Moments and random number generation for the truncated elliptical family of distributions

Katherine A. L. Valeriano1 · Christian E. Galarza2 · Larissa A. Matos1

Received: 10 February 2022 / Accepted: 20 December 2022 / Published online: 4 January 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
This paper proposes an algorithm to generate random numbers from any member of the truncated multivariate elliptical family of distributions with a strictly decreasing density generating function. Based on the ideas of Neal (Ann stat 31(3):705–767, 2003) and Ho et al. (J Stat Plan Inference 142(1):25–40, 2012), we construct an efficient sampling method by means of a slice sampling algorithm with Gibbs sampler steps. We also provide a faster approach to approximate the first and the second moment for the truncated multivariate elliptical distributions where Monte Carlo integration is used for the truncated partition and explicit expressions for the non-truncated part (Galarza et al., in J Multivar Anal 189(104):944, 2022). Examples and an application to environmental spatial data illustrate its usefulness. Methods are available for free in the new R library relliptical.

Keywords Elliptical distributions · Slice sampling algorithm · Truncated distributions · Truncated moments

1 Introduction
The use of truncated distributions arises in a wide variety of statistical models, such as survival analysis, censored data models, Bayesian models with truncated parameter space, and abound in agronomy, biology, environmental monitoring, medicine, and economics. Algorithms like Expectation-Maximization (EM) (Dempster et al. 1977) are employed frequently in multivariate censored data analysis under a likelihood-based perspective for its facility to deal with missing and partially observed data. This algorithm requires the computation of conditional truncated moments, commonly the first two moments. For example, Matos et al. (2016) and Matos et al. (2013) estimated the parameters of a censored mixed-effects model for irregularly repeated measures via the EM algorithm, which needed to compute the first two moments of a truncated multivariate t (TMVT) and a truncated multivariate normal (TMVN) distributions, respectively. Variations of the EM algorithm, such as Stochastic Approximation EM (SAEM) (Delyon et al. 1999) and Monte Carlo EM (MCEM) (Wei and Tanner 1990), replace the conditional expectations with an approximation that requires drawing independent random observations from a truncated distribution. For instance, Lachos et al. (2017) estimated the parameters of a linear spatial model for censored data using the SAEM algorithm, which needed to generate random samples from the TMVN distribution to perform the stochastic approximation step. More recently, using the SAEM algorithm, Lachos et al. (2019) proposed a robust multivariate linear mixed model for multiple censored responses based on the scale mixtures of normal (SMN) distributions. Moreover, Gelfand et al. (1992) showed how to perform Bayesian analysis for constrained parameters or truncated data problems using Gibbs sampling. Therefore, sampling random observations and computing moments from truncated distributions is a task of considerable interest.

There are several methods to generate random samples from a truncated distribution in the literature, being the rejection sampling (RS) technique the most common. For instance, in the simplest case, when the non-truncated distribution is considered the “proposal” probability density
function (pdf), the RS method draws samples from the latter and retains only the samples inside the support region. However, the procedure may be inefficient for truncated distributions, especially when the truncation interval is too small or located at a less probable area of the pdf. As an alternative, Robert (1995) proposed an accept-reject algorithm, which dramatically improves the RS method’s efficiency for simulating one-dimensional truncated normal distributions. This method is available through the R (R Core Team 2021) function rtnorm (Hadfield 2022). Besides, the Gibbs sampler algorithm is the most commonly used Markov Chain Monte Carlo (MCMC) method, which generates random observations from a multivariate density by sampling in succession from the full conditional distributions. Nevertheless, its implementation may need methods for sampling from nonstandard univariate distributions. See that an extension to the multivariate framework of the accept-reject algorithm was implemented through a Gibbs sampler (Geman and Geman 1984; Gelfand and Smith 1990).

Later, some automatic and self-tuning samplers for univariate distributions emerged, such as the adaptive rejection sampling (ARS) (Gilks and Wild 1992), the adaptive rejection Metropolis sampling (ARMS) Gilks et al. (1995), the adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2 (ARMS2) (Meyer et al. 2008), the independent doubly adaptive rejection Metropolis sampling (IA²RMS) (Martino et al. 2015a), among others. ARS is an efficient random generator for log-concave distributions, which leads to independent samples and ensures that the sequence of proposals converges to the target pdf. This method reduces the number of evaluations of the target in RS by improving the proposal after each rejection, which is piecewise exponential. To deal with non-log-concave distributions, ARMS generalizes the ARS method by incorporating a Metropolis-Hastings step. This algorithm may return correlated samples, and it cannot guarantee the convergence of the sequence of proposals to the target; even so, it has often been preferred over other MCMC techniques due to its good performance. This method is available in the R library armsspp (Bertolacci 2019). In contrast, ARMS2 is an extension of ARS and ARMS, whose proposal function is a sequence of two piecewise exponential and \( n - 2 \) truncated normal densities, with \( n \) representing the number of elements in the set of support points. On the other hand, IA²RMS is a variation of ARMS, which ensures that the sequence of proposals converges to the target leading to a reduction of the correlation. This algorithm, unlike ARMS, decouples the adaptation mechanism from the proposed construction, allowing one to consider simpler alternatives for the candidate density.

Another alternative to sampling from truncated multivariate distributions is the slice sampler algorithm (Neal 2003), which turns sampling from a truncated density into sampling repeatedly from uniform distributions instead by introducing an auxiliary variable. This approach is often easier to implement than Gibbs sampling. In general, it is easy to code, fast, and does not generate samples out of the truncation region, making it more efficient than the conventional rejection method. Using auxiliary variables was also employed by Damien and Walker (2001) to sample from the TMVN distribution. In the same way, Ho et al. (2012) used slice sampling to draw random points from the TMVT distribution.

Regarding calculating moments from truncated distributions, there are a few libraries in R that provide truncated multivariate moments. For instance, the package tmvtnorm (Wilhelm 2022) computes the mean and the variance of the TMVN distribution by deriving its moment generating function, which is an extension of the method described by Tallis (1961). In contrast, the MomTrunc library (Galarza et al. 2021a) uses a recursive approach method proposed by Kan and Robotti (2017) to compute arbitrary higher-order moments (Galarza et al. 2022b). For the TMVT distribution, the packages TTmoment (Ho et al. 2015) and MomTrunc compute its two first moments. The first library only handles integer degrees of freedom greater than 4, while the latter can compute even high-order moments for any degrees of freedom (Galarza et al. 2021c).

To the best of our knowledge, there are no proposals in the literature to generate samples from other multivariate truncated distributions in the elliptical class other than the TMVN and TMVT distributions (available in the tmvtnorm and TTmoment packages). Hence, motivated by the slice sampling algorithm, we propose a general method to obtain samples from any truncated multivariate elliptical distribution with a strictly decreasing density generating function (dgf). Using conditional expectation properties, we also construct an efficient algorithm to approximate the moments of the most common distribution of this class: the truncated multivariate normal, Student-\( t \), slash, contaminated normal, and Pearson VII distributions. This method requires less running time than the existing ones since it deals separately with the truncated and non-truncated parts of the vector. Our proposal can be reached through the R package relliptical.

Finally, it is worth mentioning that moments of truncated elliptical distributions can be used to compute truncated moments for the selection elliptical family of distributions. This wide family includes complex multivariate asymmetric versions of the elliptical distributions as the extended skew-normal and the unified skew-\( t \) distributions, among others. Therefore, our proposal opens the doors for the calculation of truncated moments of complex elliptical asymmetric distributions, which are of particular interest for the development of robust censored models with asymmetry, heavy tails, and missingness (see, for instance, De Alencar et al. 2021; Galarza et al. 2021b).
The paper is organized as follows. Section 2 shows some results related to the elliptical and truncated elliptical family of distributions and briefly describes the slice sampling algorithm. Section 3 is devoted to formulating the sampling algorithm for the truncated elliptical class of distributions, whereas Sect. 4 focuses on a method to approximate the first and the second moment. In the last two sections, we present results for some well-known members of this family, and a simulation study that compares the mean and covariance matrix for the TMVT distribution estimated through different R methods is presented as well. Section 5 displays an application of censored Gaussian spatial models throughout the analysis of the Missouri dioxin contamination dataset. Finally, Sect. 6 concludes with a discussion.

2 Preliminaries

2.1 Elliptical family of distributions

As defined in Muirhead (2009) and Fang et al. (2018), a random vector \( X \in \mathbb{R}^p \) is said to follow an elliptical distribution with location parameter \( \mu \in \mathbb{R}^p \), positive-definite scale matrix \( \Sigma \in \mathbb{R}^{p \times p} \), and density generating function \( g \), if its pdf is given by

\[
f_X(x) = c_p |\Sigma|^{-\frac{1}{2}} g \left( (x - \mu)^\top \Sigma^{-1} (x - \mu) \right),
\]

for \( x \in \mathbb{R}^p \), where \( g(t) \) is a non-negative Lebesgue measurable function on \([0, \infty)\) such that \( \int_0^\infty t^{p/2-1} g(t) \, dt < \infty \) and \( |\Sigma| \) denotes the determinant of matrix \( \Sigma \). Moreover,

\[
c_p = \frac{\Gamma(p/2)}{\pi^{p/2}} \left( \int_0^\infty t^{p/2-1} g(t) \, dt \right)^{-1},
\]

is the normalizing constant, with \( \Gamma(\cdot) \) representing the complete gamma function. We will use the notation \( X \sim E \ell_p(\mu, \Sigma; g) \).

Members of the elliptical family of distributions are characterized by their density-generating function \( g \). Some examples of the elliptical family of distributions are:

- The multivariate normal distribution, \( X \sim N_p(\mu, \Sigma) \), with mean \( \mu \) and variance-covariance matrix \( \Sigma \), arises when the dgf takes the form \( g(t) = \exp(-t/2), \ t \geq 0 \).
- The multivariate Student-\( t \)-distribution, \( X \sim t_p(\mu, \Sigma, v) \), with location parameter \( \mu \), scale matrix \( \Sigma \), and degrees of freedom \( v > 0 \), is obtained when \( g(t) = (1 + t/v)^{-\nu/2}, \ t \geq 0 \).
- The multivariate power exponential, \( X \sim PE_p(\mu, \Sigma, \beta) \), with kurtosis parameter \( \beta > 0 \). In this case, \( g(t) = \exp(-t^\beta/2), \ t \geq 0 \). A particular case of the power exponential distribution is the normal distribution, which arises when \( \beta = 1 \).
- The multivariate slash, \( X \sim SL_p(\mu, \Sigma, v) \), we get a random variable with multivariate slash distribution when \( g(t) = \int_0^1 u^{v/2-1} \exp(-ut/2) \, du, \ t \geq 0, \ v > 0 \).
- The multivariate Pearson VII distribution, \( X \sim PVII_p(\mu, \Sigma, m, v) \), with parameters \( \mu \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}, m > p/2, \) and \( v > 0 \) is obtained when \( g(t) = (1 + t/v)^{-m}, \ t \geq 0 \).

For more distributions belonging to this family, please see Fang et al. (2018).

2.2 Truncated elliptical family of distributions

Let \( A \subseteq \mathbb{R}^p \) be a measurable set. We say that a random vector \( Y \in \mathbb{R}^p \) has truncated elliptical distribution with support \( A \), location parameter \( \mu \in \mathbb{R}^p \), scale parameter \( \Sigma \in \mathbb{R}^{p \times p} \), and dgf \( g \), if its pdf is given by

\[
f_Y(y) = \frac{f_X(y)}{\Pr(X \in A)} \mathbb{1}_A(y), \ y \in A,
\]

where \( X \sim E \ell_p(\mu, \Sigma; g) \). We use the notation \( Y \sim TE(\mu, \Sigma; g, A) \). Notice that the pdf of \( Y \) is written as the ratio between the pdf of \( X \sim E \ell_p(\mu, \Sigma; g) \) and \( \Pr(X \in A) \), so the pdf of \( Y \) exists if the pdf of \( X \) does, which occurs if \( \Sigma \) is positive-definite (see, Morán-Vásquez and Ferrari 2021). The variable \( Y \) is also said to be an elliptical distribution truncated on \( A \), being represented by \( Y = X | (X \in A) \).

As in the elliptical family of distributions, the dgf \( g \) determines any distribution within the truncated elliptical class. For example, if \( g(t) = (1 + t/v)^{-(v+p)/2}, \ t \geq 0, \ v > 0 \), then \( Y \) has TMVT distribution. We will denote the different members of the truncated elliptical family defined in the subsection before as \( Y \sim T \ell_p(\mu, \Sigma; A) \) for the TMVN distribution, \( Y \sim T P(\mu, \Sigma, v; A) \) for the TMVT distribution, \( Y \sim TPVII_p(\mu, \Sigma, \beta; A) \) for the truncated multivariate power exponential, \( Y \sim TSL_p(\mu, \Sigma, v; A) \) for the truncated multivariate slash distribution, and \( Y \sim TPVII_p(\mu, \Sigma, m, v; A) \) for the truncated multivariate Pearson VII distribution.

2.3 Slice sampling algorithm

A slice sampler is a form of auxiliary variable technique in which one or more variables are introduced to facilitate the construction of an MCMC method. The idea of using auxiliary variables in MCMC methods was established for the Ising model by Swendsen and Wang (1987), and it was brought into the statistical literature by Besag and Green.
Following Neal (2003), an MCMC method can be constructed using the principle that it can sample from a given distribution by simulating uniformly from the region under the plot of its density function.

Suppose we are interest in sampling from the distribution of a random variable \( X \in \mathbb{R}^p \), whose pdf is proportional to the function \( f(x) \). The slice sampler algorithm simulates uniformly from the \((p+1)\)-dimensional region under the plot of \( f(x) \) by introducing a real auxiliary variable, \( Y \), such that the joint pdf of \( X \) and \( Y \) is uniform over the region \( V = \{(x, y) : 0 < y < f(x)\} \), i.e., \( f_{X,Y}(x, y) \propto \mathbb{I}(0 < y < f(x)) \), with \( \mathbb{I}(\cdot) \) being an indicator function. Therefore, we can obtain samples from the distribution of \( X \) by sampling jointly \((x, y)\) and then ignoring \( y \).

Note that generating independent random points uniformly distributed on \( V \) may not be easy. To overcome this problem, Neal (2003) defined a Markov Chain that converges to this uniform distribution similar to Gibbs sampler or Metropolis-Hastings algorithms. Then, considering Gibbs sampler steps, the slice sampling at iteration \( k \) works as follows: given the current value of \( x_{k-1} \) sample \( y_k \) from \( Y \mid (X = x_{k-1}) \sim U(0, f(x_{k-1})) \), then draw \( x_k \) from the conditional distribution of \( X \) given \( y_k \), which is uniform over the region \( S_k = \{x : y_k < f(x)\} \), i.e., \( X \mid (Y = y_k) \sim U(\{x : y_k < f(x)\}) \), for all \( k = 1, 2, \ldots, n \), where \( n \) is the desired sample size. We can notice that the slice sampling method is easily implemented for univariate random variables. In contrast, for the multivariate case \((p > 1)\), sampling uniformly from the region \( S_k \) may be complex, in which case we can employ some update for \( x \) that leaves the uniform distribution invariant over this slice.

Figure 1 shows the steps of the slice sampling algorithm for a univariate random variable \( X \). Given an initial value \( X = x_0 \),

1. Simulate \( y_1 \) from \( Y \mid (X = x_0) \sim U(0, f(x_0)) \).
2. Simulate \( x_1 \) from \( X \mid (Y = y_1) \sim U(S_1) \), with \( S_1 = \{x : y_1 < f(x)\} \).

These two steps are repeated \( n \) times by making \( x_0 = x_1 \) in the next iteration.

### 3 Sampling from the truncated elliptical family of distributions

This section is devoted to describe our slice sampling algorithm with Gibbs steps to generate random observations from a multivariate elliptical distribution with strictly decreasing dfg.

Without loss of generality, we first consider a \( p \)-variate truncated elliptical distribution with zero location parameter, positive-definite scale matrix \( R \in \mathbb{R}^{p \times p} \), dfg \( g \), and truncation region \( A = \{(x_1, \ldots, x_p)^T : a_1 < x_1 < b_1, \ldots, a_p < x_p < b_p\} = \{x : a < x < b\}, a, b \in \mathbb{R}^p \), in other words, we will consider \( X \sim T\mathcal{E}_p(0, R; g, A) \). Here \( R \) is a correlation matrix, such that the scale matrix can be written as \( \Sigma = A \Lambda A^T \), where \( \Lambda = \text{diag}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \ldots, \sqrt{\sigma_{pp}}) \). The pdf of \( X \) is given by

\[
f_X(x) \propto g \left( x^T R^{-1} x \right) \mathbb{I}(x \in A) .
\tag{3}
\]

Now, in order to sample uniformly from the \((p+1)\)-dimensional region under the plot of \( f_X(x) \), we introduce an auxiliary variable \( Y \), such that the joint pdf of \( X \) and \( Y \) is

\[
f_{X,Y}(x, y) \propto \mathbb{I}(0 < y < g \left( x^T R^{-1} x \right))
\times \mathbb{I}(a < x < b) .
\tag{4}
\]

It is enough to calculate the conditional distributions of \( Y \mid X \) and \( X \mid Y \) to establish our slice sampling algorithm with Gibbs steps to generate random observations from the pdf in (4). These are given by:

\[
f_{Y/X}(y \mid x) \propto \mathbb{I}(0 < y < g(x^T R^{-1} x))
\text{ and }

f_{X/Y}(x \mid y) \propto \mathbb{I}(\{x : y < g(x^T R^{-1} x) \land a < x < b\}) .
\]

Note that sampling \( y \) from the distribution of \( Y \mid X = x \) is straightforward, but sampling from \( X \mid Y = y \) is not trivial. Thus, we use the idea of Damien and Walker (2001), and Ho et al. (2012), which consists in sampling each element of \( X \) in succession from the full conditional distributions; in other words, we will apply the Gibbs sampler algorithm within
the slice sampling mechanism. Therefore, note that the full conditional distributions are \( f_{X_j \mid X_{-j}, y} \) \( y \mid (X_{-j}, y) \propto \mathbb{1}(x_j \in A_j) \), where \( X_{-j} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)^T \) and \( A_j = \{ x_j \mid y < g(X_j^T \mathbf{R}^{-1} x) \wedge a_j < x_j < b_j \} \), for all \( j = 1, \ldots, p \). To find the elements of the real set \( A_j \), let \( y \) be the value sampled at the current iteration of the algorithm and \( \kappa_y = g^{-1}(y) \). Then, we have that

**Algorithm 1** Slice sampling algorithm

**Require:** Sample size \( n \geq 1 \), initial value \( x_0 = (x_1^{(0)}, \ldots, x_p^{(0)})^T \in \mathbb{R}^p \), scale matrix \( \mathbf{R} \in \mathbb{R}^{p \times p} \), lower bound \( a \in \mathbb{R}^p \), upper bound \( b \in \mathbb{R}^p \), and strictly decreasing \( \text{df} g(t), t \geq 0 \).

1. for \( i = 1 \) to \( n \) do
2. \( x_i \leftarrow \) \( y \)
3. for \( j = 1 \) to \( p \) do
4. \( y_j \leftarrow \sum_{i \neq j} x_i \rho_{ij} \)
5. \( \lambda_j \leftarrow -\frac{1}{\rho_{jj}} \sum_{i \neq j} x_i \rho_{ij} \)
6. \( \tau_j \leftarrow \left( \lambda_j^2 + \frac{1}{\rho_{jj}} (\kappa_y - y_j) \right)^{1/2} \)
7. Sample \( x_j \leftarrow \text{U}(\max(a_j, \lambda_j - \tau_j), \min(b_j, \lambda_j + \tau_j)) \)
8. \( \mathbf{x} \leftarrow (x_1, \ldots, x_p)^T \)
9. \( \mathbf{X} \leftarrow (X_1, \ldots, X_n)^T \)
10. end for
11. end for
12. Ensure: \( \mathbf{X} \)

Since \( g \) is a strictly decreasing function, it follows that \( y < g(x_i^T \mathbf{R}^{-1} x) \) is equivalent to \( \kappa_y > x_i^T \mathbf{R}^{-1} x \). The \( i \)-th element of the inverse of \( \mathbf{R} \), \( \eta_i = \sum_{i \neq j} x_i \rho_{ij} \), and \( \lambda_j = -\frac{1}{\rho_{jj}} \sum_{i \neq j} x_i \rho_{ij} \).

Combining items 1 and 2, we obtain that \( \lambda_j < \tau_j < x_j < \lambda_j + \tau_j \), where \( \tau_j = \left( \lambda_j^2 + \frac{1}{\rho_{jj}} (\kappa_y - y_j) \right)^{1/2} \).

Because \( x_j \in (a_j, b_j) \), thereby \( a_j^* = \max(a_j, \lambda_j - \tau_j) < x_j < x_j < \min(b_j, \lambda_j + \tau_j) = b_j^* \).

It follows that \( A_j = (a_j^*, b_j^*) \). The steps to draw \( n \) samples from a \( p \)-variate truncated elliptical distribution \( \mathbf{X} \sim \text{TE}(0, \mathbf{R}, g, \mathbf{A}) \) are summarized in Algorithm 1. As seen, only univariate uniform simulations are involved in the algorithm, which are fast to compute. Note also that the assumption that the \( \text{df} g \) is strictly decreasing has been used in step 1. However, when it is not possible to find an analytical expression for \( \kappa_y = g^{-1}(y) \), a numerical method is used, leading us to a more computationally expensive algorithm. Additionally, the slice sampler method generates random observations conditioned on previous values, resulting in a sequence of correlated samples. Thus, it is essential to analyze the dependence effect. See, for instance, Robert and Casella (2010), and Section C from Appendix, where samples generated from specific bivariate distributions were observed, and the autocorrelation drops quickly, being negligibly small when lags become large, evidencing well mixing and quickly converging. The sampling method described here can be extended for a general \( \text{df} g \) by constructing an adequate “slice” at each iteration. Please refer to Section 4.1 in Neal (2003).

Moreover, members of the truncated elliptical family of distributions are closed under affine transformations (Fang et al. 2018). Hence drawing samples from \( \mathbf{Y} \sim \text{TE}(\mathbf{P}(\mu, \Sigma; g, (\mathbf{a}, \mathbf{b})) \) may be readily done by sampling first from \( \mathbf{X} \sim \text{TE}(\mathbf{P}(0, \mathbf{R}; g, (\mathbf{a}, \mathbf{b}))) \) and then recovering \( \mathbf{Y} \) by the following transformation \( \mathbf{Y} = \mu + \mathbf{A} \mathbf{X} \), such that \( \Sigma = \mathbf{A} \mathbf{A}^T \).

### 4 Moments of truncated multivariate elliptical distributions

This section describes an adaptation of the method proposed by Galarza et al. (2022a) to compute the first moment and the variance-covariance matrix of a random vector whose distribution belongs to the truncated elliptical family. Furthermore, this method will be applied to some well-known distributions.

Let \( \mathbf{X} \) be a \( p \)-variate random vector that follows a truncated multivariate elliptical distribution with location parameter \( \mu \in \mathbb{R}^p \), positive-definite scale matrix \( \Sigma \in \mathbb{R}^{p \times p} \), \( \text{df} g \), and support \( \mathbf{A} \subseteq \mathbb{R}^p \), i.e., \( \mathbf{X} \sim \text{TE}(\mu, \Sigma; g, \mathbf{A}) \). The more straightforward approach for computing the first two moments of \( \mathbf{X} \) is to use Monte Carlo integration. Following this method, the moments \( \mathbb{E} (\mathbf{X}) \), \( \mathbb{E} (\mathbf{XX}^T) \), and \( \text{Cov} (\mathbf{X}) \) can be approximated by

\[
\mathbb{E} (\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i,
\]
\[
\mathbb{E} (\mathbf{XX}^T) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T,
\]
\[
\text{Cov} (\mathbf{X}) = \mathbb{E} (\mathbf{XX}^T) - \mathbb{E} (\mathbf{X}) \mathbb{E} (\mathbf{X})^T,
\]

where \( \mathbf{x}_i \) is the \( i \)-th sample of the random vector \( \mathbf{X} \) draws independently from \( \text{TE}(\mu, \Sigma; g, \mathbf{A}) \). However, it is well-known that the execution time needed to perform Monte Carlo integration depends on the algorithm employed to draw samples, the number of random points \( n \) used in the approximation, and the length of the random vector \( p \). Then, it depends on some variables that might represent a consider-
able computational effort. Nevertheless, we can save time when the random vector $X$ has non-truncated components, following the idea of Galarza et al. (2022a). They proposed to decompose $X$ into two vectors, $X_1$ and $X_2$, in such a way that $X_1$ is the random vector of truncated components and $X_2$ is the non-truncated part. Then, the first two moments for the truncated variables are computed using any method, and the remaining moments are computed using properties of the conditional expectation. Before showing our algorithm, we state an extremely important result.

**Proposition 1** (Marginal and conditional distribution of the Elliptical family) Let $X \in \mathbb{R}^p$ be partitioned into two vectors, $X_1 \in \mathbb{R}^{p_1}$ and $X_2 \in \mathbb{R}^{p_2}$, such that $p = p_1 + p_2$ and $X = (X_1^T, X_2^T)^T$ has joint multivariate elliptical distribution as follows

$$X \equiv \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim E_{p_1+p_2} \left( \mu, \Sigma; g^{(p_1+p_2)} \right),$$

with

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

partitioned such that $\mu_1 \in \mathbb{R}^{p_1}$, $\mu_2 \in \mathbb{R}^{p_2}$ are location vectors, $\Sigma_{11} \in \mathbb{R}^{p_1 \times p_1}$, $\Sigma_{22} \in \mathbb{R}^{p_2 \times p_2}$, $\Sigma_{12} \in \mathbb{R}^{p_1 \times p_2}$, $\Sigma_{21} \in \mathbb{R}^{p_2 \times p_1}$ are dispersion matrices, and $g^{(p_1+p_2)}$ is the dgf. The distribution of $X_1$ and $X_2 (X_1 = x)$ belong to the elliptical family of distributions, as follows

$$X_1 \sim E_{p_1} \left( \mu_1, \Sigma_{11}; g_{1}^{(p_1)} \right)$$

$$X_2 | (X_1 = x) \sim E_{p_2} \left( \mu_2 + \Sigma_{12} \Sigma_{11}^{-1} (x - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right),$$

Therefore, considering that $X_1$ is the vector of truncated variables with truncation region $A_1$ and $X_2$ is the vector of non-truncated variables, such that $A = A_1 \times \mathbb{R}^{p_2}$, by Proposition 1, we have that

$$X_1 \sim TE_{p_1} \left( \mu_1, \Sigma_{11}; g_{1}^{(p_1)} , A_1 \right)$$

$$X_2 | (X_1 = x) \sim E_{p_2} \left( \mu_2 + \Sigma_{12} \Sigma_{11}^{-1} (x - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right),$$

Let $\xi_1 = \mathbb{E}(X_1 | X_1 \in A_1)$ and $\Omega_{11} = \text{Cov}(X_1 | X_1 \in A_1)$. Then, using the results exposed by Galarza et al. (2022a), the first moment of $X$ can be computed by $\mathbb{E}(X | X \in A) = \mathbb{E}(\mathbb{E}(X | X_1) | X_1 \in A_1)$, i.e.,

$$\mathbb{E}(X | X \in A) = \left( \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (\xi_1 - \mu_1) \right).$$

On the other hand, the variance-covariance matrix of $X$ is given by

$$\text{Cov}(X | X \in A) = \begin{pmatrix} \Omega_{11} & \Sigma_{12} \Sigma_{11}^{-1} - \Sigma_{11}^{-1} \Omega_{11} \\ \Sigma_{21} \Omega_{11}^{-1} - \Sigma_{11}^{-1} \Sigma_{21} & \Psi_{22} \end{pmatrix},$$

where $\Psi_{22} = \omega_{2,1} \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \left( \omega_{2,1} I_{p_1} - \Sigma_{11}^{-1} \right) \Sigma_{12}$, $\omega_{2,1} = \mathbb{E} \left( h(X_1) | X_1 \in A_1 \right)$, $h(X_1) = \text{tr} \left( \text{Cov}(X_2 | X_1) \Sigma_{21}^{-1} \right) / p_2$, and $\Sigma_{2,1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$. Note that $h(X_1)$ depends on the conditional distribution of $X_2 | X_1$, taking a different expression for each member of the elliptical family.

Thus, this work proposes to use Monte Carlo integration to approximate the truncated moments $\xi_1$, $\Omega_{11}$, and $\omega_{2,1}$ (when necessary) instead of computing them using recursion-based methods (e.g., Galarza et al. 2021c) which can be computationally more expensive for high dimensions or complex elliptical distributions. For instance, to the best of our knowledge, closed-form expressions to compute the first two moments of truncated elliptical distributions only exist for the TVMN and TMVT distributions. These expressions are not very efficient (as demonstrated in a simulation study in Sect. 4.2.2) in practice because, due to their recursive nature, they suffer from error propagation and require an intensive calculation of probabilities, which in turn depend on numerical approximation methods. A summary of how our method works is given in Algorithm 2. The first moment and the variance-covariance matrix are approximated by Equations (5)–(6).

**Algorithm 2** Mean and variance approximation

**Require:** Sample size $n \geq 1$, location parameter $\mu \in \mathbb{R}^p$, scale matrix $\Sigma \in \mathbb{R}^{p \times p}$, lower bound $a \in \mathbb{R}^p$, upper bound $b \in \mathbb{R}^p$ and dgf $g(t), t \geq 0$.

1: Identify: $\mu_1, \mu_2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, A_1 = \{x_i : a_i < x_i < b_i \}$
2: for $i \leftarrow 1 \text{ to } n$ do
3: Draw $x_i$ from $X_1 \sim TE_{p_1} (\mu_1, \Sigma_{11}; g_{1}^{(p_1)}, A_1)$ using Algorithm 1  
4: end for

5: $\hat{\xi}_1 \leftarrow \frac{1}{n} \sum_{i=1}^{n} x_i$; $\tilde{\Omega}_{11} \leftarrow \frac{1}{n} \sum_{i=1}^{n} x_i x_i^\top - \hat{\xi}_1 \hat{\xi}_1^\top$; $\tilde{\omega}_{2,1} \leftarrow \frac{1}{n} \sum_{i=1}^{n} h(x_i)$

6: $\tilde{\mathbb{E}}(X) \leftarrow \left( \mu_2 + \Sigma_{21} \tilde{\xi}_1 \right)$; $\tilde{\text{Cov}}(X) \leftarrow \begin{pmatrix} \tilde{\Omega}_{11} & \Sigma_{12} \tilde{\xi}_1 \\ \Sigma_{21} \tilde{\xi}_1 & \tilde{\Omega}_{22} \end{pmatrix}$

7: $\tilde{\mathbb{E}}(XX^\top) \leftarrow \tilde{\text{Cov}}(X) + \tilde{\mathbb{E}}(X) \tilde{\mathbb{E}}(X)^\top$

Ensure: $\hat{\mathbb{E}}(X), \hat{\mathbb{E}}(XX^\top), \hat{\text{Cov}}(X)$

© Springer
4.1 Mean and variance for the truncated elliptical distributions

This subsection is devoted to analyze how Algorithm 2 works for some specific distributions considering all the conditions used previously. Let \( \mu_2'(x) = \mu_2 + \frac{\Sigma_{11}^{-1}(x - \mu_1)}{\Sigma_{11}^{-1}} \), \( \Sigma_{21} = \Sigma_{22} - \frac{\Sigma_{21}^{-1}}{\Sigma_{11}^{-1}} \Sigma_{12} \), and \( \delta_1(x) = (x - \mu_1)^T \Sigma_{11}^{-1} (x - \mu_1) \).

- **Normal**: If \( X \sim N_p(\mu, \Sigma) \), the marginal is \( X_1 \sim N_p(\mu_1, \Sigma_{11}) \) and the conditional distribution is \( X_2|X_1 = x \sim N_p(\mu_2'(x), \Sigma_{22} - \frac{\Sigma_{21}^{-1}}{\Sigma_{11}^{-1}} \Sigma_{12}) \). Then, to compute the moments for \( X \sim TN_p(\mu, \Sigma; A) \) with the conditions above, Algorithm 2 firstly samples \( X_1 \) from the TMVN distribution with location parameter \( \mu_1 \), scale matrix \( \Sigma_{11} \), truncation region \( A_1 \), and \( \omega_{2,1} = 1 \).
- **Student-t**: If \( X \sim t_p(\mu, \Sigma, \nu) \), \( \nu > 0 \), the marginal and conditional distributions are \( X_1 \sim t_{p_1}(\mu_1, \Sigma_{11}) \) and \( X_2|X_1 = x \sim t_{p_2}(\mu_2'(x), \nu_2) \), respectively, such that \( \nu_2 = \nu/(\nu - p_1) \). For this distribution \( E(X) \) exists, if \( \nu > 1 \) and \( \text{Cov}(X) \) exists, if \( \nu > 2 \). Therefore, the moments for \( X \sim \text{Tr} t_p(\mu, \Sigma, \nu; A) \) are computed by sampling \( X_1 \) from the TMVT distribution with location parameter \( \mu_1 \), scale matrix \( \Sigma_{11} \), \( \nu \) degrees of freedom, truncation region \( A_1 \), and \( \omega_{2,1} \) given by

\[
\omega_{2,1} = \frac{\nu + E(\delta_1(X_1) \mid X_1 \in A_1)}{\nu + p_1 - 2},
\]

with \( E(\delta_1(X_1) \mid X_1 \in A_1) = \text{tr}(\Sigma_{11}^{-1}) + (\xi_1 - \mu_1)^T \Sigma_{11}^{-1} (\xi_1 - \mu_1) \). It is worth mentioning that for doubly truncated variables, the mean and the variance exist for all \( \nu > 0 \). Then, if \( X \) has at least two doubly truncated variables, the mean and the variance-covariance matrix exist for all \( \nu > 0 \). For more details about the existence of these moments, please refer to Galarza et al. (2022a).

- **Pearson VII**: If \( X \sim PVII_p(\mu, \Sigma, m, \nu) \), \( m > p/2 \), \( \nu > 0 \), then \( E(X) \) is \( \mu \) and \( \text{Cov}(X) = \frac{\nu}{2m-p-2} \Sigma \). In this case, \( E(X) \) exists, if \( m > (p + 1)/2 \) and \( \text{Cov}(X) \) exists, if \( m > (p/2)/2 \). The marginal and the conditional distributions are \( X_1 \sim PVII_{p_1}(\mu_1, \Sigma_{11}, m - p_2/2, \nu) \) and \( X_2 \mid (X_1 = x) \sim PVII_{p_2}(\mu_2'(x), \Sigma_{22}, m, \delta_1(x) + \nu) \), respectively. So, the proposed algorithm for \( X \sim \text{TPVII}_p(\mu, \Sigma, m, \nu; A) \) is implemented by sampling \( X_1 \) from the truncated multivariate Pearson VII distribution with location parameter \( \mu_1 \), scale matrix \( \Sigma_{11} \), additional parameters \( m - p_2/2 > p_1/2 \), \( \nu > 0 \), and truncation region \( A_1 \). The constant \( \omega_{2,1} \) is

\[
\omega_{2,1} = \frac{\nu + E(\delta_1(X_1) \mid X_1 \in A_1)}{2m - p_2 - 2},
\]

where \( E(\delta_1(X_1) \mid X_1 \in A_1) \) is given as in the Student-t distribution. For this distribution, first and second moments for doubly truncated variables exist for all \( m > p/2 \). Then, if \( X \) has at least two doubly truncated variables, the mean and the variance exist for all \( m > p/2 \). For more details about the existence of the moments, refer to Subsection A.1 in the Appendix section.

- **Slashed**: If \( X \sim SL_p(\mu, \Sigma, \nu, v), v > 0 \), then \( E(X) = \mu \) and \( \text{Cov}(X) = \frac{\nu}{v-1} \Sigma \). In this case, \( \text{Cov}(X) \) exists, if \( v > 1 \). The marginal distribution is \( X_1 \sim SL_{p_1}(\mu_1, \Sigma_{11}, \nu) \) and the conditional distribution is \( X_2 \mid (X_1 = x) \sim EF_{p_2}(\mu_2'(x), \Sigma_{22}, \nu) \), such that \( \gamma \), \( \nu \) distribution belongs to the elliptical family (see Appendix, Subsection A.2). So, the moments for \( X \sim TSL_p(\mu, \Sigma, v; A) \) are calculated by sampling \( X_1 \) from the truncated multivariate slash distribution with location parameter \( \mu_1 \), scale matrix \( \Sigma_{11} \), \( v \) degrees of freedom, and truncation region \( A_1 \). The constant \( \omega_{2,1} \) is

\[
\omega_{2,1} = \frac{\nu + E(\delta_1(X_1) \mid X_1 \in A_1)}{v - 1}
\]

\[
\times E(\frac{SL_{p_2}(\mu_1, \Sigma_{11}, v - 1)}{SL_{p_1}(\mu_1, \Sigma_{11}, v)} \mid X_1 \in A_1)
\]

where \( SL_p(x; \mu, \Sigma, v) \) denotes the pdf of a \( p \)-variate slash distribution with parameters \( \mu, \Sigma, \) and \( v \). As usual, this constant can also be approximated via Monte Carlo integration.

- **Contaminated Normal**: If \( X \sim CN_p(\mu, \Sigma, \nu, \rho, \omega) \), \( 0 < \nu, \rho, \omega \leq 1 \), then the marginal is \( X_1 \sim CN_p(\mu_1, \Sigma_{11}, \nu, \rho) \) and the conditional distribution is \( X_2 \mid (X_1 = x) \sim CN_{p_2}(\mu_2'(x), \Sigma_{22}, \nu_2(x), \rho_2) \), with \( \nu_2(x) = \nu \phi_p(x; \mu_1, \rho_2^{-1} \Sigma_{11})/\kappa(x) \), \( \kappa(x) = \nu \phi_p(x; \mu_1, \rho_2^{-1} \Sigma_{11}) + (1 - \nu) \phi_p(x; \mu_1, \Sigma_{11}) \), and \( \phi_p(x; \mu, \Sigma) \) denoting the pdf of the \( p \)-variate normal distribution with mean \( \mu \) and variance \( \Sigma \) evaluated at point \( x \in \mathbb{R}^p \). Thus, our algorithm for \( X \sim TNCN_p(\mu, \Sigma, v, \rho; A) \) samples \( X_1 \) from the truncated contaminated normal distribution with parameters \( \mu_1, \Sigma_{11}, \nu, \rho \), and \( \omega \). The constant is \( \omega_{2,1} = \frac{v_2}{\rho^2 + 1 - v_2} \), where \( v_2 = E(\delta_2(X_1) \mid X_1 \in A_1) \). This value is also approximated via Monte Carlo integration.

- **Power exponential**: If \( X \sim PE_p(\mu, \Sigma, \beta), \beta > 0 \), then \( E(X) = \mu \) and \( \text{Cov}(X) = \omega \Sigma \), with \( \omega = \frac{2^{1/\beta} \Gamma(\frac{n+2}{2})/n \Gamma(\frac{n}{2})}{} \). The marginal distribution of \( X_1 \) belongs to the elliptical family of distributions with dgf \( g^{(p)}(t) = t^{p-2} \int_0^1 u^{p-2} (1 - u)^{-p_2/(2p)} \exp[-(\frac{t}{2})^2] \) dw, i.e., \( X_1 \sim EF_{p_1}(\mu_1, \Sigma_{11}, g^{(p)}) \). The conditional distribution is \( X_2 \mid (X_1 = x) \sim EF_{p_2}(\mu_2'(x), \Sigma_{22}, g^{(p)}(x)) \) where \( g^{(p_2)}(t) = \exp[-(\frac{1}{2} t + \delta_1(x))^2] (Gómez et al.
Therefore to approximate the moments for \( \mathbf{X} \sim \operatorname{TP}_{p}(\mu, \Sigma, \beta; A) \), we will use a different approach that consists of drawing points from the whole random vector of length \( p \) and then approximate the moments using Monte Carlo integration. Since sampling directly from the marginal distribution of \( X_i \) could be really complicated, as well as to compute \( \varphi_{2,1} \).

4.2 Numerical examples

4.2.1 Simulation study I

We illustrate how the method works considering a random vector \( \mathbf{X} = (X_1, X_2, X_3, X_4)^\top \) of length 4 with truncated Student-\( \nu \) distribution. In this example, \( X_2 \) is not truncated, and the other components are doubly truncated. The objective is to study the performance of the estimates for the mean and the variance-covariance elements obtained through Algorithm 2, considering a different number of samples for approximation and thinning, where thinning is a factor for reducing autocorrelation between observations. After that, we compare those results with the ones obtained from the \( \text{TT.moment} \) functions and from the \( \text{meanvarTMD} \) functions, and those are comparable with those from \( \text{meanvarTMD} \) are similar to those from MC with \( n = 10^5 \) and thinning = 3 for most cases, except for \( \sigma_{11}, \sigma_{33}, \) and \( \sigma_{13} \). For these parameters, our method showed better performance.

4.2.2 Simulation study II

In the previous example, it was observed that the estimates obtained from Algorithm 2 with \( n = 10^5 \) and thinning = 3 are good enough to estimate the mean and variance of a multivariate \( (p = 4) \) variable with TMVT distribution, even though the best results were obtained through the \( \text{TT.moment} \) function. In this example, our goal is to analyze the computational time required for our method and the functions \( \text{meanvarTMD} \) and \( \text{TT.moment} \) to estimate the first two moments and the variance-covariance matrix of a \( p \)-variate random vector with TMVT distribution considering \( p = 50, 100, \) and 150. In each case, we set 10\%, 20\%, and 40\% of the variables doubly truncated. The methods were run on a Windows 10 machine using R 4.0.3 on an Intel Core i7-7700 Processor with 3.60 GHz and 32 GB of RAM.

Table 1 displays the median of the CPU time (in seconds) required for our algorithm and functions \( \text{meanvarTMD} \) and \( \text{TT.moment} \). For our proposal were considered three scenarios \( n = 10^4 \) with no thinning, \( n = 10^4 \) and \( n = 10^5 \) with thinning = 3. The results are based on 100 simulations, and they were computed through the \( \text{R} \) function \( \text{microbenchmark} \). This table also shows the relative time (R.Time) computed, taking the time used by our method with \( n = 10^5 \) and thinning = 3 as reference. We will refer to this configuration as the “reference method.” For our algorithm, we observe that the time required to estimate the moments depends on the number of random observations sampled and the number of truncated variables. Note that estimating the moments with \( n = 10^4 \) took 3.50\% of the time required for the reference method, and it is worth mentioning that the number of samples needed for the first method is 3.33\% of the number of samples used for the reference one. Our proposal with \( n = 10^4 \) and thinning = 3 already needed 10\% of the execution time used by the reference method. Observe that \( \text{meanvarTMD} \) was faster than the reference procedure for
Fig. 2 Boxplot based on 100 estimates of the truncated mean. The red line denotes the median of TT.moment estimates.

Fig. 3 Boxplot based on 100 estimates of the variance-covariance elements. The red line represents the median of the estimates obtained from function TT.moment.

Table 1 Median of the CPU time (in seconds) based on 100 simulations

| Method       | Measure  | p = 50 |       | p = 100 |       | p = 150 |       |
|--------------|----------|--------|-------|---------|-------|---------|-------|
|              |          | 10%    | 20%   | 40%     | 10%   | 20%     | 40%   | 10%   | 20%   | 40%   |
| n = 10⁴      | Median   | 0.011  | 0.030 | 0.139   | 0.030 | 0.140   | 0.952 | 0.071 | 0.382 | 2.842 |
| thinning = 1 | R.Time   | 0.035  | 0.035 | 0.034   | 0.036 | 0.035   | 0.034 | 0.036 | 0.034 | 0.034 |
| n = 10⁵      | Median   | 0.031  | 0.084 | 0.404   | 0.085 | 0.405   | 2.820 | 0.199 | 1.118 | 8.461 |
| thinning = 3 | R.Time   | 0.100  | 0.100 | 0.100   | 0.100 | 0.100   | 0.100 | 0.101 | 0.100 | 0.100 |
| n = 10⁶      | Median   | 0.314  | 0.844 | 4.042   | 0.846 | 4.044   | 28.217| 1.974 | 11.182| 84.619|
| thinning = 3 | R.Time   | –      | –     | –       | –     | –       | –     | –     | –     | –     |
| meanvarTMD   | Median   | 0.118  | 4.102 | 49.189  | 3.781 | 48.681  | 367.243 | 21.209| 157.179| 1215.630|
|              | R.Time   | 0.375  | 4.861 | 12.170  | 4.467 | 12.037  | 13.015 | 10.746| 14.056| 14.366|
| TT.moment    | Median   | 7.452  | 24.027| 94.408  | 62.026| 202.704 | 789.641| 242.701| 800.360| 3081.367|
|              | R.Time   | 23.767 | 28.047| 23.358  | 73.279| 50.122  | 27.984 | 122.974| 71.574 | 36.414|
vectors with 5 doubly truncated variables. It also seems that the time needed by \texttt{meanvarTMD} depends only on the number of truncated variables. In comparison, the \texttt{TT.moment} function is much more time-consuming in all scenarios if compared with our proposal and function \texttt{meanvarTMD}, e.g., for a random vector of length \( p = 100 \) and 40 doubly truncated variables, it needed 28 times longer than the reference method. An additional example regarding the computational time required to compute the truncated moments for other elliptical distributions can be found in Appendix B.

5 Application on spatial model for censored data

Algorithms proposed in this work have applications in several statistical models, such as censored data models, survival analysis, Bayesian models with truncated parameter space, among others. For instance, our \texttt{reelliptical} package has been applied in the \texttt{ARpLMEC} package (Oliviari et al. 2022), which estimates parameters for censored mixed-effects models with a symmetric elliptical error distribution. At the same time, Mattos et al. (2022) used our methods in semiparametric mixed models for longitudinal data with censored responses and heavy tails. In the following application, we consider the Gaussian spatial censored linear (SCL) model defined by Lachos et al. (2017) and Ordoñez et al. (2018).

In the SCL model, the data is generated from \( Z = X \beta + \xi \), with \( \xi \sim N_p(0, \Sigma) \) and \( \Sigma = [\text{Cov}(Z_i, Z_j)] = \sigma^2 R(\phi) + \tau^2 I_p \). It also has the particularity that the response variable \( Z \) is not fully observed. Instead, it is observed \( V_i \) and \( C_i \) at each location, for \( i = 1, \ldots, p \), where \( C_i = 0 \) and \( V_i = Z_i \), if \( Z_i \) is observed, and \( C_i = 1 \) and \( V_i = (V_{i1}, V_{i2}) \), if \( Z_i \) is censored or missing. Because of the difficulties in working directly with the observed log-likelihood function, Lachos et al. (2017) suggested using an EM-type algorithm to obtain the ML estimates of \( \theta \). The authors also considered the following parameterization for \( \Sigma = \sigma^2 \Psi \), with \( \Psi = R(\phi) + v^2 I_n \) and \( v^2 = \tau^2 / \sigma^2 \), to help with the identifiability of the parameters. See also Diggle and Ribeiro (2007). Hence, the EM algorithm works as follows:

- \textbf{E-step:} Let \( \hat{\theta}^{(k)} \) be the current estimate of \( \theta \), then the conditional expectation of the complete-data log-likelihood without the constant is

\[
Q_k(\theta) = \mathbb{E}(\ell(\theta | Z_c) | V, C, \hat{\theta}^{(k)}) \propto \frac{1}{2} \left[ \log | \Psi | + n + \log \frac{1}{\sigma^2} \Lambda^{(k)} \right],
\]

where \( \Lambda^{(k)} = \text{tr}(ZZ^\top \Psi^{-1}) - 2Z(k)^\top \Psi^{-1}X\beta + \beta^\top X^\top \Psi^{-1}X\beta \). Therefore, the E-step reduces only to the computation of \( \tilde{Z}^{(k)} = \mathbb{E}(ZZ^\top | V, C, \hat{\theta}^{(k)}) \) and \( \tilde{\beta}^{(k)} = \mathbb{E}(Z | V, C, \hat{\theta}^{(k)}) \). In the traditional EM algorithm, we should now evaluate the conditional expectations, which is possible through the R packages \texttt{tmvtnorm} or \texttt{MomTrunc}. However, it is computationally expensive when the proportion of censored observations is non-negligible. An alternative is to use the MCEM algorithm, which approximates the conditional expectations using MC integration. For the SCL model, the MCE-step is performed by estimating \( \tilde{Z}^{(k)} \) and \( \tilde{\beta}^{(k)} \) through Algorithm 2.

- \textbf{M-step:} The conditional maximization step is carried out, and \( \hat{\theta}^{(k)} \) is updated by maximizing \( Q_k(\theta) \) over \( \theta \) to obtain a new estimate \( \hat{\theta}^{(k+1)} \), which leads to the expressions:

\[
\begin{align*}
\hat{\beta}^{(k+1)} &= (X^\top \hat{\Psi}^{-1(k)} X)^{-1} X^\top \hat{\Psi}^{-1(k)} \tilde{Z}^{(k)} , \\
\hat{\sigma}^2(k+1) &= \frac{1}{n} \left[ \text{tr} \left( \frac{ZZ^\top(k)}{\hat{\Psi}^{-1(k)}} \right) - 2\tilde{Z}^\top(k) \Psi^{-1}X\beta^{(k+1)} \\
&\quad + \beta^\top(k+1) X^\top \Psi^{-1}X\beta^{(k+1)} \right] , \\
\hat{\alpha}^{(k+1)} &= \arg\max_{\alpha \in \mathbb{R}^+ \times \mathbb{R}^+} \left( -\frac{1}{2} \log | \Psi | - \frac{1}{2\hat{\sigma}^2(k+1)} \text{tr} (ZZ^\top(k) \Psi^{-1}) \\
&\quad - 2\tilde{Z}^\top(k) \Psi^{-1}X\beta^{(k+1)} + \beta^\top(k+1) X^\top \Psi^{-1}X\beta^{(k+1)} \right) .
\end{align*}
\]

with \( \alpha = (\phi, v^2)^\top \). Note that \( \hat{\tau}^2 \) can be recovered by \( \hat{\tau}^2(k+1) = \hat{\sigma}^2(k+1) / \hat{\sigma}^2(k+1) \). An efficient M-step can be easily accomplished by using, for instance, the \texttt{roptim} package (Pan and Pan 2022). In general, the estimates of \( \theta \) may vary slightly around the solution. Hence, one possible final estimate of the parameters may be computed as the mean of the estimates after applying a burn-in and a thinning process.

5.1 Missouri dioxin contamination data

The proposed MCEM algorithm will be applied to analyze the Missouri dioxin contamination dataset available in \texttt{CensSpatial} package. The dataset contains 127 observations distributed in an area of 3600 \( \times 65 \) \( m^2 \) on the shoulders of a country road located in Missouri, U.S.A. The observations correspond to the level of contamination by dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin or TCDD) at sampled points along the road, where 43% of the observations (55 sites) were censored, falling below some detection limit,
which ranges from 0.10 to 0.79 mg/kg. The spatial directions are the x-direction (measured in 1/100 ft) and the y-direction (in ft). Please, refer to Fridley and Dixon (2007) for more details.

This dataset was firstly analyzed by Zirschky and Harris (1986), who concluded that data appeared to be log-normally distributed. Hence, we fit the model $\log(Z_i) = \beta_0 + \xi_i$, $i = 1, \ldots, 127$. The model parameters were estimated using the MCEM algorithm and compared with the estimates from the SAEM and EM algorithms. All methods were performed using 500 iterations and an exponential correlation function to take into account the variation between spatial points. For the MCEM algorithm, we evaluated four cases; in one of those scenarios, it was considered linearly increasing sample sizes between 100 and 1000. Other scenarios considered constant sample sizes of 20, 5000, and $10^5$. In order to use the SAEM algorithm, we set two configurations; one draws points using the rmvtnorm function (from package rmvtnorm), and the optimization procedure via optimx function (Nash et al. 2020). This method is available in the CensSpatial package, and from now on, we refer to this algorithm by SAEM. The second one draws points using the proposed slice sampler, while the R function roptim executes the optimization procedure. We will refer to the latter as SAEM-SS. Lastly, moments were computed using the MomTrunc package for the EM algorithm. The functions used to estimate the parameters via MCEM, SAEM-SS, and EM are available in the RcppCensSpatial package (Valeriano et al. 2022).

The results of the ML estimates are shown in Table 2, where $n$ is the number of samples to approximate the conditional mean, and $c$ indicates the percentage of iterations without memory in the SAEM algorithm (Lachos et al. 2017; Ordoñez et al. 2018). Final estimates for MCEM and EM methods were computed as the mean of the estimates at each iteration after applying a burn-in of 250 and thinning of 3 observations. In contrast, the SAEM and SAEM-SS estimates are the ones obtained at the last iteration. We see that the SAEM-SS estimates are similar to those from MCEM with $n = 20$, while the estimates for the EM algorithm are similar to MCEM with $n = 5000$. The estimates obtained through MCEM and EM for the regression coefficient $\beta_0$ were $-2.400$, while the SAEM algorithm estimated this parameter equal to $-2.010$. The estimates achieved from MCEM and SAEM methods for the partial sill $\sigma^2$ and the nugget effect $\tau^2$ suggest that the spatial process explains 97% and 95% of the variability in data, respectively. Regarding the spatial scaling parameter $\phi$, it was around 15.05 and 14.10 for the MCEM and SAEM algorithms, respectively. These values imply that for distances greater than 45 and 42 feet, respectively, the correlation between two observations falls to less than 0.05. This table also shows the maximized log-likelihood value, information criteria AIC and BIC, and the running time in minutes. Based on the information criteria, we can conclude that MCEM with $n = 5000$ best fits the Missouri dioxin contamination data. Furthermore, it does not seem necessary to consider sample sizes as large as $n = 10^5$ because that configuration does not gain the precision of the estimates and is more time-consuming.

Figure 4 shows the convergence graphs of the parameter estimates achieved from MCEM, SAEM-SS, SAEM, and EM algorithms. Notice that the variability in the estimates for MCEM decreases when the sample size increases from 100 to 1000 (aquamarine line). As expected, the estimates obtained from MCEM with $n = 20$ (gray line) present more variability than the other three scenarios in which we considered larger sample sizes. In contrast, MCEM with $n = 10^5$ (black line) reported the lowest variability in the estimates. The estimates of the parameters computed through the EM algorithm present more variability than those from MCEM with $n = 5000$ (red line). It is probably due to the computational stability of the numerical methods involved in the MomTrunc package; this is why we decided to consider a burn-in and thinning procedure to compute the EM final estimates.

### Table 2: Missouri data—ML estimates and information criteria (AIC and BIC) obtained through MCEM, SAEM and EM algorithms considering the exponential correlation function

| Algorithm  | $n$   | $c$     | $\beta_0$ | $\sigma^2$ | $\phi$ | $\tau^2$ | Log-likelihood | AIC          | BIC          | Time (min) |
|------------|------|---------|-----------|------------|--------|----------|----------------|--------------|--------------|------------|
| MCEM       | 20   | −2.355  | 6.577     | 14.702     | 0.213  | −143.128 | 294.257        | 305.633      | 0.936        |
|            | $10^2 - 10^3$ | −2.402  | 6.808     | 15.095     | 0.207  | −143.108 | 294.216        | 305.592      | 1.819        |
|            | 5000 | −2.410  | 6.847     | 15.076     | 0.206  | −143.095 | 294.191        | 305.568      | 10.206       |
|            | $10^5$ | −2.408  | 6.845     | 15.053     | 0.205  | −143.136 | 294.272        | 305.649      | 185.341      |
| SAEM-SS    | 20   | 0.25    | −2.311    | 6.218      | 14.964 | 0.220   | −143.175       | 294.350      | 0.658        |
| SAEM       | 20   | 0.25    | −2.014    | 4.858      | 14.206 | 0.245   | −143.840       | 295.681      | 6.079        |
|            | $10^5$ | 1.00    | −2.010    | 4.829      | 14.136 | 0.245   | −143.865       | 295.729      | 915.164      |
| EM         | −     | −2.417  | 6.888     | 15.092     | 0.206  | −143.122 | 294.244        | 305.620      | 1661.472     |
6 Conclusions

This work describes an algorithm to generate random numbers from members of the truncated elliptical family of distributions with a strictly decreasing density generating function through a slice sampling algorithm with Gibbs sampler steps. In addition, we presented an efficient approach to approximate the first and the second moment for these distributions. We briefly introduce the functions available in our R package `relliptical` in order to perform sample generation and estimation of the first two moments. Simulation studies were performed to investigate the properties of estimates and the robustness of our algorithm. Moreover, we compared our approach with others available in the R software (only for the normal and Student-$t$ cases), showing that our approach over-performed others in terms of precision and computational time. We illustrate the usefulness of truncated moments on the Missouri dioxin contamination dataset, where a spatial model for censored data was fitted. The results presented in this paper can be reproduced through the R package `relliptical`, available at CRAN for download.

Future extensions of the work include considering a more general class of density generating functions (not only strictly decreasing) in the sampling method. Other methods could also be explored to sample from the truncated elliptical family, such as IA$^2$RMS (Martino et al. 2015a) or the fast universal self-tuned sampler (FUSS) (Martino et al. 2015b) within Gibbs. The first one is interesting because it returns asymptotically independent samples and tries to maintain the computational cost bounded (as in ARS and ARMS). About the latter one, it was demonstrated through simulation studies to be faster than some well-known MCMC methods for sampling from some specific bivariate distributions, besides the accuracy and generation of virtually independent samples. On the other hand, the method used to approximate moments for the truncated elliptical family can be extended...
to the context of asymmetric multivariate elliptical distributions, so the fast computation of their truncated moments may lead the way to propose more flexible and robust models relating to censored models for mixed-effects models, longitudinal data, and spatial models, among others.

7 Supplementary information

Practitioners may find a walk-through of our proposed relliptical R package.

Acknowledgements

The research of Katherine A. L. Valeriano was supported by CAPES. Larissa A. Matos acknowledges support from FAPESP-Brazil (Grant 2020/16713-0).

Appendix A: Further results for some multivariate elliptical distributions

A.1 The multivariate pearson VII distribution

A random variable $X \in \mathbb{R}^p$ is said to have a multivariate Pearson VII distribution with location parameter $\mu \in \mathbb{R}^p$, positive-definite scale matrix $\Sigma \in \mathbb{R}^{p \times p}$, extra parameters $m > p/2$ and $\nu > 0$, if its pdf is given by

$$f_X(x) = \frac{\Gamma(m)}{(\pi \nu)^{p/2} \Gamma(m-p/2)} |\Sigma|^{-1/2} \times \left(1 + \frac{1}{\nu} (x - \mu) \Sigma^{-1} (x - \mu) \right)^{-m},$$

with $x \in \mathbb{R}^p$. The random vector $X$ can also be represented as a scale mixture of normal (SMN) distributions, i.e., $X = \mu + U^{-1/2} Z$, where $Z$ has a $p$-variate normal distribution with mean $0 \in \mathbb{R}^p$ and variance-covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$. Here, $U$ follows Gamma distribution with scale parameter $m - p/2$ and rate parameter $\nu/2$, where $Z$ is independent of $U$. This implies that $X | (U = u) \sim N_p(\mu, u^{-1} \Sigma)$ and $U \sim \text{Gamma}(m - p/2, 2\nu)$.

Therefore, the mean and the variance-covariance matrix of $X$ are

$$\mathbb{E}(X) = \mu, \quad m > \frac{p + 1}{2},$$

$$\text{Cov}(X) = \text{Cov}(\mathbb{E}(X | U)) + \mathbb{E}(\text{Cov}(X | U)) = \mathbb{E}(U^{-1}) \Sigma = \frac{\nu \Sigma}{2m - p - 2}, \quad m > \frac{p + 2}{2}.$$  

A.1.1 Marginal and conditional distribution

Now suppose that the vector $X$ is partitioned into two random vectors $X_1 \in \mathbb{R}^{p_1}$ and $X_2 \in \mathbb{R}^{p_2}$, such that $p = p_1 + p_2$, and consider the partition of $\mu$ and $\Sigma$ used in Proposition 1, i.e.,

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$  

First, notice that $(X - \mu)^T \Sigma^{-1} (X - \mu) = \delta_1(X_1) + \delta_2(X_2,1)$, where $\delta_1(X_1) = (X_1 - \mu_1)^T \Sigma_{11}^{-1} (X_1 - \mu_1)$, $\delta_2(X_2,1) = (X_2 - \mu_2)^T \Sigma_{22}^{-1} (X_2 - \mu_2,1)$, $\mu_2 = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1)$ and $\Sigma_{21} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$. By the results above, the marginal pdf of $X_1$ is given by

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{p_2}} f_X(x) dx_2$$

$$= \frac{\Gamma(m)}{(\pi \nu)^{p/2} \Gamma(m-p/2)} |\Sigma_{11}|^{-1/2} \int_{\mathbb{R}^{p_2}} \left(1 + \frac{\delta_1(x_1)}{\nu} + \frac{\delta_2(x_2,1)}{\nu} \right)^{-m} dx_2$$

$$= \frac{\Gamma(m)}{(\pi \nu)^{p/2} \Gamma(m-p/2)} |\Sigma_{11}|^{-1/2} \left(1 + \frac{\delta_1(x_1)}{\nu} \right)^{-m}$$

$$= \frac{\Gamma(m-p/2)}{(\pi \nu)^{p/2} \Gamma(m-p/2)} \left(1 + \frac{\delta_1(x_1)}{\nu} \right)^{-m} |\Sigma_{11}|^{-1/2}$$

$$= \frac{\Gamma(m-p/2)}{(\pi \nu)^{p/2} \Gamma(m-p/2)} |\Sigma_{11}|^{-1/2} \left(1 + \frac{\delta_1(x_1)}{\nu} \right)^{-m}, \quad x_1 \in \mathbb{R}^{p_1}.$$  

Hence, the marginal distribution of $X_1$ is also Pearson VII with parameters $\mu_1$, $\Sigma_{11}$, $m - p_2/2$ and $\nu$, i.e., $X_1 \sim \text{PVII}_{p_1}^{1/2}(\mu_1, \Sigma_{11}, m - p/2, \nu)$. On the other hand, the conditional pdf of $X_2 | (X_1 = x_1)$ is given by

$$f_{X_2 | X_1}(x_2 | x_1) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)}$$

$$= \frac{\Gamma(m)}{(\pi (\nu + \delta_1(x_1)))^{p/2} \Gamma(m-p/2)} \left(1 + \frac{\delta_2(x_2,1)}{\nu + \delta_1(x_1)} \right)^{-m}$$

$$= \frac{\Gamma(m-p/2)}{(\pi (\nu + \delta_1(x_1)))^{p/2} \Gamma(m-p/2)} \left(1 + \frac{\delta_2(x_2,1)}{\nu + \delta_1(x_1)} \right)^{-m} |\Sigma_{11}|^{-1/2} \left(1 + \frac{\delta_1(x_1)}{\nu} \right)^{-m}, \quad x_1 \in \mathbb{R}^{p_1}, x_2 \in \mathbb{R}^{p_2}.$$  

Therefore, the conditional distribution has also a Pearson VII distribution with parameters $\mu_2, \Sigma_{22}, m$, and $\nu + \delta_1(x_1)$, i.e., $X_2 | (X_1 = x_1) \sim \text{PVII}_{p_2}^{1/2}(\mu_2, \Sigma_{22}, m, \nu + \delta_1(x_1))$.

A.1.2 Existence of its truncated moments

Let $X \sim \text{PVII}_{p}^{1/2}(\mu, \Sigma, m, \nu)$, $m > p/2$, $\nu > 0$, and let $A \subseteq \mathbb{R}^p$ be a truncation region of interest. Then, the expectation and the variance-covariance matrix of $X$ given $X \in A$ exist in the following cases:

- If $A = \mathbb{R}^p$ or $A$ is unbounded (at most one finite limit in each dimension), so the expectation exists for $m > (p+1)/2$.
If \( A \) is bounded (all truncation points are finite), then \( \mathbb{E}(X | X \in A) \) and \( \text{Cov}(X | X \in A) \) exist for all \( m > p/2 \), since the distribution is bounded.

- If \( X \) can be partitioned into two random variables \( X_1 \in \mathbb{R}^{p_1} \) and \( X_2 \in \mathbb{R}^{p_2} \) such that the truncation region associated to \( X_1 \) (say, \( A_1 \)) is bounded, from the last item we have \( \mathbb{E}(X_1 | X \in A) \) and \( \text{Cov}(X_1 | X \in A) \) exist for all \( m > p/2 \) and \( v > 0 \). On the other hand, it follows from Fubini’s theorem that \( \mathbb{E}(X_2 | X \in A) \) will exist if and only if \( \mathbb{E}(X_2 | X_1) \) exists; this occurs for all \( m > (p_2 + 1)/2 \). Note that the existence of \( \mathbb{E}(X_2 | X_1) \) also implies that \( \text{Cov}(X_1, X_2 | X \in A) \) exists. Additionally, \( \text{Cov}(X_2 | X \in A) \) exists if and only if \( \text{Cov}(X_2 | X_1) \) exists, which holds for all \( m > (p_2 + 2)/2 \).

**Remark:** It is equivalent to say that \( \mathbb{E}(X | X \in A) \) exists for all \( m \), if at least one dimension containing a finite limit exists. Besides, if at least two dimensions containing finite limits exist, we have that \( \text{Cov}(X | X \in A) \) exists for all \( m > p/2 \).

In order to illustrate the result, consider \( X \sim \text{PVII}_2(\mu, \Sigma, m, v) \), with \( v = 1 \), \( \mu = 0 \), and \( \Sigma = \begin{pmatrix} 1 & 0.20 \\ 0.20 & 1 \end{pmatrix} \).

We are interested in observing what happens with the elements of \( \mathbb{E}(X | X \in A) \) and \( \text{Cov}(X | X \in A) \) for \( A = \{ x \in \mathbb{R}^2 : a < x < b \} \) in the following three scenarios:

(a) \( m = 2 \), \( b = (\infty, \infty)^T \);
(b) \( m = 1.40 \), \( b = (0.80, \infty)^T \);
(c) \( m = 2 \), \( b = (0.80, \infty)^T \);

and lower limit \( a = (-0.80, -0.60)^T \) for all scenarios. Figure 5 displays the trace evolution of the MC estimates for the mean and variance-covariance elements \( \mu_1, \mu_2, \sigma_{11}, \sigma_{12} \) and \( \sigma_{22} \) for each case. The red dashed line represents the value for the parameter estimated via MC with \( 10^6 \) samples, and we refer to this value as the “true value”.

For the first case, we have that \( (p + 1)/2 = 3/2 < 2 \), then only the first moment exists. Therefore, we observe in the first row of Fig. 5 that only the estimates of \( \mu_1 \) and \( \mu_2 \) converge to their true values as the sample size increase. In the second scenario (middle row), we have that all elements converge except \( \sigma_{22} \). This happens because the truncation limits for the first variable are finite and \( m > (p_2 + 1)/2 = 1 \). In the last case, scenario c), convergence is attained for all parameters, since the condition \( m > (p_2 + 2)/2 = 3/2 \) holds. Note that even with 2000 MC simulations there exists a significant variability in the chains.

---

**A.2 The multivariate slash distribution**

A random vector \( X \in \mathbb{R}^p \) has multivariate slash distribution with location parameter \( \mu \in \mathbb{R}^p \), positive-definite scale matrix \( \Sigma \in \mathbb{R}^{p \times p} \), and \( v > 0 \) degrees of freedom, denoted by \( X \sim \text{SL}_p(\mu, \Sigma, v) \), if its pdf is given by

\[
f_X(x) = v \int_0^1 u^{v-1} \phi_p \left( x; \mu, u^{-1} \Sigma \right) du, \quad x \in \mathbb{R}^p,
\]

where \( \phi_p(\mathbf{x}; \mu, \Sigma) \) is the pdf of a \( p \)-variate normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \). We denote its pdf by \( \text{SL}_p(\mathbf{x}; \mu, \Sigma, v) \) which can be evaluated through numerical methods, e.g., using the R function \text{integrate}. The random vector \( X \) can also be represented in the family of the SMN distributions, this is, \( X = \mu + U^{-1/2}Z \), where the random variables \( U \) and \( Z \) are both independent and have \( \text{Beta}(v, 1) \) and \( \text{N}_p(0, \Sigma) \) distributions, respectively. Therefore, the mean and variance-covariance matrix of the random vector \( X \) are given by

\[
\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X | U)) = \mathbb{E}(\mu) = \mu.
\]

\[
\text{Cov}(X) = \text{Cov}(\mathbb{E}(X | U)) + \mathbb{E}(\text{Cov}(X | U)) = \mathbb{E}(U^{-1}) \Sigma = \frac{v}{v-1} \Sigma, \quad v > 1.
\]

**A.2.1 Marginal and conditional distribution**

Considering a partition in the same manner as used for the Pearson VII distribution, the marginal pdf of \( X_1 \) is given by

\[
f_{X_1}(x_1) = \int_{\mathbb{R}^{p_2}} f_X(x) dx_2
\]

\[
= \int_{\mathbb{R}^{p_2}} v \int_0^1 u^{v-1} \phi_p \left( x; \mu, u^{-1} \Sigma \right) du dx_2
\]

\[
= v \int_{\mathbb{R}^{p_2}} \int_0^1 u^{v-1} \phi_p \left( x_2; \mu_2, u^{-1} \Sigma_{2,2} \right) du dx_2
\]

\[
\times \phi_p \left( x_1; \mu_1, u^{-1} \Sigma_{11} \right)
\]

Thus, \( X_1 \in \mathbb{R}^{p_1} \) follows a slash distribution with location parameter \( \mu_1 \in \mathbb{R}^{p_1} \), scale matrix \( \Sigma_{11} \in \mathbb{R}^{p_1 \times p_1} \), and \( v > 0 \) degrees of freedom. On the other hand, the conditional pdf of \( X_2 | (X_1 = x_1) \) is given by

\[
f_{X_2|X_1}(x_2 | x_1) = \frac{f_{X}(x_1, x_2)}{f_{X_1}(x_1)}
\]
The evolution of the MC estimates for the mean and variance-covariance elements of $X \mid (X \in A)$ under scenarios a), b) and c). The red dashed line represents the true estimated value computed using numerical methods.

Then, it is possible to notice that the slash distribution is not closed under conditioning. Furthermore, the pdf of $X_2 \mid (X_1 = x_1)$ belongs to the elliptical family of distributions with pdf $g(t) = \int_0^1 u^{v-1/(2-1)} \exp(-u t + \delta(x_1)/2) du$, i.e., $X_2 \mid (X_1 = x_1) \sim E\mathcal{t}(\mu, \Sigma, v; g)$. To determine the mean of the random vector $X_2 \mid (X_1 = x_1)$, we compute the conditional expected value of the $i$th element of $X_2$ as follows

$$
\mathbb{E}(X_{2i} \mid X_1 = x_1) = \int_{\mathbb{R}^p} x_{2i} f_{X_2 \mid X_1}(x_2 \mid x_1) dx_2
$$

$$
= \frac{v}{f_{X_1}(x_1)} \int_{\mathbb{R}^p} x_{2i} \int_0^1 u^{v-1/(2-1)} \exp(-u t + \delta(x_1)/2) du, \quad \mu, \Sigma, v > 0,
$$

where $\mu_{2,1}^{(i)}$ represents the $i$th element of the vector $\mu_{2,1}$, and $\mathbb{E}(X_2 \mid X_1 = x_1) = \mu_{2,1}$. Now, to compute the elements of the variance-covariance matrix of the conditional random vector, we first determine $\mathbb{E}(X_{2i} X_{2j} \mid X_1 = x_1)$ for all $i, j = 1, \ldots, p_2$, as

$$
\mathbb{E}(X_{2i} X_{2j} \mid X_1 = x_1) = \int_{\mathbb{R}^p} x_{2i} x_{2j} f_{X_2 \mid X_1}(x_2 \mid x_1) dx_2
$$

$$
= \frac{v}{f_{X_1}(x_1)} \int_{\mathbb{R}^p} x_{2i} x_{2j} \int_0^1 u^{v-1/(2-1)} \exp(-u t + \delta(x_1)/2) du, \quad \mu, \Sigma, v > 0,
$$

where $\mu_{2,1}^{(i,j)}$ represents the $i$th element of the vector $\mu_{2,1}$.
Cov for each case.

consider 10%, 20%, and 40% of doubly truncated variables

family, with considering different distributions in the truncated elliptical

A complementary study of Simulation study II (Sect. 4) was

from truncated distributions

the variance-covariance matrix of a multivariate random vector

our method in order to estimate the first two moments and

conducted to examine the computational time required for

these results, we have that

\[
\text{Cov}(X_{2i}, X_{2j} | X_1 = x_1) \quad = \quad \frac{1}{\nu - 1} \left( \frac{S_{L_{p_1}}(x_1; \mu_1, \Sigma_{11}, \nu - 1)}{S_{L_{p_1}}(x_1; \mu_1, \Sigma_{11}, \nu)} \right) \sigma_{2i,j}^{(ij)}
\]

\[
\quad \quad \quad \nu > 1. \text{ Therefore, the covariance matrix of the random vector }
\]

\[
X_2 | (X_1 = x_1) \text{ will be given by }
\]

\[
\text{Cov}(X_2 | X_1 = x_1) \quad = \quad \frac{1}{\nu - 1} \left( \frac{S_{L_{p_1}}(x_1; \mu_1, \Sigma_{11}, \nu - 1)}{S_{L_{p_1}}(x_1; \mu_1, \Sigma_{11}, \nu)} \right) \Sigma_{2,1}.
\]

Appendix B: CPU time to compute moments from truncated distributions

A complementary study of Simulation study II (Sect. 4) was

conducted to examine the computational time required for

our method in order to estimate the first two moments and

the variance-covariance matrix of a p-variate random vector

considering different distributions in the truncated elliptical family, with p = 50 and 100. As in Simulation study II, we consider 10%, 20%, and 40% of doubly truncated variables for each case.

Table 3 shows the median of the CPU time (in seconds)

needed for function mvtrelliptical to compute the first two moments and the covariance matrix. We considered a TMVN, a truncated contaminated normal with \( \nu = 1/2 \) and \( \rho = 1/5 \), a truncated Pearson VII with parameters \( m = 55 \) and \( \nu = 3 \), a truncated slash with \( \nu = 2 \) degrees of freedom, and a truncated power exponential distribution with kurtosis \( \beta = 1/2 \). For each case, our method was applied setting \( n = 10^4 \) and \( 10^5 \) with \( \text{thinning} = 3 \). Notice that the time needed by the algorithm for TMVN, TMVT, and truncated Pearson VII distributions are similar and depend only on the number of truncated variables and samples used in the approximation. Our method requires more time to compute moments from the truncated contaminated normal distribution when compared to the latter results. This is because the algorithm uses a numerical method to calculate the inverse of the dgf. Besides, it is interesting noting that there is no time difference between computing the moments for a truncated slash distribution with five or ten doubly truncated variables. This occurs since the function used to approximate the integral on the dgf is more time-consuming when \( \nu + p/2 - 1 \) is not an integer. Finally, the computation of the moments for the truncated power exponential distribution required approximately the same time for random vectors of equal length regardless of the number of doubly truncated variables. For this case, the method samples values for the whole vector, leading to no time difference.

Appendix C: The rtruncated R package

The rtruncated package offers random numbers generation from members of the truncated multivariate elliptical family of distribution such as the truncated versions of the normal, Student-t, Pearson VII, slash, logistic, Kotz-type, among others. Particular distributions can be provided by specifying the density-generating function. It also computes the first two moments (covariance matrix as well) for some particular distributions. Next, we will show the functions available.

A.3 Random number generator

Its main function for random number generation is called rtruncated, which is based on the methods described in Sect. 3 of the main document, and whose signature is the following.

```r
rtruncated{
  n=1e4, mu=rep(0,length(lower)),
  Sigma=diag(length(lower)),
  lower, upper=rep(Inf,length(lower)),
  dist="Normal", nu=NULL, expr=NULL,
  gFun=NULL, ginvFun=NULL,
  burn.in=0, thinning=1
}
```

In this function, \( n \geq 1 \) is the number of observations to be sampled, \( \mu \) is the additional parameter or vector of parameters depending on the distribution of \( X \), \( \mu \) is the location parameter, \( \Sigma \) is the positive-definite scale matrix, and \( \text{lower} \) and \( \text{upper} \) are the lower and upper truncation points, respectively. The truncated normal, Student-t, power exponential, Pearson VII, slash, and contaminated normal distributions can be specified through the argument \( \text{dist} \).

The following examples illustrate the function rtruncated, for drawing samples from truncated bivariate distributions with location parameter \( \mu = (0,0)^T \), scale matrix elements \( \sigma_{11} = \sigma_{22} = 1 \), and \( \sigma_{12} = \sigma_{21} = 0.70 \), and truncation region \( A = \{ x : a < x < b \} \), with \( a = (-2, -2)^T \) and \( b = (3, 2)^T \). The distributions considered are the predefined ones in the package.

- Truncated normal

```r
rtruncated(n=1e4, mu=c(0,0),
Sigma=matrix(c(1,0.7,0.7,1),2,2),
lower=c(-2,-2), upper=c(3,2),
dist="Normal")
```
remark is that there exist closed-form expressions to compute this parameter is a non-negative scalar. An important remark is that there exist closed-form expressions to compute $\kappa_y = g^{-1}(y)$ for the normal, Student-$t$, power exponential, and Pearson VII distributions; however, the contaminated normal and slash distributions require numerical methods for this purpose. This value is calculated as the root of the function $g(t) = y - 0, t \geq 0$, through the Newton–Raphson algorithm for the contaminated normal, and using Brent’s method (Brent 2013), for the slash distribution, a mixture of linear interpolation, inverse quadratic interpolation, and the bisection method.

This function also allows generating random numbers from other truncated elliptical distributions not specified in the dist argument, by supplying the dgf through arguments either expr or gFun. The easiest way is to provide the dgf expression to argument expr as a character. The notation used in expr needs to be understood by the Ryacas package (Andersen et al. 2020), and the R environment. For instance, for the dgf $g(t) = e^{-t}$, the user must provide expr = “exp(1)ˆ(-t)”. For this case, when a character expression is provided to expr, the algorithm tries to compute a closed-form expression for the inverse function of $g(t)$; however, this is not always possible (a warning message is returned). On the other hand, if it is not possible to pass an expression to expFun, due to the complexity of the expression, the user may provide a custom R function to the ginvFun argument. By default, its inverse function is approximated numerically; however, the user may also provide its inverse to the ginvFun argument to gain some computational time. When ginvFun is provided, arguments dist and expr are ignored.

For example, to generate samples from the bivariate truncated logistic distribution with same parameters as before, and which has dgf $g(t) = e^{-t}/(1 + e^{-t})^2, t \geq 0$, we can run the following code.

\[
rtelliptical(n=1e4, mu=c(0,0),
Sigma=matrix(c(1,0.7,0.7,1), 2,2),
lower=c(-2,-2), upper=c(3,2),
dist="CN", nu=c(0.70,0.20))
\]
and $2N + p > 2$, whose dgf is $g(t) = t^{N-1}e^{-rt^2}, r \geq 0$ (Fang et al. 2018). For this distribution, $g(t)$ is not strictly decreasing for all parameter values, however, for $(2 - p)/2 < N \leq 1$, it holds. Hence, our proposal works for $r > 0, s > 0$, and $(2 - p)/2 < N \leq 1$. For this type of more complex dgf, it is advisable to pass it through the gFun argument as an R function (with other parameters as fixed values). In the following example, we draw samples from a bivariate Kotz-type distribution with settings as before, and extra parameters $r = 2, s = 1/4, N = 1/2$.

```r
rtelliptical(n=1e4,mu=c(0,0),
 Sigma=matrix(c(1,0.7,0.7,1) ,2,2),
 lower=c(-2,-2),upper=c(3,2),
gFun=function (t){ t^(-1 /2) *exp (-2 *t ^ (1/4))})
```

Figure 6 shows the scatterplot and marginal histograms for the $n = 10^4$ observations sampled from each of the truncated bivariate distributions referred above.

As mentioned by Robert and Casella (2010) and Ho et al. (2012), the slice sampling algorithm with Gibbs steps generates random samples conditioned on previous values, resulting in a sequence of correlated samples. Thus, it is essential to analyze the dependence effect of the proposed algorithm. Figure 7 displays the autocorrelation plots for each one of the distributions, where we notice that the autocorrelation drops quickly and becomes negligibly small when lags become large, evidencing well mixing and quickly converging for these examples. If necessary, initial observations can be discarded by means of the burn.in argument. Finally, autocorrelation can be decimated by setting the thinning argument. Thinning consists of picking separated points from the sample at each $k$th step. The thinning factor reduces the autocorrelation of the random points in the Gibbs sampling process. As natural, this value must be an integer greater than or equal to 1.

### A.4 Mean and variance-covariance matrix computation

Algorithm 2 for the distributions detailed in subsection 4.1 is available through the function mvtelliptical, whose signature, together with default values, is the following.

```r
mvtelliptical({
 lower,upper=rep(Inf,length(lower)),
 mu=rep(0,length(lower)),
 Sigma=diag(length(lower)),
 dist="Normal",nu=NULL,n=1e4,
 burn.in=0, thinning=3
})
```

The arguments lower and upper are the lower and upper truncation points of length $p$, respectively, $\mu$ is the location parameter of length $p$, $\Sigma$ is the $p \times p$ positive-definite scale matrix, $\nu$ is the additional parameter or vector of parameters depending on the dgf $g$. The argument dist indicates the distribution to be used. The parameters $n$, burn.in, and thinning are related to the Monte Carlo approximation, where $n$ is the number of samples to be generated, burn.in is the number of samples to be discarded as burn-in phase, and thinning is a factor for reducing autocorrelation between observations.
Fig. 7  Sample autocorrelation plots of $X_1$ and $X_2$ sampled from the bivariate truncated elliptical distributions in Fig. 6.

References

Andersen, M., Goedman, R., Grothendieck, G., et al.; Ryacas: R interface to the YACAS Computer Algebra System. R package version 1.1.3.1 (2020)

Bertolacci, M.; armspp: Adaptive rejection metropolis sampling (ARMS) via ‘Repp’ R package version 0.0.2 (2019)

Besag, J., Green, P.J.: Spatial statistics and bayesian computation. J. Roy. Stat. Soc.: Ser. B (Methodol.) 55(1), 25–37 (1993)

Brent, R.P.: Algorithms for Minimization Without Derivatives. Prentice-Hall, Englewood Cliffs, New Jersey (2013)

Damien, P., Walker, S.G.: Sampling truncated normal, beta, and gamma densities. J. Comput. Graph. Stat. 10(2), 206–215 (2001)

De Alencar, F.H., Galarza, C.E., Matos, L.A., et al.; Finite mixture modeling of censored and missing data using the multivariate skew-normal distribution. Adv. Data Anal. Class. (2021)

Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the EM algorithm. Ann. Stat. 27(1), 94–128 (1999)

Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc.: Ser. B (Methodol.) 39(1), 1–22 (1977)

Diggle, P.J., Ribeiro, P.J.: Model-based Geostatistics. Springer, New York (2007)

Fang, K.W., Kotz, S., Ng, K.W.: Symmetric Multivariate and Related Distributions. Chapman and Hall/CRC (2018)
Fridley, B.L., Dixon, P.: Data augmentation for a bayesian spatial model involving censored observations. Environmetrics 18(2), 107–123 (2007)

Galarza, C.E., Kan, R., Lachos, V.H.: MomTrunc: Moments of folded and doubly truncated multivariate distributions. R Package Version 5, 97 (2021)

Galarza, C.E., Lachos, V.H., Bourguignon, M.: A skew-t quantile regression for censored and missing data. Stat 10(1), e379 (2021)

Galarza, C.E., Lin, T.I., Wang, W.L., et al.: On moments of folded and truncated multivariate Student-t distributions based on recurrence relations. Metrika 84(6), 825–850 (2021)

Galarza, C.E., Matos, L.A., Castro, L.M., et al.: Moments of the doubly truncated selection elliptical distributions with emphasis on the unified multivariate skew-t distribution. J. Multivar. Anal. 189(104), 944 (2022)

Galarza, C.E., Matos, L.A., Lachos, V.H.: An EM algorithm for estimating the parameters of the multivariate skew-normal distribution with censored responses. METRON 80, 231–253 (2022)

Gelfand, A.E., Smith, A.F.: Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85(410), 398–409 (1990)

Gelfand, A.E., Smith, A.F., Lee, T.M.: Bayesian analysis of constrained marginal densities. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 45(4), 455–472 (1995)

Gómez, E., Gómez-Villegas, M., Marín, J.M.: A multivariate generalization of the power exponential family of distributions. Commun. Stat.-Theory Methods 27(3), 589–600 (1998)

Hadfield, J.: MCMCglmm: MCMC generalised linear mixed models. R Package Version 2, 34 (2022)

Ho, H.J., Lin, T.I., Wang, W.L., et al. TTmoment: sampling and calculating the first and second moments for the doubly truncated multivariate t distribution. R Package version 1.0 (2015)

Ho, H.J., Lin, T.I., Chen, H.Y., et al.: Some results on the truncated multivariate t distribution. J. Stat. Plan. Inference 142(1), 25–40 (2012)

Kan, R., Robotti, C.: On moments of folded and truncated multivariate normal distributions. J. Comput. Graph. Stat. 26(4), 930–934 (2017)

Lachos, V.H., Matos, L.A., Barbosa, T.S., et al.: Influence diagnostics in spatial models with censored response. Environmetrics 28(7), e2464 (2017)

Lachos, V.H., A. Matos, L., Castro, L.M., et al.: Flexible longitudinal linear mixed models for multiple censored responses data. Stat. Med. 38(6), 1074–1102 (2019)

Martino, L., Read, J., Luengo, D.: Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling. IEEE Trans. Signal Process. 63(12), 3123–3138 (2015)

Martino, L., Yang, H., Luengo, D., et al.: A fast universal self-tuned sampler within Gibbs sampling. Digit. Signal Process. 47, 68–83 (2015)

Matos, L.A., Prates, M.O., Chen, M.H., et al.: Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. Stat. Sin. 23(3), 1323–1345 (2013)

Matos, L.A., Castro, L.M., Lachos, V.H.: Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. TEST 25(4), 627–653 (2016)

Mats, T.B., Lachos, V.H., Castro, L.M., et al.: Extending multivariate Student’s-t semiparametric mixed models for longitudinal data with censored responses and heavy tails. Stat. Med. 41(19), 3696–3719 (2022)

Meyer, R., Cai, B., Perron, F.: Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. Comput. Stat. Data Anal. 52(7), 3408–3423 (2008)

Morán-Vásquez, R.A., Ferrari, S.L.: New results on truncated elliptical distributions. Commun. Math. Stat. 9, 299–313 (2021)

Muirhead, R.J.: Aspects of Multivariate Statistical Theory, vol. 197. Wiley, New York (2009)

Nash, J.C., Varadhan, R., Grotendieck, G.: optimx: Expanded replacement and extension of the ‘optim’ function. R package version 2020-4.2 (2020)

Neal, R.M.: Slice sampling. Ann. Stat. 31(3), 705–767 (2003)

Olivi, R.C., Zhong, K., Garay, A.M., et al.: ARplMEC: censored mixed-effects models with different correlation structures. R Package Version 2(4), 1 (2022)

Ordoñez, J.A., Bandyopadhyay, D., Lachos, V.H., et al.: Geostatistical estimation and prediction for censored responses. Spatial Stat. 23, 109–123 (2018)

Pan, Y., Pan, J.: roript: An R package for general purpose optimization with C++. R package version 0.1.6 (2022)

R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2021)

Robert, C.P.: Simulation of truncated normal variables. Stat. Comput. 5(2), 121–125 (1995)

Robert, C.P., Casella, G.: Introducing Monte Carlo Methods with R, vol. 18. Springer, New York (2010)

Swendsen, R.H., Wang, J.S.: Nonuniversal critical dynamics in Monte Carlo simulations. Phys. Rev. Lett. 58(2), 86 (1987)

Tallis, G.M.: The moment generating function of the truncated multivariate normal distribution. J. Roy. Stat. Soc.: Ser. B (Methodol.) 23(1), 223–229 (1961)

Valeriano, K.A., Ordoñez, A., Galarza, C.E., et al.: RcppCensSpatial: Spatial estimation and prediction for censored/missing responses. R package version 0.3.0 (2022)

Wei, G.C., Tanner, M.A.: A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. J. Am. Stat. Assoc. 85(411), 699–704 (1990)

Wilhelm, S.: tmvtnorm: Truncated multivariate normal and Student t distributions. R Package Version 1, 5 (2022)

Zirkhly, J.H., Harris, D.J.: Geostatistical analysis of hazardous waste site data. J. Environ. Eng. 112(4), 770–784 (1986)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.