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

Christian E. Galarza Morales, Larissa A. Matos, Dipak K. Dey, and Victor H. Lachos

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

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

1. Introduction

In many applications in 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. These variables are also often skewed, departing from the traditional assumption of using symmetric distributions. Thus, the need to study truncated distributions along with their properties naturally arises. For example, expectations of the form $E[Y \mid Y > y^*]$ are very common in biostatistics, survival analysis, or finance, where the above expression is used as a point estimator for observations for which we have only partial information. For the three areas above, the threshold $y^*$ could be a detection limit of a medical equipment, the survival time of a coil in an endurance test, or a quantile representing the best 5% market scenario. For multivariate responses, complex distributions must be considered to model the covariance structure, skewness, heavy-tailedness, among others. In this case, computing multivariate truncated moments can be challenging, so lately, the interest in calculating them has been awakened.

For instance, Tallis (1961) provided the formulas 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 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 formulas for moments of the truncated univariate Student-t distribution in a recursive form. Flecher, Allard, and Naveau (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 (TT) distribution under the double truncation. Arismendi (2013) provided explicit expressions for computing arbitrary order product moments of the TN 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 TN distribution, Kan and Robotti (2017) recently presented recurrence relations for integrals directly related to the density of the multivariate normal distribution for computing arbitrary order product
moments of the TN distribution. These recursions offer a fast computation of the moments of folded normal (FN) and TN distributions, which require evaluating \( p \)-dimensional integrals that involve the Normal (N) density. Explicit expressions for some low order moments of FN and TN distributions are presented cleverly, although some proposals to calculate the moments of the univariate truncated skew-normal distribution and truncated univariate skew-normal/ independent distribution (Flecher, Allard, and Naveau 2010) has recently been published. So far, to the best of our knowledge, there has been an attempt to study neither moments nor product moments of the multivariate folded extended skew-normal (FESN) and truncated multivariate extended skew-normal (TESN) distributions.

In this article, we develop recurrence relations for integrals involving the density of TESN distribution based on the idea of Kan and Robotti (2017). Moreover, as a byproduct, our proposed methods allow us to compute the product moments of folded and truncated distributions of the normal, SN (Azzalini and Dalla-Valle 1996) and their respective univariate versions. In addition to their direct applications, these moments also allow the implementation of complex estimation models for truncated or censored responses, for example, under the presence of asymmetry, multimodality (De Alencar et al. 2021), heavy-tailedness (Galarza et al. 2021), among others. The new methodology is implemented in the R package “MomTrunc” (Galarza and Lachos 2018) available on CRAN repository.

The rest of this article is organized as follows. In Section 2 we briefly discuss some preliminary results related to the multivariate SN, ESN, and TESN distributions and some of its key properties. Section 3 presents a recurrence formula of an integral to be applied in the essential evaluation of moments of the TESN distribution and explicit expressions for the first two moments of the TESN and TN distributions. A direct relation between the TESN and TN distributional moments is also presented, which is used to improve the proposed methods. In Section 4, employing approximations, we propose strategies to circumvent some numerical problems that arise on limiting distributions and extreme cases. We compare our proposal with other popular methods of the literature in Section 5. Section 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, a direct application of TESN moments is developed in the context of risk measurement in finance is presented in Section 7. Some concluding remarks are presented in Section 8.

2. Preliminaries

We start our exposition by defining some notation and presenting the basic concepts which are used throughout the development of our theory. As is usual in probability theory and its applications, we denote a random variable by an upper-case letter and its realization by the corresponding lower case and use boldface letters for vectors and matrices. Let \( I_p \) and \( J_p \) represent an identity matrix and a matrix of ones, respectively, both of dimension \( p \times p \), \( A^T \) be the transpose of \( A \), and \( |X| = (|X_1|, \ldots, |X_p|) \) mean the absolute value of each component of the vector \( X \). For multiple integrals, we use the shorthand notation

\[
\int_a^b f(x)dx = \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(x_1, \ldots, x_p)dx_1 \cdots dx_p,
\]

where \( a = (a_1, \ldots, a_p)^T \) and \( b = (b_1, \ldots, b_p)^T \).

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 \( Y \) follows a multivariate SN distribution with \( p \times 1 \) location vector \( \mu \), \( p \times p \) positive definite dispersion matrix \( \Sigma \) and \( p \times 1 \) skewness parameter vector, and we write \( Y \sim SN_p(\mu, \Sigma, \lambda) \), if its joint probability density function (pdf) is given by

\[
SN_p(y; \mu, \Sigma, \lambda) = 2\phi_p(y; \mu, \Sigma) \Phi_1 \times (\lambda^\top \Sigma^{-1/2}(y - \mu)),
\]

where \( \Phi_1(\cdot, \mu, \Sigma) \) represents the density probability distribution (pdf) of a \( p \)-variate normal distribution with vector mean \( \mu \) and variance-covariance matrix \( \Sigma \), and \( \Phi_1(\cdot) \) stands for the cumulative distribution function (cdf) of a standard univariate normal distribution. If \( \lambda = 0 \), then (1) reduces to the symmetric \( N_p(\mu, \Sigma) \) pdf. Except by a straightforward difference in the parametrization considered in (1), this model corresponds to the one introduced by Azzalini and Dalla-Valle (1996), whose properties were extensively studied in Azzalini and Capitanio (1999) (see also, Arellano-Valle and Genton 2005).

Proposition 1 (cdf of the SN). If \( Y \sim SN_p(\mu, \Sigma, \lambda) \), then for any \( y \in \mathbb{R}^p \)

\[
F_Y(y) = P(Y \leq y) = 2\Phi_{p+1}(y^\top, 0)^T; \mu^*, \Omega),
\]

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

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

2.2. The Extended Multivariate Skew-Normal Distribution

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

\[
ESN_p(y; \mu, \Sigma, \lambda, \tau) = \xi^{-1}\phi_p(y; \mu, \Sigma) \Phi_1 \times (\tau + \lambda^\top \Sigma^{-1/2}(y - \mu)),
\]

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

\[
\text{ESN}_p(y; \mu, \Sigma, \lambda, \tau) \sim \phi_p(y; \mu, \Sigma), \quad \text{as} \quad \tau \to +\infty.
\]

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

\[
E[Z] = \eta \lambda \quad \text{and} \quad \text{cov}[Z] = I_p - E[Z]\left(E[Z] - \frac{\tau}{1 + \lambda^T \lambda}\right)^T,
\]

with \( \eta = \phi_1(\tau; 0, 1 + \lambda^T \lambda) / \xi \). Then, the mean vector and variance-covariance matrix of \( Y \) can be easily computed as

\[
E[Y] = \mu + \Sigma^{1/2} E[Z] \quad \text{and} \quad \text{cov}[Y] = \Sigma^{1/2} \text{cov}[Z] \Sigma^{1/2}.
\]

The following propositions are crucial to develop our methods. The proofs can be found in Appendix A.

**Proposition 2 (Marginal and conditional distribution of the ESN).** Let \( Y \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau) \) and \( Y \) is partitioned as \( Y = (Y_1^T, Y_2^T)^T \) of dimensions \( p_1 \) and \( p_2 \) \((p_1 + p_2 = p)\), respectively. Let

\[
\Sigma = \left( \begin{array}{cc} 
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22} 
\end{array} \right), \quad \mu = (\mu_1^T, \mu_2^T)^T, \\
\lambda = (\lambda_1^T, \lambda_2^T)^T \quad \text{and} \quad \varphi = (\varphi_1^T, \varphi_2^T)^T
\]

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

\[
Y_1 \sim \text{ESN}_p(\mu_1, \Sigma_{11}, c_{12} \Sigma_{12}^{-1/2} \varphi_2, \mathbf{c}_{12} \tau) \quad \text{and} \quad 
Y_2 | Y_1 = y_1 \sim \text{ESN}_p(\mu_2, \Sigma_{22}, \Sigma_{21}, \Sigma_{22}^{-1/2} \varphi_2, \tau_2),
\]

where \( \mathbf{c}_{12} = (1 + \varphi_2^T \Sigma_{221} \varphi_2)^{-1/2}, \varphi_1 = \varphi_1 + \varphi_2^T \Sigma_{221} \varphi_2, \Sigma_{221} = \Sigma_{22} - \Sigma_{21} \Sigma_{12}^{-1} \Sigma_{12} \varphi_2, \Sigma_{222} = \Sigma_{22} - \Sigma_{221} \Sigma_{12}^{-1} \Sigma_{12} \varphi_2, \Sigma_{221} = \Sigma_{22} - \Sigma_{22} \Sigma_{12}^{-1} \Sigma_{12} \varphi_2, \mu_1 = \mu_1 + \Sigma_{221}^{-1} (y_1 - \mu_1) \) and \( \tau_2 = \tau + \varphi_2^T (y_1 - \mu_1) \).

**Proposition 3 (Stochastic representation of the ESN).** Let \( X = (X_1^T, X_2^T)^T \sim N_{p+1}(\mu^*, \Omega) \). If \( Y \overset{d}{=} (X_1 | X_2 < \tilde{r}) \), it follows that \( Y \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau) \), with \( \mu^* \) and \( \Omega \) as defined in Proposition 1, and \( \tilde{r} = \tau / (1 + \lambda^T \lambda)^{1/2} \).

The stochastic representation above can be derived from Arellano-Valle and Genton (2010, prop. 1).

**Proposition 4 (cdf of the ESN).** If \( Y \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau) \), then for any \( y \in \mathbb{R}^p \)

\[
F_Y(y) = P(Y \leq y) = \xi^{-1} \Phi_{p+1}(y^T, \tilde{r}; \mu^*, \Omega).
\]

The proof is direct from Proposition 3 by noting that \( \xi = P(X_2 < \tilde{r}) \). From now on, for \( Y \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau) \), we will denote its cdf as \( F_Y(y) \equiv \Phi_p(y; \mu, \Sigma, \lambda, \tau) \) for simplicity.

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

\[
f(y | \mu, \Sigma, \lambda, \tau) = \frac{\text{ESN}_p(y; \mu, \Sigma, \lambda, \tau)}{P(Y \in \mathcal{A})} \mathcal{I}_\mathcal{A}(y),
\]

where \( I_\mathcal{A} \) is the indicator function of \( \mathcal{A} \). We use the notation \( Y \sim \text{TESN}_p(\mu, \Sigma, \lambda, \tau; \mathcal{A}) \). If \( \mathcal{A} \) has the form

\[
\mathcal{A} = \{ x \in \mathbb{R}^p : a_i \leq x_i \leq b_i \},
\]

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

### 3. On Moments of the Doubly Truncated Multivariate ESN Distribution

#### 3.1. A Recurrence Relation

For two \( p \)-dimensional vectors \( x = (x_1, \ldots, x_p)^T \) and \( \kappa = (k_1, \ldots, k_p)^T \), let \( x^\kappa \) stand for \( (x_1^{k_1}, \ldots, x_p^{k_p}) \), and let \( a_{i0} \) be a vector \( a \) with its \( i \)-th element being removed. For a matrix \( A \), we let \( A_{(i)} \) stand for the \( i \)-th row of \( A \) with its \( i \)-th element being removed. Similarly, \( A_{(i,j)} \) stands for the matrix \( A \) with its \( i \)-th row and \( j \)-th columns being removed. Besides, let \( e_i \) denote a \( p \times 1 \) vector with its \( i \)-th element equaling one and zero otherwise. Let

\[
\mathcal{L}_p(a, b; \mu, \Sigma, \lambda, \tau) = \int_a^b \text{ESN}_p(x; \mu, \Sigma, \lambda, \tau) \, dx.
\]

We are interested in evaluating the integral

\[
\mathcal{F}_p^0(a, b; \mu, \Sigma, \lambda, \tau) = \int_a^b x^\kappa \text{ESN}_p(x; \mu, \Sigma, \lambda, \tau) \, dx.
\]

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

\[
\mathcal{F}_p^0(a, b; \mu, \Sigma, \lambda, \tau) \equiv \mathcal{F}_p^0(a, b; \mu, \Sigma) = \int_a^b \phi_p(x; \mu, \Sigma) \, dx.
\]

with boundary condition

\[
\mathcal{L}_p(a, b; \mu, \Sigma, 0, 0) \equiv \mathcal{L}_p(a, b; \mu, \Sigma) = \int_a^b \phi_p(x; \mu, \Sigma) \, dx.
\]

Note that we use a calligraphic style for the integrals of interest \( \mathcal{F}_p^0 \) and \( \mathcal{L}_p \) when we work with the skewed version. In both expressions (4) and (5), for the normal case, we are using compatible notation with the one used by Kan and Robotti (2017).
3.1.1. Univariate Case
When $p = 1$, it is straightforward to use integration by parts to show that
\[
F^1_p(a, b; \mu, \sigma^2, \lambda, \tau) = \xi^{-1} \left[ \Phi_2((b - \mu, \tau)^T; 0, \Omega) - \Phi_2((a - \mu, \tau)^T; 0, \Omega) \right],
\]
and for $k > 1$, the rest being
\[
F^1_{k+1}(a, b; \mu, \sigma^2, \lambda, \tau) = \mu F^1_p(b, \mu, \sigma^2, \lambda, \tau) + k \sigma^2 F^1_{k-1}(a, b; \mu, \sigma^2, \lambda, \tau) + \sigma^2 \left( a^T \Sigma_1 \mu \right), \quad \Sigma_1 \mu = b \Sigma_1 E \nu_1 \quad \text{for } k \geq 0,
\]
where $\Omega = \left( \begin{array}{cc} \sigma^2 & -\sigma \psi \\ -\sigma \psi & 1 \end{array} \right)$, $\psi = \lambda/\sqrt{1 + \lambda^2}$, $\mu_b = 1/\sigma$, $r = \sigma/\sqrt{1 + \lambda^2}$. When $p > 1$, we need a similar recurrence relation in order to compute $F^p_k(a, b; \mu, \sigma, \lambda, \tau)$ which is presented in the following theorem.

3.1.2. Multivariate Case

**Theorem 1.** For $p \geq 1$ and $i = 1, \ldots, p$,
\[
F^p_{k+i}(a, b; \mu, \sigma, \lambda, \tau) = \mu_i F^p_{k+i-1}(a, b; \mu, \sigma, \lambda, \tau) + \delta_i F^p_{k+i-1}(a, b; \mu - \mu_b, \Gamma) + \epsilon_i \Sigma d_k, \quad (6)
\]
depends on the $\Sigma$ and $d_k$ is a $p$-vector with $i$-th element
\[
d_k = k F^p_{k+i-1}(a, b; \mu, \sigma, \lambda, \tau)
\]
\[
+ \delta_i F^p_{k+i-1}(a, b; \mu - \mu_b, \Gamma) + \epsilon_i \Sigma d_k, \quad (6)
\]
where $\delta = (\delta_1, \ldots, \delta_p)^T = \eta \Sigma^{1/2} \lambda, \mu_b = \bar{\Delta}, \Gamma = \Sigma - \Delta \Delta^T$ and $d_k$ is a $p$-vector with $i$-th element
\[
d_k = k F^p_{k+i-1}(a, b; \mu, \sigma, \lambda, \tau)
\]
\[
+ \delta_i F^p_{k+i-1}(a, b; \mu - \mu_b, \Gamma) + \epsilon_i \Sigma d_k, \quad (6)
\]
where $\bar{\langle} \bar{\mu}_k \rangle = \mu_j + \Sigma(j) \left( \frac{a_j - \mu_j}{\sigma_j^2} \right)$, $\bar{\langle} \bar{\mu}_k \rangle = \mu_j + \Sigma(j) \left( \frac{b_j - \mu_j}{\sigma_j^2} \right)$, $\bar{\langle} \bar{\psi}_j \rangle = \psi_j + \frac{1}{\sigma_j^2} \Sigma(j) \phi(j)$, $\bar{\langle} \bar{\psi}_j \rangle = \psi_j + \frac{1}{\sigma_j^2} \Sigma(j) \phi(j)$, $\bar{\langle} \bar{\psi}_j \rangle = \psi_j + \frac{1}{\sigma_j^2} \Sigma(j) \phi(j)$, $\bar{\langle} \bar{\tau}^a \rangle = \tau + \bar{\psi}_j(a_j - \mu_j)$, and $\bar{\langle} \bar{\tau}^b \rangle = \tau + \bar{\psi}_j(b_j - \mu_j)$.

**Proof.** See Appendix A.

This delivers a simple way to compute any arbitrary moments of multivariate TSN distribution $F^p_k$ 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 $F^p_k$ that can be easily computed through our proposed $R$ package `MomTrunc` available at CRAN. When $k_j = 0$, the first term in (7) 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 $d_j$'s are $-\infty$ and all the $b_j$'s are $+\infty$, for $Y \sim ESN_p(\mu, \Sigma, \lambda, \tau)$, we have that
\[
F^p_k(-\infty, +\infty; \mu, \Sigma, \lambda, \tau) = E[Y^k],
\]
and in this case, the recursive relation is
\[
E[Y^{k+e}] = \mu_i E[Y^k] + \delta_i E[W^k] + \sum_{j=1}^p \sigma_j \epsilon_j E[Y^{k-e}], \quad i = 1, \ldots, p,
\]
with $W \sim N_p(\mu - \mu_b, \Gamma)$.

It is worth stressing that any arbitrary truncated moment of $Y$, that is,
\[
E[Y^k | a \leq Y \leq b] = \frac{F^p_k(a, b; \mu, \Sigma, \lambda, \tau)}{L_p(a, b; \mu, \Sigma, \lambda, \tau)}, \quad (8)
\]
can be computed using the recurrence relation given in Theorem 1. The following section proposes another approach to compute (8) using a unique corresponding arbitrary moment to a truncated normal vector.

3.2. Computing ESN Moments Based on Normal Moments

**Theorem 2.** We have that
\[
F^p_k(a, b; \mu, \Sigma, \lambda, \tau) = \xi^{-1} F^{p+1}_k(a, b; \mu^*, \Sigma, \lambda, \tau),
\]
where $\mu^*$ and $\Sigma$ as defined in Proposition 1, and $\kappa^* = (\kappa^T, \tau)^T$, $a^* = (a^T, -\infty)^T$ and $b^* = (b^T, \tau)^T$.

In particular, for $\kappa = 0$, then
\[
L_p(a, b; \mu, \Sigma, \lambda, \tau) = \xi^{-1} L_{p+1}(a^*, b^*; \mu^*, \Sigma, \lambda, \tau). \quad (9)
\]

**Proof.** See Appendix A.

Equation (9) offers us a very convenient manner to compute $L_p(a, b; \mu, \Sigma, \lambda, \tau)$, since efficient algorithms already exist to calculate $L_p(a, b; \mu, \Sigma)$ (see, e.g., Genz 1992), which avoids performing $p^2$ evaluations of cdf of the multivariate N distribution.

**Corollary 1.** For $Y \sim ESN_p(\mu, \Sigma, \lambda, \tau)$ and $X \sim N_{p+1}(\mu^*, \Sigma)$, it follows from Theorem 2 that
\[
E[Y^k | a \leq Y \leq b] = E[X^k | a^* < X < b^*],
\]
with $a^*, b^*, \kappa^*, \mu^*$ and $\Sigma$ as defined in Theorem 2.

3.3. Mean and Covariance Matrix of Multivariate TESN Distributions

Let us consider $Y \sim TESN_p(\mu, \Sigma, \lambda, \tau, [a, b])$. In light of Theorem 1, we have that
\[
E[Y] = \mu_i + \frac{1}{L} \left[ \delta_i L + \sum_{j=1}^p \sigma_j \left( \Sigma_j \phi(j) \right) \right],
\]
with $a_j, b_j, \bar{\phi}_j, \bar{\Sigma}_j, \bar{\varphi}_j, \bar{\tau}^a$.
for \( i = 1, \ldots, p \), where \( L \equiv L_p(a, b, \mu, \Sigma, \lambda, \tau) \) and \( L \equiv L_p(a, b, \mu - \mu_0, \Gamma) \). It follows that

\[
E[Y] = \mu + \frac{1}{L} [L_\delta + \Sigma (q_a - q_b)]. \tag{10}
\]

where the \( j \)th element of \( q_a \) and \( q_b \) are

\[
q_{a,i} = ESN_1(a_i; \mu_j, \sigma_j^2, c_j \phi_j, c_j \tau) L_{p-1}.
\]

\[
q_{b,j} = ESN_1(b_j; \mu_j, \sigma_j^2, c_j \phi_j, c_j \tau) L_{p-1}.
\]

Denoting \( D = [d_{a1}, \ldots, d_{ap}] \), we can write

\[
E[YY^\top] = E[E[Y]^2] = \mu E[Y] + \frac{1}{L} [L_\delta E[W] + \Sigma D],
\]

\[
\text{cov}[Y] = \left[ \mu - E[Y] \right] E[Y]^\top + \frac{1}{L} [L_\delta E[W] + \Sigma D],
\]

where \( W \sim TN_p(\mu - \mu_0, \Gamma, [a, b]) \), that is a \( p \)-variate truncated normal distribution on \([a, b]\). Besides, from Corollary 1, we have that the first two moments of \( Y \) can also be computed as

\[
E[Y] = E[X]_{(p+1)}, \tag{11}
\]

\[
E[YY^\top] = E[XX^\top]_{(p+1,p+1)}, \tag{12}
\]

with \( X \sim TN_{p+1}(\mu^*, \Omega_1; [a^*, b^*]) \). Note that \( \text{cov}[Y] = E[YY^\top] - E[Y]E[Y]^\top \). Equations (11) and (12) are more convenient for computing \( E[Y] \) and \( \text{cov}[Y] \) since all boils down to compute the mean and the variance-covariance matrix for a \( p+1 \)-variate TN distribution which integrals are less complex than the ESN ones.

### 3.4. Mean and Covariance Matrix of TN Distributions

Some approaches exist to compute the moments of a TN distribution. For instance, for double truncation, Manjunath and Wilhelm (2009) (method available through the \texttt{tmvtnorm} R package) computed the mean and variance of \( X \) directly deriving the MGF of the TN distribution. On the other hand, Kan and Robotti (2017) (method available through the \texttt{MomTrunc} R package) is able to compute arbitrary higher-order TN moments using a recursive approach as a result of differentiating the multivariate normal density. For right truncation, Vaida and Liu (2009) (see supplemental material) proposed a method to compute the mean and variance of \( X \) 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 and Liu (2009) algorithm to handle doubly truncation.

### 3.5. Deriving the First Two Moments of a Double TN Distribution Through Its MGF

**Theorem 3.** Let \( X \sim TN_p(0, R; [a, b]) \), with \( R \) being a correlation matrix of order \( p \times p \). Then, the first two moments of \( X \) are given by

\[
E[X] = \frac{\partial m(t)}{\partial t}^\top |_{t=0} = -\frac{1}{L} Rq,
\]

\[
E[XX^\top] = \frac{\partial^2 m(t)}{\partial t^2} |_{t=0} = R + \frac{1}{L} RHR,
\]

and consequently,

\[
\text{cov}[X] = R + \frac{1}{L^2} R(JH - qq^\top)R,
\]

where \( L \equiv L_p(a, b, 0, R) \), \( q = q_a - q_b \), with the \( j \)th element of \( q_a \) and \( q_b \) as

\[
q_{a,i} = \phi_1(a_i) L_{p-1}(a_i, b_i; a_i R(i,i), \tilde{\mu}_i) \quad \text{and}
\]

\[
q_{b,j} = \phi_1(b_i) L_{p-1}(a_i, b_i; b_i R(i,i), \tilde{\mu}_i),
\]

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

\[
h_{ij} = h_{ij}^{aa} - h_{ij}^{ba} - h_{ij}^{bb} + h_{ij}^{ab},
\]

\[
= \phi_2(a_i, a_j; \rho_j) L_{p-2}(a_i, a_j; \mu_{ij}^{aa}, \tilde{\mu}_i) - \phi_2(b_i, b_j; \rho_j) L_{p-2}(a_i, a_j; b_i, \mu_{ij}^{bb}, \tilde{\mu}_i) - \phi_2(a_i, b_j; \rho_j) L_{p-2}(a_i, a_j; b_i, \mu_{ij}^{ab}, \tilde{\mu}_i) + \phi_2(b_i, a_j; \rho_j) L_{p-2}(a_i, a_j; a_i, \mu_{ij}^{ba}, \tilde{\mu}_i),
\]

and diagonal elements

\[
h_{ii} = a_i q_{a_i} - b_i q_{b_i} - R_{(i,i)} H(i,i), \tag{13}
\]

with \( \tilde{\mu}_i =\ R_{(i,i)} - R_{(i,i)} a_i R_{(i,i)} \), and \( \tilde{\mu}_j = \ R_{(j,j)} - R_{(i,i)} a_i R_{(i,i)} \).

**Proof.** See Appendix A.

The main difference of our proposal in Theorem 3 and other approaches deriving the MGF relies on (13), where the diagonal elements are recycled using the off-diagonal elements \( h_{ij} \), \( 1 \leq i \neq j \leq p \). Furthermore, for \( W \sim TN_p(\mu; \Sigma; \tilde{a}, \tilde{b}) \), we have that

\[
E[W] = \mu - S E[X] \quad \text{and} \quad \text{cov}[W] = S \text{cov}[X] S,
\]

where \( \Sigma \) being a positive-definite matrix, \( S = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p) \), and truncation limits \( \tilde{a} \) and \( \tilde{b} \) such that \( a = S^{-1}(\tilde{a} - \mu) \) and \( b = S^{-1}(\tilde{b} - \mu) \).
4. Dealing With Limiting and Extreme Cases

Let consider \( Y \sim ESN_p(\mu, \Sigma, \lambda, \tau) \). As \( \tau \to \infty \), we have that \( \xi = \Phi(\tau) \to 1 \). Besides, as \( \tau \to -\infty \), we have that \( \xi \to 0 \) and consequently \( f_p^\mu(a, b; \mu, \Sigma, \lambda, \tau) = \xi^{-1} f_{p+1}^\mu(a^t, b^t; \mu^*, \Omega) \to \infty \). Thus, for negative \( \tau \) values small enough, we are not able to compute \( E[Y^p] \) due to computational precision. For instance, in R software, \( \Phi(\tau) = 0 \) for \( \tau < -37 \). The next proposition helps us to circumvent this problem. The proof is given in Appendix A.

Proposition 5. (Limiting distribution for the ESN) As \( \tau \to -\infty \),
\[
ESN_p(y; \mu, \Sigma, \lambda, \tau) \to \phi_p(y; \mu - \mu_b, \Gamma).
\]

4.1. Approximating the Mean and Variance-Covariance of a TN Distribution for Extreme Cases

While using the normal relation (11) and (12), we may also face numerical problems for extreme settings of \( \lambda \) and \( \tau \) due to the scale matrix \( \Omega \) depends on them. The most common problem is that the normalizing constant \( L_p(a^*, b^*; \mu^*, \Omega) \) 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 moment-generating Monte Carlo (MC) samples due to the high rejection ratio when subsampling to a small integration region.

For instance, consider a bivariate truncated normal vector \( X = (X_1, X_2)^T \), with \( X_1 \) and \( X_2 \) having zero mean and unit variance, \( \text{cov}(X_1, X_2) = -0.5 \) and truncation limits \( a = (-20, -10)^T \) and \( b = (-9, 10)^T \). Then, we have that the limits of \( X_1 \) are far from the density mass since \( P(-20 \leq X_1 \leq -9) \approx 0 \). For this case, both the \texttt{etnfunorm} function from the \texttt{etnfunorm} R package and the \texttt{Matlab} codes provided in Kan and Robotti (2017) return wrong mean values outside the truncation interval \( (a, b) \) and negative variances. Values are quite high too, with mean values greater than \( 1 \times 10^{10} \) and all the elements of the variance-covariance matrix greater than \( 1 \times 10^{20} \).

When changing the first upper limit from \( -9 \) to \( -13 \), that is \( b = (-13, 10)^T \), both routines return NaN for \( E[X_1] \) and NaN values for all the elements.

Although the above scenarios seem unusual, extreme situations that require correction are more common than expected. Actually, the development of this part was motivated as we identified this problem when we fit censored regression models with high asymmetry and the presence of outliers. Hence, we present a correction method to approximate the mean and the variance-covariance of a multivariate TN distribution even when the numerical precision of the software is a limitation.

4.1.1. Dealing With Out-of-Bounds Limits

Consider the partition \( X = (X_1^T, X_2^T)^T \) such that \( \text{dim}(X_1) = p_1, \text{dim}(X_2) = p_2 \), where \( p_1 + p_2 = p \). It is well known that
\[
E[X] = E\left[ \frac{E[X_1 | X_2]}{X_2} \right],
\]
and
\[
\text{cov}[X] = \left[ \begin{array}{ccc}
\text{cov}[E[X_1 | X_2]] + \text{cov}[E[X_1 | X_2]] & \text{cov}[E[X_1 | X_2], X_2] \\
\text{cov}[X_2, E[X_1 | X_2]] & \text{cov}[X_2]
\end{array} \right].
\]

Now, consider \( X \sim TN_p(\mu, \Sigma, \{a, b\}) \) to be partitioned as above. Also consider the corresponding partitions of \( \mu, \Sigma, \{a, b\} \).

4.1.2. Dealing With Double Infinite Limits

Let \( p_1 \) be the number of pairs in \( [a, b] \) that are both infinite. We consider the partition \( X = (X_1^T, X_2^T)^T \), such that the upper and lower truncation limits associated with \( X_1 \) are both infinite, but at least one of the truncation limits associated with \( X_2 \) is finite. Since \( a_1 = -\infty \) and \( b_1 = \infty \), it follows that \( X_1 \sim N_{p_1}(\mu_1, \Sigma_{11}) \), \( X_2 \sim TN_{p_2}(\mu_2, \Sigma_{22}, \{a_2, b_2\}) \) and \( X_1 | X_2 \sim N_{p_1}(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (\xi_2 - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}) \). This leads to
\[
E[X] = \left[ \begin{array}{c}
\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (\xi_2 - \mu_2) \\
\frac{E[X_1 | X_2]}{X_2}
\end{array} \right],
\]
\[
\text{cov}[X] = \left[ \begin{array}{ccc}
\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} & \Sigma_{12} \Sigma_{22}^{-1} \xi_2 & \psi_{22} \Sigma_{22}^{-1} \\
\psi_{22} \Sigma_{22}^{-1} & \psi_{22} & \psi_{22}
\end{array} \right].
\]
5. Comparison of Computational Times

Since this is the first attempt to compute the moments of a TESN, it is not possible to compare our approach with others methods already implemented in statistical software, for instance, R or Stata. However, this section intends to compare three possible approaches to compute the mean vector and variance-covariance matrix of a $p$-variate TESN distribution based on our results. We consider our first proposal derived from Theorem 1, which is derived directly from the ESN pdf and the normal relation given in Theorem 2. For the latter, we use different methods for computing the mean and variance-covariance of a TN distribution. The parameters setting of this simulation is described in Appendix C, and the methods that we compare are the following:

Proposal 1: Theorem 1, that is, equations (10), and (12),
Proposal 2: Normal relation (NR) in Theorem 2 using Theorem 3,
Proposal 3: NR in Theorem 2 using the Matlab routine from Kan and Robotti (2017),
Proposal 4: NR in Theorem 2 using the tmvtnorm R function from Manjunath and Wilhelm (2009).

The left panel of Figure 1 shows the number of integrals required to achieve this for different dimensions $p$. We compare 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 newly proposed method since it reduces the number of integral involved almost to half, 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 and Liu (2009) (now, a particular case of proposal 2) requires up to 4 times less integrals than when we use the proposal 3. As seen before, the normal relation proposal 2 outperforms 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 times when using the equivalent normal approach are depicted in the right panel of Figure 1. Here, we compare the absolute processing time of the mean and variance-covariance of a TN distribution under the methods in proposals 2, 3, and 4 for different dimensions $p$. In general, our proposal is the fastest one, as expected. Proposal 3 resulted better only for $p \leq 2$, which confirms the necessity for a faster algorithm in order to deal with high-dimensional problems. Proposal 4 resulted in being the slowest one by far.

5.1. Computational Time in Real Life

For applications where a unique truncated expectation is required (e.g., conditional tail expectations as a measure of risk in finance), the computation cost may seem insignificant, however, iterative algorithms depending on these quantities become computationally intensive. For instance, in longitudinal censored models under a frequentist point of view, an EM algorithm reduces to the computation of the moments of multivariate truncated moments (Lachos et al. 2017) at each iteration and for all censored observations along subjects.

Under controlled settings (please refer to the Appendix Section C), our method resulted 11 and 121 times faster than using MC simulation with a number of replications between $10^4$ and $10^5$, respectively. For instance, 125K integrals will be required for an algorithm that converges in 250 iterations and a modest dataset with 100 subjects and only four censored observations, we achieved a reduction in computation from 9 to 100 hours to only 50 minutes. Other models, such as geostatistical models, are even more demanding, so small differences in times may make the difference between a tractable and a nontractable problem. It is worth remembering that, without these expectations, these must be approximated invoking Monte Carlo methods.

6. On Moments of Multivariate Folded ESN Distributions

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

![Figure 1](image-url). Number of integrals required and absolute processing time (in seconds) for computing the mean vector and variance-covariance matrix for a $p$-variate TESN distribution for three different approaches under double truncation.
Theorem 4 (pdf and cdf of a MFD). Let $X \in \mathbb{R}^p$ be a $p$-variate random vector with pdf $f_X(x; \theta)$ and cdf $F_X(x; \theta)$, with $\theta$ being a set of parameters characterizing such a distribution. If $Y = |X|$, then the joint pdf and cdf of $Y$ that follows a folded distribution of $X$ are given, respectively, by

\[
f_Y(y) = \sum_{s \in S(p)} f_X(A_s y; \theta) \quad \text{and} \quad F_Y(y) = \sum_{s \in S(p)} \pi_s F_X(A_s y; \theta), \quad \text{for } y \geq 0,
\]

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

Corollary 2. If $X \sim f_X(x; \xi, \Psi)$ belongs to the location-scale family of distributions, with location and scale parameters $\xi$ and $\Psi$ respectively, then $Z = \lambda X \sim f_X(z; \lambda \xi, \lambda \Psi A_s)$ and consequently the joint pdf and cdf of $Y = |X|$ are given by

\[
f_Y(y) = \sum_{s \in S(p)} f_Y(y; A_s \xi, A_s \Psi A_s) \quad \text{and} \quad F_Y(y) = \sum_{s \in S(p)} \pi_s F_Y(A_s y; \xi, \Psi), \quad \text{for } y \geq 0.
\]

Hence, the $k$th moment of $Y$ as follows:

\[
E[Y^k] = \sum_{s \in S(p)} E[(Z_s^+)^k],
\]

where $X^+$ denotes the positive component of the random vector $X$.

Let $X \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau)$. We now turn our attention to discuss the computation of any arbitrary order moment of $|X|$, a FESN distribution. Let define the $I^p_{E}(\mu, \Sigma, \lambda, \tau)$ function as

\[
I^p_{E}(\mu, \Sigma, \lambda, \tau) = \int_0^{\infty} y^k \text{ESN}_p(y; \mu, \Sigma, \lambda, \tau) dy.
\]

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

\[
I^p_{E}(\mu, \Sigma, \lambda, \tau) = \mathcal{F}_F^p(0, +\infty; \mu, \Sigma, \lambda, \tau).
\]

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

Proposition 6. If $X \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau)$, then $Z = \lambda X \sim \text{ESN}_p(\mu_s, \Sigma_s, \lambda_s, \tau)$ and consequently the joint pdf, cdf, and the $k$th raw moment of $|X|$ are, respectively, given by

\[
f_Y(y) = \sum_{s \in S(p)} \text{ESN}_p(y; \mu_s, \Sigma_s, \lambda_s, \tau), \quad \text{and} \quad F_Y(y) = L_p(-y; \mu, \Sigma, \lambda, \tau)
\]

and

\[
E[Y^k] = \sum_{s \in S(p)} I^p_{E}(\mu_s, \Sigma_s, \lambda_s, \tau),
\]

where $y_s = \lambda_s y, \mu_s = \lambda_s \mu, \Sigma_s = \lambda_s \Sigma \lambda_s$, and $\lambda_s = \lambda_s \lambda$.

Remark 1. As a consequence of Proposition 6, we also have the new vectors $\delta_s = \lambda_s \delta_s, \mu_{bs} = \lambda_s \mu_b, \phi_s = \lambda_s \phi, \tilde{\mu}_{js}^a = \lambda_s (\mu_{js})^a$ and $\tilde{\mu}_{js}^b = \lambda_s (\mu_{js})^b$, and matrix $\Gamma_s = \lambda_s \Gamma_s$, while the constants $\xi_s, \eta_s, \xi, \eta_s, \Sigma_s$, and $\tilde{\tau}_s$ remain invariant with respect to $s$.

From Proposition 6, we can compute any arbitrary moment of an FESN distribution as a sum of $I^p_{E}$ integrals. In light of Theorem 1, the recurrence relation for $I^p_{E}$ can be written as

\[
I^p_{E}(\mu + \delta, \Sigma, \lambda, \tau) = \mu_j I^p_{E}(\mu, \Sigma, \lambda, \tau) + \delta_j I^p_{E}(\mu - \mu_b, \Gamma)
\]

\[
+ \sum_{i=1}^p \sigma_i d_{k,i}, \quad i = 1, \ldots, p,
\]

where

\[
d_{k,i} = \begin{cases} 
\frac{\delta_j I^p_{E}(\mu, \Sigma, \lambda, \tau)}{\Gamma_{ij}} ; & \text{for } k_j > 0 \\
\frac{\delta_j I^p_{E}(\mu, \Sigma, \lambda, \tau)}{\Gamma_{ij}} ; & \text{for } k_j = 0
\end{cases}
\]

with $\tilde{\mu}_j = \mu_{ij} - \frac{\sigma_j}{\lambda_{ij}} \Sigma_{ij}$ and $\tilde{\tau}_j = \tau - \tilde{\phi}_j \mu_j$.

It is also possible to use the normal relation in Theorem 2 to compute $E[|X|^k]$ in a more straightforward manner, as in the following proposition.

Proposition 7. Let $Y = |X|$, with $X \sim \text{ESN}_p(\mu, \Sigma, \lambda, \tau)$. In light of Theorem 4, it follows that

\[
E[Y^k] = \xi^{-1} \sum_{s \in S(p)} \rho^{-1} \; (\mu_s^+, \Omega_s^-),
\]

where $\rho_s(\mu, \Sigma) \equiv E^p_0(0, \infty; \mu, \Sigma), \mu_s^+ = (\mu_s^+ \tilde{\tau})^T$ and $\Omega_s^- = \begin{pmatrix} \Sigma_s^+ & -\Delta_s \\ -\Delta_s & 1 \end{pmatrix}$, with $\mu_s = \lambda_s \mu, \Sigma_s = \lambda_s \Sigma \lambda_s, \Delta_s = \lambda_s \Delta_s, \text{and } \Omega_s^- \text{ standing for the block matrix } \Omega_s$ with all its off-diagonal block elements signs changed.

The proof is direct from Theorem 2 as $I^p_{E}$ is a particular case of $I^p_{E}$. From Proposition 2, we have that the mean and variance-covariance matrix can be calculated as a sum of $2^p$ terms as well, that is

\[
E[Y] = \sum_{s \in S(p)} E[Z_s^+] \quad \text{and} \quad \text{cov}[Y] = \sum_{s \in S(p)} E[Z_s^+ Z_s^{+\top}] - E[Y] E[Y]^T,
\]

where $Z_s^+$ is the positive component of $Z_s = \lambda_s X \sim \text{ESN}_p(\mu_s, \Sigma_s, \lambda_s, \tau)$. Note that there are $2^p$ times more integrals to be calculated than the nonfolded case, representing a substantial computational effort for high-dimensional problems.

In order to circumvent this, we can use the fact that $E[Y] = (E[Y_1], \ldots, E[Y_p])^T$ and the elements of $E[Y Y^\top]$ are given by the second moments $E[Y_i^2]$ and $E[Y_i Y_j], 1 \leq i \neq j \leq p$. 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 $Y_i$ and the covariance terms $\text{cov}(Y_i, Y_j)$. 
In Appendix B, we circumvent this situation by proposing explicit expressions for the mean and the variance-covariance of the multivariate FESN distribution. As expected, this approach is much faster than the one using equations in (19). For instance, when we consider a trivariate folded ESN distribution, we have that it is about 56x times faster than using MC methods and 10x times faster than using equations in (19). Time comparison (summarized in the figure in the supplementary material, right panel), as well as sample codes of our MomTrunc R package, are provided in Appendices C and D, respectively.

7. Application of SN 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 \( y_\alpha \) be the \((1 - \alpha)\)th quantile of \( Y \), that is, \( P(Y > y_\alpha) = \alpha \). Hence, the tail conditional expectation (TCE) (see, e.g., Denuit et al. 2006) is denoted by

\[
\text{TCE}_Y(y_\alpha) = E[Y \mid Y > y_\alpha].
\]

This can be interpreted as the expected value of the \( \alpha \)\% worse losses. The quantile \( y_\alpha \) is usually chosen to be high in order to be pessimistic, for instance, \( \alpha = 0.05 \). 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 is said to be a coherent measure, holding desirable mathematical properties in the context of risk measurement and is a convex function of the selection weights (Artzner et al. 1999; Pflug 2000). In the multivariate context, if we consider having a portfolio of \( p \) assets, credit scores, etc., then the sum of risks arises as a natural and simple measure of total risk. Then, its TCE is given by \( \text{TCE}_S(s_\alpha) = E[S \mid S > s_\alpha] \), however, this measure disregards the covariance structure among the different assets or business lines. To this end, Landsman and Valdez (2003) extended the TCE to the multivariate framework. The multivariate TCE (MTCE) is given by

\[
\text{MTCE}_Y(y_\alpha) = E[Y \mid Y > y_\alpha] = E[Y \mid Y_1 > y_{1,\alpha}, \ldots, Y_p > y_{p,\alpha}],
\]

where \( \alpha = (\alpha_1, \ldots, \alpha_p)^T \) be a vector of quantiles of interest which may be different per each dimension. Furthermore, considering a multivariate approach, let us compute the TCE of \( \text{TCE}_S(s_\alpha) \) as a decomposed sum, that is

\[
E[S \mid S > s_\alpha] = \sum_{i=1}^{p} \mathbf{1}^T E[Y \mid S > s_\alpha] = E[Y_i \mid S > s_\alpha],
\]

where each term \( E[Y_i \mid S > s_\alpha] \) 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.

Next, we propose the computation of the MTCE when \( Y \) follows a \( p \)-variate SN distribution. Even we could have considered a more flexible distribution for \( Y \) as the ESN, this last suffers from a risk of over-parametrization, in the sense that its log-likelihood may be nearly flat with respect to the extension parameter, which could lead to an erratic estimation.

7.1. MTCE Considering a Multivariate SN Distribution

For practical purposes, suppose that a set of risks \( Y \) are distributed as \( Y \sim SN_p(\mu, \Sigma, \lambda) \). Let \( y \) represents a realization of \( Y \). Based on \( y \), it is possible to estimate the set of parameters \( \theta = (\mu, \Sigma, \lambda)^T \), for instance, through maximum likelihood estimation. It follows from Equation (11) that MTCE\(_Y(y_\alpha) = E[X_1 \mid X_1 > y_\alpha, X_2 > 0] = E[X_1 \mid X > x_\alpha] \), with \( x_\alpha = (y_\alpha^T, 0)^T \), and

\[
X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{p+1}(\mu^*, \Omega^*) = \begin{pmatrix} \mu^T \\ \Omega^* \end{pmatrix} = \begin{pmatrix} \Sigma & \Delta \\ \Delta^T & 1 \end{pmatrix}.
\]

Note that, to write the MTCE as a left truncated normal expectation, we have conveniently multiplied \( X_2 \) in Theorem 2 by -1, which results equivalent since \( X_2 \) is symmetric with respect to zero.

On the other hand, by noticing that \( S = \mathbf{1}^T Y \), it is easy to obtain that \( S \sim SN_1(1^T \mu, 1^T \Sigma, 1^T \Gamma, 1^T \Sigma 1^T / \Omega) \). Thus, Corollary 1 let us compute the TCE of the sum as \( \text{TCE}_S(s_\alpha) = E[U_2 \mid U_1 > 0, U_2 > s_\alpha] = (U_1, U_2)^T \sim N_2(\xi_U, \Omega_U) \), where

\[
\xi_U = \begin{pmatrix} 1^T \mu \\ 0 \end{pmatrix}, \quad \text{and} \quad \Omega_U = \begin{pmatrix} 1^T \Sigma 1 & 1^T \Delta \\ \Delta^T & 1 \end{pmatrix}.
\]

Even though we can compute \( \text{TCE}_S(s_\alpha) \) directly, it is of interest to compute it as a sum of individual risks as in (22). To this end, consider \( V = AY \), with \( A \) being a \((p + 1) \times p \) real matrix with rank \( r \leq p \). Since \( V \) is also SN distributed, in light of Corollary 1, we have that \( E[V \mid a \leq V \leq b] = E[X \mid a^* \leq X \leq b^*] \), where \( X \) follows a \( p + 2 \)-variate multivariate normal with corresponding parameters and truncation limits \( a^* \) and \( b^* \) as in Theorem 2. By setting \( A = (I_p, 1)^T \), we obtain that \( V = (Y, S)^T \) and thus \( E[V \mid S > s_\alpha] = E[V_1 \mid V \geq (-\infty, s_\alpha)^T] = E[W_1 \mid W \geq (-\infty, s_\alpha, 0)^T] \) with \( W = (W_1^T, U_1, U_2)^T \) being a partitioned normal vector,

\[
W = \begin{pmatrix} W_1 \\ U \end{pmatrix} \sim N_{p+2}\left(\begin{pmatrix} \mu \\ \xi_U \end{pmatrix}, \begin{pmatrix} \Sigma & \Delta U \\ \Delta^T U & \Omega_U \end{pmatrix}\right),
\]

where \( \Delta U = (\Sigma 1, \Delta) \). Observe that the expectation \( E[V \mid S > s_\alpha] \) involves a nontruncated partition \( W_1 \) and thus it can be simplified by means of our results in Section 4. Letting \( E = E[U \mid U_1 > 0, U_2 > s_\alpha] \), after some algebra, it finally follows from Equation (16) that

\[
E[V \mid S > s_\alpha] = \mu + \Delta U \Omega_U^{-1}(E - \xi_U) = (1^T \Gamma)^{-1}\left[\xi_1(\Delta 1^T \Sigma - \Sigma 1 \Delta^T) + \varepsilon_2 \Gamma\right].
\]
8. Conclusions

In this article, we have developed a recurrence approach for computing order product moments of TESN and FESN distributions and explicit expressions for the first two moments as a byproduct, generalizing results obtained by Kan and Robotti (2017) for the normal case. The proposed methods also include the moments of the well-known truncated multivariate SN distribution introduced by Azzalini and Dalla-Valle (1996). For the TESN, we have proposed a robust, optimized algorithm based only on normal integrals, which for the normal limiting case outperforms the existing popular method for computing the first two moments, even for extreme cases where all available algorithms fail. The proposed method (including its limiting and particular 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 and 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 new 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).

Finally, we anticipate in the near future extending these results to the extended skew-t distribution (Azzalini-Valle and Genton 2010). We conjecture that our method can be extended to the context of the family of other scale mixtures of skew-normal distributions (Branco and Dey 2001). An in-depth investigation of such extension is beyond the scope of the present article, but it is an interesting topic for further research.

Supplementary Material

The Supplementary Materials contains the following two files:

1. A zip compressed file, which contains the codes to reproduce the Figures in the article as well as a walk-through of our proposed package.
2. A pdf file which is the online appendix with the following sections:
   A. Proofs of propositions and theorems;
   B. Explicit expressions for moments of some folded ESN distributions;
   C. Simulation study;
   D. The R MomTrunc package.

Acknowledgments

We want to express our deepest gratitude to the editor, associate editor, and reviewers for all precious and insightful suggestions, which certainly improve the quality of the article.

Funding

C. Galarza acknowledges support from FAPESP-Brazil (Grant nos. 2015/17110-9 and Grant 2018/11580-1).

ORCID

Christian E. Galarza Morales http://orcid.org/0000-0002-4818-6006
Larissa A. Matos http://orcid.org/0000-0002-2635-0901
Dipak K. Dey http://orcid.org/0000-0002-9367-9731
Victor H. Lachos http://orcid.org/0000-0002-7239-2459

References

Azzalini-Valle, R. B., and Azzalini, A. (2006), “On the Unification of Families of Skew-Normal Distributions,” Scandinavian Journal of Statistics, 33, 561–574. [457]
Azzalini-Valle, R. B., and Genton, M. G. (2005), “On Fundamental Skew Distributions,” Journal of Multivariate Analysis, 96, 93–116. [456]
Ho, H. J., Lin, T.-I., Chen, H. Y., and Wang, W. L. (2012), “Some Results on the Truncated Multivariate t Distribution,” Journal of Statistical Planning and Inference, 142, 25–40. [455]
Houthakker, H. (1959), The Scope and Limits of Futures Trading. Cowles Foundation for Research in Economics at Yale University. [455]
Jawitz, J. W. (2004), “Moments of Truncated Continuous Univariate Distributions,” Advances in Water Resources, 27, 269–281. [455]
Kan, R., and Robotti, C. (2017), “On Moments of Folded and Truncated Multivariate Normal Distributions,” Journal of Computational and Graphical Statistics, 25, 930–934. [455,456,457,459,460,461,464]
Kim, H. M. (2008), “A Note on Scale Mixtures of Skew Normal Distribution,” Statistics and Probability Letters, 78, 1694–1701. [455]
Lachos, V. H., Moreno, E. J. L., Chen, K., and Cabral, C. R. B. (2017), “Finite Mixture Modeling of Censored Data Using the Multivariate Student-t Distribution,” Journal of Multivariate Analysis, 159, 151–167. [461,464]
Landsman, Z., and Valdez, E. A. (2003), “Tail Conditional Expectations for Elliptical Distributions,” North American Actuarial Journal, 7, 55–71. [463]

Lien, D.-H. D. (1985), “Moments of Truncated Bivariate Log-Normal Distributions,” Economics Letters, 19, 243–247. [455]

Manjunath, B., and Wilhelm, S. (2009), “Moments Calculation for the Double Truncated Multivariate Normal Density,” available at SSRN 1472153. [459, 461]

Matos, L. A., Prates, M. O., Chen, M. H., and Lachos, V. H. (2013), “Likelihood-Based Inference for Mixed-Effects Models With Censored Response Using the Multivariate-t Distribution,” Statistica Sinica, 23, 1323–1342. [464]

Pflug, G. C. (2000), “Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk,” in Probabilistic Constrained Optimization, ed. Stanislav P. Uryasev, Springer, pp. 272–281. [463]

Tallis, G. M. (1961), “The Moment Generating Function of the Truncated Multi-Normal Distribution,” Journal of the Royal Statistical Society, Series B, 23, 223–229. [455]

Vaida, F., and Liu, L. (2009), “Fast Implementation for Normal Mixed Effects Models With Censored Response,” Journal of Computational and Graphical Statistics, 18, 797–817. [459, 461, 464]