu}_{b2.1}, \boldsymbol{\Gamma}_{22.1})$, $\mathcal{L}_{2.1} = \mathcal{L}_{p_2}(\mathbf{a}_2, \mathbf{b}_2; \boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1})$ *and* $\mathbf{W}_2 \sim TN_p(\boldsymbol{\mu}_{2.1} - \boldsymbol{\mu}_{b2.1}, \boldsymbol{\Gamma}_{22.1}, [\mathbf{a}_2, \mathbf{b}_2])$, *with* $\boldsymbol{\lambda}_{2.1} = \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2$, $\boldsymbol{\mu}_{2.1}$, $\boldsymbol{\Sigma}_{22.1}$ *and* $\tau_{2.1}$ *as in Proposition 4.2; and* $\eta_{2.1}$, $\boldsymbol{\mu}_{b2.1}$ *and* $\boldsymbol{\Gamma}_{22.1}$ *can be computed as expressions* η, $\boldsymbol{\mu}_b$ *and* $\boldsymbol{\Gamma}$ *in Proposition 4.6 but using the new set of parameters* $\boldsymbol{\mu}_{2.1}$, $\boldsymbol{\Sigma}_{22.1}$, $\boldsymbol{\lambda}_{2.1}$ *and* $\tau_{2.1}$ *(instead of* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ *and* τ*).*

Proof. From Proposition 2, it is known that $\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) = \tau_{2.1} + \boldsymbol{\lambda}_{2.1}^\top \boldsymbol{\Sigma}_{22.1}^{-1/2}(\mathbf{Y}_2 - \boldsymbol{\mu}_{2.1})$, then it is enough to apply Proposition 4.6 by considering $\mathbf{Y}$ as $\mathbf{Y}_2 \sim \text{TESN}_{p_2}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, [\mathbf{a}_2, \mathbf{b}_2])$. □

4.3 Multivariate SN censored responses

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a $p \times 1$ response vector for the ith sample unit, for $i \in \{1, \ldots, n\}$, and consider the set of random samples (independent and identically distributed):

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim \text{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}), \tag{4.15}$$

with location vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$, dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ depending on an unknown and reduced parameter vector $\boldsymbol{\alpha}$ and skewness parameter $\boldsymbol{\lambda}$. However, the response vector $\mathbf{Y}_i$ may not be fully observed due to censoring, so we define $(\mathbf{V}_i, \mathbf{C}_i)$ the observed data for the ith sample, where $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$ with elements being either an uncensored observation ($V_{ik} = V_{0i}$) or the interval censoring level ($V_{ik} \in [V_{1ik}, V_{2ik}]$), and $\mathbf{C}_i = (C_{i1}, \ldots, C_{ip})^\top$ is the vector of censoring indicators, satisfying

$$C_{ik} = \begin{cases} 1 & \text{if } V_{1ik} \leqslant Y_{ik} \leqslant V_{2ik}\,, \\ 0 & \text{if } Y_{ik} = V_{0i}\,, \end{cases} \tag{4.16}$$

for all $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, p\}$, i.e., $C_{ik} = 1$ if Y_{ik} is located within a specific interval. In this case, (4.15) along with (4.16) defines the multivariate skew-normal interval censored model (hereafter, the SN-C model). For instance, left censoring structure causes truncation from the lower limit of the support of the distribution, since we only know that the true observation Y_{ik} is greater than or equal to the observed quantity V_{1ik}. Moreover, missing observations can be handled by considering $V_{1ik} = -\infty$ and $V_{2ik} = +\infty$.

4.3.1 The likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$ is a realization of $\mathbf{Y}_i \sim \text{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. In order to obtain the likelihood function of the SN-C model, first we treat, separately, the observed and censored components of $\mathbf{y}_i$, i.e., $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $C_{ik} = 0$ for all elements in the p_i^o-dimensional vector $\mathbf{y}_i^o$, and $C_{ik} = 1$ for all elements in the p_i^c-dimensional vector $\mathbf{y}_i^c$. On according to that, we write $\mathbf{V}_i = \text{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $\mathbf{V}_i^c = (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c)$ with

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

Then, using Proposition 4.2, we have that $\mathbf{Y}_i^o \sim \text{SN}_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, c_i^{oc} \boldsymbol{\Sigma}_i^{oo\,1/2} \tilde{\boldsymbol{\varphi}}_i^o)$ and $\mathbf{Y}_i^c \mid \mathbf{Y}_i^o = \mathbf{y}_i^o \sim \text{ESN}_{p_i^c}(\boldsymbol{\mu}_i^{co}, \boldsymbol{\Sigma}_i^{cc.o}, \boldsymbol{\Sigma}_i^{cc.o\,1/2} \boldsymbol{\varphi}_i^c, \tau_i^{co})$, where

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co} \boldsymbol{\Sigma}_i^{oo\,-1} (\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \quad \boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co} (\boldsymbol{\Sigma}_i^{oo})^{-1} \boldsymbol{\Sigma}_i^{oc}, \quad \tilde{\boldsymbol{\varphi}}_i^o = \boldsymbol{\varphi}_i^o + \boldsymbol{\Sigma}_i^{oo\,-1} \boldsymbol{\Sigma}_i^{oc} \boldsymbol{\varphi}_i^c, \tag{4.17}$$

$$c_i^{oc} = (1 + \boldsymbol{\varphi}_i^{c\top}\boldsymbol{\Sigma}_i^{cc.o}\boldsymbol{\varphi}_i^c)^{-1/2} \quad \text{and} \quad \tau_i^{c.o} = \tilde{\boldsymbol{\varphi}}_i^{o\top}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o). \tag{4.18}$$

Let $\mathbf{V} = \text{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \text{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ denote the observed data. Therefore, the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}_\Sigma^\top, \boldsymbol{\lambda}^\top)^\top$, where $\boldsymbol{\alpha}_\Sigma$ denotes a minimal set of parameters such that $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ is well defined (e.g. the upper triangular elements of $\boldsymbol{\Sigma}$ in the unstructured case), for the observed data $(\mathbf{V}, \mathbf{C})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln L_i, \tag{4.19}$$

where $L_i \equiv L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i)$ represents the likelihood function of $\boldsymbol{\theta}$ for the ith sample, say

$$\begin{aligned} L_i &= f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{V}_{1i}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{V}_{2i}^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ &= \mathcal{L}_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \boldsymbol{\mu}_i^{co}, \boldsymbol{\Sigma}_i^{cc.o}, \boldsymbol{\Sigma}_i^{cc.o\,1/2}\boldsymbol{\varphi}_i^c, \tau_i^{co})\, SN_{p_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, c_i^{oc}\boldsymbol{\Sigma}_i^{oo\,1/2}\tilde{\boldsymbol{\varphi}}_i^o). \end{aligned}$$

4.3.2 Parameter estimation via the EM algorithm

In this subsection, we describe how to carry out ML estimation for the SN-C model. The EM algorithm, originally proposed by Dempster *et al.* (1977), is a very popular iterative optimization strategy commonly used to obtain ML estimates for incomplete-data problems. This algorithm has many attractive features such as the numerical stability, the simplicity of implementation and quite reasonable memory requirements (McLachlan & Krishnan, 2008).

From the stochastic representation of multivariate ESN distribution in Proposition 4.3, setting $\tau = 0$, we can write $\mathbf{Y}_i|(T_i = t_i) \sim N_p(\boldsymbol{\mu} + \boldsymbol{\Delta}t_i, \boldsymbol{\Gamma})$ and $T_i \sim \text{HN}(0, 1)$, with HN referring to a Half normal distribution The complete data log-likelihood function of an equivalent set of parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\Delta}^\top, \boldsymbol{\alpha}_\Gamma^\top)^\top$, where $\boldsymbol{\alpha}_\Gamma = \text{vech}(\boldsymbol{\Gamma})$, is given by $\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta})$, where the individual complete data log-likelihood is

$$\ell_{ic}(\boldsymbol{\theta}) = -\frac{1}{2}\big\{\ln|\boldsymbol{\Gamma}| + (\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta}t_i)\big\} + c,$$

with c being a constant that does not depend on $\boldsymbol{\theta}$. Subsequently, the EM algorithm for the SN-C model can be summarized as follows:

E-step: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Delta}}^{(k)}, \hat{\boldsymbol{\alpha}}_\Gamma^{(k)})$ at the kth step of the algorithm, the E-step provides the conditional expectation of the complete data log-likelihood function

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \text{E}\Big[\ell_c(\boldsymbol{\theta}) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\Big] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) \propto -\frac{1}{2}\ln|\boldsymbol{\Gamma}| - \frac{1}{2}\text{tr}\Bigg[\Big\{\widehat{\mathbf{y}_i^2}^{(k)} + \boldsymbol{\mu}\boldsymbol{\mu}^\top + \widehat{t_i^2}^{(k)}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top - 2\boldsymbol{\mu}\widehat{\mathbf{y}}_i^{(k)\top} - 2\widehat{t\mathbf{y}}_i^{(k)}\boldsymbol{\Delta}^\top$$

$$+2\widehat{t_i}^{(k)}\boldsymbol{\Delta}\boldsymbol{\mu}^\top\Big\}\boldsymbol{\Gamma}^{-1}\Big],$$

with $\widehat{\mathbf{y}_i^r}^{(k)} = \mathbb{E}_{T_i\mathbf{Y}_i}[\mathbf{Y}_i^r \,|\, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{t_i^r}^{(k)} = \mathbb{E}_{T_i\mathbf{Y}_i}[T_i^r \,|\, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ (for $r = \{1, 2\}$, with $\mathbf{Y}_i^1 = \mathbf{Y}_i$ and $\mathbf{Y}_i^2 = \mathbf{Y}_i\mathbf{Y}_i^\top$) and $\widehat{t\mathbf{y}_i}^{(k)} = \mathbb{E}_{T_i\mathbf{Y}_i}[T_i\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$.

M-step: Conditionally maximizing $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ with respect to each entry of $\boldsymbol{\theta}$, we update the estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\alpha}}_\Gamma^{(k)})$ by

$$\widehat{\boldsymbol{\mu}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\widehat{\mathbf{y}}_i^{(k)} - \widehat{t}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\Big\}, \tag{4.20}$$

$$\widehat{\boldsymbol{\Delta}}^{(k+1)} = \left\{\sum_{i=1}^{n}\widehat{t}_i^{2(k)}\right\}^{-1}\sum_{i=1}^{n}\Big\{\widehat{t\mathbf{y}_i}^{(k)} - \widehat{t}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)}\Big\}, \tag{4.21}$$

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\widehat{\mathbf{y}_i^2}^{(k)} + \widehat{\boldsymbol{\mu}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top} + \widehat{t_i^2}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} - 2\widehat{\boldsymbol{\mu}}^{(k+1)}\widehat{\mathbf{y}}_i^{(k)\top}\\ &-2\widehat{t\mathbf{y}_i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} + 2\widehat{t}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top}\Big\}.\end{aligned} \tag{4.22}$$

Algorithm is iterated until a suitable convergence rule is satisfied. In the later analysis, the algorithm stops when the relative distance between two successive evaluations of the log-likelihood defined in (4.19) is less than a tolerance, i.e., $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)} \mid \mathbf{V}, \mathbf{C})/\ell(\widehat{\boldsymbol{\theta}}^{(k)} \mid \mathbf{V}, \mathbf{C}) - 1| < \epsilon$, for example, $\epsilon = 10^{-6}$. Once converged, we can recover $\widehat{\boldsymbol{\lambda}}$ and $\widehat{\boldsymbol{\Sigma}}$ using the expressions

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Gamma}} + \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^\top \qquad \text{and} \qquad \widehat{\boldsymbol{\lambda}} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\Delta}}}{(1 - \widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Delta}})^{1/2}}.$$

It is important to stress that, from equations (4.20) and (4.22), the E-step reduces to the computation of $\widehat{\mathbf{y}}_i^{(k)}$, $\widehat{\mathbf{y}_i^2}^{(k)}$ $\widehat{t}_i^{(k)}$, $\widehat{t_i^2}^{(k)}$ and $\widehat{t\mathbf{y}_i}^{(k)}$. To compute these expected values, first note that for any measurable function of T_i and $\mathbf{Y}_i$, such that $g(T_i, \mathbf{Y}_i) = g_1(T_i)g_2(\mathbf{Y}_i)$, we have that

$$\mathbb{E}_{T_i\mathbf{Y}_i}[g(T_i, \mathbf{Y}_i)|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[g_1(\mathbf{Y}_i)\mathbb{E}_{T_i}[g_2(T_i)|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i]. \tag{4.23}$$

Then,

$$\begin{aligned}\widehat{\mathbf{y}_i^r} &= \mathbb{E}_{T_i\mathbf{Y}_i}[\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i],\\ \widehat{t_i^r} &= \mathbb{E}_{T_i\mathbf{Y}_i}[T_i^r|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbb{E}_{T_i}[T_i^r|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i],\\ \widehat{t\mathbf{y}_i} &= \mathbb{E}_{T_i\mathbf{Y}_i}[T_i\mathbf{Y}_i|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i\mathbb{E}_{T_i}[T_i|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i].\end{aligned}$$

From Cabral *et al.* (2012), we know that $T_i|\mathbf{Y}_i \sim TN_1(M^2(\boldsymbol{\theta})\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}), M^2(\boldsymbol{\theta}), (0, \infty))$, having that

$$\mathbb{E}_{T_i}[T_i|\mathbf{Y}_i] = M^2(\boldsymbol{\theta})\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) + M(\boldsymbol{\theta})\frac{\phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))}{\Phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))}, \tag{4.24}$$

$$\mathbb{E}_{T_i}[T_i^2|\mathbf{Y}_i] = [M^2(\boldsymbol{\theta})\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})]^2 + M^3(\boldsymbol{\theta})\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})\frac{\phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i-\boldsymbol{\mu}))}{\Phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i-\boldsymbol{\mu}))} + M^2(\boldsymbol{\theta}), \tag{4.25}$$

where $M(\boldsymbol{\theta}) = (1+\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1/2}$.

Subsequently, on according to expressions (4.23), (4.24) and (4.25), and Corollary 4.2, we have the implementable expressions to the conditional expectations as follows:

1. If the ith subject has only non-censored components, then

$$\begin{aligned}
\widehat{\mathbf{y}_i^r}^{(k)} &= \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}] = \mathbf{y}_i^r,\\
\widehat{t_i^r}^{(k)} &= \mathbb{E}_{T_i\mathbf{Y}_i}[T_i^r|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}] = \mathbb{E}_{T_i}[T_i^r|\mathbf{Y}_i,\widehat{\boldsymbol{\theta}}^{(k)}],\\
\widehat{t\mathbf{y}_i}^{(k)} &= \mathbb{E}_{T_i\mathbf{Y}_i}[T_i\mathbf{Y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}] = \mathbf{y}_i\mathbb{E}_{T_i}[T_i|\mathbf{Y}_i,\widehat{\boldsymbol{\theta}}^{(k)}],
\end{aligned}$$

with $\mathbf{y}_i^0 = 1$, $\mathbf{y}_i^1 = \mathbf{y}_i$ and $\mathbf{y}_i^2 = \mathbf{y}_i\mathbf{y}_i^\top$ and $\mathbb{E}_{T_i}[T_i^r|\mathbf{Y}_i,\widehat{\boldsymbol{\theta}}^{(k)}] = \mathbb{E}_{T_i}[T_i^r|\mathbf{Y}_i]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(k)}}$ for $r=\{1,2\}$.

2. If the ith subject has only censored components, we have

$$\begin{aligned}
\widehat{\mathbf{y}_i^r}^{(k)} &= \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\mathbf{w}_i^r}^{(k)},\\
\widehat{t_i}^{(k)} &= M^2(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\widehat{\mathbf{w}_i}^{(k)} - \widehat{\boldsymbol{\mu}}^{(k)}) + \widehat{\gamma}_i^{(k)}M(\widehat{\boldsymbol{\theta}}^{(k)}),\\
\widehat{t_i^2}^{(k)} &= M^4(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\widehat{\mathbf{w}_i^2}^{(k)} - 2\widehat{\mathbf{w}_i}^{(k)}\widehat{\boldsymbol{\mu}}^{(k)\top} + \widehat{\boldsymbol{\mu}}^{(k)}\widehat{\boldsymbol{\mu}}^{(k)\top})\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Delta}}^{(k)} + M^2(\widehat{\boldsymbol{\theta}}^{(k)})\\
&\quad + \widehat{\gamma}_i^{(k)}M^3(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\widehat{\mathbf{w}_{0i}}^{(k)} - \widehat{\boldsymbol{\mu}}^{(k)}),\\
\widehat{t\mathbf{y}_i}^{(k)} &= M^2(\widehat{\boldsymbol{\theta}}^{(k)})(\widehat{\mathbf{w}_i}^{2(k)} - \widehat{\mathbf{w}_i}^{(k)}\widehat{\boldsymbol{\mu}}^{(k)\top})\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Delta}}^{(k)} + \widehat{\gamma}_i^{(k)}M(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\mathbf{w}_{0i}}^{(k)},
\end{aligned}$$

where

$$\widehat{\mathbf{w}}_i^{(k)} = \mathbb{E}[\mathbf{W}_i \mid \widehat{\boldsymbol{\theta}}^{(k)}], \quad \widehat{\mathbf{w}}_i^{2(k)} = \mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}] \quad \text{and} \quad \widehat{\mathbf{w}_{0i}}^{(k)} = \mathbb{E}[\mathbf{W}_{0i} \mid \widehat{\boldsymbol{\theta}}^{(k)}],$$

with $\mathbf{W}_i \sim \text{TSN}_p(\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \widehat{\boldsymbol{\lambda}}^{(k)}, [\mathbf{v}_{1i}, \mathbf{v}_{2i}])$, $\mathbf{W}_{0i} \sim \text{TN}_p(\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Gamma}}^{(k)}, [\mathbf{v}_{1i}, \mathbf{v}_{2i}])$ and

$$\widehat{\gamma}_i^{(k)} = \frac{1}{\sqrt{\frac{\pi}{2}\left(1+\widehat{\boldsymbol{\lambda}}^{(k)\top}\widehat{\boldsymbol{\lambda}}^{(k)}\right)}}\frac{L_p(\mathbf{v}_{1i},\mathbf{v}_{2i},\widehat{\boldsymbol{\mu}}^{(k)},\widehat{\boldsymbol{\Gamma}}^{(k)})}{\mathcal{L}_p(\mathbf{v}_{1i},\mathbf{v}_{2i},\widehat{\boldsymbol{\mu}}^{(k)},\widehat{\boldsymbol{\Sigma}}^{(k)},\widehat{\boldsymbol{\lambda}}^{(k)},0)}.$$

3. If the ith subject has both censored and uncensored components and given that $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$, and $(\mathbf{Y}_i^c \,|\, \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$ are equivalent processes, we have

$$\begin{aligned}
\widehat{\mathbf{y}}_i^{(k)} &= \text{E}(\mathbf{Y}_i \,|\, \mathbf{Y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \text{vec}(\mathbf{y}_i^o, \widehat{\mathbf{w}}_i^{c(k)}),\\
\widehat{\mathbf{y}_i^2}^{(k)} &= \text{E}(\mathbf{Y}_i\mathbf{Y}_i^\top \,|\, \mathbf{Y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top} & \mathbf{y}_i^o\widehat{\mathbf{w}}_i^{c(k)\top} \\ \widehat{\mathbf{w}}_i^{c(k)}\mathbf{y}_i^{o\top} & \widehat{\mathbf{w}}_i^{2c(k)} \end{pmatrix},
\end{aligned}$$

$$
\begin{aligned}
\widehat{\mathbf{y}}_{0i}^{(k)} &= \text{vec}(\mathbf{y}_i^o, \widehat{\mathbf{w}}_{0i}^{c(k)}),\\
\widehat{t}_i^{(k)} &= M^2(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\widehat{\mathbf{y}}_i^{(k)} - \widehat{\boldsymbol{\mu}}^{(k)}) + \widehat{\gamma}_i^{(k)} M(\widehat{\boldsymbol{\theta}}^{(k)}),\\
\widehat{t_i^2}^{(k)} &= M^4(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\widehat{\mathbf{y}_i^2}^{(k)} - 2\widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k)\top} + \widehat{\boldsymbol{\mu}}^{(k)}\widehat{\boldsymbol{\mu}}^{(k)\top})\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Delta}}^{(k)} + M^2(\widehat{\boldsymbol{\theta}}^{(k)})\\
&\quad + \widehat{\gamma}_i^{(k)} M^3(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Gamma}}^{-1(k)}(\widehat{\mathbf{y}_{0i}}^{(k)} - \widehat{\boldsymbol{\mu}}^{(k)}),\\
\widehat{t\mathbf{y}_i}^{(k)} &= M^2(\widehat{\boldsymbol{\theta}}^{(k)})(\widehat{\mathbf{y}_i}^{2(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k)\top})\widehat{\boldsymbol{\Gamma}}^{-1(k)}\widehat{\boldsymbol{\Delta}}^{(k)} + \widehat{\gamma}_i^{(k)} M(\widehat{\boldsymbol{\theta}}^{(k)})\widehat{\mathbf{y}_{0i}}^{(k)},
\end{aligned}
$$

where

$$
\widehat{\mathbf{w}}_i^{c(k)} = \mathbb{E}[\mathbf{W}_i^c \mid \widehat{\boldsymbol{\theta}}^{(k)}], \quad \widehat{\mathbf{w}}_i^{2c(k)} = \mathbb{E}[\mathbf{W}_i^c\mathbf{W}_i^{c\top} \mid \widehat{\boldsymbol{\theta}}^{(k)}] \quad \text{and} \quad \widehat{\mathbf{w}_{0i}}^{c(k)} = \mathbb{E}[\mathbf{W}_{0i}^c \mid \widehat{\boldsymbol{\theta}}^{(k)}],
$$

with $\mathbf{W}_i^c \sim \text{TESN}_{p_i^c}\big(\widehat{\boldsymbol{\mu}}_i^{co(k)}, \widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \widehat{\boldsymbol{\lambda}}_i^{co(k)}, \widehat{\tau}_i^{co(k)}, [\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c]\big)$, $\mathbf{W}_{0i}^c \sim \text{TN}_p(\widehat{\mathbf{m}}_i^{co(k)}, \widehat{\boldsymbol{\Gamma}}_i^{cc.o(k)}, [\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c])$ and

$$
\widehat{\gamma}_i^{(k)} = \frac{\eta_i^{co}\, L_p(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \widehat{\mathbf{m}}_i^{co(k)}, \widehat{\boldsymbol{\Gamma}}_i^{cc.o(k)})}{\mathcal{L}_p(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \widehat{\boldsymbol{\lambda}}_i^{co(k)}, \widehat{\tau}_i^{co(k)})},
$$

with $\boldsymbol{\lambda}_i^{co} = \boldsymbol{\Sigma}_i^{cc.o^{1/2}}\boldsymbol{\varphi}_i^c$, $\mathbf{m}_i^{co} = \boldsymbol{\mu}_i^{co} - \boldsymbol{\mu}_{bi}^{co}$, where η_i^{co}, $\boldsymbol{\mu}_{bi}^{co}$ and $\boldsymbol{\Gamma}_i^{cc.o}$ can be computed as expressions η, $\boldsymbol{\mu}_b$ and $\boldsymbol{\Gamma}$ in Proposition 4.6, but using the new set of parameters $\boldsymbol{\mu}_i^{co}$, $\boldsymbol{\Sigma}_i^{cc.o}$, $\boldsymbol{\lambda}_i^{co}$ and τ_i^{co} (instead of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ and τ).

To compute the truncated moments $\mathbb{E}[\mathbf{W}_{0i}]$, $\mathbb{E}[\mathbf{W}_i]$ and $\mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top]$ given in items 2 and 3, we use the `MomTrunc R` package.

4.3.3 Regression setting

Suppose that we have observations on n independent individuals, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{Y}_i \sim SN_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, $i = 1, \ldots, n$. Associated with individual i we assume a known $p \times q$ covariate matrix $\mathbf{X}_i$, which we use to specify the linear predictor $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a q-dimensional vector of unknown regression coefficients. In this case, the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top)^\top$. The E-Step of the EM algorithm updates $\boldsymbol{\beta}$ as follows

$$
\widehat{\boldsymbol{\beta}}^{(k+1)} = (\sum_{i=1}^n \mathbf{X}_i^\top\mathbf{X}_i)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top(\widehat{\mathbf{y}}_i^{(k)} - \widehat{t}_i^{(k)}\boldsymbol{\Delta}^{(k)}), \tag{4.26}
$$

and the necessary quantities of the E and M-steps found in Subsection 4.3.2 remain the same once we plug-in $\widehat{\boldsymbol{\mu}}^{(k)}$ by $\widehat{\boldsymbol{\mu}}_i^{(k)} = \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}$.

4.4 Applications

To exemplify the method developed in this work, we considered all three datasets introduced in chapter 1: (a) Apple data: a bivariate example with missing data, (b) Concentration levels data: an interval-censoring data; and (c) Wine data: a skew normal censored regression.

4.4.1 Apple data: A bivariate example with missing data

First, we apply our methodology to the apple data introduced in subsection 1.2.2. We consider our proposed SN-C model with $p = 2$ dimension to fit the data, that is, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}) \sim \text{SN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. In order to fit the models via the EM-based ML estimation developed in Subsection 4.3.2, we employed different sets of initial values and chose the fitting result with the largest maximized log-likelihood value to be the global maxima. For the sake of comparison, we also fit a multivariate N-C model, which can be treated as a reduced multivariate SN-C model for $\boldsymbol{\lambda} = \mathbf{0}$.

A graphical representation for these fitted models is displayed in Figure 9, with the scatter for the observed data, predicted points using both models, and overlaid contours of the fitted SN and N densities.

Results are summarized on Tables 3 and 4. In Table 3, we can see that the estimates for the skewness parameter $\boldsymbol{\lambda}$ are quite high (due to the small tolerance used for the stopping rule of the algorithm) evidencing a significant departure from symmetry. From Figure 9, note that these high values lead to a truncated effect for the response region. As expected, the SN model outperforms the N model in terms of log-likelihood and AIC. Predicted missing values are shown in Table 4, where we can see that not considering the asymmetry in the model leads to underestimation. Comparing our results with Lin *et al.* (2009), which studies the same dataset without considering any restriction (as censoring) on the missing data, we have that our proposed model presents similar results in terms of log-likelihood and AIC ($\ell(\hat{\boldsymbol{\theta}}|\mathbf{Y}) = -98.47$ and AIC $= 208.94$). It is worth to mention that Lin *et al.* (2009) works with a different version of the SN distribution, the one introduced by Sahu *et al.* (2003).

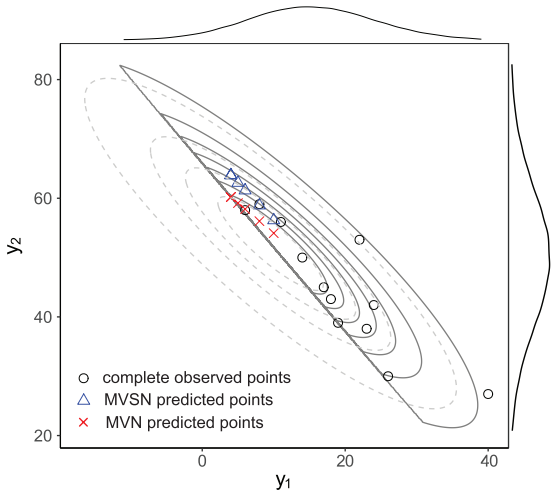


Figure 9 – Apple data. Scatter and predictive plot of the apple data, overlaid on the contours of fitted SN model (solid lines) and N model (dashed lines). ESN marginal densities are shown on borders.

Table 3 – Apple data. Comparison of ML estimates between the two models

Model	μ_1	μ_2	σ_{11}	σ_{12}	σ_{22}	λ_1	λ_2	$\ell(\hat{\boldsymbol{\theta}}\|\mathbf{Y})$	AIC
Normal	14.72	49.33	89.53	−90.69	114.68	-	-	−101.79	213.57
Skew-normal	9.56	52.36	118.31	−111.09	135.70	1163.26	317.91	−98.23	210.45

Table 4 – Apple data. Comparisons of EM predictions of the six missing values

Model	$\hat{y}_{13,2}$	$\hat{y}_{14,2}$	$\hat{y}_{15,2}$	$\hat{y}_{16,2}$	$\hat{y}_{17,2}$	$\hat{y}_{18,2}$
Normal	60.19	60.19	59.18	58.17	56.14	54.12
Skew-normal	63.88	63.88	62.59	61.32	58.79	56.29

4.4.2 Concentration levels data: interval-censoring to fit positive left-censored data

In the second application, we consider the concentration levels data as in Chapter 2. These data were previously analyzed by Hoffman & Johnson (2015), where they proposed a pseudo-likelihood approach for estimating parameters of multivariate normal and log-normal models.

Censored responses in addition to asymmetric behavior of the data, lead us to propose a SN-C model to fit the data, now with dimension $p = 5$, that is, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i5}) \sim \text{SN}_5(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. For the sake of comparison, we also fit a multivariate N-C model as in the previous application.

To guarantee strictly positive concentration levels, we consider an interval-censoring analysis by setting all lower limit of detection equal to 0 for all trace metals. Again, we standardize the dataset to have zero mean and variance equal to one as in Wang *et al.* (2019). The ML estimates of the parameters were obtained using the EM algorithm described in Subsection 4.3.2. The estimated skewness parameter $\hat{\boldsymbol{\lambda}}$ as well as the log-likelihood and AIC are shown in Table 5. Here, we can see that the estimates of $\hat{\boldsymbol{\lambda}}$ are quite different from zero, indicating a lack of adequacy of the symmetry assumption for the VDEQ data. The AIC value is lower for our SN-C model as expected.

Table 5 – VDEQ data. ML estimates for the skewness parameter and model comparison.

Model	λ_1	λ_2	λ_3	λ_4	λ_5	$\ell(\hat{\boldsymbol{\theta}}\|\mathbf{Y})$	AIC
Normal	-	-	-	-	-	−1351.596	2743.192
Skew-normal	5.693	16.442	28.579	−1.382	0.488	−1269.078	2588.157

Figure 10 shows the histograms on diagonal and pair-wise scatter plots for the concentration levels study. From the histograms we can see how censored observations (taking values over the dashed lines) are distributed adequately to the left (blue bins)

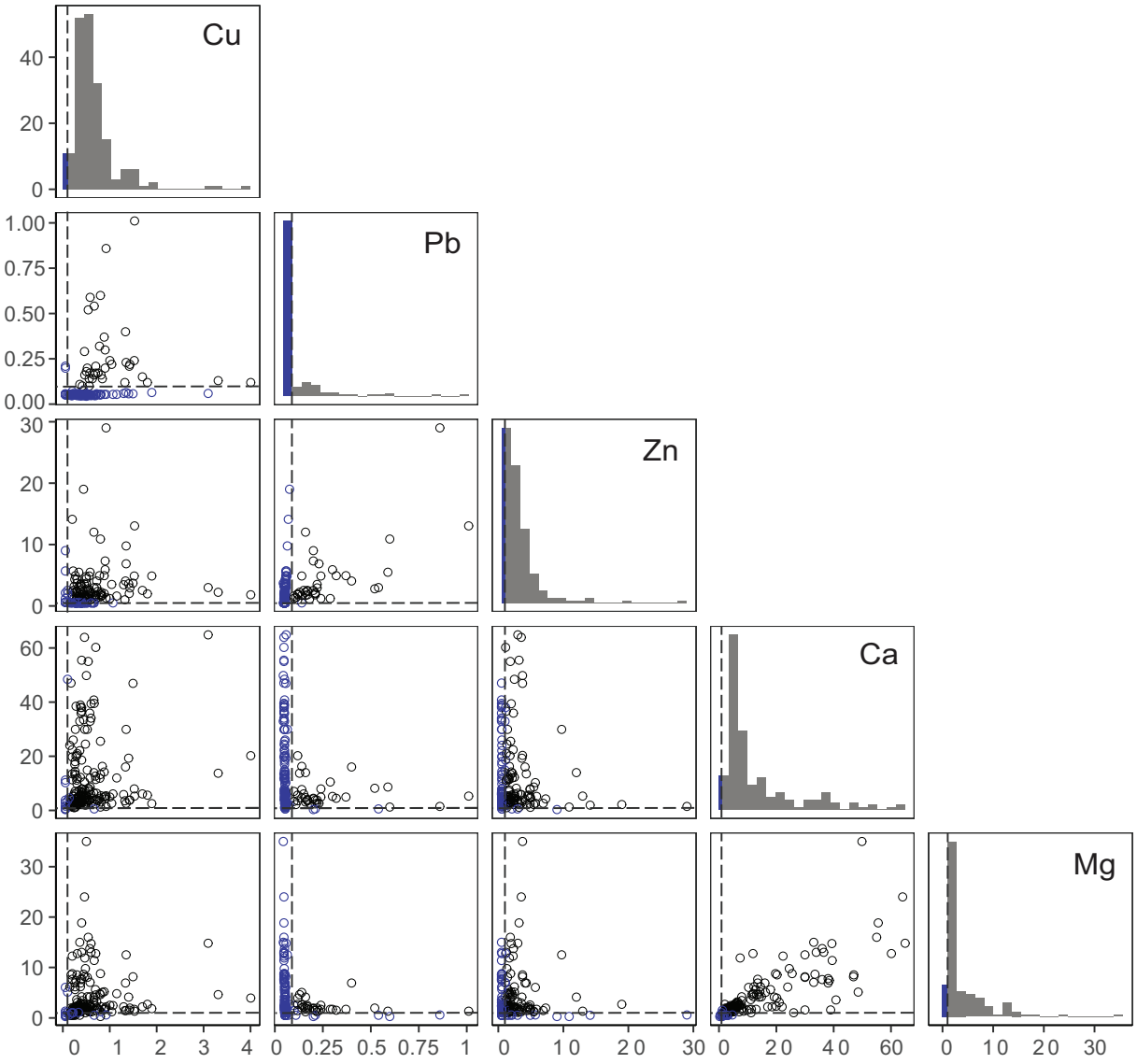
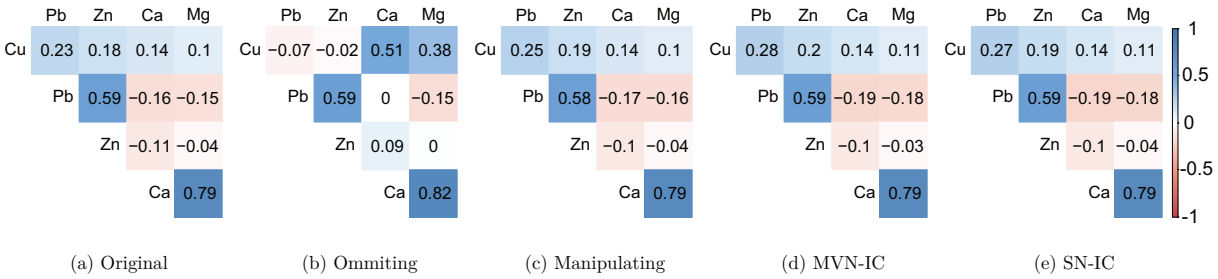


Figure 10 – VDEQ data. Histograms (diagonal) and pair-wise scatter plots (lower-triangle) for the concentration levels study. Complete observed points are represented in black points (gray bins) and SN predicted observations in blue points (bins). Limits of detection are represented in dashed lines.

after fitting our proposed model, while gray bins represent complete observed points. On the other hand, the scatter plots of the show complete observed (black) points and the predicted observations using the multivariate SN-C model (blue points).

Finally, with the aim of validating the proposed censored model approach, we compare the correlation matrices of the data by considering 5 strategies:

a) *Original:* original data
b) *Omitting:* zeros are not considered
c) *Manipulating:* multiplying the limit of detection by the factor 0.75
d) N-C model
e) SN-C model



(a) Original (b) Ommiting (c) Manipulating (d) MVN-IC (e) SN-IC

Figure 11 – VDEQ data. Correlation matrices of the concentration levels for 5 different strategies.

The results are depicted in Figure 11. From this figure we can see that the correlation matrices for the N and SN models are similar. Based on the AIC, we consider the second one as a reference. We can get very decent results for this study by using the original data (a) or even manipulating the data (c), with both tending to underestimate the correlations. Omitting (b) is by far the worst strategy. For example, the correlation between the Pb and Cu is poorly estimated to the point that they have the sign changed. Similar problems arise for the correlations between Zn and other three elements. Given the large number of censored observations, omitting leads to loss of information (as is the case of the correlation between Ca and Pb, as well as between Ca and Mg, where correlation was estimated to be zero).

4.4.3 Wine data: A skew normal censored regression with censored and missing values

For this data, we propose the following simultaneous model:

$$acidity_i = \beta_{10} + \beta_{11} sugar_i + \beta_{12} flavonoids_i + \beta_{13} pH_i + \beta_{14} ODdw_i + \varepsilon_{1i}, \quad (4.27)$$

$$alcohol_i = \beta_{20} + \beta_{21} sugar_i + \beta_{22} flavonoids_i + \beta_{23} pH_i + \beta_{24} ODdw_i + \varepsilon_{2i}, \quad (4.28)$$

where we consider a correlation structure between the *acidity* and *alcohol*, that is, $\text{cov}(\varepsilon_l, \varepsilon_b) \neq 0$, and $\mathbb{E}[\varepsilon_{2i}] = \mathbb{E}[\varepsilon_{2i}] = 0$. The proposed model can be written in a matrix form as

$$\mathbf{y}_i = [\mathbf{I}_2 \otimes \mathbf{x}_i]\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, n, \quad (4.29)$$

where for the object i, $\mathbf{y}_i = (acidity, alcohol)_i^\top$ is a bivariate response of interest, $\mathbf{x}_i = (1, sugar, flavanoids, pH, ODdw)_i^\top$ is a covariate vector, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ is a 10×1 vector, with $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ being a 5×1 vector of regression coefficient for acidity and alcohol, respectively; and finally $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \varepsilon_{2i})^\top$ is the zero mean error term considered to be *i.i.d.* as $\boldsymbol{\varepsilon}_i \sim \text{SN}_2(-\sqrt{2/\pi}\,\boldsymbol{\Delta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, where $\boldsymbol{\Sigma}$ is a 2×2 dispersion matrix, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ is the skewness parameters and $\boldsymbol{\Delta}$ as defined in Proposition 4.1.

Furthermore, $flavonoids$ are complicated compounds responsible for the color and flavor of grapes and consequently of the wine, while $ODdw$ is a measure of protein

content. The units for the variables is not registered in the database, however they seem to be positive quantities.

In order to validate our methodology, we alter the data by creating censoring as well as missing values at random. Hence, we create 15% of right-censored values for *acidity*, 15% of left-censored values for *alcohol* and an additionally 10% of missing values, which were selected randomly along the remaining non-censored points.

This setting led to a total of 24.4% of censored/missing points. Note that, only 57.3% of the measures had 0 missing/censored responses, 36.5% had exactly one missing/censored characteristic and 6.2% observations with no information at all, that is, both responses are missing/censored.

Furthermore, since quantities are strictly positive measures, to guarantee this, we consider this feature by setting the lower limit of detection always greater than 0 for both responses. This lead us to propose a skew-normal censored regression (SN-CR) model defined in (4.29) to fit the data. For the sake of comparison, we also fit a multivariate normal censored regression (N-CR) model and the skew-normal regression (SN-R) model for the original non-disturbed data.

Table 6 – Wine data. Model comparison criteria for fitting the N-CR and SN-CR models in the disturbed data.

Model	λ_1	λ_2	$\ell(\hat{\boldsymbol{\theta}}\|\mathbf{Y})$	AIC	BIC
N-CR	-	-	-726.68	1479.35	1529.73
SN-CR	2.96	-2.07	-718.16	1466.32	1524.44

Table 7 – Wine data. Estimated regression coefficients using the SN-CR model for the original and disturbed data.

	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{20}	β_{21}	β_{22}	β_{23}	β_{24}	$\bar{y}_1$	$\bar{y}_2$
Original	257.14	1.86	-5.97	-58.31	-5.21	10.01	0.13	0.26	-0.03	-0.29	85.64	13.00
Disturbed	259.38	2.17	-6.53	-60.72	-5.32	9.34	0.15	0.22	0.02	-0.23	86.25	12.95

For model selection, we consider the log-likelihood ($\ell(\hat{\boldsymbol{\theta}}|\mathbf{Y})$), Akaike information criterion (AIC, Akaike, 1974) as well as the Bayesian Information Criterion (BIC, Schwarz *et al.*, 1978), displayed in Table 6. From this table, we can see that the estimates of $\boldsymbol{\lambda}$ are not near from zero, indicating a significant skewness and consequently a lack of adequacy of the symmetry assumption for this dataset. All criteria point out to our SN-CR model, as expected.

We also analyzed the parameter estimation when we have the original and the disturbed (missing/censored) data, for both datasets we fitted the SN-CR model. Estimated values for the regression coefficients $\boldsymbol{\beta}$, dispersion matrix $\boldsymbol{\Sigma}$, skewness parameter

$\boldsymbol{\lambda}$ can be found in Table 7, we can see that the estimated values are closer, showing that it is reasonable to accommodate a mechanism for censoring/missing into the model.

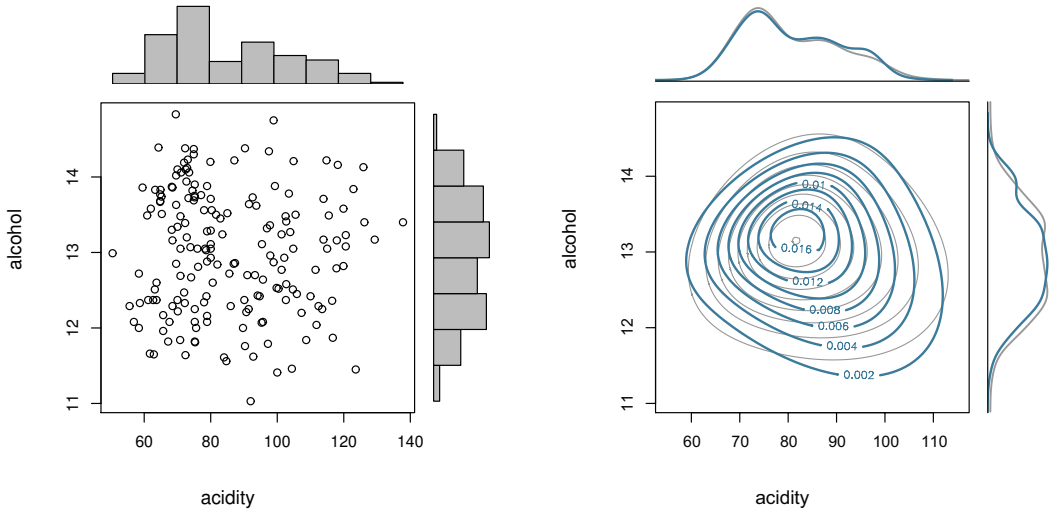


Figure 12 – Wine data. Scatter plots with marginal histograms (left panel) and estimated densities (right panel) for the original data (in black) and disturbed data (in blue) using our proposed SN-CR model.

Note from Figure 12 that the estimated densities for the original dataset (in gray) and the estimated densities using our model (in blue) for the disturbed data are almost indistinguishable, showing that our proposal offers a good performance in prediction, even when the censoring/missing levels are high.

4.5 Conclusions

In this paper, a novel exact EM algorithm for skew-normal censored responses has been developed. Our proposal uses closed-form expressions at the E-step, that rely on formulas of the mean and variance of a multivariate truncated skew-normal distribution. These formulas are available in closed form and have been derived recently (chapter 3). Our approach includes some previously proposed solutions, such as, the skew-normal linear regression model proposed by Lachos *et al.* (2007), the classic Tobit linear models in which the error terms are assumed to follow a Gaussian distribution and the multivariate skew-normal models with incomplete data proposed by Lin *et al.* (2009), among others.

We applied our methods to three real datasets containing missing and censored components, where we demonstrate the superiority of the SN-C model by providing more adequate results when the available data have asymmetric behavior. Furthermore, our results reveal that our method has very competitive performance in terms of imputation when the skew-normal model is imposed. Therefore, it is noteworthy that the use of the SN-C model can offer a better fit and more precise inferences. It is important to remark

that we assumed the dropout/censoring mechanism to be "missing at random" (MAR), (see, Diggle *et al.*, 2002, p 283). However, in the case where MAR with ignorability is not realistic, the relationship between the unobserved measurements and censoring process should be further investigated. The proposed method (including the limiting normal symmetric case) has been coded and implemented in the `R MomTrunc` package, which is available for the users on CRAN repository.

5 Moments of the doubly truncated selection elliptical distributions with emphasis on the unified multivariate skew-t distribution

5.1 Introduction

Truncated moments have been a topic of high interest in the statistical literature, whose possible applications are wide, from simple to complex statistical models as survival analysis, censored data models, and in the most varied areas of applications such as agronomy, insurance, finance, biology, among others. These areas have data whose inherent characteristics lead to the use of methods that involve these truncated moments, such as restricted responses to a certain interval, partial information such as censoring (which may be left, right or interval), among others. The need to have more flexible models that incorporate features such as asymmetry and robustness, has led to the exploration of this area in last years. From the first two one-sided truncated moments for the normal distribution, useful in Tobin's model Tobin (1958), its evolution led to its extension to the multivariate case Tallis (1961), double truncation Manjunath & Wilhelm (2009), heavy tails when considering the Student's t bivariate case in Nadarajah (2007), and finally the first two moments for the multivariate Student's t case in Ho *et al.* (2012). Besides the interval-type truncation in cases before, Arismendi & Broda (2017) considers an interesting non-centered ellipsoid elliptical truncation of the form $\mathbf{a} \leqslant (\mathbf{x}-\boldsymbol{\mu}_{\mathbf{A}})^{\top}\mathbf{A}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{A}})$ on well known distributions as the multivariate normal, Student's t, and generalized hyperbolic distribution. On the other hand, Kan & Robotti (2017) recently proposed a recursive approach that allows calculating arbitrary product moments for the normal multivariate case. Based on the latter, Roozegar *et al.* (2020) proposes the calculation of doubly truncated moments for the normal mean-variance mixture distributions which includes several well-known complex asymmetric multivariate distributions as the generalized hyperbolic distribution.

Unlike Roozegar *et al.* (2020), we focus our efforts to the general class of asymmetric distributions called the selection elliptical family multivariate. This large family of distributions includes complex multivariate asymmetric versions of well-known elliptical distributions as the normal, Student's t, exponential power, hyperbolic, Slash, Pearson type II, contaminated normal, among others. We go further in details for the unified skew-t (SUT) distribution, a complex multivariate asymmetric heavy-tailed distribution which includes the extended skew-t (EST) distribution (Arellano-Valle & Genton, 2010),

the skew-t (ST) distribution (Azzalini & Capitanio, 2003) and naturally its analogous normal cases when $\nu \to \infty$.

The rest of the paper is organized as follows. In Section 5.2 we present some preliminaries results, most of them being definitions of the class of distributions and its special cases of interest along the document. Section 5.3, the addresses the moments for the doubly truncated selection elliptical distributions. We establish formulas for high order moments as well as its first two moments. We present a methodology to deal with some limiting cases of interest and when a non-truncated partition exists, and we establish sufficient and necessary conditions for the existence of these truncated moments. Section 5.4 bases results from Section 5.3 to the SUT case. In Section 5.5, a brief numerical study is presented in order to validate the methodology. In Section 5.6, a direct application of ST truncated moments is developed in the context of risk measurement in Finance. Section 5.7 presents some lemmas and corollaries useful in censored modeling framework. These are given for the SUT distribution and its particular cases EST (ESN) and ST (SN) distributions. Finally, Section 5.8 proposes estimation on interval-censored models for skew-t responses. We last conclude with some comments and future research.

5.2 Preliminaries

5.2.1 Selection distributions

First, we start our exposition defining a selection distribution as in Arellano-Valle *et al.* (2006b).

Definition 5.1 (selection distribution). *Let $\mathbf{X}_1 \in \mathbb{R}^q$ and $\mathbf{X}_2 \in \mathbb{R}^p$ be two random vectors, and denote by C a measurable subset of $\mathbb{R}^q$. We define a selection distribution as the conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1 \in C$, that is, as the distribution of $(\mathbf{X}_2 \mid \mathbf{X}_1 \in C)$. We say that a random vector $\mathbf{Y} \in \mathbb{R}^p$ has a selection distribution if $\mathbf{Y} \stackrel{d}{=} (\mathbf{X}_2 \mid \mathbf{X}_1 \in C)$.*

We use the notation $\mathbf{Y} \sim SLCT_{p,q}$ with parameters depending on the characteristics of $\mathbf{X}_1$, $\mathbf{X}_2$, and C. Furthermore, for $\mathbf{X}_2$ having a probability density function (pdf) $f_{\mathbf{X}_2}$ say, then $\mathbf{Y}$ has a pdf $f_{\mathbf{Y}}$ given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}_2}(\mathbf{y}) \frac{\mathbb{P}(\mathbf{X}_1 \in C \mid \mathbf{X}_2 = \mathbf{y})}{\mathbb{P}(\mathbf{X}_1 \in C)}. \tag{5.1}$$

Since selection distribution depends on the subset $C \in \mathbb{R}^q$, particular cases are obtained. One of the most important case is when the selection subset has the form

$$C(\mathbf{c}) = \{\mathbf{x}_1 \in \mathbb{R}^q \mid \mathbf{x}_1 > \mathbf{c}\}. \tag{5.2}$$

In particular, when $\mathbf{a} = \mathbf{0}$, the distribution of $\mathbf{Y}$ is called to be a simple selection distribution.

In this work, we are mainly interested in the case where $(\mathbf{X}_1, \mathbf{X}_2)$ has a joint density following an arbitrary symmetric multivariate distribution $f_{\mathbf{X}_1,\mathbf{X}_2}$. For $\mathbf{Y} \overset{d}{=} (\mathbf{X}_2 \mid \mathbf{X}_1 \in C)$, this setting leads to a $\mathbf{Y}$ p-variate random vector following a skewed version of f, which its pdf can be computed in a simpler manner as

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\int_C f_{\mathbf{X}_1,\mathbf{X}_2}(\mathbf{x}_1, \mathbf{y})\, \mathrm{d}\mathbf{x}_1}{\int_C f_{\mathbf{X}_1}(\mathbf{x}_1)\, \mathrm{d}\mathbf{x}_1}. \tag{5.3}$$

5.2.2 Selection elliptical (SE) distributions

A quite popular family of selection distributions arises when $\mathbf{X}_1$ and $\mathbf{X}_2$ have a joint multivariate elliptically contoured (EC) distribution, as follows:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim EC_{q+p}\left(\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, h^{(q+p)}\right), \tag{5.4}$$

where $\boldsymbol{\xi}_1 \in \mathbb{R}^q$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^p$ are location vectors, $\boldsymbol{\Omega}_{11} \in \mathbb{R}^{q\times q}$, $\boldsymbol{\Omega}_{22} \in \mathbb{R}^{p\times p}$, and $\boldsymbol{\Omega}_{21} \in \mathbb{R}^{p\times q}$ are dispersion matrices, and, in addition to these parameters, $h^{(q+p)}$ is a density generator function. We denote the selection distribution resulting from (5.4) by $SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$. They typically result in skew-elliptical distributions, except for two cases: $\boldsymbol{\Omega}_{21} = \mathbf{0}_{p\times q}$ and $C = C(\boldsymbol{\xi}_1)$ (for more details, see Arellano-Valle *et al.* (2006b)). Given that the elliptical family of distributions is closed under marginalization and conditioning, the distribution of $\mathbf{X}_2$ and $(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x})$ are also elliptical, where their respective pdfs are given by

$$\mathbf{X}_2 \sim EC_p(\boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, h^{(p)}), \tag{5.5}$$

$$\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x} \sim EC_q(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{x} - \boldsymbol{\xi}_2), \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}, h_{\mathbf{x}}^{(q)}), \tag{5.6}$$

with induced conditional generator

$$h_{\mathbf{x}}^{(q)}(u) = \frac{h^{(q+p)}(u + \delta_2(\mathbf{x}))}{h^{(p)}\delta_2(\mathbf{x})},$$

with $\delta_2(\mathbf{x}) \triangleq (\mathbf{x} - \boldsymbol{\xi}_2)^\top \boldsymbol{\Omega}_{22}^{-1}(\mathbf{x} - \boldsymbol{\xi}_2)$. These last equations imply that the selection elliptical distributions are also closed under marginalization and conditioning. Furthermore, it is well-know that the SE family is closed under linear transformations. For $\mathbf{A} \in \mathbb{R}^{r\times p}$ and $\mathbf{b} \in \mathbb{R}^r$ being a matrix of rank $r \leqslant p$ and a vector, respectively, it holds that the linear transformation $\mathbf{AY} + \mathbf{b} \overset{d}{=} (\mathbf{AX}_2 + \mathbf{b}) \mid (\mathbf{X}_1 > \mathbf{0})$, where $\overset{d}{=}$ is an acronym that stands for identically distributed, and then

$$\mathbf{AY} + \mathbf{b} \sim SLCT\text{-}EC_{r,q}\left(\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \mathbf{A}\boldsymbol{\xi}_2 + \mathbf{b} \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12}\mathbf{A}^\top \\ \mathbf{A}\boldsymbol{\Omega}_{21} & \mathbf{A}\boldsymbol{\Omega}_{22}\mathbf{A}^\top \end{pmatrix}, h^{(q+r)}\right). \tag{5.7}$$

Notice from Equation (5.3), that alternatively we can write

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\int_C f_{q+p}(\mathbf{x}_1, \mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})\, \mathrm{d}\mathbf{x}_1}{\int_C f_q(\mathbf{x}_1; \boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}, h^{(q)})\, \mathrm{d}\mathbf{x}_1}. \tag{5.8}$$

5.2.3 Particular cases for the SE distribution

Some particular cases, useful for our purposes, are detailed next. For further details, we refer to Arellano-Valle *et al.* (2006b).

Unified-skew elliptical (SUE) distribution

Let $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$. $\mathbf{Y}$ is said to follow the unified skew-elliptical distribution introduced by Arellano-Valle & Azzalini (2006b) when the truncation subset $C = C(\mathbf{0})$. From (5.8), it follows that

$$f_{\mathbf{Y}}(\mathbf{y}) = f_p(\mathbf{y}; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, h^{(p)}) \frac{F_q(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{y} - \boldsymbol{\xi}_2); \mathbf{0}, \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}, h_{\mathbf{y}}^{(q)})}{F_q(\boldsymbol{\xi}_1; \boldsymbol{\Omega}_{11}, h^{(q)})}, \tag{5.9}$$

where $f_p(\mathbf{y}; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, h^{(p)}) = |\boldsymbol{\Omega}_{22}|^{-1/2} h^{(p)}(\delta_{\mathbf{X}_2}(\mathbf{y}))$, and $F_q(\mathbf{z}; \mathbf{0}, \boldsymbol{\Theta}, g^{(q)})$ denote the cumulative distribution function (cdf) of the $EC_q(\mathbf{0}, \boldsymbol{\Theta}, g^{(q)})$. Note that the density in (5.9) extends the family of skew elliptical distributions proposed by Branco & Dey (2001) (see also, Azzalini & Capitanio, 2003), which consider $q = 1$ and $\xi_1 = 0$.

Scale-mixture of unified-skew normal (SMSUN) distribution

Let W being a nonnegative random variable with cdf G. For a generator function $h^{(p+q)}(u) = \int_0^\infty (2\pi\zeta(w))^{-(p+q)/2} e^{-u/2\zeta(w)} \mathrm{d}G(w)$, several skewed and thick-tailed distributions can be obtained from different specifications of the weight function $\zeta(\cdot)$ and G. It is said that $\mathbf{Y}$ follows a SMSUN distribution, if its probability density function (pdf) takes the general form

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\xi}_2, \zeta(w)\boldsymbol{\Omega}_{22}) \frac{\Phi_q(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{y} - \boldsymbol{\xi}_2); \zeta(w)\{\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}\})}{\Phi_q(\boldsymbol{\xi}_1; \zeta(w)\boldsymbol{\Omega}_{11})} \mathrm{d}G(w), \tag{5.10}$$

where $\Phi_r(\cdot; \boldsymbol{\Sigma})$ represents the cdf of a r-variate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Here $\mathbf{Y} \mid (W = w)$ follow a unified skew-normal (SUN) distribution, where we write $\mathbf{Y} \mid (W = w) \sim SUN(\boldsymbol{\xi}, \zeta(w)\boldsymbol{\Omega})$.

- Unified skew-normal (SUN) distribution

  Setting W as a degenerated r.v. in 1 ($\mathbb{P}(W = 1) = 1$) and $\zeta(w) = w$, then $h^{(p+q)}(u) = (2\pi)^{-(p+q)/2}e^{-u/2}$, $u \geqslant 0$, for which $h^{(p)}(u) = (2\pi)^{-p/2}e^{-u/2}$. Then, $\mathbf{Y}$ follow a SUN distribution, that is, $\mathbf{Y} \sim SUN_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega})$, with pdf as

$$f_{\mathbf{Y}}(\mathbf{y}) = \phi_p(\mathbf{y}; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}) \frac{\Phi_q(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{y} - \boldsymbol{\xi}_2); \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21})}{\Phi_q(\boldsymbol{\xi}_1; \boldsymbol{\Omega}_{11})}. \tag{5.11}$$

- Unified skew-t (SUT) distribution

  For $W \sim G(\nu/2, \nu/2)$ and weight function $\zeta(w) = 1/w$, we obtain $h^{(p+q)}(u) = \frac{\Gamma((p+q+\nu)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\pi^{(p+q)/2}}\{1 + u\}^{-(p+q+\nu)/2}$ and hence (5.10) becomes

$$f_{\mathbf{Y}}(\mathbf{y}) = t_p(\mathbf{y}; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_2, \nu) \frac{T_q(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{y} - \boldsymbol{\xi}_2); \frac{\nu + \delta_2(\mathbf{y})}{\nu + p}\{\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}\}, \nu + p)}{T_q(\boldsymbol{\xi}_1; \boldsymbol{\Omega}_{11}, \nu)}, \tag{5.12}$$

  where $T_r(\cdot; \boldsymbol{\Sigma}, \nu)$ represents the cdf of a r-variate Student's t distribution with location vector $\mathbf{0}$, scale matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν. For $\mathbf{Y}$ with pdf as in (5.12) is said to follow a SUT distribution, which is denoted by $\mathbf{Y} \sim SUT_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ and was introduced by Arellano-Valle & Azzalini (2006b). It is well-know that (5.12) reduces to a SUN pdf (5.11) as $\nu \to \infty$ and to an unified skew-Cauchy (SUC) distribution, when $\nu = 1$.

  Furthermore, using the following parametrization:

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\mu} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top\boldsymbol{\Lambda} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Sigma} \end{pmatrix}, \tag{5.13}$$

  where $\boldsymbol{\Omega}_{21} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}$, with $\boldsymbol{\Sigma}^{1/2}$ being the square root matrix of $\boldsymbol{\Sigma}$ such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$, we use the notation $\mathbf{Y} \sim SUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi})$, to stand for a p-variate EST distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, positive-definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, shape matrix parameter $\boldsymbol{\lambda} \in \mathbb{R}^{p\times q}$, extension vector parameter $\boldsymbol{\tau} \in \mathbb{R}^q$ and positive-definite correlation matrix $\boldsymbol{\Psi} \in \mathbb{R}^{q\times q}$. The pdf $\mathbf{Y}$ is now simplified to

$$SUT_{p,q}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi}) = t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \frac{T_q\big((\boldsymbol{\tau} + \boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))\nu(\mathbf{y}), \boldsymbol{\Psi}; \nu + p\big)}{T_q(\boldsymbol{\tau}; \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top\boldsymbol{\Lambda}, \nu)}, \tag{5.14}$$

  with $\nu^2(\mathbf{x}) \equiv \nu^2_{\mathbf{X}}(\mathbf{x}) \triangleq (\nu + dim(\mathbf{x}))/(\nu + \delta(\mathbf{x}))$ and $\delta(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^\top\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$ being the Mahalanobis distance. The pdf in (5.14) is equivalent to the one found in Arellano-Valle & Genton (2010), with a different parametrization. Although the unified skew-t distribution above is appealing from a theoretical point of view, the particular case, when $q = 1$, leads to simpler but flexible enough distribution of interest for practical purposes.

Extended skew-t (EST) distribution

For $q = 1$, we have that $\boldsymbol{\Psi} = 1$, $\boldsymbol{\Lambda} = \boldsymbol{\lambda}$ and $T_q(\mathbf{x}; \boldsymbol{\Psi}, \nu) = T_1(x/\sqrt{\psi}, \nu)$, hence (5.14) reduces to the pdf of a EST distribution, denoted by $EST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, that is,

$$EST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau) = t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \frac{T_1\big((\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))\nu(\mathbf{y}); \nu + p\big)}{T_1(\tilde{\tau}; \nu)}. \quad (5.15)$$

with $\tilde{\tau} = \tau/\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}$ .Here, $\boldsymbol{\lambda} \in \mathbb{R}^p$ is a shape parameter which regulates the skewness of $\mathbf{Y}$, and $\tau \in \mathbb{R}$ is a scalar. Location and scale parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ remains as before. Here, we write $\mathbf{Y} \sim EST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ Notice that, $SUT_{p,1} = EST_p$. Besides, it is straightforward to see that

$$EST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu) \longrightarrow t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu), \ \text{as} \ \tau \to \infty,$$

where $t_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ corresponds to the pdf of a multivariate Student's t distribution with location parameter $\boldsymbol{\mu}$, scale parameter $\boldsymbol{\Sigma}$ and degrees of freedom ν. On the other hand, when $\tau = 0$, we retrieve the skew-t distribution $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ say, which density function is given by

$$ST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu) = 2t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\, T_1\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})\, \nu(\mathbf{y}); \nu + p\big), \quad (5.16)$$

that is, $EST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0, \nu) = ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. Other properties are studied in Arellano-Valle & Genton (2010), with a slightly different parametrization.

Six different densities for special cases of the truncated SUT distribution are shown in Figure 13. Symmetrical cases normal and Student's t are shown at first row ($\boldsymbol{\lambda} = \mathbf{0}$), skew cases: skew-normal (SN) and ST at second row ($\tau = 0$) and extended skew cases: extended skew-normal (ESN) and EST at the third row. Location vector $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$ remains fixed for all cases.

- ### Others unified skewed distributions

Others unified members are given by different combinations of the weight function $\zeta(W)$ and the mixture cdf G. For instance, we obtain an *unified skew-slash* distribution when $\zeta(w) = 1/w$ and $W \sim \text{Beta}(\nu, 1)$; an *unified skew-contaminated-normal* distribution when $\zeta(W) = 1/W$ and W is a discrete r.v. with probability mass function (pmf) $g(w; \nu, \gamma) = \nu \mathbb{I}_{\{w=\gamma\}} + (1-\nu)\mathbb{I}_{\{w=1\}}$, with $\mathbb{I}$ being the identity function. Besides, Branco & Dey (2001) mentions some other distributions as the skew-logistic, skew-stable, skew-exponential power, skew-Pearson type II and finite mixture of skew-normal distribution. It is worth mentioning that even though Branco & Dey (2001) works with a subclass of the SMSUN, when $q = 1$ and $\xi_1 = 0$, unified versions of these are readily computed by considering the same respective weight function $\zeta(\cdot)$ and mixture distribution G.

5.3 On moments of the doubly truncated selection elliptical distribution

Let $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$ with pdf as in (5.8) and let also $\mathbb{A}$ be a Borel set in $\mathbb{R}^p$. We say that a random vector $\mathbf{W}$ has a truncated selection elliptical (TSE) distribution on $\mathbb{A}$ when $\mathbf{W} \overset{d}{=} \mathbf{Y}|(\mathbf{Y} \in \mathbb{A})$. In this case, the pdf of $\mathbf{W}$ is given by

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{f_{\mathbf{Y}}(\mathbf{w})}{P(\mathbf{Y} \in \mathbb{A})} \mathbf{1}_{\mathbb{A}}(\mathbf{w}),$$

where $\mathbf{1}_{\mathbb{A}}$ is the indicator function of $\mathbb{A}$. We use the notation $\mathbf{W} \sim TSLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C; \mathbb{A})$. If $\mathbb{A}$ has the form

$$\mathbb{A} = \{(y_1, \ldots, y_p) \in \mathbb{R}^p : \ a_1 \leqslant y_1 \leqslant b_1, \ldots, a_p \leqslant y_p \leqslant b_p\} = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{a} \leqslant \mathbf{y} \leqslant \mathbf{b}\}, \quad (5.17)$$

hence we use the notation $\{\mathbf{Y} \in \mathbb{A}\} = \{\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}\}$, where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$, where a_i and b_i values may be infinite, by convention. Here, we say that the distribution of $\mathbf{W}$ is doubly truncated. Analogously we define $\{\mathbf{Y} \geqslant \mathbf{a}\}$ and $\{\mathbf{Y} \leqslant \mathbf{b}\}$. Thus,

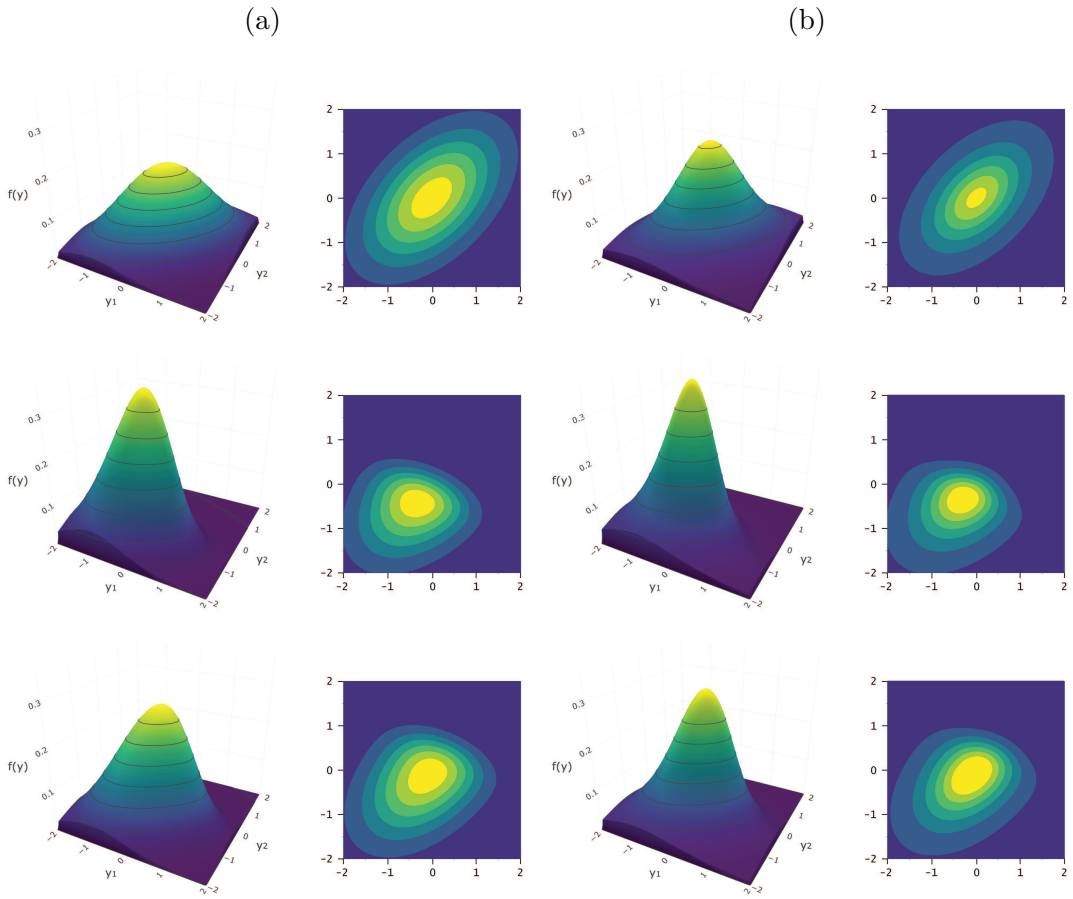


Figure 13 – Densities for particular cases of $\mathbf{Y}$ being a truncated SUT distribution. (a) Normal cases at left column (normal, SN and ESN from top to bottom) and (b) Student's-t cases at right (Student's t, ST and EST from top to bottom).

we say that the distribution of $\mathbf{W}$ is truncated from below and truncated from above, respectively. For convenience, we also use the notation $\mathbf{W} \sim TSLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C; (\mathbf{a}, \mathbf{b}))$ with the last parameter indicating the truncation interval. Analogously, we do denote $TEC_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(p)}; (\mathbf{a}, \mathbf{b}))$ to refer to a p-variate (doubly) truncated elliptical (TE) distribution on $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$. Some characterizations of the doubly TE have been recently discussed in Morán-Vásquez & Ferrari (2019).

5.3.1 Moments of a TSE distribution

For two p-dimensional vectors $\mathbf{y} = (y_1, \ldots, y_p)^\top$ and $\mathbf{k} = (k_1, \ldots, k_p)^\top$, let $\mathbf{y}^{\mathbf{k}}$ stand for $(y_1^{k_1}, \ldots, y_p^{k_p})$, that is, we use a pointwise notation. Next, we present a formulation to compute arbitrary product moments of a TSE distribution.

Theorem 5.1 (moments of a TSE). *Let $\mathbf{X} \sim EC_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})$ as in (5.6.1). Let C be a truncation subset of the form $C(\mathbf{c}, \mathbf{d}) = \{\mathbf{x}_1 \in \mathbb{R}^q \mid \mathbf{c} \leqslant \mathbf{x}_1 \leqslant \mathbf{d}\}$. For $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C(\mathbf{c}, \mathbf{d}))$, it holds that $\mathbb{E}[\mathbf{Y}^{\mathbf{k}}] = \mathbb{E}[Y_1^{k_1} Y_2^{k_2} \ldots Y_p^{k_p}]$ can be computed as*

$$\mathbb{E}[\mathbf{Y}^{\mathbf{k}} \mid \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}], \tag{5.18}$$

with $\boldsymbol{\kappa} = (\mathbf{0}_q^\top, \mathbf{k}^\top)^\top$, $\boldsymbol{\alpha} = (\mathbf{c}^\top, \mathbf{a}^\top)^\top$ and $\boldsymbol{\beta} = (\mathbf{d}^\top, \mathbf{b}^\top)^\top$, where $\mathbf{k} = (k_1, k_2, \ldots, k_p)^\top$, with $k_i \in \mathbb{N}$, for $i = 1, \ldots, p$.

Proof. Since $\mathbf{Y} \overset{d}{=} \mathbf{X}_2 \mid (\mathbf{c} \leqslant \mathbf{X}_1 \leqslant \mathbf{d})$, the proof is direct by noting that

$$\begin{aligned} \mathbf{Y} \mid (\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}) &\overset{d}{=} \mathbf{X}_2 \mid (\mathbf{c} \leqslant \mathbf{X}_1 \leqslant \mathbf{d} \cap \mathbf{a} \leqslant \mathbf{X}_2 \leqslant \mathbf{b}) \\ &\overset{d}{=} \mathbf{X}_2 \mid (\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}). \end{aligned}$$

□

Corollary 5.1 (first two moments of a TSE). *Under the same conditions of Theorem 5.1, let $\mathbf{m} = \mathbb{E}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$ and $\mathbf{M} = \mathbb{E}[\mathbf{X}\mathbf{X}^\top \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$, both partitioned as*

$$\mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \qquad \text{and} \qquad \mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix},$$

respectively. Then, the first two moments of $\mathbf{Y} \mid (\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b})$ are given by

$$\mathbb{E}[\mathbf{Y} \mid \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \mathbf{m}_2, \tag{5.19}$$

$$\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top \mid \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \mathbf{M}_{22}, \tag{5.20}$$

where $\mathbf{m}_2 \in \mathbb{R}^p$ and $\mathbf{M}_{22} \in \mathbb{R}^{p \times p}$.

For the particular truncation subset $C(\mathbf{c})$ as in (5.2), theorem 5.1 and corollary 5.1 holds considering $\boldsymbol{\alpha} = (\mathbf{c}^\top, \mathbf{a}^\top)^\top$ and $\boldsymbol{\beta} = (\infty^\top, \mathbf{b}^\top)^\top$. Notice that, theorem 5.1 and

corollary 5.1 state that we are able to compute any arbitrary moment of $\mathbf{Y} \mid (\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b})$, that is, a TSE distribution just using an unique corresponding moment of a doubly TE distribution $\mathbf{X} \mid (\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta})$.

This is highly convenient since doubly truncated moments for some members of the elliptical family of distributions are already available in the literature and statistical softwares.

5.3.2 Dealing with limiting and extreme cases

Consider $\mathbf{X} \sim EC_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})$ and $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$ as in Theorem 5.1 with truncation subset $C = C(\mathbf{0})$. As $\boldsymbol{\xi}_1 \to \infty$, we have that $\mathbb{P}(\mathbf{X}_1 \geqslant \mathbf{0}) \to 1$. Besides, as $\boldsymbol{\xi}_1 \to -\infty$, we have that $\mathbb{P}(\mathbf{X}_1 \geqslant \mathbf{0}) \to 0$ and consequently $\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}) = \mathbb{P}(\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta})/\mathbb{P}(\mathbf{X}_1 \geqslant \mathbf{0}) \to \infty$. Thus, for $\boldsymbol{\xi}_1$ containing high negative values small enough, sometimes we are not able to compute $\mathbb{E}[\mathbf{Y}^{\mathbf{k}}]$ due to computation precision, mainly when we work with distributions with lighter tails densities. For instance, for a normal univariate case, $\Phi_1(\xi_1) = 0$ for $\xi_1 \leqslant -38$ in `R` software. The next proposition helps us to circumvent this problem.

Proposition 5.1 (limiting case of a SE). *As* $\boldsymbol{\xi}_1 \to -\infty$ $(\xi_{1i} \to -\infty,\ i = 1, \ldots, q)$,

$$SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C(\mathbf{0})) \longrightarrow EC_p(\boldsymbol{\xi}_2 - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}, h_{\mathbf{0}}^{(p)}). \quad (5.21)$$

Proof. Let $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim EC_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})$ and $\mathbf{Y} \sim TSLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C(\mathbf{0}); (\mathbf{a}, \mathbf{b}))$. As $\boldsymbol{\xi}_1 \to -\infty$, we have that $\mathbb{P}(\mathbf{X}_1 \geqslant \mathbf{0}) \to 0$, $\mathbb{E}[\mathbf{X}_1 | \mathbf{X}_1 \geqslant \mathbf{0}] \to \mathbf{0}$ and $\mathrm{Var}[\mathbf{X}_1 | \mathbf{X}_1 \geqslant \mathbf{0}] \to \mathbf{0}$, hence $\mathbf{X}_1 | \mathbf{X}_1 \geqslant \mathbf{0}$ becomes degenerated on $\mathbf{0}$. From Definition 5.1, $\mathbf{Y} \xrightarrow{d} (\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{0})$, and by the conditional distribution in Equation (5.6), it is straightforward to show that $\mathbf{X}_2 | \mathbf{X}_1 \sim EC_p(\boldsymbol{\xi}_2 + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\xi}_1), \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}, h_{\mathbf{X}_1}^{(p)})$. Evaluating $\mathbf{X}_1 = \mathbf{0}$ we achieve (5.21) concluding the proof. □

5.3.3 Approximating the mean and variance-covariance of a TE distribution for extreme cases

While using the relation (5.19) and (5.20), we may face numerical problems trying to compute $\mathbf{m} = \mathbb{E}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$ and $\mathbf{M} = \mathbb{E}[\mathbf{X}\mathbf{X}^\top \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$ for extreme settings of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$. Usually, it occurs when $\mathbb{P}(\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}) \approx 0$ because the probability density is far from the integration region $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. It is worth mentioning that, for these cases, it is not even possible to estimate the moments generating Monte Carlo (MC) samples via rejection sample due to the high rejection ratio when subsetting to a small integration region. Other methods as Gibbs sampling are preferable under this situation.

Hence, we present correction method in order to approximate the mean and the variance-covariance of a multivariate TE distribution even when the numerical precision of the software is a limitation.

5.3.3.1 Dealing with out-of-bounds limits

Consider $\mathbf{X} \sim EC_r\big(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(r)}\big)$ to be partitioned as $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^\top)^\top$ such that $dim(\mathbf{X}_1) = r_1$, $dim(\mathbf{X}_2) = r_2$, where $r_1 + r_2 = r$. Also, consider $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ partitioned as before. Suppose that we are not able to compute $\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} | \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$, because there exists a partition $\mathbf{X}_2$ of $\mathbf{X}$ of dimension r_2 that is out-of-bounds, that is $P(\boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2) \approx 0$. Notice that this happens because $\mathbb{P}(\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}) \leqslant P(\boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2) \approx 0$. Besides, we suppose that $P(\boldsymbol{\alpha}_1 \leqslant \mathbf{X}_1 \leqslant \boldsymbol{\beta}_1) > 0$. Since the limits of $\mathbf{X}_2$ are out-of-bounds (and $\boldsymbol{\alpha}_2 < \boldsymbol{\beta}_2$), we have two possible cases: $\boldsymbol{\beta}_2 \to -\infty$ or $\boldsymbol{\alpha}_2 \to \infty$. For convenience, let $\boldsymbol{\mu}_2 = \mathbb{E}[\mathbf{X}_2 \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2]$ and $\boldsymbol{\Sigma}_{22} = \text{cov}[\mathbf{X}_2 \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2]$. For the first case, as $\boldsymbol{\beta}_2 \to -\infty$, we have that $\boldsymbol{\mu}_2 \to \boldsymbol{\beta}_2$ and $\boldsymbol{\Sigma}_{22} \to \mathbf{0}_{r_2 \times r_2}$. Analogously, we have that $\boldsymbol{\mu}_2 \to \boldsymbol{\alpha}_2$ and $\boldsymbol{\Sigma}_{22} \to \mathbf{0}_{r_2 \times r_2}$ as $\boldsymbol{\alpha}_2 \to \infty$. Hence, $\mathbf{X}_2 \mid (\boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2)$ is degenerated on $\boldsymbol{\mu}_2$ and then $\mathbf{X}_{1.2} \overset{d}{=} \mathbf{X}_1 \mid (\mathbf{X}_2 = \boldsymbol{\mu}_2) \sim EC_{r_1}(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\xi}_2), \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}, h_{\boldsymbol{\mu}_2}^{(r_1)})$. Given that $\text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \mathbf{0}$ and $\text{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2], \mathbf{X}_2] = \mathbf{0}$, it follows that

$$\mathbb{E}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}] = \begin{bmatrix} \boldsymbol{\mu}_{1.2} \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \text{and} \qquad \text{cov}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}] = \begin{bmatrix} \boldsymbol{\Sigma}_{11.2} & \mathbf{0}_{r_1 \times r_2} \\ \mathbf{0}_{r_2 \times r_1} & \mathbf{0}_{r_2 \times r_2} \end{bmatrix}, \tag{5.22}$$

with $\boldsymbol{\mu}_{1.2} = \mathbb{E}[\mathbf{X}_{1.2} \mid \boldsymbol{\alpha}_1 \leqslant \mathbf{X}_{1.2} \leqslant \boldsymbol{\beta}_1]$ and $\boldsymbol{\Sigma}_{11.2} = \text{cov}[\mathbf{X}_{1.2} \mid \boldsymbol{\alpha}_1 \leqslant \mathbf{X}_{1.2} \leqslant \boldsymbol{\beta}_1]$ being the mean and variance-covariance matrix of a r_1-variate TE distribution.

In the event that there are double infinite limits, we can part the vector as well, in order to avoid unnecessary calculation of these integrals.

5.3.3.2 Dealing with a non-truncated partition

Now, consider $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ to be partitioned such that the upper and lower truncation limits associated with $\mathbf{X}_1$ are both infinite, but at least one of the truncation limits associated with $\mathbf{X}_2$ is finite. Then r_1 be the number of pairs in $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ that are both infinite, that is, $dim(\mathbf{X}_1) = r_1$ and $dim(\mathbf{X}_2) = r_2$, by complement. Since $\boldsymbol{\alpha}_1 = -\infty$ and $\boldsymbol{\beta}_1 = \infty$, it follows that $\mathbf{X}_2 \mid (\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}) \sim TEC_{r_2}\big(\boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, h^{(r_2)}; [\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2]\big)$ and $\mathbf{X}_1|\mathbf{X}_2 \sim EC_{r_1}\big(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\xi}_2), \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}, h_{\mathbf{X}_2}^{(r_1)}\big)$. Let $\boldsymbol{\mu}_2 = \mathbb{E}[\mathbf{X}_2 \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2]$ and $\boldsymbol{\Sigma}_{22} = \text{cov}[\mathbf{X}_2 \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2]$. Hence, it follows that $\mathbb{E}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}] = \mathbb{E}[\mathbb{E}[\mathbf{X}_1 \mid \mathbf{X}_2] \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2]$, that is

$$\mathbb{E}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}] = \mathbb{E}\left[\left.\begin{pmatrix} \boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\xi}_2) \\ \mathbf{X}_2 \end{pmatrix}\right| \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2\right]$$

$$= \begin{bmatrix} \boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\xi}_2) \\ \boldsymbol{\mu}_2 \end{bmatrix}. \tag{5.23}$$

On the other hand, we have that $\operatorname{cov}[\mathbf{X}_2, \mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \operatorname{cov}[\mathbf{X}_2, \mathbf{X}_2\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}] = \boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}$, $\operatorname{cov}[\mathbb{E}[\mathbf{X}_1|\mathbf{X}_2]] = \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}$ and $\mathbb{E}[\operatorname{cov}[\mathbf{X}_1|\mathbf{X}_2]] = \omega_{1.2}(\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21})$, with $\omega_{1.2}$ being a constant depending of the conditional generating function $h_{\mathbf{X}_2}^{(r_1)}$. Finally,

$$\operatorname{cov}[\mathbf{X} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}] = \begin{bmatrix} \omega_{1.2}\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\left(\omega_{1.2}\mathbf{I}_{p_2} - \boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{22}^{-1}\right)\boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, \tag{5.24}$$

where $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_{22}$ are the mean vector and variance-covariance matrix of a TE distribution, so we can use (5.19) and (5.20) as well.

Note that $\mathbf{X}_1 \mid (\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}) \nsim EC_{r_1}\left(\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}, h^{(r_1)}\right)$ even though $-\infty \leqslant \mathbf{X}_1 \leqslant \infty$ since $\mathbf{X}_1 \mid (\boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}) = \mathbf{X}_1 \mid (\boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2)$. In general, the marginal distributions of a TE distribution are not TE, however this holds for $\mathbf{X}_2$ due to the particular case $\boldsymbol{\alpha}_1 = -\infty$ and $\boldsymbol{\beta}_1 = \infty$.

Particular cases

Notice that the constant $\omega_{1.2}$ will vary depending of the elliptical distribution we are using. For instance, if $\mathbf{X} \sim t_{r_1+r_2}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ then it follows that $\mathbf{X}_2 \sim t_{r_2}\left(\boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \nu\right)$ and $\mathbf{X}_1|\mathbf{X}_2 \sim t_{r_1}\left(\boldsymbol{\xi}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\xi}_2), (\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21})/\nu^2(\mathbf{X}_2), \nu + r_2\right)$. In this case, it takes the form $\omega_{1.2} = \mathbb{E}[(\nu + r_2)/(\nu + r_2 - 2)\nu^2(\mathbf{X}_2) \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2]$, which is given by

$$\begin{aligned} \omega_{1.2} &= \mathbb{E}\left[\frac{\nu + \delta(\mathbf{X}_2)}{\nu + r_2 - 2} \mid \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2\right], \\ &= \left(\frac{\nu}{\nu - 2}\right) \frac{L_{r_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\xi}_2, \nu\boldsymbol{\Omega}_{22}/(\nu - 2), \nu - 2)}{L_{r_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\xi}_2, \boldsymbol{\Omega}_{22}, \nu)}, \end{aligned} \tag{5.25}$$

where $L_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ denotes the integral

$$L_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = \int_{\boldsymbol{\alpha}}^{\boldsymbol{\beta}} t_r(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)\mathrm{d}\mathbf{y}, \tag{5.26}$$

that is, $L_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \nu) = \mathbb{P}(\boldsymbol{\alpha} \leqslant \mathbf{Y} \leqslant \boldsymbol{\beta})$ for $\mathbf{Y} \sim t_r(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$. Probabilities in (5.25) are involved in the calculation of $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_{22}$ so they are recycled. For the normal case, it is straightforward to see that $\omega_{1.2} = 1$, by taking $\nu \to \infty$.

As can be seen, we can use equations (5.23) and (5.24) to deal with double infinite limits, where the truncated moments are computed only over a r_2-variate partition, avoiding some unnecessary integrals and saving significant computational effort. On the other hand, expression (5.22) let us to approximate the mean and the variance-covariance matrix for cases where the computational precision is a limitation.

5.3.4 Existence of the moments of a TE and TSE distribution

It is well know that for some members of EC family of distributions, their moments do not exist, however, this could be different depending of the truncation limits.

Let $\mathbf{X} \sim EC_r(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(r)})$ be partitioned as in Subsection 5.3.3.2, with r_1 being the number of pairs in $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ that are both finite and $r_2 = r - r_1$. Similarly, $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1^\top, \boldsymbol{\kappa}_2^\top)^\top$ is partitioned as well. If $r_1 = r$, then the truncation limits $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ contains only finite elements, and hence $\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$ exists for all $\boldsymbol{\kappa} \in \mathbb{N}^r$ because the distribution is bounded. When $r_2 \geqslant 1$, there exists at least one pair in $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ containing infinite values, and the expectation may not exist. Given that $\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}] = \mathbb{E}[\mathbf{X}_1^{\boldsymbol{\kappa}_1} \mathbb{E}[\mathbf{X}_2^{\boldsymbol{\kappa}_2} \mid \mathbf{X}_1, \boldsymbol{\alpha}_2 \leqslant \mathbf{X}_2 \leqslant \boldsymbol{\beta}_2] \mid \boldsymbol{\alpha}_1 \leqslant \mathbf{X}_1 \leqslant \boldsymbol{\beta}_1]$, for any measurable function g, $\mathbb{E}[g(\mathbf{X}_1) \mid \boldsymbol{\alpha}_1 \leqslant \mathbf{X}_1 \leqslant \boldsymbol{\beta}_1]$ always exists, and $(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2)$ is not bounded, it is straightforward to see that $\mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$ exist if and only if (*iff*) the inner expectation $\mathbb{E}[\mathbf{X}_2^{\boldsymbol{\kappa}_2} \mid \mathbf{X}_1]$ exists.

As seen, the existence only depends of the order of the moment $\boldsymbol{\kappa}_2$ and the distribution of $\mathbf{X}_2|\mathbf{X}_1$, this last depending on the conditional generating function $h_{\mathbf{X}_1}^{(r_2)}$.

If $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$, with truncation subset of the form $C(\mathbf{c}, \mathbf{d})$ and $r = p+q$ say. It follows from Theorem 5.1, that $\mathbb{E}[\mathbf{Y}^{\mathbf{k}} \mid \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}]$. Hence, the same condition holds taking in account that $\boldsymbol{\kappa} = (\mathbf{0}_q^\top, \mathbf{k}^\top)^\top$, $\boldsymbol{\alpha} = (\mathbf{c}^\top, \mathbf{a}^\top)^\top$ and $\boldsymbol{\beta} = (\mathbf{d}^\top, \mathbf{b}^\top)^\top$. Next, we present a result for a particular case.

5.4 The doubly truncated SUT distribution

For the rest of the paper we shall focus on the computation of the moments of the doubly truncated unified skew-t (TSUT) distribution, denoted by $\mathbf{W} \sim TSUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Sigma}; (\mathbf{a}, \mathbf{b}))$. Besides, we shall study some of its properties and for its particular case (when $q = 1$), the doubly truncated extended skew-t distribution, say $\mathbf{W} \sim TEST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu; (\mathbf{a}, \mathbf{b}))$. For the limiting symmetrical case, we shall use the notation $\mathbf{W} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; (\mathbf{a}, \mathbf{b}))$ to refer to a p-variate truncated Student-t (TT) distribution on $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p$. Finally, $\mathbf{W} \sim TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; (\mathbf{a}, \mathbf{b}))$ will stand for a p-variate truncated normal distribution on the interval $(\mathbf{a}, \mathbf{b})$. Hereinafter we shall omit the expression *doubly* due to we only work with intervalar truncation.

Corollary 5.2 (moments of a TSUT). *If* $\mathbf{Y} \sim SUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi})$, *it follows from Theorem 5.1 that*

$$\mathbb{E}[\mathbf{Y}^{\mathbf{k}} \mid \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}] = \mathbb{E}[\mathbf{X}^{\boldsymbol{\kappa}} \mid \boldsymbol{\alpha} \leqslant \mathbf{X} \leqslant \boldsymbol{\beta}],$$

where $\mathbf{X} \sim t_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ *with* $\boldsymbol{\xi}$ *and* $\boldsymbol{\Omega}$ *as defined in expression* (5.13) *and* $\boldsymbol{\kappa} = (\mathbf{0}_q^\top, \mathbf{k}^\top)^\top$, $\boldsymbol{\alpha} = (\mathbf{0}_q^\top, \mathbf{a}^\top)^\top$ *and* $\boldsymbol{\beta} = (\infty_q^\top, \mathbf{b}^\top)^\top$.

5.4.1 Mean and covariance matrix of the TSUT distribution

Let $\mathbf{Y} \sim TSUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi}; (\mathbf{a}, \mathbf{b}))$ and $\mathbf{X} \sim Tt_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu; (\boldsymbol{\alpha}, \boldsymbol{\beta}))$. From Corollary 5.2, we have that the first two moments of $\mathbf{Y}$ can be computed as

$$\mathbb{E}[\mathbf{Y}] = \mathbf{m}_2, \tag{5.27}$$

$$\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] = \mathbf{M}_{22}, \tag{5.28}$$

where $\mathbf{m} = \mathbb{E}[\mathbf{X}]$ and $\mathbf{M} = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ are partitioned as in Corollary 5.1. Notice that $\text{cov}[\mathbf{Y}] = \mathbb{E}[\mathbf{Y}\mathbf{Y}^\top] - \mathbb{E}[\mathbf{Y}]\mathbb{E}[\mathbf{Y}^\top]$.

Equations (5.27) and (5.28) are convenient for computing $\mathbb{E}[\mathbf{Y}]$ and $\text{cov}[\mathbf{Y}]$ since all boils down to compute the mean and the variance-covariance matrix for a $q+p$-variate TT distribution which can be calculated using the our `MomTrunc R` package available on CRAN.

Existence of the moments of a TSUT

Let also p_1 be the number of pairs in $(\mathbf{a}, \mathbf{b})$ that are both finite. Without loss of generality, we assume $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, where the upper and lower truncation limits associated with $\mathbf{Y}_1$ are both finite, but at least one of the truncation limits associated with $\mathbf{Y}_2$ is not finite, say $dim(\mathbf{Y}_1) = p_1$ and $dim(\mathbf{Y}_2) = p_2$, with $p_1 + p_2 = p$. Consider the partitions of $\mathbf{a} = (\mathbf{a}_1^\top, \mathbf{a}_2^\top)^\top$,$\mathbf{b} = (\mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$ and $\mathbf{k} = (\mathbf{k}_1^\top, \mathbf{k}_2^\top)^\top$ as well. The next proposition gives a sufficient condition for the existence of the moment of a TSUT distribution.

Proposition 5.2 (existence of the moments of a TSUT). *Under the conditions above,* $\mathbb{E}[\mathbf{Y}^{\mathbf{k}} \mid \mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}]$ *exists* iff $sum(\mathbf{k}_2) < \nu + p_1$.

Proof. From subsection 5.3.4, it is suffices to demonstrate that $\mathbb{E}[\mathbf{X}_2^{\boldsymbol{\kappa}_2} | \mathbf{X}_1]$ exists. Since $\boldsymbol{\alpha} = (\mathbf{0}_q^\top, \mathbf{a}_1^\top, \mathbf{a}_2^\top)^\top$ and $\boldsymbol{\beta} = (\infty_q^\top, \mathbf{b}_1^\top, \mathbf{b}_2^\top)^\top$, it follows that $r_1 = p_1$, $r_2 = q + p_2$, $\boldsymbol{\kappa}_1 = \mathbf{k}_1$ and $\boldsymbol{\kappa}_2 = (\mathbf{0}_q^\top, \mathbf{k}_2^\top)^\top$. It is easy to show that the distribution of $\mathbf{X}_2 | \mathbf{X}_1$ is a $(q + p_2)$-variate Student-t distribution with $\nu + p_1$ degrees of freedom. Hence, the above expectation exists *iff* $sum(\mathbf{k}_2) < \nu + p_1$. □

From Proposition 5.2, see that $\mathbb{E}[\mathbf{Y}]$ and $\mathbb{E}[\mathbf{Y}\mathbf{Y}^\top]$ exist *iff* $\nu + p_1 > 1$ and $\nu + p_1 > 2$ respectively. Since $\nu > 0$, this is equivalent to say that, (5.23) *exists if at least one dimension containing a finite limit exists. Besides,* (5.24) *exists if at least two dimensions containing a finite limit exist.*

This sufficient condition for the existence of the first two moments of a truncated SUT distribution holds for the truncated Student-t ($q = 0$) and for the truncated EST distribution ($q = 1$) due to the condition does not depend on q.

Corollary 5.3 (Proposition 1 for a SUT). *As* $\boldsymbol{\tau} \to -\infty$ $(\tau_i \to -\infty,\ i = 1, \ldots, q)$,

$$SUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi}) \longrightarrow t_p(\boldsymbol{\gamma}, \omega_\tau \boldsymbol{\Gamma}, \nu + q), \tag{5.29}$$

with $\boldsymbol{\gamma} = \boldsymbol{\mu} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\tau}$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}$ *and* $\omega_\tau = \nu^2_{\mathbf{x}_1}(\mathbf{0}) = (\nu + \boldsymbol{\tau}^\top \boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\tau})/(\nu + q)$ *with* $\boldsymbol{\Omega}_{11} = \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top\boldsymbol{\Lambda}$.

In particular, for $q = 1$,

$$EST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu) \longrightarrow t_p(\boldsymbol{\gamma}, (\nu + \tilde{\tau}^2)/(\nu + 1)\boldsymbol{\Gamma}, \nu + 1), \tag{5.30}$$

with $\boldsymbol{\gamma} = \boldsymbol{\mu} - \tilde{\tau}\boldsymbol{\Delta}$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, *and* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$.

5.5 Numerical example

In order to illustrate our method, we performed a simple Monte Carlo (MC) simulation study to show how MC estimators for the mean vector and variance-covariance matrix elements converge to the real values computed by our method.

We consider a bivariate TSUT distribution $\mathbf{Y} \sim TSUT_{2,2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi}; (\mathbf{a}, \mathbf{b}))$ with lower and upper truncation limits $\mathbf{a} = (-0.8, -0.6)^\top$ and $\mathbf{b} = (0.5, 0.7)^\top$ respectively, null location vector $\boldsymbol{\mu} = \mathbf{0}$, degrees of freedom $\nu = 4$,

$$\boldsymbol{\tau} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 4 \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} 1 & 3 \\ -3 & -2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Psi} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}.$$

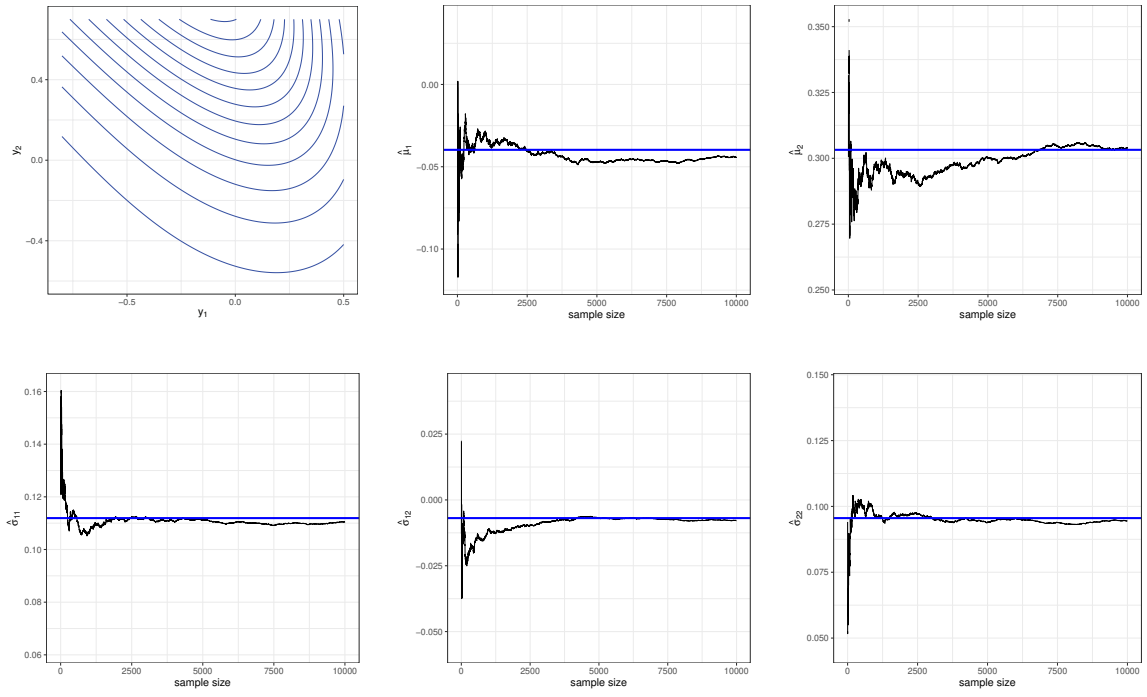


Figure 14 – Contour plot for the TSUT density (upper left corner) and trace plots of the evolution of the MC estimates for the mean and variance-covariance elements of $\mathbf{Y}$. The solid line represent the true estimated value by our proposal.

Figure 14 shows the contour plot for the TSUT density (upper left corner) as well as the evolution trace of the MC estimates for the mean (first row) and variance-covariance (last row) elements μ_1, μ_2, σ_{11}, σ_{12} and σ_{22}. Estimated true values for the mean vector and the variance-covariance matrix were computed using equations (5.27) and (5.28), being

$$\mathbb{E}[\mathbf{Y}] = \begin{pmatrix} -0.039 \\ 0.303 \end{pmatrix} \quad \text{and} \quad \text{cov}[\mathbf{Y}] = \begin{pmatrix} 0.112 & -0.007 \\ -0.007 & 0.096 \end{pmatrix},$$

which are depicted as a blue solid line in Figure 14. Note that even with 1000 MC simulations there exists a significant variation in the chains.

5.6 Application of SE truncated moments on tail conditional expectation

Let Y be a random variable representing in this context, the total loss in a portfolio investment, a credit score, etc. Let y_α be the $(1-\alpha)$th quantile of Y, that is, $\mathbb{P}(Y > y_\alpha) = \alpha$. Hence, the tail conditional expectation (TCE) (see, e.g., Denuit *et al.* (2006)) is denoted by

$$TCE_Y(y_\alpha) = \mathbb{E}[Y \mid Y > y_\alpha]. \tag{5.31}$$

This can be interpreted as the expected value of the $\alpha\%$ worse losses. The quantile y_α is usually chosen to be high in order to be pessimistic, for instance, $\alpha = 0.05$. Notice that, if we consider a variable Y which we are interested on maximizing, for example, the pay-off of a portfolio, we simply compute $TCE_{-Y}(-y_\alpha) = -\mathbb{E}[Y \mid Y \leqslant -y_\alpha]$, being a measure of worst expected income.

Main applications of TCE are in actuarial science and financial economics: market risk, credit risk of a portfolio, insurance, capital requirements for financial institutions, among others. TCE (also known as tail value at risk, TVaR) represents an alternative to the traditional value at risk (VaR) that is more sensitive to the shape of the tail of the loss distribution. Furthermore, if Y is a continuous r.v., TCE coincides with the well-known risk measure expected shortfall (Acerbi & Tasche, 2002). In contrast with VaR, TCE is said to be a coherent measure, holding desirable mathematical properties in the context of risk measurement and and is a convex function of the selection weights (Artzner *et al.*, 1999; Pflug, 2000). A good reference to several risk measures and their properties can be found in Sereda *et al.* (2010).

Multivariate framework

Let consider a set of p assets, business lines, credit scores, $\mathbf{Y} = (Y_1, \cdots, Y_p)^\top$. In the multivariate case, the sum of risks arises as a natural and simple measure of total

risk. Hence, the sum $S = Y_1 + Y_2 + \cdots + Y_p$ follows a univariate distribution and from (5.31), we have that the TCE for S is given by

$$TCE_S(s_\alpha) = \mathbb{E}[S \mid S > s_\alpha]. \tag{5.32}$$

Even though we may know the marginal distribution of S, it is preferable to compute the total risk TCE of S as a decomposed sum, that is

$$\mathbb{E}[S \mid S > s_\alpha] = \sum_{i=1}^{p} \mathbb{E}[Y_i \mid S > s_\alpha], \tag{5.33}$$

where each term $\mathbb{E}[Y_i \mid S > s_\alpha]$ represents the average amount of risk due to Y_i. This decomposed sum offers a way to study the individual impact of the elements of the set, being an improvement to (5.32).

In order to model combinations of correlated risks, Landsman & Valdez (2003) extended the TCE to the multivariate framework. The multivariate TCE (MTCE) is given by

$$MTCE_{\mathbf{Y}}(\mathbf{y}_{\boldsymbol{\alpha}}) = \mathbb{E}[\mathbf{Y} \mid \mathbf{Y} > \mathbf{y}_{\boldsymbol{\alpha}}] = \mathbb{E}[\mathbf{Y} \mid Y_1 > y_{1\alpha_1}, \ldots, Y_p > y_{p\alpha_p}], \tag{5.34}$$

with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ be a vector of quantiles of interest. Notice that the quantile-level for the MTCE is fixed per each risk $i = 1, \ldots, p$, in contrast with the TCE of the sum, which is fixed over all the sum of risk S.

5.6.1 MTCE for selection elliptical distributions

Let consider $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$. Without loss of generality, we consider the selection subset $C = C(\mathbf{0})$. It follows from Theorem 5.1 that

$$MTCE_{\mathbf{Y}}(\mathbf{y}_{\boldsymbol{\alpha}}) = \mathbb{E}[\mathbf{X}_2 \mid \mathbf{X} > \mathbf{x}_{\boldsymbol{\alpha}}], \tag{5.35}$$

where $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim EC_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})$ and $\mathbf{x}_{\boldsymbol{\alpha}} = (\mathbf{0}_q^\top, \mathbf{y}_{\boldsymbol{\alpha}}^\top)^\top$. It is noteworthy that the computation of the MTCE for $\mathbf{Y}$ following a SE distribution relies on the calculation of truncated moments for its symmetrical elliptical multivariate case.

On the other hand, by noticing that $S = \mathbf{1}^\top \mathbf{Y}$, it follows from (5.7) that S is an univariate SE distribution given by $S \sim SLCT\text{-}EC_{1,q}(\boldsymbol{\xi}_s, \boldsymbol{\Omega}_s, h^{(q+1)}, C)$, with

$$\boldsymbol{\xi}_S = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \mathbf{1}^\top \boldsymbol{\xi}_2 \end{pmatrix}, \qquad \text{and} \qquad \boldsymbol{\Omega}_S = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12}\mathbf{1} \\ \mathbf{1}^\top \boldsymbol{\Omega}_{21} & \mathbf{1}^\top \boldsymbol{\Omega}_{22} \mathbf{1} \end{pmatrix}.$$

Hence, its TCE in (5.32) can be easily computed as $\mathbb{E}[S \mid S > s_\alpha] = \mathbb{E}[W_2 \mid \mathbf{W}_1 > \mathbf{0}, W_2 > s_\alpha]$, $\mathbf{W} = (\mathbf{W}_1^\top, W_2)^\top \sim EC_{q+1}(\boldsymbol{\xi}_s, \boldsymbol{\Omega}_s, h^{(1+q)})$, due to $S \overset{d}{=} W_2 \mid (\mathbf{W}_1 > \mathbf{0})$. Next, we establish a general proposition for computing $\mathbb{E}[S \mid S > \alpha_s]$ in matrix form as a decomposed sum.

Proposition 5.3. *Let* $\mathbf{Y} \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C)$*, with* $\boldsymbol{\xi}$ *and* $\boldsymbol{\Omega}$ *as in* (5.4)*, and* $\mathbf{W} = (\mathbf{W}_1^\top, W_2)^\top \sim EC_{q+1}(\boldsymbol{\xi}_S, \boldsymbol{\Omega}_S, h^{(1+q)})$ *as before. It follows that*

$$\mathbb{E}[S \mid S > s_\alpha] = \mathbf{1}^\top \mathbf{s}, \tag{5.36}$$

with $\mathbf{s} = \boldsymbol{\xi}_2 + \boldsymbol{\Omega}_{2S}\boldsymbol{\Omega}_S^{-1}(\boldsymbol{\mathcal{E}}_S - \boldsymbol{\xi}_S)$*, where* $\boldsymbol{\Omega}_{2S} = (\boldsymbol{\Omega}_{21}, \boldsymbol{\Omega}_{22}\mathbf{1})$ *and* $\boldsymbol{\mathcal{E}}_S = \mathbb{E}[\mathbf{W} \mid \mathbf{W}_1 > \mathbf{0}, W_2 > s_\alpha]$.

Proof. Let $\mathbf{A} = (\mathbf{1}, \mathbf{I}_p)^\top$ be a real matrix of dimensions $(p+1) \times p$. For $\mathbf{V} = \mathbf{AY}$, it follows that

$$\mathbf{V} = \begin{pmatrix} V_1 \\ \mathbf{V}_2 \end{pmatrix} \sim SLCT\text{-}EC_{p+1,q}\left(\boldsymbol{\xi}_V = \begin{pmatrix} \boldsymbol{\xi}_S \\ \boldsymbol{\xi}_2 \end{pmatrix}, \boldsymbol{\Omega}_V = \begin{pmatrix} \boldsymbol{\Omega}_S & \boldsymbol{\Omega}_{2S}^\top \\ \boldsymbol{\Omega}_{2S} & \boldsymbol{\Omega}_{22} \end{pmatrix}, h^{(q+1+p)}, C\right),$$

where $\mathbf{V} = (S, \mathbf{Y}^\top)^\top$. It comes from the definition of selection distribution that $\mathbf{V} \overset{d}{=} (X_2, \mathbf{X}_3^\top)^\top | (\mathbf{X}_1 > 0)$, where $\mathbf{X} = (\mathbf{X}_1^\top, X_2, \mathbf{X}_3^\top)^\top$ is a partitioned random vector with elements of dimensions q, 1 and p respectively, where $\mathbf{X} \sim EC_{p+q+1}(\boldsymbol{\xi}_V, \boldsymbol{\Omega}_V; h^{(q+1+p)})$. Hence, it is straightforward to see that

$$\mathbf{s} = \mathbb{E}[\mathbf{Y} \mid S > s_\alpha] = \mathbb{E}[\mathbf{X}_3 \mid \mathbf{X}_1 > \mathbf{0}, X_2 > s_\alpha, -\infty \leqslant \mathbf{X}_3 \leqslant \infty].$$

Since there exists a non-truncated partition, the result in (5.36) immediately follows from equation (5.23), where $\mathbf{W} = (\mathbf{X}_1, X_2)^\top$. □

Observation 5.1. *It is noteworthy that, the* i*th element of vector* $\mathbf{s}$*, say* $s_i = \mathbf{e}_i^\top \mathbf{s}$*, is equal to* $\mathbb{E}[Y_i \mid S > \alpha_s]$*, representing the contribution to the total risk due to the* i*th risk.*

Observation 5.2. *Since* $S \overset{d}{=} W_2 \mid (\mathbf{W}_1 > \mathbf{0})$*, it follows that the last element of the vector* $\boldsymbol{\mathcal{E}}_s$ *is equivalent to* $\mathbb{E}[S \mid S > s_\alpha] = \mathbb{E}[W_2 \mid \mathbf{W}_1 > \mathbf{0}, W_2 > s_\alpha]$.

5.6.2 Application of MTCE using a ST distribution

Suppose that a set of risks $\mathbf{Y}$ are distributed as $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. Let $\mathbf{y}$ represents a realization of $\mathbf{Y}$. Based on $\mathbf{y}$, the set of parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)^\top$ can be estimated through maximum likelihood estimation. It follows that

$$MTCE_{\mathbf{Y}}(\mathbf{y}_{\boldsymbol{\alpha}}) = \mathbb{E}[\mathbf{X}_2 \mid X_1 > 0, \mathbf{X}_2 > \mathbf{y}_{\boldsymbol{\alpha}}], \tag{5.37}$$

where $\mathbf{X} = (X_1, \mathbf{X}_2^\top)^\top \sim t_{1+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$ with

$$\boldsymbol{\xi} = \begin{pmatrix} 0 \\ \boldsymbol{\mu} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega} = \begin{pmatrix} 1 & \boldsymbol{\Delta}^\top \\ \boldsymbol{\Delta} & \boldsymbol{\Sigma} \end{pmatrix}. \tag{5.38}$$

Additionally, using simple algebraic manipulation, it follows from (5.7) that

$$S \sim ST_1\left(\mu_S = \sum_{i=1}^{p} \mu_i,\ \sigma_S^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} \sigma_{ij},\ \lambda_S = \frac{\Delta_S}{\sqrt{\sigma_S^2 - \Delta_S^2}},\ \nu\right), \tag{5.39}$$

with $\Delta_S = \sum_{i=1}^{p} \Delta_i$. Besides, the TCE of the sum is given by $TCE_S(s_\alpha) = \mathbb{E}[W_2 \mid W_1 > 0, W_2 > s_\alpha]$, $\mathbf{W} = (W_1^\top, W_2)^\top \sim t_2(\boldsymbol{\xi}_S, \boldsymbol{\Omega}_S, \nu)$, where

$$\boldsymbol{\xi}_S = \begin{pmatrix} 0 \\ \mu_S \end{pmatrix}, \qquad \text{and} \qquad \boldsymbol{\Omega}_S = \begin{pmatrix} 1 & \Delta_S \\ \Delta_S & \sigma_S^2 \end{pmatrix}.$$

We have from Proposition 5.3 that

$$\begin{aligned} \mathbb{E}[Y_i \mid S > \alpha_s], &= \mathbf{e}_i^\top \left[\boldsymbol{\mu} + (\boldsymbol{\Delta}, \boldsymbol{\Sigma}\mathbf{1})\, \boldsymbol{\Omega}_S^{-1}(\boldsymbol{\mathcal{E}}_S - \boldsymbol{\xi}_S)\right], \\ &= \mu_i + \mathcal{E}_{S1}(\Delta_i \sigma_S^2 + \sigma_{iS}\Delta_S) - (TCE_S(s_\alpha) - \mu_S)(\Delta_i \Delta_S + \sigma_{iS}), \end{aligned} \tag{5.40}$$

with $\mathcal{E}_{S1} = \mathbb{E}[W_1 \mid W_1 > 0, W_2 > s_\alpha]$ and $\sigma_{iS} = \sum_{j=1}^{p} \sigma_{ij}$. Besides, summing (5.40) over $i = 1, \ldots, p$, and after some straightforward algebra we obtain that $TCE_S(s_\alpha) = \mathbb{E}[S \mid S > s_\alpha]$ can be written as

$$\begin{aligned} TCE_S(s_\alpha) &= \mu_S + \mathcal{E}_{S1} \sum_{i=1}^{p} \left\{\Delta_i \sigma_S^2 + \sigma_{iS}\Delta_S\right\} - (TCE_S(s_\alpha) - \mu_S) \sum_{i=1}^{p} \left\{\Delta_i \Delta_S + \sigma_{iS}\right\} \\ &= \mu_S + \frac{2\Delta_S \sigma_S^2}{1 + \Delta_S^2 + \sigma_S^2} \mathcal{E}_{S1}. \end{aligned} \tag{5.41}$$

Finally, plugging (5.41) in (5.40), we obtain a explicit expression for $\mathbb{E}[Y_i \mid S > \alpha_s]$ that does not depends on $TCE_S(s_\alpha)$, that is

$$\mathbb{E}[Y_i \mid S > \alpha_s] = = \mu_i + \left(\Delta_i \sigma_S^2 + \sigma_{iS}\Delta_S - \frac{2\Delta_S \sigma_S^2}{1 + \Delta_S^2 + \sigma_S^2}(\Delta_i \Delta_S + \sigma_{iS})\right) \mathcal{E}_{S1}. \tag{5.42}$$

5.7 Additional results related to interval censored mechanism

Under interval censoring mechanism the implementation of inferences depends on the computation of certain marginal and conditional expectations (Matos *et al.*, 2013). For instance, for $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim \phi_{1+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \nu)$, as in (5.13), with $\boldsymbol{\Psi} = 1$, $\boldsymbol{\Lambda} = \boldsymbol{\lambda}$ and $\boldsymbol{\tau} = 0$, it holds that $f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y}) = \phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})\big)$. Then,

$$\mathbb{E}\left[g(\mathbf{Y}) \frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}\right] = \mathbb{E}\left[g(\mathbf{Y}) \frac{\phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})\big)}{\Phi\big(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})\big)}\right], \tag{5.43}$$

where $g(\cdot)$ is a measurable function. The expectation in the right side of the expression (5.43) is highly used to perform inferences under SN censored models from a likelihood-based perspective, such as the E-Step of the EM-algorithm (Dempster *et al.*, 1977).

Next, we derive general expressions that are involved in interval censored modeling, specifically, in the E-step of the EM algorithm. These expressions arise, when we consider the responses $\mathbf{Y}_i$, $i = 1, \ldots, n$, to be i.i.d. realizations from a selection elliptical distribution or any of its particular cases. For instance, a SUT, EST or ST distribution or any normal limiting case as the SUN, ESN or SN distribution as the example in (5.43).

Lemma 5.1. *Let* $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top \sim EC_{q+p}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)})$ *and* $\mathbf{Y} \sim TSLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C; (\mathbf{a}, \mathbf{b}))$ *with truncation subset* $C = C(\mathbf{0})$*. For any measurable function* $g(\mathbf{y}) : \mathbb{R}^p \to \mathbb{R}$*, we have that*

$$\mathbb{E}\left[g(\mathbf{Y})\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}\right] = \frac{\mathbb{P}(\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})}\frac{\mathbb{E}[g(\mathbf{W})]}{\mathbb{P}(\mathbf{X}_1 \geqslant \mathbf{0})} f_{\mathbf{X}_1}(\mathbf{0}), \tag{5.44}$$

where $\mathbf{X}_1 \sim EC_p(\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}, h^{(q)})$*,* $\mathbf{Y}_0 \sim SLCT\text{-}EC_{p,q}(\boldsymbol{\xi}, \boldsymbol{\Omega}, h^{(q+p)}, C(\mathbf{0}))$*,* $\mathbf{W}_0 \sim EC_p(\boldsymbol{\xi}_2 - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{21}, h_{\mathbf{0}}^{(p)})$ *and* $\mathbf{W} \overset{d}{=} \mathbf{W}_0 \mid (\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})$*.*

Proof. Using basic probability theory, we have

$$\begin{aligned}
&= \mathbb{E}\left[g(\mathbf{Y})\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}\right] \\
&= \frac{1}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})}\int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{y})\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})} f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y}, \\
&= \frac{1}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})}\int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{y})\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}\frac{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{y}) f_{\mathbf{X}_2}(\mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0})}\mathrm{d}\mathbf{y}, \\
&= \frac{1}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})}\int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{y})\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{y}) f_{\mathbf{X}_2}(\mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0})}\mathrm{d}\mathbf{y}, \\
&= \frac{1}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})}\frac{f_{\mathbf{X}_1}(\mathbf{0})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0})}\int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{y}) f_{\mathbf{X}_2}(\mathbf{y} \mid \mathbf{X}_1 = \mathbf{0})\,\mathrm{d}\mathbf{y}, \\
&= \frac{\mathbb{P}(\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})}\frac{\mathbb{E}[g(\mathbf{W})]}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0})} f_{\mathbf{X}_1}(\mathbf{0}),
\end{aligned}$$

where $\mathbf{W}_0 \overset{d}{=} \mathbf{X}_2 | (\mathbf{X}_1 = \mathbf{0})$ and $\mathbf{W} \overset{d}{=} \mathbf{W}_0 \mid (\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})$. □

Lemma 5.2. *Consider* $\mathbf{X}$*,* $\mathbf{Y}$ *and* g *as in Lemma 5.1. Now, consider* $\mathbf{Y}$ *to be partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 *(*$p_1 + p_2 = p$*). For a given random variable* $\mathbf{U}$*, let* $\mathbf{U}^*$ *stands for* $\mathbf{U}^* \overset{d}{=} \mathbf{U} \mid \mathbf{Y}_1$*. It follows that*

$$\mathbb{E}\left[g(\mathbf{Y}_2)\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}\middle| \mathbf{Y}_1\right] = \frac{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{W}_0^* \leqslant \mathbf{b}_2)}{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{Y}_0^* \leqslant \mathbf{b}_2)}\frac{\mathbb{E}[g(\mathbf{W}_2)]}{\mathbb{P}(\mathbf{X}_1^* > \mathbf{0})} f_{\mathbf{X}_1^*}(\mathbf{0}) \tag{5.45}$$

with $\mathbf{X}_1$*,* $\mathbf{Y}_0$*, and* $\mathbf{W}_0$ *as defined in Lemma 5.1, and* $\mathbf{W}_2 \overset{d}{=} \mathbf{W}_0^* \mid (\mathbf{a}_2 \leqslant \mathbf{W}_0^* \leqslant \mathbf{b}_2)$*.*

Proof. Consider $\mathbf{X}_2$ partitioned as $\mathbf{X}_2 = (\mathbf{X}_{21}^\top, \mathbf{X}_{22}^\top)^\top$ such that $dim(\mathbf{X}_{21}) = dim(\mathbf{Y}_1)$ and $dim(\mathbf{X}_{22}) = dim(\mathbf{Y}_2)$. Since $f_{\mathbf{Y}_2}(\mathbf{y}_2|\mathbf{Y}_1 = \mathbf{y}_1) = f_{\mathbf{Y}}(\mathbf{y})/f_{\mathbf{Y}_1}(\mathbf{y}_1)$, it follows (in a similar manner that the proof of Lemma 5.1) that

$$\begin{aligned}
&= \mathbb{E}\left[g(\mathbf{Y}_2)\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{Y})}\middle| \mathbf{Y}_1\right] \\
&= \frac{1}{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{Y}_0^* \leqslant \mathbf{b}_2)}\int_{\mathbf{a}_2}^{\mathbf{b}_2} g(\mathbf{y}_2)\frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}\frac{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_{21} = \mathbf{y}_1)}\frac{f_{\mathbf{X}_2}(\mathbf{y})}{f_{\mathbf{X}_{21}}(\mathbf{y}_1)}\mathrm{d}\mathbf{y}_2,
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{Y}_0^* \leqslant \mathbf{b}_2)} \int_{\mathbf{a}_2}^{\mathbf{b}_2} g(\mathbf{y}_2) \frac{f_{\mathbf{X}_1}(\mathbf{0} \mid \mathbf{X}_2 = \mathbf{y})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_{21} = \mathbf{y}_1)} \frac{f_{\mathbf{X}_2}(\mathbf{y})}{f_{\mathbf{X}_{21}}(\mathbf{y}_1)} \mathrm{d}\mathbf{y}_2, \\
&= \frac{1}{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{Y}_0^* \leqslant \mathbf{b}_2)} \frac{f_{\mathbf{X}_1}(\mathbf{0})}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_{21} = \mathbf{y}_1)} \int_{\mathbf{a}_2}^{\mathbf{b}_2} g(\mathbf{y}_2) \frac{f_{\mathbf{X}_2}(\mathbf{y} \mid \mathbf{X}_1 = \mathbf{0})}{f_{\mathbf{X}_{21}}(\mathbf{y}_1)} \, \mathrm{d}\mathbf{y}_2, \\
&= \frac{1}{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{Y}_0^* \leqslant \mathbf{b}_2)} \frac{f_{\mathbf{X}_1}(\mathbf{0} | \mathbf{X}_{21} = \mathbf{y}_1)}{\mathbb{P}(\mathbf{X}_1 > \mathbf{0} \mid \mathbf{X}_{21} = \mathbf{y}_1)} \int_{\mathbf{a}_2}^{\mathbf{b}_2} g(\mathbf{y}_2) f_{\mathbf{X}_{22}}(\mathbf{y}_2 \mid \mathbf{X}_{21} = \mathbf{y}_1, \mathbf{X}_1 = \mathbf{0}) \, \mathrm{d}\mathbf{y}_2, \\
&= \frac{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{W}_0^* \leqslant \mathbf{b}_2)}{\mathbb{P}(\mathbf{a}_2 \leqslant \mathbf{Y}_0^* \leqslant \mathbf{b}_2)} \frac{\mathbb{E}[g(\mathbf{W}_2)]}{\mathbb{P}(\mathbf{X}_1^* > \mathbf{0})} f_{\mathbf{X}_1^*}(\mathbf{0}),
\end{aligned}
$$

where $\mathbf{W}_0^* \overset{d}{=} \mathbf{X}_{22} | (\mathbf{X}_{21} = \mathbf{y}_1, \mathbf{X}_1 = \mathbf{0})$ and $\mathbf{W}_2 \overset{d}{=} \mathbf{W}_0^* \mid (\mathbf{a}_2 \leqslant \mathbf{W}_0^* \leqslant \mathbf{b}_2)$. □

In the next corollaries we particularize the aforementioned lemmas to the truncated SUT, EST, SUN and ESN distributions.

Corollary 5.4 (Lemma 5.1 for a SUT). *Let* $\mathbf{Y} \sim TSUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi}, (\mathbf{a}, \mathbf{b}))$. *For any measurable function* $g(\mathbf{y}) : \mathbb{R}^p \to \mathbb{R}$, *we have that*

$$
\mathbb{E}\left[g(\mathbf{Y}) \frac{t_q\big((\boldsymbol{\tau} + \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}))\, \nu(\mathbf{Y}), \boldsymbol{\Psi}; \nu + p\big)}{T_q\big((\boldsymbol{\tau} + \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}))\, \nu(\mathbf{Y}), \boldsymbol{\Psi}; \nu + p\big)} \right] = \frac{\mathbb{P}(\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})} \mathbb{E}[g(\mathbf{W})] \boldsymbol{\eta}, \tag{5.46}
$$

where $\boldsymbol{\eta} = t_q(\boldsymbol{\tau}; \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}, \nu) / T_q(\boldsymbol{\tau}; \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}, \nu)$, $\mathbf{Y}_0 \sim SUT_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \nu, \boldsymbol{\Psi})$, $\mathbf{W}_0 \sim t_p(\boldsymbol{\gamma}, \omega_\tau \boldsymbol{\Gamma}, \nu + q)$ *and* $\mathbf{W} \overset{d}{=} \mathbf{W}_0 \mid (\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})$. *When* $\boldsymbol{\tau} = \mathbf{0}$, *we have that* $\boldsymbol{\eta} = 2\, t_q(\boldsymbol{\tau}; \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}, \nu)$ *and* $\mathbf{W}_0 \sim t_p(\boldsymbol{\mu}, \nu \boldsymbol{\Gamma}/(\nu + q), \nu + q)$.

In particular for $q = 1$, $\mathbf{Y} \sim TEST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu; (\mathbf{a}, \mathbf{b}))$, *and*

$$
\mathbb{E}\left[g(\mathbf{Y}) \frac{t_1\big((\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}))\, \nu(\mathbf{Y}); \nu + p\big)}{T_1\big((\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}))\, \nu(\mathbf{Y}); \nu + p\big)} \right] = \frac{\mathbb{P}(\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})} \eta\, \mathbb{E}[g(\mathbf{W})], \tag{5.47}
$$

with $\eta = t_1(\tau; 1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}, \nu) / T_1(\tilde{\tau}; \nu)$, $\mathbf{Y}_0 \sim EST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\tau}, \nu)$, $\mathbf{W}_0 \sim t_p(\boldsymbol{\gamma}, (\nu + \tilde{\tau}^2)\boldsymbol{\Gamma}/(\nu + 1), \nu + 1)$, *and* $\mathbf{W} \overset{d}{=} \mathbf{W}_0 \mid (\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})$. *Similarly, when* $\tau = 0$, *we have that* $\eta = 2\, t_1(0; 1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}, \nu)$ *and* $\mathbf{W}_0 \sim t_p(\boldsymbol{\mu}, \nu \boldsymbol{\Gamma}/(\nu + 1), \nu + 1)$.

Corollary 5.5 (Lemma 5.1 for a SUN). *Taking* $\nu \to \infty$, $\mathbf{Y} \sim TSUN_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \boldsymbol{\Psi}, (\mathbf{a}, \mathbf{b}))$ *and hence from Lemma 5.1 it follows that*

$$
\mathbb{E}\left[g(\mathbf{Y}) \frac{\phi_q\big(\boldsymbol{\tau} + \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}), \boldsymbol{\Psi}\big)}{\Phi_q\big(\boldsymbol{\tau} + \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}), \boldsymbol{\Psi}\big)} \right] = \frac{\mathbb{P}(\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})} \mathbb{E}[g(\mathbf{W})] \boldsymbol{\eta}, \tag{5.48}
$$

where $\boldsymbol{\eta} = \phi_q(\boldsymbol{\tau}; \mathbf{0}, \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}) / \Phi_q(\boldsymbol{\tau}; \mathbf{0}, \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})$, $\mathbf{Y}_0 \sim SUN_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \boldsymbol{\Psi})$, $\mathbf{W}_0 \sim N_p(\boldsymbol{\gamma}, \boldsymbol{\Gamma})$, *and* $\mathbf{W} \overset{d}{=} \mathbf{W}_0 \mid (\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})$. *When* $\boldsymbol{\tau} = \mathbf{0}$, *we have that* $\boldsymbol{\eta} = 2\, \phi_q(\mathbf{0}; \boldsymbol{\Psi} + \boldsymbol{\Lambda}^\top \boldsymbol{\Lambda})$ *and* $\mathbf{W}_0 \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Gamma})$.

In particular for $q = 1$, $\mathbf{Y} \sim TESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; (\mathbf{a}, \mathbf{b}))$, *and*

$$
\mathbb{E}\left[g(\mathbf{Y}) \frac{\phi\big(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})\big)}{\Phi\big(\tau + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})\big)} \right] = \frac{\mathbb{P}(\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})}{\mathbb{P}(\mathbf{a} \leqslant \mathbf{Y}_0 \leqslant \mathbf{b})} \eta\, \mathbb{E}[g(\mathbf{W})], \tag{5.49}
$$

with $\eta = \phi(\tau; 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})/\Phi(\tilde{\tau})$, $\mathbf{Y}_0 \sim ESN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \boldsymbol{\tau})$, $\mathbf{W}_0 \sim N_p(\boldsymbol{\gamma}, \boldsymbol{\Gamma})$, *and* $\mathbf{W} \overset{d}{=} \mathbf{W}_0 \mid (\mathbf{a} \leqslant \mathbf{W}_0 \leqslant \mathbf{b})$. *Similarly, when* $\tau = 0$, *we have that* $\eta = \sqrt{2/\pi(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})}$ *and* $\mathbf{W}_0 \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Gamma})$.

5.8 Multivariate ST censored responses

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a $p \times 1$ response vector for the ith sample unit, for $i \in \{1, \ldots, n\}$, and consider the set of random samples (independent and identically distributed):

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim \mathrm{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu), \tag{5.50}$$

with location vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$, dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ depending on an unknown and reduced parameter vector $\boldsymbol{\alpha}$, skewness parameter $\boldsymbol{\lambda}$ and degrees of freedom ν. However, the response vector $\mathbf{Y}_i$ may not be fully observed due to censoring, so we define $(\mathbf{V}_i, \mathbf{C}_i)$ the observed data for the ith sample, where $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$ with elements being either an uncensored observation ($V_{ik} = V_{0i}$) or the interval censoring level ($V_{ik} \in [V_{1ik}, V_{2ik}]$), and $\mathbf{C}_i = (C_{i1}, \ldots, C_{ip})^\top$ is the vector of censoring indicators, satisfying

$$C_{ik} = \begin{cases} 1 & \text{if } V_{1ik} \leqslant Y_{ik} \leqslant V_{2ik}\,, \\ 0 & \text{if } Y_{ik} = V_{0i}\,, \end{cases} \tag{5.51}$$

for all $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, p\}$, i.e., $C_{ik} = 1$ if Y_{ik} is located within a specific interval. In this case, (5.50) along with (5.51) defines the multivariate skew-t interval censored model (hereafter, the ST-C model). For instance, left censoring structure causes truncation from the lower limit of the support of the distribution, since we only know that the true observation Y_{ik} is greater than or equal to the observed quantity V_{1ik}. Moreover, missing observations can be handled by considering $V_{1ik} = -\infty$ and $V_{2ik} = +\infty$.

5.8.1 The likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$ is a realization of $\mathbf{Y}_i \sim \mathrm{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. In order to obtain the likelihood function of the ST-C model, first we treat, separately, the observed and censored components of $\mathbf{y}_i$, i.e., $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $C_{ik} = 0$ for all elements in the p_i^o-dimensional vector $\mathbf{y}_i^o$, and $C_{ik} = 1$ for all elements in the p_i^c-dimensional vector $\mathbf{y}_i^c$. On according to that, we write $\mathbf{V}_i = \mathrm{vec}(\mathbf{V}_i^o, \mathbf{V}_i^c)$, where $\mathbf{V}_i^c = (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c)$ with

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

See that, we must rely on the marginal and conditional distribution of a ST variate. Next, we propose a general result for the EST variate in a similar manner than Arellano-Valle & Genton (2010).

Proposition 5.4 (*Marginal and conditional distribution of the EST*). *Let* $\mathbf{Y} \sim EST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$ *and* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 *($p_1 + p_2 = p$), respectively. Let*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \quad \text{and} \quad \boldsymbol{\varphi} = (\boldsymbol{\varphi}_1^\top, \boldsymbol{\varphi}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$ *and* $\boldsymbol{\varphi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. *Then,*

$$\mathbf{Y}_1 \sim EST_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \tau_1, \nu),$$
$$\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1 \sim EST_{p_2}(\boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu + p_1)$$

with $\tilde{\boldsymbol{\lambda}}_1 = c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1$,$\tau_1 = c_{12}\tau$, $\boldsymbol{\lambda}_{2.1} = \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2$, $\tau_{2.1} = \nu(\mathbf{y}_1)(\tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))$ *where* $c_{12} = (1 + \boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\tilde{\boldsymbol{\Sigma}}_{22.1} = \boldsymbol{\Sigma}_{22.1}/\nu^2(\mathbf{y}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$ *and* $\nu^2(\mathbf{y}_1) = (\nu + p_1)/(\nu + \delta(\mathbf{y}_1))$.

Proof. See Appendix section C. □

Then, from Proposition 5.4, we have that $\mathbf{Y}_i^o \sim \mathrm{ST}_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \tilde{\boldsymbol{\lambda}}_i^o, \nu)$ and $\mathbf{Y}_i^c \mid (\mathbf{Y}_i^o = \mathbf{y}_i^o) \sim \mathrm{EST}_{p_i^c}(\boldsymbol{\mu}_i^{co}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \tau_i^{co}, \nu_i^{co})$, with $\tilde{\boldsymbol{\lambda}}_i^o = c_i^{oc}\boldsymbol{\Sigma}_i^{oo\,1/2}\tilde{\boldsymbol{\varphi}}_i^o$, $\boldsymbol{\lambda}_i^{c.o} = \boldsymbol{\Sigma}_i^{cc.o\,1/2}\boldsymbol{\varphi}_i^c$, $\tilde{\boldsymbol{\Sigma}}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc.o}/\nu^2(\mathbf{y}_i^o)$ and $\nu_i^{co} = \nu + p_i^o$, where

$$\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo\,-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \quad \boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}, \quad \tilde{\boldsymbol{\varphi}}_i^o = \boldsymbol{\varphi}_i^o + \boldsymbol{\Sigma}_i^{oo\,-1}\boldsymbol{\Sigma}_i^{oc}\boldsymbol{\varphi}_i^c,$$
$$c_i^{oc} = (1 + \boldsymbol{\varphi}_i^{c\top}\boldsymbol{\Sigma}_i^{cc.o}\boldsymbol{\varphi}_i^c)^{-1/2} \quad \text{and} \quad \tau_i^{co} = \nu(\mathbf{y}_i^o)\,\tilde{\boldsymbol{\varphi}}_i^{o\top}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o). \tag{5.52}$$

Let $\mathbf{V} = \mathrm{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \mathrm{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ denote the observed data. Therefore, the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\alpha}_\Sigma^\top, \boldsymbol{\lambda}^\top)^\top$, where $\boldsymbol{\alpha}_\Sigma$ denotes a minimal set of parameters such that $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ is well defined (e.g. the upper triangular elements of $\boldsymbol{\Sigma}$ in the unstructured case), for the observed data $(\mathbf{V}, \mathbf{C})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln L_i, \tag{5.53}$$

where L_i represents the likelihood function of $\boldsymbol{\theta}$ for the ith sample, given by

$$\begin{aligned} L_i \equiv L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i) &= f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}) = f(\mathbf{v}_{1i}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{v}_{2i}^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ &= \mathcal{L}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \boldsymbol{\mu}_i^{co}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{c.o}, \tilde{\tau}_i^{co}, \nu + p_i^o)\, ST_{p_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \tilde{\boldsymbol{\lambda}}_i^o, \nu), \end{aligned}$$

where $\mathcal{L}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \tau, \nu)$ denotes the integral

$$\mathcal{L}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \tau, \nu) = \int_{\boldsymbol{\alpha}}^{\boldsymbol{\beta}} EST_r(\mathbf{w}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \tau, \nu)\mathrm{d}\mathbf{w}, \tag{5.54}$$

that is, $\mathcal{L}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \tau, \nu) = \mathbb{P}(\boldsymbol{\alpha} \leqslant \mathbf{W} \leqslant \boldsymbol{\beta})$ for $\mathbf{W} \sim EST_r(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \tau, \nu)$. For the ST case ($\tau = 0$), we simply omit the τ parameter, that is, $\mathcal{L}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu) = \mathcal{L}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, 0, \nu)$.

5.8.2 Parameter estimation via the EM algorithm

In this subsection, we describe how to carry out ML estimation for the ST-C model. The EM algorithm, originally proposed by Dempster *et al.* (1977), is a very popular iterative optimization strategy commonly used to obtain ML estimates for incomplete-data problems. This algorithm has many attractive features such as the numerical stability, the simplicity of implementation and quite reasonable memory requirements (McLachlan & Krishnan, 2008).

From the stochastic representation of the multivariate ST distribution, it can be hierarchical represented as,

$$\mathbf{Y}_i \mid (U_i = u_i, T_i = t_i) \sim N_p(\boldsymbol{\mu} + \boldsymbol{\Delta} t_i, u_i^{-1}\boldsymbol{\Gamma}) \tag{5.55}$$

$$U_i \sim \text{Gamma}(\nu/2, \nu/2) \tag{5.56}$$

$$T_i \sim \text{HT}(\nu), \tag{5.57}$$

with $\text{HT}(\nu)$ referring to a Half standard Student's t distribution with degrees of freedom ν, with U_i and T_i being mutually independent, and $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$ as in proposition 5.3. The complete data log-likelihood function of an equivalent set of parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\Delta}^\top, \boldsymbol{\alpha}_\Gamma^\top, \nu)^\top$, where $\boldsymbol{\alpha}_\Gamma = \text{vech}(\boldsymbol{\Gamma})$, is given by $\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta})$, where the individual complete data log-likelihood is

$$\ell_{ic}(\boldsymbol{\theta}) = -\frac{1}{2}\big\{\ln|\boldsymbol{\Gamma}| + u_i(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta} t_i)^\top \boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta} t_i)\big\} + c,$$

with c being a constant that does not depend on $\boldsymbol{\theta}$. Subsequently, the EM algorithm for the ST-C model can be summarized as follows:

E-step: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\alpha}}_\Gamma^{(k)}, \widehat{\nu}^{(k)})$ at the kth step of the algorithm, the E-step provides the conditional expectation of the complete data log-likelihood function

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \text{E}\Big[\ell_c(\boldsymbol{\theta}) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\Big] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) \propto -\frac{1}{2}\ln|\boldsymbol{\Gamma}| - \frac{1}{2}\text{tr}\left[\left\{\widehat{u\mathbf{y}_i^2}^{(k)} + \widehat{u}_i^{(k)}\boldsymbol{\mu}\boldsymbol{\mu}^\top + \widehat{ut_i^2}^{(k)}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top - 2\widehat{u\mathbf{y}_i}^{(k)}\boldsymbol{\mu}^\top - 2\widehat{ut\mathbf{y}_i}^{(k)}\boldsymbol{\Delta}^\top + 2\widehat{ut_i}^{(k)}\boldsymbol{\Delta}\boldsymbol{\mu}^\top\right\}\boldsymbol{\Gamma}^{-1}\right],$$

with $\widehat{u\mathbf{y}_i^r}^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_i\mathbf{Y}_i^r \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{ut_i^r}^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i^r \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ (for $r = \{1, 2\}$, with $\mathbf{Y}_i^1 = \mathbf{Y}_i$ and $\mathbf{Y}_i^2 = \mathbf{Y}_i\mathbf{Y}_i^\top$), $\widehat{ut\mathbf{y}_i}^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{u}_i^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$.

M-step: Conditionally maximizing $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ with respect to each entry of $\boldsymbol{\theta}$, we update the estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\alpha}}_{\Gamma}^{(k)}, \widehat{\nu}^{(k)})$ by

$$\widehat{\boldsymbol{\mu}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{u\mathbf{y}}_i^{(k)} - \widehat{ut}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\right\}, \tag{5.58}$$

$$\widehat{\boldsymbol{\Delta}}^{(k+1)} = \left\{\sum_{i=1}^{n}\widehat{ut}_i^{2(k)}\right\}^{-1}\sum_{i=1}^{n}\left\{\widehat{ut\mathbf{y}}_i^{(k)} - \widehat{ut}_i\widehat{\boldsymbol{\mu}}^{(k+1)}\right\}, \tag{5.59}$$

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\mu}}^{(k+1)\top} - 2\widehat{ut\mathbf{y}}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} + 2\widehat{ut}_i^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top}\\ &+ \widehat{ut}_i^{2(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} + \widehat{u}_i\widehat{\boldsymbol{\mu}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top}\Big\}\end{aligned} \tag{5.60}$$

Then we update the parameter ν by maximizing the marginal log-likelihood function for $\mathbf{y}$, that is, $\widehat{\nu}^{(k+1)} = \arg\max_{\nu}\sum_{i=1}^{n}\log f(\mathbf{V}_i \mid \mathbf{C}_i, \boldsymbol{\theta}^{(k+1)}; \nu^{(k)})$.

Algorithm is iterated until a suitable convergence rule is satisfied. In the later analysis, the algorithm stops when the relative distance between two successive evaluations of the log-likelihood defined in (5.53) is less than a tolerance, i.e., $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)} \mid \mathbf{V}, \mathbf{C})/\ell(\widehat{\boldsymbol{\theta}}^{(k)} \mid \mathbf{V}, \mathbf{C}) - 1| < \epsilon$, for example, $\epsilon = 10^{-6}$. Once converged, we can recover $\widehat{\boldsymbol{\lambda}}$ and $\widehat{\boldsymbol{\Sigma}}$ using the expressions

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Gamma}} + \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^{\top} \qquad \text{and} \qquad \widehat{\boldsymbol{\lambda}} = \frac{\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\Delta}}}{(1 - \widehat{\boldsymbol{\Delta}}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Delta}})^{1/2}}.$$

It is important to stress that, from equations (5.58) to (5.60), the E-step reduces to the computation of $\widehat{u}_i^{(k)}$, $\widehat{u\mathbf{y}}_i^{(k)}$, $\widehat{u\mathbf{y}_i^2}^{(k)}$ $\widehat{ut}_i^{(k)}$, $\widehat{ut}_i^{2(k)}$ and $\widehat{ut\mathbf{y}}_i^{(k)}$. Details of these expectations can be found in Appendix C.

5.8.3 Regression setting

Suppose that we have observations on n independent individuals, $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, where $\mathbf{Y}_i \sim ST_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, $i = 1, \ldots, n$. Associated with individual i we assume a known $p \times q$ covariate matrix $\mathbf{X}_i$, which we use to specify the linear predictor $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a q-dimensional vector of unknown regression coefficients. In this case, the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\alpha}^{\top}, \boldsymbol{\lambda}^{\top})^{\top}$. The E-Step of the EM algorithm updates $\boldsymbol{\beta}$ as follows

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = (\sum_{i=1}^{n}\mathbf{X}_i^{\top}\mathbf{X}_i)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^{\top}(\widehat{u\mathbf{y}}_i^{(k)} - \widehat{ut}_i^{(k)}\boldsymbol{\Delta}^{(k)}), \tag{5.61}$$

and the necessary quantities of the E and M-steps found in Subsection 5.8.2 remain the same once we plug-in $\widehat{\boldsymbol{\mu}}^{(k)}$ by $\widehat{\boldsymbol{\mu}}_i^{(k)} = \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}$.

5.9 Conclusions

In this paper, we proposed expressions to compute product moment of truncated multivariate distributions belonging to the selection elliptical family, showing in a clever way that their moments can be computed using an unique moment for their respective elliptical symmetric case. In contrast with other works, we avoid cumbersome expressions, having neat formulas for high-order truncated moments. To the best of our knowledge, this is the first proposal discussing the conditions of existence of the truncated moments for members of the selection elliptical family. Also, we propose optimized methods able to deal with extreme setting of the parameters, partitions with almost zero volume or no truncation. In order to show the applicability of this work, we have developed an application of truncated ST moments in risk measurement in Finance context as well as a ST censored model, a robust model capable to deal with missing data, outliers and skewness.

6 Concluding remarks

In this last chapter, we present the scientific production resulting from this thesis: original articles and software. In addition to the articles of our authorship, we present articles by other authors that have been based on the results of this thesis. A summary of the main functions of the proposed `MomTrunc` package are presented in subsection 6.1.2. Finally, we close the chapter with two sections concluding the results of this thesis, as well as sketching some future research.

6.1 Technical production

6.1.1 Submitted papers

As result of the present work, we have written four articles with three of them being already submitted to high impact journals. Resulting works presented in chapters 2-5 are respectively:

1. **Galarza, C.**, Lachos V. Lin, T.-I., & Wang, W.-L. (2020+) "Moments of doubly truncated multivariate student-t distribution: A recurrence approach". *Submitted to Statistica Sinica.*

2. **Galarza, C.**, Matos, L. Dey, D. & Lachos V. (2019+) "On moments of folded and doubly truncated multivariate extended skew-normal distributions". *Submitted to Journal of Computational and Graphical Statistics.*

3. **Galarza, C.**, Matos, L. & Lachos, V. (2020+) "Likelihood-based inference for multivariate skew-normal censored regression models". *Submitted to METRON.*

4. **Galarza, C.**, Matos, L. & Lachos, V. (2020+) "Moments of the doubly truncated selection elliptical distributions with emphasis on the unified multivariate skew-t distribution". *To be submitted.*

5. **Galarza, C.**, Matos, L. & Lachos, V. (2020+) "Likelihood-based inference for multivariate skew-t censored regression models with missing data". *Under construction.*

Other works based on the results in this document are:

6. Mattos, T.B., Matos, L. & Lachos, V. (2019) "A semiparametric mixed-effects model for censored longitudinal data". *Technical report, RT-UConn 15, University of Connecticut.*

7. De Alencar, F., **Galarza, C.**, Matos, L. & Lachos, V. (2019) "Finite mixture modeling of censored and missing data using the multivariate skew-normal distribution". *Technical report, RT-UConn 31, University of Connecticut.*

6.1.2 R package implementation

`MomTrunc`: Moments of Folded and Doubly Truncated Multivariate Distributions

It computes arbitrary products moments (mean vector and variance-covariance matrix), for some doubly truncated (and folded) multivariate distributions. These distributions belong to the family of selection elliptical distributions, which includes well known skewed distributions as the unified skew-t distribution (SUT) and its particular cases as the extended skew-t (EST), skew-t (ST) and the symmetric student-t (MVT) distribution. Analogous normal cases unified skew-normal (SUN), extended skew-normal (ESN), skew-normal (SN), and symmetric normal (MVN) are also included. Density, probabilities and random deviates are also offered for these members.

Probabilities can be computed using the functions `pmvSN()` and `pmvESN()` for the normal cases SN and ESN and, `pmvST()` and `pmvEST()` for the t cases ST and EST respectively, which offer the option to return the logarithm in base 2 of the probability, useful when the true probability is too small for the machine precision. These functions above use methods in Genz & Bretz (2009) through the `mvtnorm` package (linked direclty to our `C++` functions) and Cao *et al.* (2019b) through the package `tlrmvnmvt`. For the double truncated Student-t cases SUT, EST, ST and T, decimal degrees of freedom are supported. Computation of arbitrary moments are based in this thesis. Reference for the family of selection-elliptical distributions in this package can be found in Arellano-Valle & Genton (2005).

Next, we show part of the `MomTrunc R` manual (also available on CRAN) for the three most important functions.

`meanvarTMD` Mean and variance for doubly truncated multivariate distributions

---

Description

It computes the mean vector and variance-covariance matrix for some doubly truncated skewelliptical distributions. It supports the p-variate Normal, Skew-normal (SN), Extended Skew-normal (ESN) and Unified Skew-normal (SUN) as well as the Student's-t, Skew-t (ST), Extended Skew-t (EST) and Unified Skew-t (SUT) distribution.

Usage

```
meanvarTMD(lower = rep(-Inf,length(mu)),upper = rep(Inf,length(mu)),mu,
Sigma,lambda = NULL,tau = NULL,Gamma = NULL,nu = NULL,dist)
```

Arguments

`lower`	the vector of lower limits of length p
`upper`	the vector of upper limits of length p
`mu`	a numeric vector of length p representing the location parameter
`Sigma`	a numeric positive definite matrix with dimension $p \times p$ representing the scale parameter
`lambda`	a numeric matrix of dimension $p \times q$ representing the skewness/shape matrix parameter for the SUN and SUT distribution. For the ESN and EST distributions ($q = 1$), `lambda` is a numeric vector of dimension p (see examples at the end of this help). If `all(lambda == 0)`, the SUN/ESN/SN (SUT/EST/ST) reduces to a normal (t) symmetric distribution.
`tau`	a numeric vector of length q representing the extension parameter for the SUN and SUT distribution. For the ESN and EST distributions, tau is a positive scalar ($q = 1$). Furthermore, if `tau == 0`, the ESN (EST) reduces to a SN (ST) distribution.
`Gamma`	a correlation matrix with dimension $q \times q$. It must be provided only for the SUN and SUT cases. For particular cases SN, ESN, ST and EST, we have that `Gamma == 1`.
`nu`	It represents the degrees of freedom for the Student's t-distribution
`log2`	a boolean variable, indicating if the `log2` result should be returned. This is useful when the true probability is too small for the machine precision.

Details

Univariate case is also considered, where `Sigma` will be the variance σ^2. Normal case code is an R adaptation of the Matlab available function `dtmvnmom.m` from Kan & Robotti (2017) and it is used for `p<=3`. For higher dimensions we use the extension of the algorithm in Vaida & Liu (2009) proposed in Chapter 3.

Value

It returns a list with three elements:

`mean`	the mean vector of length p
`EYY`	the second moment matrix of dimensions $p \times p$
`varcov`	the variance-covariance matrix of dimensions $p \times p$

Warning

For the t cases, the algorithm supports degrees of freedom `nu <= 2`, however, it may take more time than usual.

Note

If `nu >= 300`, Normal case is considered.

Examples

```
a = c(-0.8,-0.7,-0.6)
b = c(0.5,0.6,0.7)
mu = c(0.1,0.2,0.3)
Sigma = matrix(data = c(1,0.2,0.3,0.2,1,0.4,0.3,0.4,1),
nrow = length(mu),ncol = length(mu),byrow = TRUE)
lambda = c(-2,0,1)

# Theoretical value
value1 = meanvarTMD(a,b,mu,Sigma,dist="normal")

#MC estimates
MC11 = MCmeanvarTMD(a,b,mu,Sigma,dist="normal") #by defalut n = 10000
MC12 = MCmeanvarTMD(a,b,mu,Sigma,dist="normal",n = 10^5) #more precision



# Now works for for any nu>0
value2 = meanvarTMD(a,b,mu,Sigma,dist = "t",nu = 0.87)
value3 = meanvarTMD(a,b,mu,Sigma,,dist = "SN")
value4 = meanvarTMD(a,b,mu,Sigma,lambda,nu = 4,dist = "ST")
value5 = meanvarTMD(a,b,mu,Sigma,lambda,tau = 1,dist = "ESN")
value6 = meanvarTMD(a,b,mu,Sigma,lambda,tau = 1,nu = 4,dist = "EST")

#Skew-unified Normal (SUN) and Skew-unified t (SUT) distributions
Lambda = matrix(c(1,0,2,-3,0,-1),3,2) #A skewness matrix p times q
Gamma = matrix(c(1,-0.5,-0.5,1),2,2) #A correlation matrix q times q
tau = c(-1,2) #A vector of extension parameters of dim q
value7 = meanvarTMD(a,b,mu,Sigma,Lambda,tau,Gamma,dist = "SUN")
value8 = meanvarTMD(a,b,mu,Sigma,Lambda,tau,Gamma,nu = 4,dist = "SUT")

#The ESN and EST as particular cases of the SUN and SUT for q == 1
Lambda = matrix(c(-2,0,1),3,1)
```

```
Gamma = 1
tau = 1
value9 = meanvarTMD(a,b,mu,Sigma,Lambda,tau,Gamma,dist = "SUN")
value10 = meanvarTMD(a,b,mu,Sigma,Lambda,tau,Gamma,nu = 4,dist = "SUT")
round(value5$varcov,2) == round(value9$varcov,2)
round(value6$varcov,2) == round(value10$varcov,2)
```

`momentsTMD` Moments for doubly truncated multivariate distributions

Description

It computes `kappa`-th order moments for for some doubly truncated skew-elliptical distributions. It supports the p-variate Normal, Skew-normal (SN) and Extended Skew-normal (ESN), as well as the Student's t, Skew-t (ST) and the Extended Skew-t (EST) distribution.

Usage

```
momentsTMD(kappa,lower = rep(-Inf,length(mu)),upper = rep(Inf,length(mu)),
mu,Sigma,lambda = NULL,tau = NULL,nu = NULL,dist)
```

Arguments

Details

Univariate case is also considered, where `Sigma` will be the variance σ^2.

Value

A data frame containing $p + 1$ columns. The p first containing the set of combinations of exponents summing up to `sum(kappa)` and the last column containing the the expected value. Normal cases (ESN, SN and normal) return `prod(kappa)+1` moments while the Student's t cases return all moments of order up to `kappa`. See example section.

Note

If `nu >= 300`, the Normal case is considered.

Examples

```
a = c(-0.8,-0.7,-0.6)
b = c(0.5,0.6,0.7)
```

`kappa`	moments vector of length p. All its elements must be integers greater or equal to 0. For the Student's-t case, kappa can be a scalar representing the order of the moment.
`lower`	the vector of lower limits of length p
`upper`	the vector of upper limits of length p
`mu`	a numeric vector of length p representing the location parameter
`Sigma`	a numeric positive definite matrix with dimension $p \times p$ representing the scale parameter
`lambda`	a numeric vector of length p representing the skewness parameter for ST and EST cases. If `lambda == 0`, the EST/ST reduces to a t (symmetric) distribution.
`tau`	It represents the extension parameter for the EST distribution. If `tau == 0`, the EST reduces to a ST distribution.
`nu`	It represents the degrees of freedom for the Student's t-distribution.
`dist`	represents the truncated distribution to be used. The values are `normal`, `SN` and `ESN` for the doubly truncated Normal, Skew-normal and Extended Skew-normal distributions and, `t`, `ST` and `EST` for the for the doubly truncated Student-t, Skew-t and Extended Skew-t distributions.

```
mu = c(0.1,0.2,0.3)
Sigma = matrix(data = c(1,0.2,0.3,0.2,1,0.4,0.3,0.4,1),
nrow = length(mu),ncol = length(mu),byrow = TRUE)

kp     = c(2,0,1)
lambda = c(-2,0,1)
value1 = momentsTMD(kp,a,b,mu,Sigma,dist="normal")
value2 = momentsTMD(kp,a,b,mu,Sigma,dist = "t",nu = 7)
value3 = momentsTMD(kp,a,b,mu,Sigma,lambda,dist = "SN")
value4 = momentsTMD(kp,a,b,mu,Sigma,lambda,tau = 1,dist = "ESN")

#T cases with kappa scalar (all moments up to 3)
value5 = momentsTMD(3,a,b,mu,Sigma,nu = 7,dist = "t")
value6 = momentsTMD(3,a,b,mu,Sigma,lambda,nu = 7,dist = "ST")
value7 = momentsTMD(3,a,b,mu,Sigma,lambda,tau = 1,nu = 7,dist = "EST")
```

dprmvEST	Multivariate Extended-Skew t Density, Probablilities and Random Deviates Generator

Description

These functions provide the density function, probabilities and a random number generator for the multivariate extended-skew t (EST) distribution with mean vector `mu`, scale matrix `Sigma`, skewness parameter `lambda`, extension parameter `tau` and degrees of freedom `nu`.

Usage

```
dmvEST(x,mu=rep(0,length(lambda)),Sigma=diag(length(lambda)),lambda,tau,nu)
pmvEST(lower = rep(-Inf,length(lambda)),upper=rep(Inf,length(lambda)),
mu = rep(0,length(lambda)),Sigma,lambda,tau,nu,log2 = FALSE)
rmvEST(n,mu=rep(0,length(lambda)),Sigma=diag(length(lambda)),lambda,tau,nu)
```

Arguments

`x`	vector or matrix of quantiles. If `x` is a matrix, each row is taken to be a quantile.
`n`	number of observations.
`lower`	the vector of lower limits of length p
`upper`	the vector of upper limits of length p
`mu`	a numeric vector of length p representing the location parameter
`Sigma`	a numeric positive definite matrix with dimension $p \times p$ representing the scale parameter
`lambda`	a numeric vector of length p representing the skewness parameter for ST and EST cases. If `lambda == 0`, the EST/ST reduces to a t (symmetric) distribution.
`tau`	It represents the extension parameter for the EST distribution. If `tau == 0`, the EST reduces to a ST distribution.
`nu`	It represents the degrees of freedom for the Student's t distribution
`log2`	a boolean variable, indicating if the `log2` result should be returned. This is useful when the true probability is too small for the machine precision.

Examples

```
#Univariate case
dmvEST(x = -1,mu = 2,Sigma = 5,lambda = -2,tau = 0.5,nu=4)
rmvEST(n = 100,mu = 2,Sigma = 5,lambda = -2,tau = 0.5,nu=4)
#Multivariate case
mu = c(0.1,0.2,0.3,0.4)
Sigma = matrix(c(1,0.2,0.3,0.1,0.2,1,0.4,-0.1,0.3,0.4,1,0.2,0.1,-0.1,0.2,
```

```
1),nrow = length(mu),ncol = length(mu),byrow = TRUE)
lambda = c(-2,0,1,2)
tau = 2
#One observation
dmvEST(x = c(-2,-1,0,1),mu,Sigma,lambda,tau,nu=4)
rmvEST(n = 100,mu,Sigma,lambda,tau,nu=4)
#Many observations as matrix
x = matrix(rnorm(4*10),ncol = 4,byrow = TRUE)
dmvEST(x = x,mu,Sigma,lambda,tau,nu=4)
lower = rep(-Inf,4)
upper = c(-1,0,2,5)
pmvEST(lower,upper,mu,Sigma,lambda,tau,nu=4)
```

Other functions: MC estimates for the first two moments of a truncated multivariate distribution (TMD) can be reach through the function `MCmeanvarTMD()`. Functions to compute the mean and variance-covariance matrix, as well as product moments for folded multivariate distributions (FMDs) are also available through the analogous `meanvarFMD()` and `momentsFMD()`, which arguments are the same for functions `meanvarFMD()` and `momentsFMD()`, except for arguments `lower` and `upper` that are not longer neeeded. Finally, A function `cdfFMD()` is provided to compute the cdf of several FMDs.

Some **R MomTrunc** package output

```
# All moments up to 3 for an 5-variate folded Student-t distribution

> momentsFMD(3,mu,S,nu)

      [k1] [k2] [k3] [k4] [k5]    Moment
 [1,]    0    0    0    0    0    1.0000
 [2,]    0    0    0    0    1    0.9598
 [3,]    0    0    0    0    2    1.5000
 [4,]    0    0    0    0    3    3.1311
 [5,]    0    0    0    1    0    0.9260
 [6,]    0    0    0    1    1    1.1925
 [7,]    0    0    0    1    2    2.3439
 [8,]    0    0    0    2    0    1.4100
 [9,]    0    0    0    2    1    2.2836
[10,]    0    0    0    3    0    2.8902
[11,]    0    0    1    0    0    0.8994
```

```
[12,]    0    0    1    0    1    0.9001
[13,]    0    0    1    0    2    1.4851
[14,]    0    0    1    1    0    0.8878
[15,]    0    0    1    1    1    1.1914
[16,]    0    0    1    2    0    1.4502
[17,]    0    0    2    0    0    1.3400
[18,]    0    0    2    0    1    1.4142
[19,]    0    0    2    1    0    1.4188
[20,]    0    0    3    0    0    2.7055
[21,]    0    1    0    0    0    0.8803
[22,]    0    1    0    0    1    0.9144
[23,]    0    1    0    0    2    1.5559
[24,]    0    1    0    1    0    0.8585
[25,]    0    1    0    1    1    1.1785
[26,]    0    1    0    2    0    1.3919
[27,]    0    1    1    0    0    0.8831
[28,]    0    1    1    0    1    0.9604
[29,]    0    1    1    1    0    0.9207
[30,]    0    1    2    0    0    1.4725
[31,]    0    2    0    0    0    1.2900
[32,]    0    2    0    0    1    1.4584
[33,]    0    2    0    1    0    1.3388
[34,]    0    2    1    0    0    1.4483
[35,]    0    3    0    0    0    2.5749
[36,]    1    0    0    0    0    0.8686
[37,]    1    0    0    0    1    0.9191
[38,]    1    0    0    0    2    1.5934
[39,]    1    0    0    1    0    0.9570
[40,]    1    0    0    1    1    1.3782
[41,]    1    0    0    2    0    1.7208
[42,]    1    0    1    0    0    0.8392
[43,]    1    0    1    0    1    0.9340
[44,]    1    0    1    1    0    0.9739
[45,]    1    0    2    0    0    1.3567
[46,]    1    1    0    0    0    0.8224
[47,]    1    1    0    0    1    0.9552
[48,]    1    1    0    1    0    0.9576
[49,]    1    1    1    0    0    0.8830
[50,]    1    2    0    0    0    1.3073
```

```
[51,]    2    0    0    0    0    1.2600
[52,]    2    0    0    0    1    1.4780
[53,]    2    0    0    1    0    1.6387
[54,]    2    0    1    0    0    1.3205
[55,]    2    1    0    0    0    1.2940
[56,]    3    0    0    0    0    2.4966
```

6.2 Conclusions

In this thesis, we proposed a methodology to calculate the truncated moments of several elliptical distributions and its skewed extended versions belonging to the family of SE distributions. High order moments are achieved using a recurrence approach, plus a 1-1 relation which let us write in a neat manner, any product moment of a member of the SE class as a moment of its respective symmetric case. Expressions for the first two moments, conditions of existence and useful expectations in the context of censored interval models, are presented in general. Various estimation and regression applications in censored models are proposed in order to show the usefulness of our proposal, considering Student's t, SN and ST deviates as well as an application of ST truncated moments in Finance. All proposed methodology has been implemented and is available in the `MomTrunc` package of the `R` software, a highly optimized package that provides truncated moments and other functions of interest for various symmetrical and asymmetric distributions.

6.3 Future research

A natural extension for this work is to calculate the moments other members of the elliptical and consequently to the SE class of distributions, as their probabilities are implemented efficiently. Multimodality can be easily handled by considering mixtures of censored regression models, extending works as Lachos *et al.* (2017) and De Alencar *et al.* (2019a) to the regression framework with interval-censored responses. Mixed effects models with skewed heavy-tailed random effects or error terms are also a natural extension. For all models above is also possible to include a semi-parametric structure for modeling the any nonlinear behavior as in Mattos *et al.* (2019), or using a spatial covariance structure for spatially correlated data. Experimental studies include covariates that often comes with substantial measurement errors (Liu & Wu, 2007). How to incorporate measurement error in covariates within our robust framework can also be part of future research. An in-depth investigation of such extensions is beyond the scope of the present work, but certainly an interesting topic for future research.

Bibliography

Acerbi, C. & Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, **31**(2), 379–388. Cited in page 109.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Cont.*, **19**, 716–723. Cited 2 times in pages 51 and 92.

Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 99–102. Cited in page 26.

Arellano-Valle, R. B. & Azzalini, A. (2006a). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, **33**(3), 561–574. Cited 3 times in pages 56, 57, and 79.

Arellano-Valle, R. B. & Azzalini, A. (2006b). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, **33**(3), 561–574. Cited 2 times in pages 98 and 99.

Arellano-Valle, R. B. & Bolfarine, H. (1995). On some characterizations of the t-distribution. *Statistics & Probability Letters*, **25**(1), 79–85. Cited in page 37.

Arellano-Valle, R. B. & Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96**, 93–116. Cited 2 times in pages 56 and 121.

Arellano-Valle, R. B. & Genton, M. G. (2010). Multivariate extended skew-t distributions and related families. *Metron*, **68**(3), 201–234. Cited 9 times in pages 20, 56, 57, 79, 80, 95, 99, 100, and 115.

Arellano-Valle, R. B., Branco, M. D. & Genton, M. G. (2006a). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics*, **34**. Cited in page 19.

Arellano-Valle, R. B., Branco, M. D. & Genton, M. G. (2006b). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics*, **34**(4), 581–601. Cited 3 times in pages 96, 97, and 98.

Arismendi, J. C. (2013). Multivariate truncated moments. *Journal of Multivariate Analysis*, **117**, 41–75. Cited 2 times in pages 36 and 54.

Arismendi, J. C. & Broda, S. (2017). Multivariate elliptical truncated moments. *Journal of Multivariate Analysis*, **157**, 29 – 44. Cited in page 95.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228. Cited in page 109.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178. Cited 2 times in pages 35 and 54.

Azzalini, A. (2020). *SN: The Skew-Normal and Related Distributions such as the Skew-t (version 1.5-5).*. Università di Padova, Italia. Cited in page 34.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, **61**, 579–602. Cited 2 times in pages 56 and 77.

Azzalini, A. & Capitanio, A. (2003). Distributions generated and perturbation of symmetry with emphasis on the multivariate skew-t distribution. *Journal of the Royal Statistical Society, Series B*, **61**, 367–389. Cited 3 times in pages 21, 96, and 98.

Azzalini, A. & Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**(4), 715–726. Cited 5 times in pages 22, 55, 56, 75, and 78.

Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113. Cited 2 times in pages 98 and 100.

Cabral, C. R. B., Lachos, V. H. & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis*, **56**(1), 126–142. Cited 2 times in pages 85 and 148.

Cao, J., Genton, M., Keyes, D. & Turkiyyah, G. (2019a). *tlrmvnmvt: Low-Rank Methods for MVN and MVT Probabilities*. R package version 2.0. Cited in page 20.

Cao, J., Genton, M. G., Keyes, D. E. & Turkiyyah, G. M. (2019b). Exploiting low rank covariance structures for computing high-dimensional normal and student-t probabilities. Cited 2 times in pages 20 and 121.

Chakraborty, A. K. & Chatterjee, M. (2013). On multivariate folded normal distribution. *Sankhya B*, **75**(1), 1–15. Cited 2 times in pages 26 and 69.

De Alencar, F., Galarza, C., Matos, L. & Lachos, V. (2019a). Finite mixture modeling of censored and missing data using the multivariate skew-normal distribution. RT-UConn 31, University of Connecticut. Cited in page 129.

De Alencar, H., Avila, L., Galarza, C. & Lachos., V. H. (2019b). *CensMFM: Finite Mixture of Multivariate Censored/Missing Data*. R package version 2.0. Cited in page 19.

De Bastiani, F., de Aquino Cysneiros, A. H. M., Uribe-Opazo, M. A. & Galea, M. (2015). Influence diagnostics in elliptical spatial linear models. *Test*, **24**(2), 322–340. Cited in page 35.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38. Cited 5 times in pages 19, 47, 84, 112, and 117.

Denuit, M., Dhaene, J., Goovaerts, M. & Kaas, R. (2006). *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons. Cited in page 109.

Diggle, P., Diggle, P. J., Heagerty, P., Heagerty, P. J., Liang, K.-Y., Zeger, S. *et al.* (2002). *Analysis of Longitudinal Data*. Oxford University Press. Cited in page 94.

Fang, K. T., Kotz, S. & Ng, K. W. (1990). *Simetric multivariate and related distribuitions*. Chapman & Hall, London. Cited in page 35.

Flecher, C., Allard, D. & Naveau, P. (2010). Truncated skew-normal distributions: moments, estimation by weighted moments and application to climatic data. *Metron*, **68**, 331–345. Cited 3 times in pages 35, 54, and 55.

Fonseca, T. C., Ferreira, M. A. & Migon, H. S. (2008). Objective bayesian analysis for the student-t regression model. *Biometrika*, **95**(2), 325–333. Cited in page 35.

Galarza, C., Kan, R. & Lachos., V. H. (2018). *MomTrunc: Moments of Folded and Doubly Truncated Multivariate Distributions*. R package version 5.69. Cited 4 times in pages 21, 36, 77, and 81.

Garay, A. M., Castro, L. M., Leskow, J. & Lachos, V. H. (2017). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, **26**(2), 542–566. Cited in page 77.

Genç, A. (2013). Moments of truncated normal/independent distributions. *Statistical Papers*, **54**, 741–764. Cited 2 times in pages 35 and 54.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, **1**(2), 141–149. Cited in page 63.

Genz, A. & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg. ISBN 978-3-642-01688-2. Cited 2 times in pages 20 and 121.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. & Hothorn, T. (2020). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-12. Cited in page 20.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*, volume 1. Springer. Cited in page 60.

Ho, H., Lin, T.-I., Wang, W.-L., Garay, A. M., Lachos, V. H. & Castro, M. (2015). *TTmoment: Sampling and Calculating the First and Second Moments for the Doubly Truncated Multivariate t Distribution*. R package version 1.0. Cited in page 21.

Ho, H. J., Lin, T. I., Chen, H. Y. & Wang, W. L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*, **142**, 25–40. Cited 4 times in pages 19, 36, 54, and 95.

Hoffman, H. J. & Johnson, R. E. (2015). Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. *Journal of Agricultural, Biological, and Environmental Statistics*, **20**(1), 156–171. Cited in page 89.

Houthakker, H. (1959). *The scope and limits of futures trading*. Cowles Foundation for Research in Economics at Yale University. Cited in page 54.

Jawitz, J. W. (2004). Moments of truncated continuous univariate distributions. *Advances in Water Resources*, **27**(3), 269–281. Cited 2 times in pages 35 and 54.

Kan, R. & Robotti, C. (2017). On moments of folded and truncated multivariate normal distributions. *Journal of Computational and Graphical Statistics*, **25**(1), 930–934. Cited 16 times in pages 20, 21, 28, 30, 31, 36, 39, 54, 55, 58, 64, 66, 68, 75, 95, and 122.

Kim, H. M. (2008). A note on scale mixtures of skew normal distribution. *Statistics and Probability Letters*, **78**. 1694-1701. Cited 3 times in pages 35, 36, and 54.

Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. & Montenegro, L. C. (2007). Likelihood based inference for multivariate skew-normal regression models. *Communications in Statistics–Theory and Methods*, **36**, 1769–1786. Cited in page 93.

Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**(1), 303. Cited in page 148.

Lachos, V. H., Moreno, E. J. L., Chen, K. & Cabral, C. R. B. (2017). Finite mixture modeling of censored data using the multivariate student-t distribution. *Journal of Multivariate Analysis*, **159**. 151-167. Cited 3 times in pages 46, 75, and 129.

Landsman, Z. & Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, **7**(4), 55–71. Cited in page 110.

Lange, K. & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, **2**(2), 175–198. Cited in page 26.

Lien, D.-H. D. (1985). Moments of truncated bivariate log-normal distributions. *Economics Letters*, **19**(3), 243–247. Cited 2 times in pages 35 and 54.

Lin, T. I., Ho, H. J. & Chen, C. L. (2009). Analysis of multivariate skew normal models with incomplete data. *Journal of Multivariate Analysis*, **100**(10), 2337–2351. Cited 3 times in pages 53, 88, and 93.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis With Missing Data*. John Wiley & Sons, New York. Cited in page 33.

Liu, W. & Wu, L. (2007). Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics*, **63**(2), 342–350. Cited in page 129.

Manjunath, B. G. & Wilhelm, S. (2009). Moments calculation for the double truncated multivariate normal density. *Available at SSRN 1472153*. Cited 4 times in pages 19, 64, 68, and 95.

Massuia, M. B., Cabral, C. R. B., Matos, L. A. & Lachos, V. H. (2015). Influence diagnostics for Student-t censored linear regression models. *Statistics*, **49**(5), 1074–1094. Cited in page 77.

Matos, L. A., Prates, M. O., Chen, M. H. & Lachos, V. H. (2013). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**, 1323–1342. Cited 5 times in pages 38, 75, 76, 77, and 112.

Mattos, T., Matos, L. & Lachos, V. (2019). A semiparametric mixed-effects model for censored longitudinal data. RT-UConn 15, University of Connecticut. Cited in page 129.

McLachlan, G. J. & Krishnan, T. (2008). *The EM Algorithm and Extensions*. Wiley, second edition. Cited 3 times in pages 47, 84, and 117.

Morán-Vásquez, R. A. & Ferrari, S. L. P. (2019). New results on truncated elliptical distributions. *Communications in Mathematics and Statistics*, (53). Cited in page 102.

Nadarajah, S. (2007). A truncated bivariate t distribution. *Economic Quality Control*, **22**(2), 303–313. Cited 2 times in pages 19 and 95.

Peel, D. & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, **10**(4), 339–348. Cited in page 35.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic Constrained Optimization*, pages 272–281. Springer. Cited in page 109.

Pinheiro, J. C., Liu, C. H. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using a multivariate t-distribution. *Journal of Computational and Graphical Statistics*, **10**, 249–276. Cited in page 35.

Roozegar, R., Balakrishnan, N. & Jamalizadeh, A. (2020). On moments of doubly truncated multivariate normal mean–variance mixture distributions with application to multivariate tail conditional expectation. *Journal of Multivariate Analysis*, **177**, 104586. Cited in page 95.

Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate distributions with applications to bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129–150. Cited in page 88.

Savalli, C., Paula, G. A. & Cysneiros, F. J. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, **6**(1), 59–76. Cited in page 35.

Schwarz, G. *et al.* (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464. Cited in page 92.

Sereda, E. N., Bronshtein, E. M., Rachev, S. T., Fabozzi, F. J., Sun, W. & Stoyanov, S. V. (2010). Distortion risk measures in portfolio optimization. In *Handbook of portfolio construction*, pages 649–673. Springer. Cited in page 109.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **23**(1), 223–229. Cited 5 times in pages 19, 35, 54, 95, and 140.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, pages 24–36. Cited 2 times in pages 19 and 95.

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**, 797–817. Cited 5 times in pages 64, 68, 75, 122, and 140.

VDEQ (2003). The quality of virginia non-tidal streams: First year report. richmond, virginia. vdeq technical bulletin. wqa/2002-001. Cited in page 33.

Wang, W. L. & Fan, T. H. (2011). Estimation in multivariate t linear mixed models for multiple longitudinal data. *Statistica Sinica*, **21**, 1857–1880. Cited in page 35.

Wang, W. L. & Lin, T. I. (2014). Multivariate t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. *Statistics in Medicine*, **33**, 3029–3046. Cited in page 35.

Wang, W. L. & Lin, T. I. (2015). Bayesian analysis of multivariate t linear mixed models with missing responses at random. *Journal of Statistical Computation and Simulation*, **85**, 3594–3612. Cited in page 35.

Wang, W.-L., Castro, L. M. & Lin, T.-I. (2017). Automated learning of t factor analysis models with complete and incomplete data. *Journal of Multivariate Analysis*, **161**, 157–171. Cited in page 35.

Wang, W.-L., Lin, T.-I. & Lachos, V. H. (2018). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical Methods in Medical Research*, **27**(1), 48–64. Cited in page 77.

Wang, W.-L., Castro, L. M., Lachos, V. H. & Lin, T.-I. (2019). Model-based clustering of censored data via mixtures of factor analyzers. *Computational Statistics & Data Analysis*. Cited 2 times in pages 50 and 89.

Wilhelm, S. & Manjunath, B. (2015). *tmvtnorm: Truncated Multivariate Normal and Student t Distribution*. R package version 1.4-10. Cited in page 21.

APPENDIX A: Appendix for chapter 2

Appendix A.1: Details for the expectations in EM algorithm

To compute the required expected values of all latent data, we find that most of them can be written in terms of $\mathbb{E}(U_i \mid \mathbf{Y}_i)$, and thereby we write $\widehat{u}_i = \mathbb{E}\{\mathbb{E}(U_i \mid \mathbf{Y}_i) \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\}$, where $\mathbb{E}(U_i \mid \mathbf{Y}_i) = (\nu + p)/(\nu + \delta)$ with $\delta = (\mathbf{Y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu})$. Subsequently, we discuss the closed-form expressions of conditional expectations as follows:

1. If the ith subject has only non-censored components, then

$$\widehat{u\mathbf{y}_i^2}^{(k)} = \widehat{u}_i^{(k)} \mathbf{y}_i \mathbf{y}_i^\top, \quad \widehat{u\mathbf{y}}_i^{(k)} = \widehat{u}_i^{(k)} \mathbf{y}_i, \quad \text{and} \quad \widehat{u}_i^{(k)} = \frac{\nu + p}{\nu + \widehat{\delta}^{(k)}(\mathbf{y}_i)},$$

   where $\widehat{\delta}^{(k)}(\mathbf{y}_i) = (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k)})^\top (\widehat{\boldsymbol{\Sigma}}^{(k)})^{-1} (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k)})$.

2. If the ith subject has only censored components, from Proposition 3 with $r = 1$, we have

$$\begin{aligned}
\widehat{u\mathbf{y}_i^2}^{(k)} &= \mathbb{E}[U_i \mathbf{Y}_i \mathbf{Y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\varphi}^{(k)}(\mathbf{V}_i) \widehat{\mathbf{w}}_i^{2^{c(k)}}, \\
\widehat{u\mathbf{y}}_i^{(k)} &= \mathbb{E}[U_i \mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\varphi}^{(k)}(\mathbf{V}_i) \widehat{\mathbf{w}}_i^{c(k)}, \\
\widehat{u}_i^{(k)} &= \mathbb{E}[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\varphi}^{(k)}(\mathbf{V}_i),
\end{aligned}$$

   where

$$\widehat{\varphi}^{(k)}(\mathbf{V}_i) = \frac{L_p(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu + 2)}{L_p(\mathbf{V}_{1i}, \mathbf{V}_{2i}; \widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{(k)}, \nu)},$$

$$\widehat{\mathbf{w}}_i^{c(k)} = \mathbb{E}[\mathbf{W}_i \mid \widehat{\boldsymbol{\theta}}^{(k)}], \quad \widehat{\mathbf{w}}_i^{2^{c(k)}} = \mathbb{E}[\mathbf{W}_i \mathbf{W}_i^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}] \tag{A.1}$$

   with $\mathbf{W}_i \sim Tt_p(\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Sigma}}^{*(k)}, \nu + 2; (\mathbf{V}_{1i}, \mathbf{V}_{2i}))$ and $\widehat{\boldsymbol{\Sigma}}^{*(k)} = \dfrac{\nu}{\nu + 2} \widehat{\boldsymbol{\Sigma}}^{(k)}$. To compute $\mathbb{E}[\mathbf{W}_i]$ and $\mathbb{E}[\mathbf{W}_i \mathbf{W}_i^\top]$ we use the results given in Subsection 3.1.

3. If the ith subject has both censored and uncensored components, then $(\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o)$, and $(\mathbf{Y}_i^c \mid \mathbf{V}_i, \mathbf{C}_i, \mathbf{y}_i^o)$ are equivalent processes. We obtain

$$\widehat{u\mathbf{y}_i^2}^{(k)} = \mathbb{E}(U_i \mathbf{Y}_i \mathbf{Y}_i^\top \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \begin{pmatrix} \mathbf{y}_i^o \mathbf{y}_i^{o\top} \widehat{u}_i^{(k)} & \widehat{u}_i^{(k)} \mathbf{y}_i^o \widehat{\mathbf{w}}_i^{c(k)\top} \\ \widehat{u}_i^{(k)} \widehat{\mathbf{w}}_i^{c(k)} \mathbf{y}_i^{o\top} & \widehat{u}_i^{(k)} \widehat{\mathbf{w}}_i^{2^{c(k)}} \end{pmatrix},$$

$$\widehat{u\mathbf{y}}_i^{(k)} = \mathbb{E}(U_i \mathbf{Y}_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \mathrm{vec}(\mathbf{y}_i^o \widehat{u}_i^{(k)}, \widehat{u}_i^{(k)} \widehat{\mathbf{w}}_i^{c(k)}),$$

$$\widehat{u}_i^{(k)} = \mathbb{E}(U_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}) = \frac{p_i^o + \nu}{\nu + \widehat{\delta}^{(k)}(\mathbf{y}_i^o)} \frac{L_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu + p_i^o + 2)}{L_{p_i^c}(\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \widehat{S}_i^{cc.o(k)}, \nu + p_i^o)},$$

   where

$$\widetilde{\mathbf{S}}_i^{co(k)} = \left\{ \frac{\nu + \widehat{\delta}^{(k)}(\mathbf{y}_i^o)}{\nu + 2 + p_i^o} \right\} \widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \quad \widehat{\delta}^{(k)}(\mathbf{y}_i^o) = (\mathbf{y}_i^o - \widehat{\boldsymbol{\mu}}_i^{o(k)})^\top (\widehat{\boldsymbol{\Sigma}}_i^{oo(k)})^{-1} (\mathbf{y}_i^o - \widehat{\boldsymbol{\mu}}_i^{o(k)}),$$

$\widehat{\boldsymbol{\Sigma}}_i^{cc.o(k)}$ is defined as in equation (4.22) in the main document, $\widehat{\mathbf{w}}_i^{c(k)}$ and $\widehat{\mathbf{w}}_i^{2^{c(k)}}$ are defined in (A.1) with $\mathbf{W}_i \sim Tt_{p_i^c}(\widehat{\boldsymbol{\mu}}_i^{co(k)}, \widetilde{\mathbf{S}}_i^{co(k)}, \nu + p_i^o + 2; (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c))$. Similarly, to compute $\mathbb{E}[\mathbf{W}_i]$ and $\mathbb{E}[\mathbf{W}_i\mathbf{W}_i^\top]$, we use the results given in Subsection 3.1.

Appendix A.2: Some illustrations using the R MomTrunc package

```
> momentsTMD(kappa=c(2,2,2),lower,upper,mu,Sigma,nu,dist = "t")

Call:
momentsTMD(kappa = c(2, 2, 2), lower, upper, mu,Sigma, dist = "t", nu)

k1 k2 k3    F(k)    E[k]
1  2  2  2  0.0002  0.0017
2  1  2  2 -0.0003 -0.0021
3  0  2  2  0.0021  0.0172
4  0  1  2 -0.0002 -0.0019
5  0  0  2  0.0161  0.1346
6  0  0  1  0.0089  0.0743
7  0  0  0  0.1194  1.0000

> meanvarTMD(lower,upper,mu,Sigma,nu,dist = "t") #Using 5000 MC sims

> means
              mean1   mean2   mean3   mean4   mean5   mean6
Proposed    -0.3587 -0.0837 -0.0781  0.2745  0.8097  0.9313
MonteCarlo  -0.3465 -0.0744 -0.0730  0.2912  0.8022  0.9327

> variances
               var1    var2    var3    var4    var5    var6
Proposed     0.0807  0.0863  0.1018  0.1340  0.0962  0.1459
MonteCarlo   0.0787  0.0888  0.0992  0.1393  0.0890  0.1464

> times
Proposed  MonteCarlo
3.50       11.89     seconds
```

APPENDIX B: Appendix for chapter 3

Appendix B.1: Proofs of propositions and theorems

Proof of Proposition 3.2. Consider the partition $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ and the corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\varphi}$. We based our proof on the factorization of $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1,\mathbf{y}_2)$ as $f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1,\mathbf{y}_2) = f_{\mathbf{Y}_1}(\mathbf{y}_1) f_{\mathbf{Y}_2|\mathbf{Y}_1=\mathbf{y}_1}(\mathbf{y}_2)$. First, for the symmetric part, we have that

$$\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \phi_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\phi_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\boldsymbol{\Sigma}_{22.1}). \tag{B.2}$$

Let now $c_{12} = (1+\boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$ and $\tau_{2.1} = \tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. By noting after some straightforward algebra that $\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}) = \boldsymbol{\varphi}^\top(\mathbf{y}-\boldsymbol{\mu}) = \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1-\boldsymbol{\mu}_1) + \boldsymbol{\varphi}_2^\top(\mathbf{y}_2-\boldsymbol{\mu}_{2.1})$ and $\boldsymbol{\lambda}^\top\boldsymbol{\lambda} = \boldsymbol{\varphi}^\top\boldsymbol{\Sigma}\boldsymbol{\varphi} = \tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}\tilde{\boldsymbol{\varphi}}_1 + \boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2$, we obtain

$$\Phi_1\left(c_{12}\tau + c_{12}\tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))\right) = \Phi_1\left(\frac{\tau_{2.1}}{(1+\boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{1/2}}\right). \tag{B.3}$$

Hence, using (B.2) and (B.3), we can rewrite the density of $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ as

$$\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\frac{\Phi_1(\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))}{\Phi_1(\tau/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2})} \\
&= \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\frac{\Phi_1\left(\tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1-\boldsymbol{\mu}_1) + \boldsymbol{\varphi}_2^\top(\mathbf{y}_2-\boldsymbol{\mu}_{2.1})\right)}{\Phi_1\left(\tau/(1+\boldsymbol{\varphi}^\top\boldsymbol{\Sigma}\boldsymbol{\varphi})^{1/2}\right)} \\
&= \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\frac{\Phi_1\left(\tau_{2.1} + \boldsymbol{\varphi}_2^\top(\mathbf{y}_2-\boldsymbol{\mu}_{2.1})\right)}{\Phi_1\left(\tau/(1+\tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}\tilde{\boldsymbol{\varphi}}_1 + \boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{1/2}\right)} \\
&= \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\frac{\Phi_1\left(\tau_{2.1} + \boldsymbol{\varphi}_2^\top(\mathbf{y}_2-\boldsymbol{\mu}_{2.1})\right)}{\Phi_1\left(c_{12}\tau/(1+c_{12}^2\tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}\tilde{\boldsymbol{\varphi}}_1)^{1/2}\right)}\frac{\Phi_1(c_{12}\tau + c_{12}\tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))}{\Phi_1\left(\tau_{2.1}/(1+\boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{1/2}\right)} \\
&= \phi_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11})\frac{\Phi_1(c_{12}\tau + c_{12}\tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}^{1/2}\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_1-\boldsymbol{\mu}_1))}{\Phi_1\left(c_{12}\tau/(1+c_{12}^2\tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}\tilde{\boldsymbol{\varphi}}_1)^{1/2}\right)} \\
&\quad\times \phi_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\boldsymbol{\Sigma}_{22.1})\frac{\Phi_1\left(\tau_{2.1} + \boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\Sigma}_{22.1}^{-1/2}(\mathbf{y}_2-\boldsymbol{\mu}_{2.1})\right)}{\Phi_1\left(\tau_{2.1}/(1+\boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{1/2}\right)}, \\
&= ESN_{p_1}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1,c_{12}\tau)\times ESN_{p_2}(\boldsymbol{\mu}_{2.1},\boldsymbol{\Sigma}_{22.1},\boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2,\tau_{2.1}).
\end{aligned}$$

□

Proof of Theorem 3.3. For $\mathbf{X} \sim TN_p(\mathbf{0},\mathbf{R};(\mathbf{a},\mathbf{b}))$, we have that its MGF is given by

$$\begin{aligned}
m(\mathbf{t}) = \mathbb{E}[\exp\{\mathbf{t}^\top\mathbf{X}\}] &= \frac{1}{L}\int_{\mathbf{a}}^{\mathbf{b}}\frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}}\exp\left\{-\tfrac{1}{2}(\mathbf{x}^\top\mathbf{R}^{-1}\mathbf{x} - 2\mathbf{t}^\top\mathbf{x})\right\}\mathrm{d}\mathbf{x}, \\
&= L^{-1}\exp\{\mathbf{t}^\top\mathbf{R}\mathbf{t}/2\}\int_{\mathbf{a}}^{\mathbf{b}}\phi_p(\mathbf{x};\mathbf{R}\mathbf{t},\mathbf{R})\mathrm{d}\mathbf{x}, \\
&= L^{-1}\exp\{\mathbf{t}^\top\mathbf{R}\mathbf{t}/2\}\,L_p(\mathbf{a},\mathbf{b};\mathbf{R}\mathbf{t},\mathbf{R}),
\end{aligned} \tag{B.4}$$

with normalizing constant $L \equiv L_p(\mathbf{a}, \mathbf{b}; \mathbf{0}, \mathbf{R})$. From Tallis (1961), we can compute the first two moments of $\mathbf{X}$ differentiating (B.4). Hence,

$$\frac{\partial m(\mathbf{t})}{\partial \mathbf{t}} = m(\mathbf{t})\mathbf{t}^\top \mathbf{R} + L^{-1} \exp\{\mathbf{t}^\top \mathbf{R}\mathbf{t}/2\} \left[\frac{\partial}{\partial \mathbf{t}} L_p(\mathbf{a}, \mathbf{b}; \mathbf{R}\mathbf{t}, \mathbf{R}) \right].$$

After a change of variable $\mathbf{w} = \mathbf{x} - \mathbf{R}\mathbf{t}$,

$$\frac{\partial}{\partial \mathbf{t}} L_p(\mathbf{a}, \mathbf{b}; \mathbf{R}\mathbf{t}, \mathbf{R}) = \frac{\partial}{\partial \mathbf{w}} \frac{\partial \mathbf{w}}{\partial \mathbf{t}} \int_{\mathbf{a}}^{\mathbf{b}} \phi_p(\mathbf{w}; \mathbf{R}) \mathrm{d}\mathbf{w} = -\mathbf{q}(\mathbf{t})^\top \mathbf{R}. \tag{B.5}$$

For $\mathbf{t} = \mathbf{0}$, we denote $\mathbf{q} \equiv \mathbf{q}(\mathbf{0})$, which is given by $\mathbf{q} = \mathbf{q}_a - \mathbf{q}_b$, with the i-th element of $\mathbf{q}_a$ and $\mathbf{q}_b$ as

$$\begin{aligned} q_{a,i} &= \phi_1(a_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; a_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i), \\ q_{b,i} &= \phi_1(b_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; b_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i), \end{aligned}$$

with $\tilde{\mathbf{R}}_i = \mathbf{R}_{(i),(i)} - \mathbf{R}_{(i),i}\mathbf{R}_{i,(i)}$. Additionally, it is straightforward that

$$\frac{\partial^2 m(\mathbf{t})}{\partial \mathbf{t} \partial \mathbf{t}^\top} = \frac{\partial m(\mathbf{t})}{\partial \mathbf{t}}^\top \mathbf{t}\mathbf{R} + m(\mathbf{t})\mathbf{R} - L^{-1} \exp\{\mathbf{t}^\top \mathbf{R}\mathbf{t}/2\}\, \mathbf{R}(\mathbf{t}\, \mathbf{q}(\mathbf{t})^\top - \mathbf{H}(\mathbf{t}))\mathbf{R}, \tag{B.6}$$

with $\mathbf{H}(\mathbf{t}) = -\frac{\partial}{\partial \mathbf{t}} \mathbf{q}(\mathbf{t})$. For $\mathbf{t} = \mathbf{0}$, we have that $\mathbf{H} \equiv \mathbf{H}(\mathbf{0}) - \frac{\partial}{\partial \mathbf{x}} \mathbf{q}$, with off-diagonal elements h_{ij} given by

$$\begin{aligned} h_{ij} &= h_{ij}^{aa} - h_{ij}^{ba} - h_{ij}^{ab} + h_{ij}^{bb} \\ &= \phi_2(a_i, a_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{aa}, \tilde{\mathbf{R}}_{ij}) - \phi_2(b_i, a_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{ba}, \tilde{\mathbf{R}}_{ij}) \\ &\quad - \phi_2(a_i, b_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{ab}, \tilde{\mathbf{R}}_{ij}) + \phi_2(b_i, b_j; \rho_{ij})\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{bb}, \tilde{\mathbf{R}}_{ij}), \end{aligned}$$

with $\boldsymbol{\mu}_{ij}^{\alpha\beta} = \mathbf{R}_{(ij),[i,j]}(\alpha_i, \beta_j)^\top$ and $\tilde{\mathbf{R}}_{ij} = \mathbf{R}_{(i,j),(i,j)} - \mathbf{R}_{(i,j),[i,j]}\mathbf{R}_{[i,j],(i,j)}$.

Finally, following Vaida & Liu (2009), we can derive the diagonal elements h_{ii} as linear combinations of the elements h_{ik} for $i \neq k$. This can be achieved as

$$\begin{aligned} h_{ii} &= -\frac{\partial}{\partial x_i} q_{a_i} + \frac{\partial}{\partial x_i} q_{b_i} \\ &= \frac{\partial}{\partial x_i} \left\{ \phi_1(a_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; a_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) - \phi_1(b_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; b_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) \right\}, \\ &= a_i \phi_1(a_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; a_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) - \phi_1(a_i) \cdot \frac{\partial}{\partial x_i} L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; a_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) \\ &\quad - b_i \phi_1(b_i)\, L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; b_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) + \phi_1(b_i) \cdot \frac{\partial}{\partial x_i} L_{p-1}(\mathbf{a}_{(i)}, \mathbf{b}_{(i)}; b_i \mathbf{R}_{(i),i}, \tilde{\mathbf{R}}_i) \\ &= a_i q_{a_i} - b_i q_{b_i} - \mathbf{R}_{i,(i)} \left\{ \phi_1(a_i) \left[\phi_1(a_j | a_i)\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{aa}, \tilde{\mathbf{R}}_{ij}) - \phi_1(b_j | a_i) \right.\right. \\ &\quad \times \left. L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{ab}, \tilde{\mathbf{R}}_{ij}) \right]_{j \neq i=1}^{p} - \phi_1(b_i) \left[\phi_1(a_j | b_i)\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{ba}, \tilde{\mathbf{R}}_{ij}) \right. \\ &\quad - \left.\left. \phi_1(b_j | b_i)\, L_{p-2}(\mathbf{a}_{(i,j)}, \mathbf{b}_{(i,j)}; \boldsymbol{\mu}_{ij}^{bb}, \tilde{\mathbf{R}}_{ij}) \right]_{j \neq i=1}^{p} \right\} \end{aligned}$$

$$= a_i q_{a_i} - b_i q_{b_i} - \mathbf{R}_{i,(i)} \left[h_{ij}^{aa} - h_{ij}^{ab} - h_{ij}^{ba} - h_{ij}^{bb} \right]_{j \neq i=1}^{p}$$
$$= a_i q_{a_i} - b_i q_{b_i} - \mathbf{R}_{i,(i)} \mathbf{H}_{(i),i}.$$

Finally, evaluating equations (B.5) and (B.6) on $\mathbf{t} = \mathbf{0}$, we obtain the expressions for $\mathbb{E}[\mathbf{X}]$ and $\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$. This ends the proof. □

Proof of Theorem 3.4. It follows that

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{y}) &= P(-\mathbf{y} \leqslant \mathbf{X} \leqslant \mathbf{y}) \\ &= P(-y_1 \leqslant X_1 \leqslant y_1, -y_2 \leqslant X_2 \leqslant y_2, \ldots, -y_p \leqslant X_p \leqslant y_p) \\ &= F_{\mathbf{X}}(\mathbf{y}) - \sum_i F_{\mathbf{X}}(\mathbf{y}_{\text{-}(i)}) + \sum_{i<j} F_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j)}) - \sum_{i<j<k} F_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j,k)}) + \ldots + (-1)^p F_{\mathbf{X}}(-\mathbf{y}), \end{aligned} \tag{B.7}$$

where $\mathbf{y}_{-(\mathbf{i})}$ denotes the $\mathbf{y}$ vector with its $\mathbf{i}th$ elements multiplied by -1. For instance, we have that $\mathbf{y}_{\text{-}(i)} = (y_1, y_2, \ldots, y_{i-1}, -y_i, y_{i+1}, \ldots, y_p)$. It is easy to see that $F_{\mathbf{Y}}(\mathbf{y})$ can be written as

$$F_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{s} \in S(p)} \pi_s F_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y}; \boldsymbol{\theta}),$$

with the constant $\pi_s = \prod_{i=1}^{p} s_i$ providing the signs $\{-1, 1\}$ correctly for each summand in (B.7).

By the other side, differentiating $F_{\mathbf{Y}}(\mathbf{y})$ in expression (B.7), we have the joint pdf of $\mathbf{Y} = |\mathbf{X}|$ given by

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{\partial^p}{\partial y_1 \partial y_2 \ldots \partial y_p} F_{\mathbf{Y}}(\mathbf{y}) \\ &= f_{\mathbf{X}}(\mathbf{y}) - (-1) \sum_i f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i)}) + (-1)^2 \sum_{i<j} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j)}) - (-1)^3 \sum_{i<j<k} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j,k)}) \\ &\quad + \ldots + (-1)^{2p} f_{\mathbf{X}}(-\mathbf{y}) \\ &= f_{\mathbf{X}}(\mathbf{y}) + \sum_i f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i)}) + \sum_{i<j} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j)}) + \sum_{i<j<k} f_{\mathbf{X}}(\mathbf{y}_{\text{-}(i,j,k)}) + \ldots + f_{\mathbf{X}}(-\mathbf{y}) \\ &= \sum_{\mathbf{s} \in S(p)} f_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y}; \boldsymbol{\theta}), \end{aligned}$$

where we have conveniently used $f_{\mathbf{X}}(\mathbf{x})$ instead of $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ for simplicity. □

Proof of Corollary 3.2. By the method of change-of-variable for $\mathbf{Z}_s = \mathbf{\Lambda}_s \mathbf{X}$, then $f_{\mathbf{Z}_s}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{\Lambda}_s \mathbf{y})$ since $\mathbf{\Lambda}_s^{-1} = \mathbf{\Lambda}_s$, $\mathbf{J} = \mathbf{\Lambda}_s$ and $|\det(\mathbf{J})| = 1$, where $\mathbf{J}$ is the Jacobian matrix of the transformation and $det(\mathbf{A})$ is the determinat of the matrix $\mathbf{A}$. Additionally, if $\mathbf{X} \sim f_{\mathbf{X}}(\cdot; \boldsymbol{\xi}, \mathbf{\Psi})$ belongs to the location-scale family of distributions with location and

scale parameters $\boldsymbol{\xi}$ and $\boldsymbol{\Psi}$, respectively, then $\mathbf{Z}_s \sim f_{\mathbf{X}}(\cdot; \boldsymbol{\Lambda}_s\boldsymbol{\xi}, \boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s)$. The $\boldsymbol{\kappa}$-th moment of $\mathbf{Y}$ can be obtained by the basic integration as

$$\begin{aligned}\int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}\mathbf{y} &= \sum_{\mathbf{s}\in S(p)} \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{X}}(\mathbf{y}; \boldsymbol{\Lambda}_s\boldsymbol{\xi}, \boldsymbol{\Lambda}_s\boldsymbol{\Psi}\boldsymbol{\Lambda}_s)\mathrm{d}\mathbf{y} \\ &= \sum_{\mathbf{s}\in S(p)} \int_{\mathbf{0}}^{\infty} \mathbf{y}^{\boldsymbol{\kappa}} f_{\mathbf{Z}_s}(\mathbf{y})\mathrm{d}\mathbf{y} = \sum_{\mathbf{s}\in S(p)} \mathbb{E}[(\mathbf{Z}_s^{\boldsymbol{\kappa}})^+].\end{aligned}$$

This concludes the proof. □

Appendix B.2: Explicit expressions for moments of some folded univariate distributions

Let $X \sim \mathrm{ESN}(\mu, \sigma^2, \lambda, \tau)$, $Y \sim \mathrm{SN}(\mu, \sigma^2, \lambda)$, $Z \sim \mathrm{N}(\mu, \sigma^2)$ and W follow a univariate half normal distribution denoted by $W \sim \mathrm{HN}(\sigma^2)$. The first four raw moments for $|X|$, $|Y|$, $|Z|$ and W are given by

$$\begin{aligned}
\mathbb{E}[|X|] &= \mu(1-2p_1) + 2\alpha\sigma^2 + \lambda\eta\sigma(1 - 2\Phi_1(0; m, \gamma^2)), \\
\mathbb{E}[|X|^2] &= \mu^2 + \sigma^2 + \lambda\eta\sigma(m + \mu), \\
\mathbb{E}[|X|^3] &= (\mu^3 + 3\mu\sigma^2)(1-2p_1) + 2\alpha(\mu^2\sigma^2 + 2\sigma^4) \\
&\quad + \lambda\eta\sigma\left\{2\gamma(m+\mu)\phi_1(m/\gamma) + (m^2 + \gamma^2 + \mu(m+\mu) + 2\sigma^2)(1 - 2\Phi_1(0; m, \gamma^2))\right\}, \\
\mathbb{E}[|X|^4] &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 + \lambda\eta\sigma\left\{m^3 + 3m\gamma^2 + m^2\mu + \gamma^2\mu + m\mu^2 + \mu^3 + (3m + 5\mu)\sigma^2\right\},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[|Y|] &= \mu(1-2p_1) + 2\alpha\sigma^2 - \lambda\eta\sigma(1 - 2\Phi_1(\mu/\gamma)), \\
\mathbb{E}[|Y|^2] &= \mu^2 + \sigma^2 + 2\mu\lambda\eta\sigma, \\
\mathbb{E}[|Y|^3] &= (\mu^3 + 3\mu\sigma^2)(1-2p_1) + 2\alpha(\mu^2\sigma^2 + 2\sigma^4) \\
&\quad + \lambda\eta\sigma\left\{4\gamma\mu\phi_1(\mu/\gamma) - (3\mu^2 + \gamma^2 + 2\sigma^2)(1 - 2\Phi_1(\mu/\gamma))\right\}, \\
\mathbb{E}[|Y|^4] &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 + 4\mu\lambda\eta\sigma\left\{\mu^2 + \gamma^2 + 2\sigma^2\right\},
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[|Z|] &= -\mu(1 - 2\Phi_1(\mu/\sigma)) + 2\sigma\phi_1(\mu/\sigma), \\
\mathbb{E}[|Z|^2] &= \mu^2 + \sigma^2, \\
\mathbb{E}[|Z|^3] &= -\mu(\mu^2 + 3\sigma^2)(1 - 2\Phi_1(\mu/\sigma)) + 2\sigma(\mu^2 + 2\sigma^2)\phi_1(\mu/\sigma), \\
\mathbb{E}[|Z|^4] &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4,
\end{aligned}$$

and

$$\mathbb{E}[W] = \frac{\sigma\sqrt{2}}{\sqrt{\pi}}, \qquad \mathbb{E}[W^2] = \sigma^2, \qquad \mathbb{E}[W^3] = \frac{4\sigma^3}{\sqrt{2\pi}} \qquad \text{and} \qquad \mathbb{E}[W^4] = 3\sigma^4,$$

with $m = \mu - \mu_b$, $p_1 = \tilde{\Phi}_1(0; \mu, \sigma^2, \lambda, \tau)$ and $\alpha = ESN_1(0; \mu, \sigma^2, \lambda, \tau)$.

Appendix B.3: Useful approximations.

Proposition B.1. *As* $\tau \to -\infty$ *,*

$$\eta \longrightarrow -\frac{\tau}{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}. \tag{B.8}$$

Proof. Let h denote the complimentary inverse Mill's ratio (CIMR) of a random variable X, given by $h(x) \triangleq f(x)/F(x)$. For $X \sim N_1(\mu, \sigma^2)$, it follows from L'Hôpital that

$$h(x) \longrightarrow -\frac{x-\mu}{\sigma^2}, \quad \text{as} \quad x \longrightarrow -\infty.$$

Setting $X \sim N_1(0, 1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})$, it follows that $h(\tau) = \eta$, ending the proof. □

Proposition B.2. *As* $\lambda \to -\infty$*,*

$$ESN_1(y;\mu,\sigma^2,\lambda,\tau) \longrightarrow TN_1(y;\mu,\sigma^2,[\mu,\infty)), \tag{B.9}$$

and as $\lambda \to +\infty$*,*

$$ESN_1(y;\mu,\sigma^2,\lambda,\tau) \longrightarrow TN_1(y;\mu,\sigma^2,(-\infty,\mu]). \tag{B.10}$$

Proof. For $\lambda \to -\infty$, it is straightforward that

$$\frac{\Phi_1(\tau+\lambda(Y-\mu)/\sigma)}{\Phi_1(\tau/(1+\lambda^2)^{1/2})} \longrightarrow \frac{\mathbb{1}\{Y<\mu\}}{1/2},$$

and for $\lambda \to +\infty$,

$$\frac{\Phi_1(\tau+\lambda(Y-\mu)/\sigma)}{\Phi_1(\tau/(1+\lambda^2)^{1/2})} \longrightarrow \frac{\mathbb{1}\{Y>\mu\}}{1/2},$$

where $\mathbb{1}\{E\}$ represents the indicator function. This completes the proof. □

Corollary B.1. *Let* M *and* N *be two large positive real numbers. If* $\lambda = -M$ *and* $\tau = \pm N$*, then*

$$ESN_1(y;\mu,\sigma^2,\lambda,\tau) \approx TN_1(y;\mu,\sigma^2,[\mu \pm \sigma N/M, \infty)),$$

and for $\lambda = M$*, then*

$$ESN_1(y;\mu,\sigma^2,\lambda,\tau) \approx TN_1(y;\mu,\sigma^2,(-\infty,\mu - \pm\sigma N/M]).$$

Appendix B.4: The `MomTrunc` R package

The methods proposed this work have been implemented in the package `MomTrunc`, which is available on CRAN repository (version 4.51). It computes the first two moments, as well as arbitrary moments for some multivariate truncated distributions (TMD) using the functions `meanvarTMD` and `momentsTMD`, respectively. Another possible distributions includes the Student-t and the ESN along with its limiting cases, say, the

SN and N distribution. These moments can be accessed by setting the `dist` parameter as `"t"`, `"ESN"`, `"SN"` and `"N"` respectively. For folded one can use the analogous functions `meanvarFMD` and `momentsFMD` and their *cdf*s through the `cdfFMD` function. Densities, probabilities and random generator functions are also offered for the multivariate ESN distribution trough the functions `dmvESN`, `pmvESN` and `rmvESN`, respectively. In the following, we present some sample codes useful for practitioners.

```
# Univariate ESN case
> dmvESN(x = -1,mu = 2,Sigma = 5,lambda = -2,tau = 0.5)
> rmvESN(n = 100,mu = 2,Sigma = 5,lambda = -2,tau = 0.5)
> pmvESN(lower = -5,upper = 2,mu = 2,Sigma = 5,lambda = -2,tau = 0.5)

# Multivariate ESN case
> mu = c(0.1,0.2,0.3,0.4)
> Sigma = matrix(data = c(1,0.2,0.3,0.1,0.2,1,0.4,-0.1,0.3,0.4,1,0.2,0.1,
-0.1,0.2,1),nrow = length(mu),ncol = length(mu),byrow = TRUE)
> lambda = c(-2,0,1,2)
> tau = 1

> dmvESN(x = c(-2,-1,0,1),mu,Sigma,lambda,tau)  #One observation
> dmvESN(x = matrix(rnorm(4*10),ncol = 4),mu,Sigma,lambda,tau)
> rmvESN(n = 100,mu,Sigma,lambda,tau)
> pmvESN(lower = rep(-Inf,4),upper = c(-1,0,2,5),mu,Sigma,lambda,tau)

# Truncated case
# First two moments
> a = c(-0.8,-0.7,-0.6) #lower bound
> b = c(0.5,0.6,0.7)    #upper bound
> mu = c(0.1,0.2,0.3)
> Sigma = matrix(data = c(1,0.2,0.3,0.2,1,0.4,0.3,0.4,1),
nrow = length(mu),ncol = length(mu),byrow = TRUE)
> lambda = c(-2,0,1)
> meanvarTMD(a,b,mu,Sigma,dist="normal")
> meanvarTMD(a,b,mu,Sigma,dist = "t",nu = 4)
> meanvarTMD(a,b,mu,Sigma,lambda,dist = "SN")
> meanvarTMD(a,b,mu,Sigma,lambda,tau = 1,dist = "ESN")

# Arbitrary moment (2,0,1)
> momentsTMD(kappa = c(2,0,1),a,b,mu,Sigma,dist="normal")
```

```
> momentsTMD(kappa = c(2,0,1),a,b,mu,Sigma,dist = "t",nu = 7)
> momentsTMD(kappa = c(2,0,1),a,b,mu,Sigma,lambda,dist = "SN")
> momentsTMD(kappa = c(2,0,1),a,b,mu,Sigma,lambda,tau = 1,dist = "ESN")

# Folded ESN case
> meanvarFMD(mu,Sigma,lambda,tau = 1,dist = "ESN")
> momentsFMD(kappa = c(2,0,1),mu,Sigma,lambda,tau = 1,dist = "ESN")
> cdfFMD(x = c(0.5,0.2,1.0,1.3),mu,Sigma,lambda,tau = 1,dist = "ESN")
```

Appendix B.5: Figures

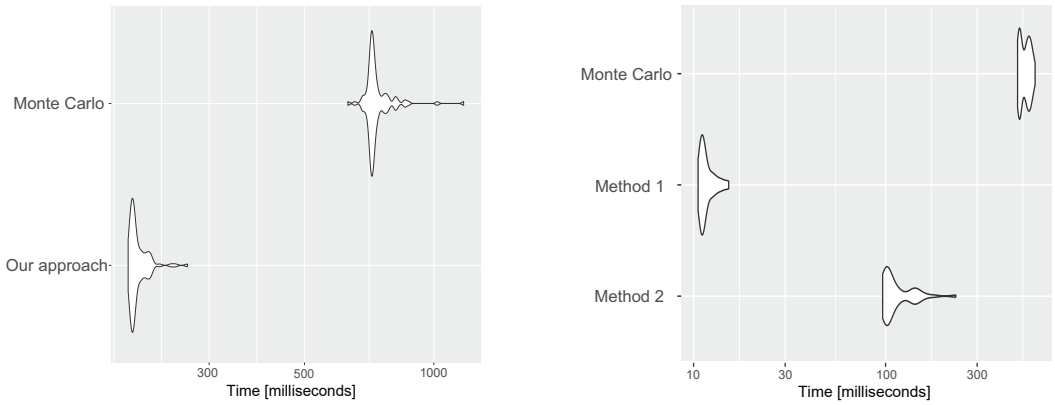
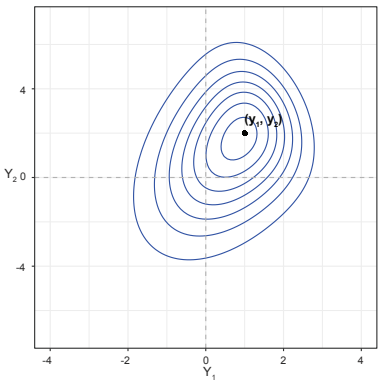
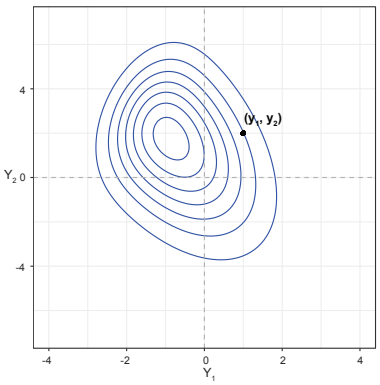


Figure 15 – Simulation study. Violin plots for the processing time to compute the mean and the variance for a 4-variate doubly TESN (left panel) and a 3-variate FESN distribution (right panel). For the FESN case, Method 1 refers to the approach in Subsection 3.6.1 and Method 2 when using equations (3.40) and (3.41).

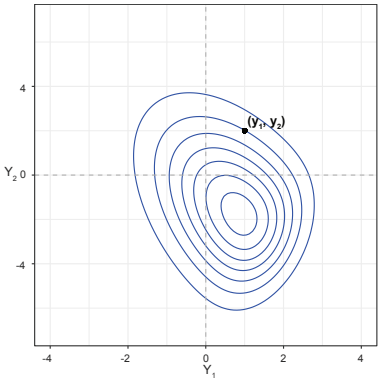
(a) Density of $\mathbf{X}_1$



(b) Density of $\mathbf{X}_2$



(c) Density of $\mathbf{X}_3$



(d) Density of $\mathbf{X}_4$

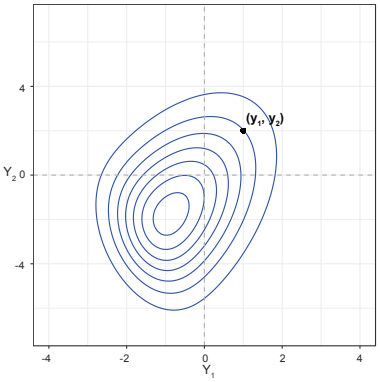


Figure 16 – Densities of X_i, $i = 1, \ldots, 4$.

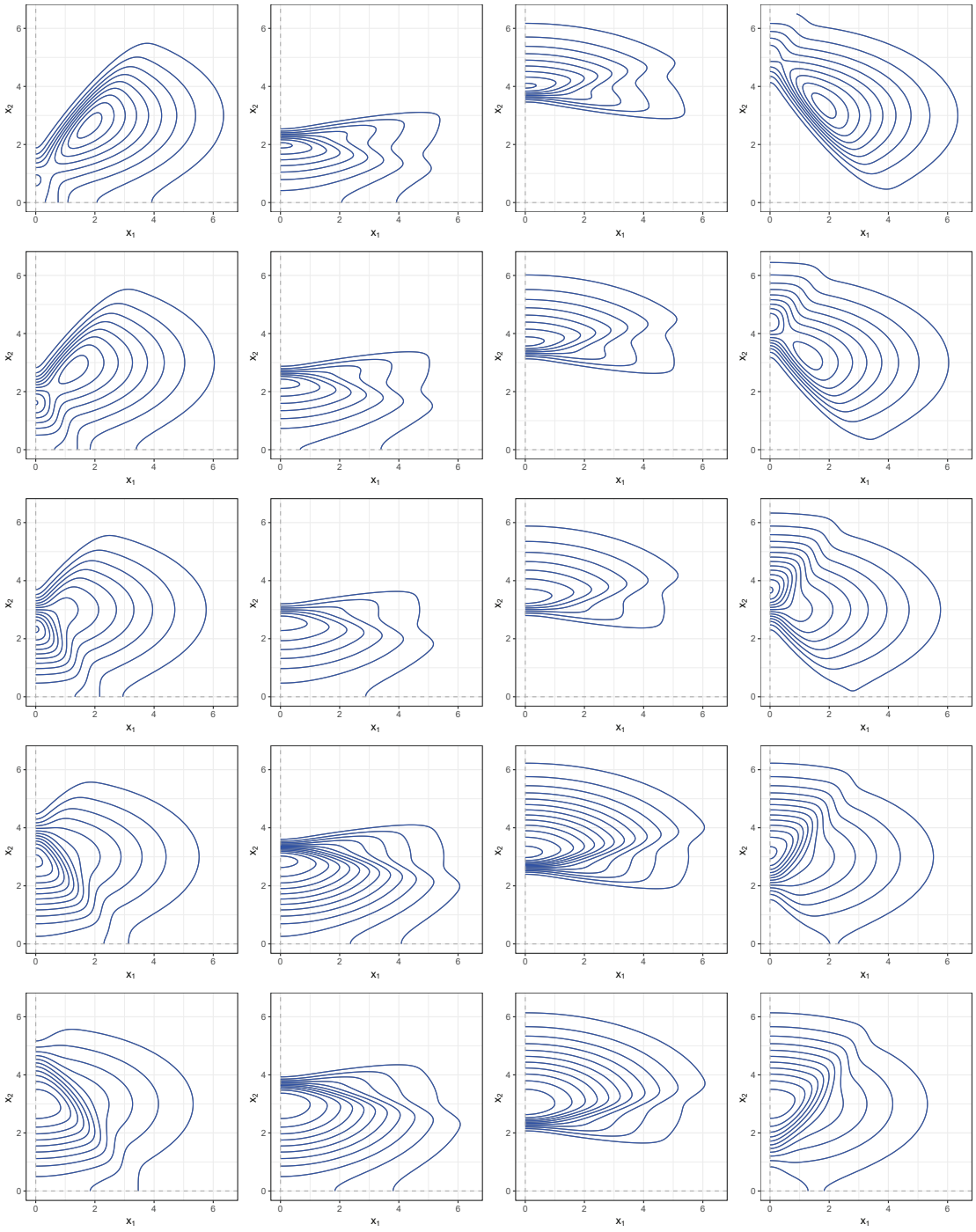


Figure 17 – Contour plots of bivariate FESN densities with same location and scale parameters, and different skewness $\boldsymbol{\lambda} = \{(8,3),(3,8),(-3,-8),(-8,-3)\}$ (from left to right) and extension $\tau = \{-4,-2,0,2,4\}$ (from top to bottom) parameters.

APPENDIX C: Appendix for chapter 5

Appendix C.1: Details for the expectations in EM algorithm

To compute the expected values, first note that for any multiplicatively separable measurable function of U_i, T_i and $\mathbf{Y}_i$, such that $g(U_i, T_i, \mathbf{Y}_i) = g_1(\mathbf{Y}_i)g_2(U_i)g_3(T_i)$, we have that

$$\begin{aligned}\mathbb{E}_{U_iT_i\mathbf{Y}_i}[g(U_i, T_i, \mathbf{Y}_i)|\mathbf{V}_i, \mathbf{C}_i] &= \mathbb{E}_{\mathbf{Y}_i}[g_1(\mathbf{Y}_i)\,\mathbb{E}_{U_iT_i}[g_2(U_i)g_3(T_i)|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i] \\ &= \mathbb{E}_{\mathbf{Y}_i}[g_1(\mathbf{Y}_i)\,\mathbb{E}_{U_i}[g_2(U_i)|\mathbf{Y}_i]\,\mathbb{E}_{T_i}[g_3(T_i)|U_i, \mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i].\end{aligned}$$

Hence,

$$\begin{aligned}\widehat{u\mathbf{y}_i^r} &= \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_i\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i], \\ \widehat{ut_i^r} &= \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i^r|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbb{E}_{U_iT_i}[U_iT_i^r|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i], \\ \widehat{ut\mathbf{y}_i^r} &= \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i]|\mathbf{V}_i, \mathbf{C}_i],\end{aligned}$$

for $r = \{0, 1, 2\}$. From Cabral *et al.* (2012), we know that $T_i \mid (\mathbf{Y}_i, U_i) \sim TN_1(\varrho^2\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}), U_i^{-1}\varrho^2, (0, \infty))$. Multiplying the first and second moment of $T_i|\mathbf{Y}_i$ by U_i and taking expectation with respect to this last, it follows that

$$\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i] = \varrho^2\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] + \varrho\phi(\boldsymbol{\theta}, \mathbf{y}_i), \tag{C.11}$$

$$\mathbb{E}_{U_iT_i}[U_iT_i^2|\mathbf{Y}_i] = \varrho^2[\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i] + 1], \tag{C.12}$$

with $\varrho = (1 + \boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1/2}$ and

$$\phi(\boldsymbol{\theta}, \mathbf{y}_i) = \mathbb{E}_{U_i}\left[U_i^{1/2}\frac{\phi_1(U_i^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))}{\Phi_1(U_i^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))}\middle|\mathbf{Y}_i\right].$$

As noted, both expectations $\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i]$ and $\mathbb{E}_{U_iT_i}[U_iT_i^2|\mathbf{Y}_i]$ depend on $\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i]$ and $\phi(\boldsymbol{\theta}, \mathbf{Y}_i)$. Lachos *et al.* (2010) states that

$$\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] = \frac{2\nu^2(\mathbf{y}_i)\,t_p(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{ST_p(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)}T_1\left(\sqrt{\frac{\nu + p + 2}{\nu + \delta_i}}A_i; \nu + p + 2\right)$$

and

$$\phi(\boldsymbol{\theta}, \mathbf{y}_i) = \frac{2\,t_p(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)}{ST_p(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)}\frac{\Gamma((\nu + p + 1)/2)}{\sqrt{\pi}\Gamma((\nu + p)/2)}\frac{(\nu + \delta_i)^{(\nu+p)/2}}{(\nu + \delta_i + A_i^2)^{(\nu+p+1)/2}},$$

where $\delta_i = \delta(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $A_i = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu})$.

By using the fact that $t_p(\mathbf{y}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = t_p(\mathbf{y}_i; \boldsymbol{\mu}, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}, \nu + 2)/\nu^2(\mathbf{y}_i)$, $\delta_i = \frac{\nu}{\nu+2}\delta(\mathbf{y}_i; \boldsymbol{\mu}, \frac{\nu}{\nu+2}\boldsymbol{\Sigma})$, $\delta_i + A_i^2 = \frac{\nu}{\nu+1}\delta(\mathbf{y}_i; \boldsymbol{\mu}, \frac{\nu}{\nu+1}\boldsymbol{\Gamma})$, $\det(\boldsymbol{\Sigma})^{1/2} = \sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}\det(\boldsymbol{\Gamma})^{1/2}$ and

equation (C.24), we can propose simplified versions of equations above in a neat manner. After some straightforward algebra, we obtain

$$\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] = \frac{ST_p(\mathbf{y}_i;\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{ST_p(\mathbf{y}_i;\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \tag{C.13}$$

and

$$\phi(\boldsymbol{\theta},\mathbf{y}_i) = \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{t_p(\mathbf{y}_i;\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{ST_p(\mathbf{y}_i;\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}. \tag{C.14}$$

Let us define the expectation of interest $\widehat{\phi\mathbf{y}_i^r} = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r\phi(\boldsymbol{\theta},\mathbf{Y}_i)|\mathbf{V}_i,\mathbf{C}_i]$, for $r=\{0,1,2\}$. Next, we present two crucial propositions to compute these expectations. Proofs can be found in next subsection C.2.

Proposition C.1. *Let* $\mathbf{Z}\sim N_p(\mathbf{0},\boldsymbol{\Gamma})$, $U\sim\text{Gamma}(\nu/2,\nu/2)$ *and* $T\sim\text{HT}(\nu)$. *Then* $\mathbf{Y}\overset{d}{=}\boldsymbol{\mu}+\boldsymbol{\Delta}T+U^{-1/2}\mathbf{Z}\sim ST_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$. *For any measurable function* $g(\mathbf{y})$, *it holds that*

$$\begin{aligned}\mathbb{E}[\phi(\boldsymbol{\theta},\mathbf{Y})g(\mathbf{Y})|\,\boldsymbol{\alpha}\leqslant\mathbf{Y}\leqslant\boldsymbol{\beta}] = &\frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\\ &\times\frac{L(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\mathbb{E}[g(\mathbf{W}_1)],\end{aligned} \tag{C.15}$$

and

$$\mathbb{E}[U\,g(\mathbf{Y})|\boldsymbol{\alpha}\leqslant\mathbf{Y}\leqslant\boldsymbol{\beta}] = \frac{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\mathbb{E}[g(\mathbf{W}_2)], \tag{C.16}$$

where $A=\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu})$, $\mathbf{W}_1\sim Tt_p(\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1;(\boldsymbol{\alpha},\boldsymbol{\beta}))$ *and* $\mathbf{W}_2\sim TST_p(\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2;(\boldsymbol{\alpha},\boldsymbol{\beta}))$.

Proposition C.2. *Consider* $\mathbf{Y}$, U *and* T *as in Proposition C.1. Now, consider* $\mathbf{Y}$ *to be partitioned as* $\mathbf{Y}=(\mathbf{Y}_1^\top,\mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 $(p_1+p_2=p)$, *respectively. Let*

$$\boldsymbol{\Gamma}=\begin{pmatrix}\boldsymbol{\Gamma}_{11}&\boldsymbol{\Gamma}_{12}\\ \boldsymbol{\Gamma}_{21}&\boldsymbol{\Gamma}_{22}\end{pmatrix},\quad \boldsymbol{\alpha}=(\boldsymbol{\alpha}_1^\top,\boldsymbol{\alpha}_2^\top)^\top,\quad \textit{and}\quad \boldsymbol{\beta}=(\boldsymbol{\beta}_1^\top,\boldsymbol{\beta}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Gamma}$, $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$. *For a multiplicatively separable measurable function* g, *it follows that*

$$\mathbb{E}[\phi(\boldsymbol{\theta},\mathbf{Y})g(\mathbf{Y})|\,\mathbf{Y}_1,\boldsymbol{\alpha}_2\leqslant\mathbf{Y}_2\leqslant\boldsymbol{\beta}_2] \tag{C.17}$$

$$\begin{aligned}=&g_1(\mathbf{Y}_1)\frac{t_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+1}\boldsymbol{\Gamma}_{11},\nu+1)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)}\\ &\times\frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{L(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1},\nu_{2.1}+1)}{\mathcal{L}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}\mathbb{E}[g_2(\mathbf{W}_1^*)],\end{aligned} \tag{C.18}$$

and

$$\mathbb{E}[U\,g(\mathbf{Y})|\mathbf{Y}_1,\boldsymbol{\alpha}_2\leqslant\mathbf{Y}_2\leqslant\boldsymbol{\beta}_2] = g_1(\mathbf{Y}_1)\frac{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+2}\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu+2)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)}$$

$$\times \frac{\mathcal{L}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1}+2)}{\mathcal{L}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})} \mathbb{E}[g_2(\mathbf{W}_2^*)], \tag{C.19}$$

where $g(\mathbf{Y}) = g_1(\mathbf{Y}_1)g_2(\mathbf{Y}_2)$, $\mathbf{W}_1^* \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1}, \nu_{2.1}+1; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$ *and* $\mathbf{W}_2^* \sim TEST_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1}+2; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$, *with* $\tau_{2.1} = \nu(\mathbf{y}_1)(\tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))$, $\nu_{2.1} = \nu + p_1$, $\tilde{\boldsymbol{\Gamma}}_{22.1} = (\boldsymbol{\Gamma}_{22} - \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12})/\nu^2(\mathbf{y}_1)$ *and remaining parameters as in Proposition 5.4.*

Subsequently, on according to expressions (C.11) - (C.19), we have the implementable expressions to the conditional expectations for three possible scenarios:

1. If the ith subject has only non-censored components, $\mathbb{E}_{U_iT_i\mathbf{Y}_i}[\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i] = \mathbf{y}_i^r$; then

$$\begin{aligned}
\widehat{u\mathbf{y}_i^r}^{(k)} &= \hat{u}_i^{(k)}\mathbf{y}_i^r, \\
\hat{u}_i^{(k)} &= \mathbb{E}_{U_i}[U_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}] \\
\widehat{ut_i^r}^{(k)} &= \mathbb{E}_{U_iT_i}[U_iT_i^r|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}], \\
\widehat{ut\mathbf{y}_i^r}^{(k)} &= \mathbf{y}_i^r\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i, \hat{\boldsymbol{\theta}}^{(k)}],
\end{aligned}$$

   with $\widehat{\phi\mathbf{y}_i^r} = \mathbf{y}_i^r\phi(\boldsymbol{\theta}, \mathbf{y}_i)$, where $\mathbf{y}_i^0 = 1$, $\mathbf{y}_i^1 = \mathbf{y}_i$ and $\mathbf{y}_i^2 = \mathbf{y}_i\mathbf{y}_i^\top$.

2. If the ith subject has only censored components, we have

$$\begin{aligned}
\widehat{u\mathbf{y}_i^r}^{(k)} &= \hat{u}_i{}^{(k)}\widehat{\mathbf{w}_{2i}^r}^{(k)}, \\
\hat{u}_i{}^{(k)} &= \frac{\mathcal{L}(\mathbf{v}_{1i}, \mathbf{v}_{2i}; \hat{\boldsymbol{\mu}}^{(k)}, \frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+2}\hat{\boldsymbol{\Sigma}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)}, \hat{\nu}^{(k)}+2)}{\mathcal{L}(\mathbf{v}_{1i}, \mathbf{v}_{2i}; \hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)}, \hat{\nu}^{(k)})}, \\
\widehat{ut_i}^{(k)} &= \hat{\varrho}^{2(k)}\hat{\boldsymbol{\Delta}}^{(k)\top}\hat{\boldsymbol{\Gamma}}^{-1(k)}\left(\widehat{u\mathbf{y}_i}^{(k)} - \hat{\boldsymbol{\mu}}^{(k)}\hat{u}_i{}^{(k)}\right) + \hat{\varrho}^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)}, \\
\widehat{ut_i^2}^{(k)} &= \hat{\varrho}^{2(k)}\hat{\boldsymbol{\Delta}}^{(k)\top}\hat{\boldsymbol{\Gamma}}^{-1(k)}\Big[(\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}_i}^{(k)}\hat{\boldsymbol{\mu}}^{\top(k)} + \hat{u}_i{}^{(k)}\hat{\boldsymbol{\mu}}^{(k)}\hat{\boldsymbol{\mu}}^{\top(k)})\hat{\boldsymbol{\Gamma}}^{-1(k)}\hat{\boldsymbol{\Delta}}^{(k)} \\
&\quad + \hat{\varrho}^{(k)}(\widehat{\tau\mathbf{y}_i^1}^{(k)} - \hat{\boldsymbol{\mu}}^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)})\Big] + \hat{\varrho}^{2(k)}, \\
\widehat{ut\mathbf{y}_i}^{(k)} &= \hat{\varrho}^{2(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}_i}^{(k)}\hat{\boldsymbol{\mu}}^{(k)\top})\hat{\boldsymbol{\Gamma}}^{-1(k)}\hat{\boldsymbol{\Delta}}^{(k)} + \hat{\varrho}^{(k)}\widehat{\phi\mathbf{y}_i^1}^{(k)},
\end{aligned}$$

   with

$$\widehat{\phi\mathbf{y}_i^r}^{(k)} = \frac{2}{\sqrt{\pi\hat{\nu}^{(k)}(1 + \hat{\boldsymbol{\lambda}}^{(k)\top}\hat{\boldsymbol{\lambda}}^{(k)})}} \frac{\Gamma(\frac{\hat{\nu}^{(k)}+1}{2})}{\Gamma(\frac{\hat{\nu}^{(k)}}{2})} \frac{L(\mathbf{v}_{1i}, \mathbf{v}_{2i}; \hat{\boldsymbol{\mu}}^{(k)}, \frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+1}\hat{\boldsymbol{\Gamma}}^{(k)}, \hat{\nu}^{(k)}+1)}{\mathcal{L}(\mathbf{v}_{1i}, \mathbf{v}_{2i}; \hat{\boldsymbol{\mu}}^{(k)}, \hat{\boldsymbol{\Sigma}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)}, \hat{\nu}^{(k)})} \widehat{\mathbf{w}}_{1i}^{r(k)}, \tag{C.20}$$

   where

$$\widehat{\mathbf{w}}_{si}^{(k)} = \mathbb{E}[\mathbf{W}_{si} \mid \hat{\boldsymbol{\theta}}^{(k)}], \quad \text{and} \quad \widehat{\mathbf{w}}_{si}^{2(k)} = \mathbb{E}[\mathbf{W}_{si}\mathbf{W}_{si}^\top \mid \hat{\boldsymbol{\theta}}^{(k)}], \tag{C.21}$$

   for $s = \{1, 2\}$, with $\mathbf{W}_{1i} \sim Tt_{p_i}(\boldsymbol{\mu}, \frac{\nu}{\nu+1}\boldsymbol{\Gamma}, \nu+1; (\mathbf{v}_{1i}, \mathbf{v}_{2i}))$ and $\mathbf{W}_{2i} \sim TST_{p_i}(\boldsymbol{\mu}, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu+2; (\mathbf{v}_{1i}, \mathbf{v}_{2i}))$.

3. If the ith subject has both censored and uncensored components and given that $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i)$, $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$, and $(\mathbf{Y}_i^c \,|\, \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$ are equivalent processes, we have

$$\begin{aligned}
\widehat{u\mathbf{y}}_i^{(k)} &= \mathbb{E}[U_i\mathbf{Y}_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{u}_i^{(k)}\,\mathrm{vec}(\mathbf{y}_i^o, \widehat{\mathbf{w}}_{2i}^{c(k)}),\\
\widehat{u\mathbf{y}_i^2}^{(k)} &= \mathbb{E}[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \begin{pmatrix} \mathbf{y}_i^o\mathbf{y}_i^{o\top}\widehat{u}_i^{(k)} & \widehat{u}_i^{(k)}\mathbf{y}_i^o\widehat{\mathbf{w}}_{2i}^{c(k)\top} \\ \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_{2i}^{c(k)}\mathbf{y}_i^{o\top} & \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_{2i}^{2c(k)} \end{pmatrix},\\
\widehat{u}_i^{(k)} &= \mathbb{E}[U_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \frac{ST_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+2}\hat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \hat{\nu}^{(k)}+2)}{ST_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \hat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \hat{\nu}^{(k)})}\\
&\quad\times \frac{\mathcal{L}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \hat{\boldsymbol{\mu}}_i^{co(k)}, \frac{\hat{\nu}_i^{co(k)}}{\hat{\nu}_i^{co(k)}+2}\tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \hat{\boldsymbol{\lambda}}_i^{co(k)}, \hat{\tau}_i^{co(k)}, \hat{\nu}_i^{co(k)}+2)}{\mathcal{L}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \hat{\boldsymbol{\mu}}_i^{co(k)}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \hat{\boldsymbol{\lambda}}_i^{co(k)}, \hat{\tau}_i^{co(k)}, \hat{\nu}_i^{co(k)})},
\end{aligned}$$

with $\widehat{ut}_i^{(k)}$, $\widehat{ut}_i^{2(k)}$ and $\widehat{ut\mathbf{y}}_i^{(k)}$ as in item 2, and

$$\begin{aligned}
\widehat{\phi\mathbf{y}}_i^{r(k)} &= \frac{2}{\sqrt{\pi\hat{\nu}^{(k)}(1+\hat{\boldsymbol{\lambda}}^{(k)\top}\hat{\boldsymbol{\lambda}}^{(k)})}}\frac{\Gamma(\frac{\hat{\nu}^{(k)}+1}{2})}{\Gamma(\frac{\hat{\nu}^{(k)}}{2})}\frac{L_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \hat{\boldsymbol{\mu}}_i^{co(k)}, \frac{\hat{\nu}_i^{co(k)}}{\hat{\nu}_i^{co(k)}+1}\tilde{\boldsymbol{\Gamma}}_i^{cc.o(k)}, \hat{\nu}_i^{co(k)}+1)}{\mathcal{L}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \hat{\boldsymbol{\mu}}_i^{co(k)}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \hat{\boldsymbol{\lambda}}_i^{co(k)}, \hat{\tau}_i^{co(k)}, \hat{\nu}_i^{co(k)})}\\
&\quad\times \frac{t_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+1}\hat{\boldsymbol{\Gamma}}_i^{oo(k)}, \hat{\nu}^{(k)}+1)}{ST_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \hat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \hat{\nu}^{(k)})}\widehat{\mathbf{w}}_{1i}^{r(k)},
\end{aligned}$$

where

$$\widehat{\mathbf{w}}_{si}^{(k)} = \mathbb{E}[\mathbf{W}_{si}^* \mid \widehat{\boldsymbol{\theta}}^{(k)}], \quad \text{and} \quad \widehat{\mathbf{w}}_{si}^{2(k)} = \mathbb{E}[\mathbf{W}_{si}^*\mathbf{W}_{si}^{*\top} \mid \widehat{\boldsymbol{\theta}}^{(k)}], \tag{C.22}$$

for $s = \{1, 2\}$, where $\mathbf{W}_{1i}^* \sim Tt_{p_i^c}(\boldsymbol{\mu}_i^{co(k)}, \frac{\nu_i^{co}}{\nu_i^{co}+1}\tilde{\boldsymbol{\Gamma}}_i^{cc.o}, \nu_i^{co}+1; (\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c))$ and $\mathbf{W}_{2i}^* \sim TEST_{p_i^c}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1}+2; (\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c))$, with $\boldsymbol{\Gamma}_i$ being partitioned like $\boldsymbol{\Sigma}_i$, $\tau_{co} = \nu(\mathbf{y}_i^c)(\tilde{\boldsymbol{\varphi}}_i^{c\top}(\mathbf{y}_i^c - \boldsymbol{\mu}_i^c))$, $\nu_i^{co} = \nu + p_i^o$ and $\tilde{\boldsymbol{\Gamma}}_i^{cc.o} = (\boldsymbol{\Gamma}_i^{cc} - \boldsymbol{\Gamma}_i^{co}\boldsymbol{\Gamma}_i^{oo-1}\boldsymbol{\Gamma}_i^{oc})/\nu^2(\mathbf{y}_i^o)$.

To compute the truncated moments $\widehat{\mathbf{w}}_{si}^{(k)}$ and $\widehat{\mathbf{w}}_{si}^{2(k)}$ given in items 2 and 3, we use our `MomTrunc R` package.

Appendix C.2: Proofs of propositions

Proof of Proposition 5.4. Consider the partition $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ and the corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\varphi}$. We based our proof on the factorization of $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1, \mathbf{y}_2)$ as $f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1, \mathbf{y}_2) = f_{\mathbf{Y}_1}(\mathbf{y}_1)f_{\mathbf{Y}_2|\mathbf{Y}_1=\mathbf{y}_1}(\mathbf{y}_2)$. First, for the symmetric part, we have that

$$t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)t_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1), \tag{C.23}$$

with $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\tilde{\boldsymbol{\Sigma}}_{22.1} = \boldsymbol{\Sigma}_{22.1}/\nu^2(\mathbf{y}_1)$ and $\nu^2(\mathbf{y}_1) = (\nu + p_1)/(\nu + \delta(\mathbf{y}_1))$.

Let now $c_{12} = (1 + \boldsymbol{\varphi}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\tau_{2.1} = \nu(\mathbf{y}_1)(\tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))$, and $\nu_{2.1} = \nu + p_1$. By noting after some straightforward algebra that $\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) = \boldsymbol{\varphi}^\top(\mathbf{y} - \boldsymbol{\mu}) = \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1) + \boldsymbol{\varphi}_2^\top(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1})$ and $\boldsymbol{\lambda}^\top\boldsymbol{\lambda} = \boldsymbol{\varphi}^\top\boldsymbol{\Sigma}\boldsymbol{\varphi} = \tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}\tilde{\boldsymbol{\varphi}}_1 + \boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2$, we obtain

$$T_1\left((\tau_1 + \tilde{\boldsymbol{\lambda}}_1^\top\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_1 - \boldsymbol{\mu}_1))\nu(\mathbf{y}_1); \nu + p_1\right) = T_1\left(\frac{\tau_{2.1}}{(1 + \boldsymbol{\lambda}_{2.1}^\top\boldsymbol{\lambda}_{2.1})^{1/2}}; \nu_{2.1}\right), \tag{C.24}$$

and

$$T_1\left(\frac{\tau}{(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}}; \nu\right) = T_1\left(\frac{\tau_1}{(1 + \tilde{\boldsymbol{\lambda}}_1^\top\tilde{\boldsymbol{\lambda}}_1)^{1/2}}; \nu\right), \tag{C.25}$$

where $\tilde{\boldsymbol{\lambda}}_1 = c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1$, $\tau_1 = c_{12}\tau$ and $\boldsymbol{\lambda}_{2.1} = \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2$. Additionally, it is easy to see that

$$\begin{aligned}
\nu^2(\mathbf{y}) &= \frac{\nu + p}{\nu + \delta(\mathbf{y})} \\
&= \frac{\nu + p_1}{\nu + \delta(\mathbf{y}_1)}\left(\frac{\nu_{2.1} + p_2}{\nu_{2.1} + \delta(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1})}\right) \\
&= \nu^2(\mathbf{y}_1)\nu^2_{\mathbf{Y}_{2.1}}(\mathbf{y}_2).
\end{aligned} \tag{C.26}$$

From this last equation, it holds that

$$T_1\big((A + \tau)\nu(\mathbf{y}); \nu + p\big) = T_1\big((\tau_{2.1} + \boldsymbol{\lambda}_{2.1}^\top\tilde{\boldsymbol{\Sigma}}_{22.1}^{-1/2}(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}))\nu_{\mathbf{Y}_{2.1}}(\mathbf{y}_2); \nu_{2.1} + p_2\big), \tag{C.27}$$

with $A = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})$, Hence, using (C.23), (C.24) and (C.25), we can rewrite the density of $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ as

$$\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\frac{T_1((\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))\nu(\mathbf{y}); \nu + p)}{T_1(\tau/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}; \nu)} \\
&= t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\frac{T_1\big((\tau_{2.1} + \boldsymbol{\lambda}_{2.1}^\top\tilde{\boldsymbol{\Sigma}}_{22.1}^{-1/2}(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}))\nu_{\mathbf{Y}_{2.1}}(\mathbf{y}_2); \nu_{2.1} + p_2\big)}{T_1\left(\tau_1/(1 + \tilde{\boldsymbol{\lambda}}_1^\top\tilde{\boldsymbol{\lambda}}_1)^{1/2}; \nu\right)} \\
&= t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)\frac{T_1\big((\tau_1 + \tilde{\boldsymbol{\lambda}}_1^\top\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_1 - \boldsymbol{\mu}_1))\nu(\mathbf{y}_1); \nu + p_1\big)}{T_1\left(\tau_1/(1 + \tilde{\boldsymbol{\lambda}}_1^\top\tilde{\boldsymbol{\lambda}}_1)^{1/2}; \nu\right)} \\
&\quad \times t_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \nu_{2.1})\frac{T_1\big((\tau_{2.1} + \boldsymbol{\lambda}_{2.1}^\top\tilde{\boldsymbol{\Sigma}}_{22.1}^{-1/2}(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}))\nu_{\mathbf{Y}_{2.1}}(\mathbf{y}_2); \nu_{2.1} + p_2\big)}{T_1\left(\tau_{2.1}/(1 + \boldsymbol{\lambda}_{2.1}^\top\boldsymbol{\lambda}_{2.1})^{1/2}; \nu_{2.1}\right)} \\
&= EST_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \tau_1, \nu) \times EST_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu + p_1).
\end{aligned}$$

□

Proof of Proposition C.1. First note that $\mathbf{Y} \mid (\boldsymbol{\alpha} \leqslant \mathbf{Y} \leqslant \boldsymbol{\beta}) \sim TST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu; (\boldsymbol{\alpha}, \boldsymbol{\beta}))$. By direct integration of the simplified expressions (C.13) and (C.14), it is readily that

$$\begin{aligned}
&\mathbb{E}[\phi(\boldsymbol{\theta}, \mathbf{Y})g(\mathbf{Y})|\, \boldsymbol{\alpha} \leqslant \mathbf{Y} \leqslant \boldsymbol{\beta}] \\
=&\frac{2}{\sqrt{\pi\nu(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\int_\alpha^\beta \frac{t_p(\mathbf{y}; \boldsymbol{\mu}, \frac{\nu}{\nu+1}\boldsymbol{\Gamma}, \nu + 1)}{ST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)}\frac{ST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)}{\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)}g(\mathbf{y})\mathrm{d}\mathbf{y}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \int_\alpha^\beta g(\mathbf{y}) t_p(\mathbf{y};\boldsymbol{\mu},\tfrac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)\mathrm{d}\mathbf{y} \\
&= \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{L(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \mathbb{E}[g(\mathbf{W}_1)]
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}_{UT\mathbf{Y}}[U\, g(\mathbf{Y}) | \boldsymbol{\alpha} \leqslant \mathbf{Y} \leqslant \boldsymbol{\beta}] &= \int_\alpha^\beta \frac{ST_p(\mathbf{y};\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \frac{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} g(\mathbf{y})\mathrm{d}\mathbf{y} \\
&= \frac{1}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \int_\alpha^\beta g(\mathbf{y}) ST_p(\mathbf{y};\boldsymbol{\mu},\tfrac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)\mathrm{d}\mathbf{y} \\
&= \frac{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \mathbb{E}[g(\mathbf{W}_2)],
\end{aligned}
$$

$\mathbf{W}_1 \sim Tt_p(\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1;(\boldsymbol{\alpha},\boldsymbol{\beta}))$ and $\mathbf{W}_2 \sim TST_p(\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2;(\boldsymbol{\alpha},\boldsymbol{\beta}))$. □

Proof of Proposition C.2. It follows from the conditional distribution of a ST distribution that $\mathbf{Y}_2 \mid (\mathbf{Y}_1, \boldsymbol{\alpha}_2 \leqslant \mathbf{Y}_2 \leqslant \boldsymbol{\beta}_2) \sim TEST_{p_2}(\boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1}; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$, with conditional parameters and in proposition 5.4 . It is straightforward that

$$
\begin{aligned}
&\mathbb{E}[\phi(\boldsymbol{\theta},\mathbf{Y})g(\mathbf{Y})|\,\mathbf{Y}_1, \boldsymbol{\alpha}_2 \leqslant \mathbf{Y}_2 \leqslant \boldsymbol{\beta}_2] \\
=&\frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \\
&\times \int_{\alpha_2}^{\beta_2} \frac{t_p(\mathbf{y};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \frac{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}{\mathcal{L}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})} g_2(\mathbf{y}_2)\mathrm{d}\mathbf{y}_2 \\
=&\frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{t_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+1}\boldsymbol{\Gamma}_{11},\nu+1)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)} g_1(\mathbf{Y}_1) \\
&\times \int_{\alpha_2}^{\beta_2} \frac{t_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1},\nu_{2.1}+1)}{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})} \frac{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}{\mathcal{L}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})} g(\mathbf{y})\mathrm{d}\mathbf{y}_2 \\
=&\frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{t_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+1}\boldsymbol{\Gamma}_{11},\nu+1)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)} \frac{L_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1},\nu_{2.1}+1)}{\mathcal{L}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})} \\
&\times g_1(\mathbf{Y}_1)\mathbb{E}[g_2(\mathbf{W}_1^*)]
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}_{UT\mathbf{Y}}[U\, g(\mathbf{Y}) | \boldsymbol{\alpha} \leqslant \mathbf{Y} \leqslant \boldsymbol{\beta}] \\
=&\int_{\alpha_2}^{\beta_2} \frac{ST_p(\mathbf{y};\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \frac{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}{\mathcal{L}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})} g(\mathbf{y})\mathrm{d}\mathbf{y}_2 \\
=&\frac{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+2}\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu+2)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)} g_1(\mathbf{Y}_1) \\
&\times \int_{\alpha_2}^{\beta_2} \frac{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1}+2)}{\mathcal{L}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})} g_2(\mathbf{y}_2)\mathrm{d}\mathbf{y}_2 \\
=&\frac{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+2}\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu+2)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)} \frac{\mathcal{L}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1}+2)}{\mathcal{L}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}
\end{aligned}
$$

$$\times\, g_1(\mathbf{Y}_1)\mathbb{E}[g_2(\mathbf{W}_2^*)],$$

where $g(\mathbf{Y}) = g_1(\mathbf{Y}_1)g_2(\mathbf{Y}_2)$, $\mathbf{W}_1^* \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1}, \nu_{2.1}+1; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$ and $\mathbf{W}_2^* \sim TEST_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1}+2; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$. □

Appendix C.3: ML estimation via the EM algorithm for ST responses

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a $p \times 1$ response vector for the ith sample unit, for $i \in \{1, \ldots, n\}$, considered to be a realization from $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim \mathrm{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. In the case that $\mathbf{Y}$ is fully observed, in order to estimate the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, we can propose a EM algorithm for ML estimation as a special case of the one proposed in subsection 5.8.2. For the equivalent set of parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Delta}, \boldsymbol{\alpha}_\Gamma, \nu)$, the algorithm can be summarized as follows:

E-step: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\alpha}}_\Gamma^{(k)}, \nu^{(k)})$ at the kth step of the algorithm, compute the expectations

$$\widehat{ut}_i^{r\,(k)} = \mathbb{E}_{U_iT_i}[U_iT_i^r \mid \mathbf{Y}_i, \widehat{\boldsymbol{\theta}}^{(k)}],$$

for $r = \{0, 1, 2\}$, using expression (C.11) and (C.12).

M-step: Update the estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\mu}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\alpha}}_\Gamma^{(k)}, \nu^{(k)})$ by

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{u}^{(k)}\mathbf{y}_i - \widehat{ut}_i^{\,(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\right\},\\
\widehat{\boldsymbol{\Delta}}^{(k+1)} &= \left\{\sum_{i=1}^{n}\widehat{ut}_i^{2\,(k)}\right\}^{-1}\sum_{i=1}^{n}\left\{\widehat{ut}_i^{\,(k)}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k+1)})\right\},\\
\widehat{\boldsymbol{\Gamma}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{u}_i^{(k)}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k+1)})(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k+1)})^\top - 2\widehat{ut}_i^{\,(k)}(\mathbf{y}_i\widehat{\boldsymbol{\Delta}}^{(k+1)\top} - \widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\mu}}^{(k+1)\top})\right.\\
&\quad \left. + \widehat{ut}_i^{2\,(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top}\right\}
\end{aligned}$$

As before, we recover $\widehat{\boldsymbol{\lambda}}$ and $\widehat{\boldsymbol{\Sigma}}$ using the expressions in (5.8.2) and we update the parameter ν by maximizing the marginal log-likelihood function for $\mathbf{y}$, that is,

$$\widehat{\nu}^{(k+1)} = \arg\max_{\nu}\sum_{i=1}^{n}\log f(\mathbf{y}_i \mid \widehat{\boldsymbol{\mu}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}^{(k+1)}, \widehat{\boldsymbol{\lambda}}^{(k+1)}; \nu^{(k)}).$$

Algorithm is iterated until a suitable convergence rule is satisfied, i.e., $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)} \mid \mathbf{Y})/\ell(\widehat{\boldsymbol{\theta}}^{(k)} \mid \mathbf{Y}) - 1| < \epsilon$,for ϵ small enough.