*Journal of Behavioral Data Science*, 2021, *1* (1), 17–33.
DOI: https://doi.org/10.35566/jbds/v1n1/p2

# Moments Calculation for the Doubly Truncated Multivariate Normal Density

B. G. Manjunath[1] and Stefan Wilhelm[2]

[1] School of Mathematics and Statistics,
University of Hyderabad, Hyderabad, India.
bgmanjunath@gmail.com
[2] Department of Finance, University of Basel, Switzerland
Stefan.Wilhelm@financial.com

**Abstract.** In the present article, we derive an explicit expression for the truncated mean and variance for the multivariate normal distribution with arbitrary rectangular double truncation. We use the moment generating approach of Tallis (1961) and extend it to general $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and all combinations of truncation. As part of the solution, we also give a formula for the bivariate marginal density of truncated multinormal variates. We also prove an invariance property of some elements of the inverse covariance after truncation. Computer algorithms for computing the truncated mean, variance and the bivariate marginal probabilities for doubly truncated multivariate normal variates have been written in R and are presented along with three examples.

*Keywords:* Multivariate normal · Double truncation · Moment generating function · Bivariate marginal density function · Graphical models · Conditional independence

## 1 Introduction

The multivariate normal distribution arises frequently and has a wide range of applications in fields such as multivariate regression, Bayesian statistics and the analysis of Brownian motion. One motivation to deal with moments of the truncated multivariate normal distribution comes from the analysis of special financial derivatives ("auto-callables" or "Expresszertifikate") in Germany. These products can expire early depending on some restrictions of the underlying trajectory, if the underlying value is above or below certain call levels. In the framework of Brownian motion the finite-dimensional distributions for log returns at any $d$ points in time are multivariate normal. When some of the multinormal variates $\mathbf{X} = (x_1, \ldots, x_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are subject to inequality constraints (e.g. $a_i \leq x_i \leq b_i$), this results in truncated multivariate normal distributions.

Several types of truncations and their moment calculation have been described so far, for example the one-sided rectangular truncation $\boldsymbol{x} \geq \boldsymbol{a}$ (Tallis,

1961), the rather unusual elliptical and radial truncations $\boldsymbol{a} \leq \boldsymbol{x}'\mathbf{R}\boldsymbol{x} \leq \boldsymbol{b}$ (Tallis, 1963) and the plane truncation $\mathbf{C}\boldsymbol{x} \geq \boldsymbol{p}$ (Tallis, 1965). Linear constraints like $\boldsymbol{a} \leq \mathbf{C}\boldsymbol{x} \leq \boldsymbol{b}$ can often be reduced to rectangular truncation by transformation of the variables (in case of a full rank matrix $\mathbf{C} : \boldsymbol{a}^* = \mathbf{C}^{-1}\boldsymbol{a} \leq \boldsymbol{x} \leq \mathbf{C}^{-1}\boldsymbol{b} = \boldsymbol{b}^*$), which makes the double rectangular truncation $\boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}$ especially important.

The existing works on moment calculations differ in the number of variables they consider (univariate, bivariate, multivariate) and the types of rectangular truncation they allow (single vs. double truncation). Single or one-sided truncation can be either from above ($\boldsymbol{x} \leq \boldsymbol{a}$) or below ($\boldsymbol{x} \geq \boldsymbol{a}$), but only on one side for all variables, whereas double truncation $\boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}$ can have both lower and upper truncation points. Other distinguishing features of previous works are further limitations or restrictions they impose on the type of distribution (e.g. zero mean) and the methods they use to derive the results (e.g. direct integration or moment-generating function).

Rosenbaum (1961) gave an explicit formula for the moments of the bivariate case with single truncation from below in both variables by direct integration. His results for the bivariate normal distribution have been extended by Shah and Parikh (1964), Regier and Hamdan (1971) and Muthén (1990) to double truncation.

For the multivariate case, Tallis (1961) derived an explicit expression for the first two moments in case of a singly truncated multivariate normal density with zero mean vector and the correlation matrix $\mathbf{R}$ using the moment generating function. Amemiya (1974) and Lee (1979) extended the Tallis (1961) derivation to a general covariance matrix $\boldsymbol{\Sigma}$ and also evaluated the relationship between the first two moments. Gupta and Tracy (1976) and Lee (1983) gave very simple recursive relationships between moments of any order for the doubly truncated case. However, except for the mean, there are fewer equations than parameters. Therefore, these recurrent conditions do not uniquely identify moments of order $\geq 2$ and are not sufficient for the computation of the variance and other higher order moments.

Table 1 summarizes our survey of existing publications dealing with the computation of truncated moments and their limitations. Even though the rectangular truncation $\boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}$ can be found in many situations, no explicit moment formulas for the truncated mean and variance in the general multivariate case of double truncation from below and/or above have been presented so far in the literature and are readily apparent. The contribution of this paper is to derive these formulas for the first two truncated moments and to extend and generalize existing results on moment calculations from especially Tallis (1961), Lee (1983), Leppard and Tallis (1989), and Muthén (1990). Besides, we also refer Kan and Robotti (2017) and Arismendi (2013) for the moment computation of folded and truncated multivariate normal distribution. However, for moments computation for skewed and extended skew normal distribution, we refer Kan and Robotti (2017) and Arellano-Valle and Genton (2005). In the sequel, we also make a note on the existing R package "MomTrunc" (see Galarza C.E. & V.H., 2021) for numerical computation of moments of folded and truncated multivariate normal

**Table 1.** Survey of previous works on the moments for the truncated multivariate normal distribution

| Author | #Variates | Truncation | Focus |
|---|---|---|---|
| Rosenbaum (1961) | bivariate | single | moments for bivariate normal variates with single truncation, $b_1 < y_1 < \infty, b_2 < y_2 < \infty$ |
| Tallis (1961) | multivariate | single | moments for multivariate normal variates with single truncation from below |
| Shah and Parikh (1964) | bivariate | double | recurrence relations between moments |
| Regier and Hamdan (1971) | bivariate | double | an explicit formula only for the case of truncation from below at the same point in both variables |
| Amemiya (1974) | multivariate | single | relationship between first and second moments |
| Gupta and Tracy (1976) | multivariate | double | recurrence relations between moments |
| Lee (1979) | multivariate | single | recurrence relations between moments |
| Lee (1983) | multivariate | double | recurrence relations between moments |
| Leppard and Tallis (1989) | multivariate | single | moments for multivariate normal distribution with single truncation |
| Muthén (1990) | bivariate | double | moments for bivariate normal distribution with double truncation, $b_1 < y_1 < a_1, b_2 < y_2 < a_2$ |
| Manjunath and Wilhelm | multivariate | double | moments for multivariate normal distribution with double truncation in all variables $\boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}$ |

and Student's t-distribution. However, the aforementioned package also suggests the "tmvtnorm" package (e.g., Wilhelm & Manjunath, 2012), which is solely based on the results presented in this note. Finally, we also refer Genz (1992) for the numerical computation for the multivariate normal probabilities.

The rest of this paper is organized as follows. Section 2 presents the moment generating function (m.g.f) for the doubly truncated multivariate normal case. In Section 3, we derive the first and second moments by differentiating the m.g.f. These results are completed in Section 4 by giving a formula for computing the bivariate marginal density. In Section 5, we present two numerical examples and compare our results with simulation results. Section 6 links our results to the theory of graphical models and derives some properties of the inverse covariance matrix. Finally, Section 7 summarizes our results and gives an outlook for further research.

## 2 Moment Generating Function

The $d$–dimensional normal density with location parameter vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and non-singular covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\varphi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\}, \ \boldsymbol{x} \in \mathbb{R}^d. \quad (1)$$

The pertaining distribution function is denoted by $\Phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x})$. Correspondingly, the multivariate truncated normal density, truncated at $\boldsymbol{a}$ and $\boldsymbol{b}$, in $\mathbb{R}^d$, is defined as

$$\varphi_{\alpha\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \begin{cases} \frac{\varphi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x})}{P\{\boldsymbol{a}\leq\mathbf{X}\leq\boldsymbol{b}\}}, & \text{for } \boldsymbol{a} \leq \boldsymbol{x} \leq \boldsymbol{b}, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Denote $\alpha = P\{\boldsymbol{a} \leq \mathbf{X} \leq \boldsymbol{b}\}$ as the fraction after truncation.

The moment generating function of a $d$–dimensional truncated random variable $\mathbf{X}$, truncated at $\boldsymbol{a}$ and $\boldsymbol{b}$, in $\mathbb{R}^d$, having the density $f(\boldsymbol{x})$ is defined as the $d$–fold integral of the form

$$m(\boldsymbol{t}) = E\left(e^{\boldsymbol{t}'\mathbf{X}}\right) = \int_{\boldsymbol{a}}^{\boldsymbol{b}} e^{\boldsymbol{t}'\boldsymbol{x}} f(\boldsymbol{x}) d\boldsymbol{x}.$$

Therefore, the m.g.f for the density in (2) is

$$m(\boldsymbol{t}) = \frac{1}{\alpha(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_{\boldsymbol{a}}^{\boldsymbol{b}} \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{x}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right) - 2\boldsymbol{t}'\boldsymbol{x}\right]\right\} d\boldsymbol{x}. \quad (3)$$

In the following, the moments are first derived for the special case $\boldsymbol{\mu} = \mathbf{0}$. Later, the results will be generalized to all $\boldsymbol{\mu}$ by applying a location transformation.

Now, consider only the exponent term in 3 for the case $\boldsymbol{\mu} = \mathbf{0}$. Then we have

$$-\frac{1}{2}\left[\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - 2\boldsymbol{t}'\boldsymbol{x}\right]$$

which can also be written as

$$\frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t} - \frac{1}{2}\left[\left(\boldsymbol{x}-\boldsymbol{\xi}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\xi}\right)\right],$$

where $\boldsymbol{\xi} = \boldsymbol{\Sigma}\boldsymbol{t}$.

Consequently, the m.g.f of the rectangularly doubly truncated multivariate normal is

$$m(\boldsymbol{t}) = \frac{e^T}{\alpha(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_{\boldsymbol{a}}^{\boldsymbol{b}} \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{x}-\boldsymbol{\xi}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\xi}\right)\right]\right\} d\boldsymbol{x}, \quad (4)$$

where $T = \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}$.

The above equation can be further reduced to

$$m(\boldsymbol{t}) = \frac{e^T}{\alpha(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_{\boldsymbol{a}-\boldsymbol{\xi}}^{\boldsymbol{b}-\boldsymbol{\xi}} \exp\left\{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right\} d\boldsymbol{x}. \tag{5}$$

For notational convenience, we write equation 5 as

$$m(\boldsymbol{t}) = e^T \Phi_{\alpha\boldsymbol{\Sigma}} \tag{6}$$

where

$$\Phi_{\alpha\boldsymbol{\Sigma}} = \frac{1}{\alpha(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_{\boldsymbol{a}-\boldsymbol{\xi}}^{\boldsymbol{b}-\boldsymbol{\xi}} \exp\left\{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right\} d\boldsymbol{x}.$$

## 3 First and Second Moment Calculation

In this section, we derive the first and second moments of the rectangularly doubly truncated multivariate normal density by differentiating the m.g.f..

Consequently, by taking the partial derivative of (6) with respect to $t_i$ we have

$$\frac{\partial m(\boldsymbol{t})}{\partial t_i} = e^T \frac{\partial \Phi_{\alpha\boldsymbol{\Sigma}}}{\partial t_i} + \Phi_{\alpha\boldsymbol{\Sigma}} \frac{\partial e^T}{\partial t_i}. \tag{7}$$

In the above equation the only essential terms that will be simplified are

$$\frac{\partial e^T}{\partial t_i} = e^T \sum_{k=1}^{d} \sigma_{i,k} t_k$$

and

$$\frac{\partial \Phi_{\alpha\boldsymbol{\Sigma}}}{\partial t_i} = \frac{\partial}{\partial t_i} \int_{a_1^*}^{b_1^*} ... \int_{a_d^*}^{b_d^*} \varphi_{\alpha\boldsymbol{\Sigma}}(\boldsymbol{x}) dx_d ... dx_1, \tag{8}$$

where $a_i^* = a_i - \sum_{k=1}^{d} \sigma_{i,k} t_k$ and $b_i^* = b_i - \sum_{k=1}^{d} \sigma_{i,k} t_k$. Subsequently, (8) is

$$\frac{\partial \Phi_{\alpha\boldsymbol{\Sigma}}}{\partial t_i} = \sum_{k=1}^{d} \sigma_{i,k} \left(F_k(a_k^*) - F_k(b_k^*)\right), \tag{9}$$

where

$$F_i(x) = \int_{a_1^*}^{b_1^*} ... \int_{a_{i-1}^*}^{b_{i-1}^*} \int_{a_{i+1}^*}^{b_{i+1}^*} ... \int_{a_d^*}^{b_d^*} \varphi_{\alpha\boldsymbol{\Sigma}}(x_1, .., x_{i-1}, x, x_{i+1}, ..x_d) dx_d ... dx_{i+1} dx_{i-1} ... dx_1. \tag{10}$$

Note that at $t_k = 0$, for all $k = 1, 2, ..., d$, we have $a_i^* = a_i$ and $b_i^* = b_i$. Therefore, $F_i(x)$ will be the $i$–th marginal density. An especially convenient way of computing these one-dimensional marginals is given in Cartinhour (1990).

From (7) – (9) for $k = 1, 2, ..., d$ all $t_k = 0$. Hence, the first moment is

$$E(X_i) = \frac{\partial m(\boldsymbol{t})}{\partial t_i}|_{\boldsymbol{t}=\mathbf{0}} = \sum_{k=1}^{d} \sigma_{i,k} \left( F_k(a_k) - F_k(b_k) \right). \tag{11}$$

Now, by taking the partial derivative of (7) with respect to $t_j$, we have

$$\frac{\partial^2 m(\boldsymbol{t})}{\partial t_j \partial t_i} = e^T \frac{\partial^2 \Phi_{\alpha \boldsymbol{\Sigma}}}{\partial t_j \partial t_i} + \frac{\partial \Phi_{\alpha \boldsymbol{\Sigma}}}{\partial t_i} \frac{\partial e^T}{\partial t_j} + \Phi_{\alpha \boldsymbol{\Sigma}} \frac{\partial^2 e^T}{\partial t_j \partial t_i} + \frac{\partial e^T}{\partial t_i} \frac{\partial \Phi_{\alpha \boldsymbol{\Sigma}}}{\partial t_j}. \tag{12}$$

The essential terms for simplification are

$$\frac{\partial^2 e^T}{\partial t_j \partial t_i} = \sigma_{i,j}$$

and clearly, the partial derivative of 9 with respect to $t_j$ gives

$$\frac{\partial^2 \Phi_{\alpha \boldsymbol{\Sigma}}}{\partial t_j \partial t_i} = \sum_{k=1}^{d} \left( \sigma_{i,k} \frac{\partial F_k(a_k^*)}{\partial t_j} \right) - \sum_{k=1}^{d} \left( \sigma_{i,k} \frac{\partial F_k(b_k^*)}{\partial t_j} \right). \tag{13}$$

In the above equation, merely consider the partial derivative of the marginal density $F_k(a_k^*)$ with respect to $t_j$. With further simplification, it reduces to

$$\begin{aligned} \frac{\partial F_k(a_k^*)}{\partial t_j} &= \frac{\partial}{\partial t_j} \int_{a_1^*}^{b_1^*} ... \int_{a_{k-1}^*}^{b_{k-1}^*} \int_{a_{k+1}^*}^{b_{k+1}^*} ... \int_{a_d^*}^{b_d^*} \varphi_{\alpha \boldsymbol{\Sigma}}(x_1, .., x_{k-1}, a_k^*, x_{k+1}, ..x_d) d\boldsymbol{x}_{-k} \\ &= \frac{\sigma_{j,k} a_k^* F_k(a_k^*)}{\sigma_{k,k}} \\ &\quad + \sum_{q \neq k} \left( \sigma_{j,q} - \frac{\sigma_{k,q} \sigma_{j,k}}{\sigma_{k,k}} \right) \left( F_{k,q}(a_k^*, a_q^*) - F_{k,q}(a_k^*, b_q^*) \right), \end{aligned} \tag{14}$$

where

$$\begin{aligned} &F_{k,q}(x, y) = \\ &\int_{a_1^*}^{b_1^*} ... \int_{a_{k-1}^*}^{b_{k-1}^*} \int_{a_{k+1}^*}^{b_{k+1}^*} ... \int_{a_{q-1}^*}^{b_{q-1}^*} \int_{a_{q+1}^*}^{b_{q+1}^*} ... \int_{a_d^*}^{b_d^*} \varphi_{\alpha \boldsymbol{\Sigma}}(x, y, \boldsymbol{x}_{-k,-q}) d\boldsymbol{x}_{-k,-q}, \end{aligned} \tag{15}$$

and the short form $\boldsymbol{x}_{-k}$ denotes the vector $(x_1, .., x_{k-1}, x_{k+1}, ..x_d)'$ in $(d-1)$–dimensions and $\boldsymbol{x}_{-k,-q}$ denotes the $(d-2)$–dimensional vector $(x_1, ..., x_{k-1}, x_{k+1}, ..., x_{q-1}, x_{q+1}, ..., x_d)'$ for $k \neq q$. The above equation (14) is deduced from Lee (1979, pp. 167). Note that for all $t_k = 0$ the term $F_{k,q}(x, y)$ will be the bivariate marginal density for which we will give a formula in the next section.

Subsequently, $\frac{\partial F_k(b_k^*)}{\partial t_j}$ can be obtained by substituting $a_k^*$ by $b_k^*$. From 12 – (15) at all $t_k = 0$, $k = 1, 2, ..., d$, the second moment is

$$\begin{aligned} E(X_i X_j) &= \frac{\partial^2 m(\boldsymbol{t})}{\partial t_j \partial t_i} | \boldsymbol{t}=\mathbf{0} \\ &= \sigma_{i,j} + \sum_{k=1}^{d} \sigma_{i,k} \frac{\sigma_{j,k} \left(a_k F_k(a_k) - b_k F_k(b_k)\right)}{\sigma_{k,k}} \\ &\quad + \sum_{k=1}^{d} \sigma_{i,k} \sum_{q \neq k} \left( \sigma_{j,q} - \frac{\sigma_{k,q}\sigma_{j,k}}{\sigma_{k,k}} \right) \left[ (F_{k,q}(a_k, a_q) - F_{k,q}(a_k, b_q)) \right. \\ &\quad \left. - (F_{k,q}(b_k, a_q) - F_{k,q}(b_k, b_q)) \right]. \end{aligned} \tag{16}$$

Having derived expressions for the first and second moments for double truncation in case of $\boldsymbol{\mu} = \mathbf{0}$, we will now generalize to all $\boldsymbol{\mu}$. If $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{a}^* \leq \boldsymbol{y} \leq \boldsymbol{b}^*$, then $\mathbf{X} = \boldsymbol{Y} - \boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{a} = \boldsymbol{a}^* - \boldsymbol{\mu} \leq \boldsymbol{x} \leq \boldsymbol{b}^* - \boldsymbol{\mu} = \boldsymbol{b}$ and $E(\boldsymbol{Y}) = E(\mathbf{X}) + \boldsymbol{\mu}$ and $Cov(\boldsymbol{Y}) = Cov(\mathbf{X})$. Equations 11 and 16 can then be used to compute $E(\mathbf{X})$ and $Cov(\mathbf{X})$. Hence, for general $\boldsymbol{\mu}$, the first moment is

$$E(Y_i) = \sum_{k=1}^{d} \sigma_{i,k} \left(F_k(a_k) - F_k(b_k)\right) + \mu_i. \tag{17}$$

The covariance matrix

$$Cov(Y_i, Y_j) = Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) \tag{18}$$

is invariant to the shift in location.

The equations 17 and 18 in combination with 11 and 16 form our derived result allow the calculation of the truncated mean and truncated variance for general double truncation. A formula for the term $F_{k,q}(x_k, x_q)$, the bivariate marginal density, will be given in the next section.

We have implemented the moment calculation for mean vector `mean`, covariance matrix `sigma` and truncation vectors `lower` and `upper` as a function `mtmvnorm(mean, sigma, lower, upper)` in the R package `tmvtnorm` (Wilhelm & Manjunath, 2010, 2012), where the code is open source. In Section 5, we will show an example of this function.

## 4 Bivariate Marginal Density Computation

In order to compute the bivariate marginal density in this section, we follow Tallis (1961, p. 223) and Leppard and Tallis (1989) that implicitly used the bivariate marginal density as part of the moments calculation for single truncation, evaluated at the integration bounds. However, we extend it to the doubly truncated case and state the function for all points within the support region.

Without loss of generality we use a z-transformation for all variates $\boldsymbol{x} = (x_1, \ldots, x_d)'$ as well as for all lower and upper truncation points $\boldsymbol{a} = (a_1, \ldots, a_d)'$ and $\boldsymbol{b} = (b_1, \ldots, b_d)'$, resulting in a $N(0, \mathbf{R})$ distribution with correlation matrix $\mathbf{R}$ for the standardized untruncated variates. In this section we treat all variables as if they are z-transformed, leaving the notation unchanged.

For computing the bivariate marginal density $F_{q,r}(x_q, x_r)$ with $a_q \leq x_q \leq b_q, a_r \leq x_r \leq b_r, q \neq r$, we use the fact that for truncated normal densities the conditional densities are also truncated normal. The following relationship holds for $x_s, z_s \in \mathbb{R}^{d-2}$ conditionally on $x_q = c_q$ and $x_r = c_r$ $(s \neq q \neq r)$:

$$\alpha^{-1}\varphi_d(x_s, x_q = c_q, x_r = c_r; \mathbf{R}) = \alpha^{-1}\varphi(c_q, c_r; \rho_{qr})\varphi_{d-2}(z_s; \mathbf{R}_{qr}), \tag{19}$$

where

$$z_s = (x_s - \beta_{sq.r}c_q - \beta_{sr.q}c_r)/\sqrt{(1-\rho^2_{sq})(1-\rho^2_{sr.q})} \tag{20}$$

and $\mathbf{R}_{qr}$ is the matrix of the second-order partial correlation coefficients for $s \neq q \neq r$. $\beta_{sq.r}$ and $\beta_{sr.q}$ are the partial regression coefficients of $x_s$ on $x_q$ and $x_r$ , respectively, and $\rho_{sr.q}$ is the partial correlation coefficient between $x_s$ and $x_r$ for fixed $x_q$.

Integrating out $(d-2)$ variables $x_s$ leads to $F_{q,r}(x_q, x_r)$ as a product of a bivariate normal density $\varphi(x_q, x_r)$ and a $(d-2)$-dimension normal integral $\Phi_{d-2}$:

$$\begin{aligned} F_{q,r}(x_q = c_q, x_r = c_r) = &\int_{a_1}^{b_1} \ldots \int_{a_{q-1}}^{b_{q-1}} \int_{a_{q+1}}^{b_{q+1}} \ldots \int_{a_{r-1}}^{b_{r-1}} \\ &\int_{a_{r+1}}^{b_{r+1}} \ldots \int_{a_d}^{b_d} \varphi_{\alpha R}(x_s, c_q, c_r) dx_s \\ &= \alpha^{-1}\varphi(c_q, c_r; \rho_{qr})\Phi_{d-2}(A^q_{rs}; B^q_{rs}; \mathbf{R}_{qr}) \end{aligned} \tag{21}$$

where $A^q_{rs}$ and $B^q_{rs}$ denote the lower and upper integration bounds of $\Phi_{d-2}$ given $x_q = c_q$ and $x_r = c_r$:

$$A^q_{rs} = (a_s - \beta_{sq.r}c_q - \beta_{sr.q}c_r)/\sqrt{(1-\rho^2_{sq})(1-\rho^2_{sr.q})} \tag{22}$$

$$B^q_{rs} = (b_s - \beta_{sq.r}c_q - \beta_{sr.q}c_r)/\sqrt{(1-\rho^2_{sq})(1-\rho^2_{sr.q})}. \tag{23}$$

The computation of $F_{q,r}(x_q, x_r)$ just needs the evaluation of the normal integral $\Phi_{d-2}$ in $d-2$ dimensions, which is readily available in most statistics software packages, for example, as the function `pmvnorm()` in the R package `mvtnorm` (Genz et al., 2012). The bivariate marginal density function `dtmvnorm(x, mean, sigma, lower, upper, margin=c(q,r))` is also part of the R package `tmvtnorm` (Wilhelm & Manjunath, 2010, 2012), where readers can find the source code as well as help files and additional examples.

## 5 Numerical Examples

### Example 1

We will use the following bivariate example with $\boldsymbol{\mu} = (0.5, 0.5)'$ and covariance matrix $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1.2 \\ 1.2 & 2 \end{pmatrix}$$

as well as lower and upper truncation points $\boldsymbol{a} = (-1, -\infty)', \boldsymbol{b} = (0.5, 1)'$, i.e. $x_1$ is doubly, while $x_2$ is singly truncated. The bivariate marginal density $F_{q,r}(x, y)$ is the density function itself and is shown in figure 1, and the one-dimensional densities $F_k(x)$ $(k = 1, 2)$ are shown in in figure 2.

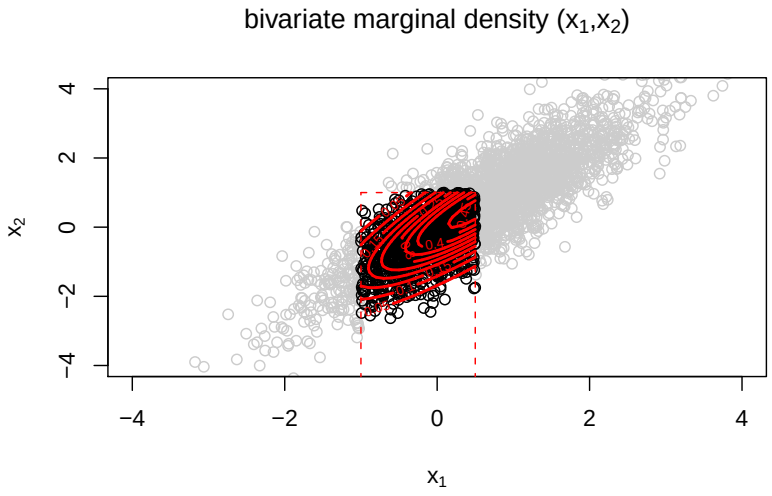



**Figure 1.** Contour plot for the bivariate truncated density function

The moment calculation for our example can be performed in R as

```
> library(tmvtnorm)
> mu    <- c(0.5, 0.5)
> sigma <- matrix(c(1, 1.2, 1.2, 2), 2, 2)
> a     <- c(-1, -Inf)
> b     <- c(0.5, 1)
> moments <- mtmvnorm(mean=mu, sigma=sigma,
>             lower=a, upper=b)
```

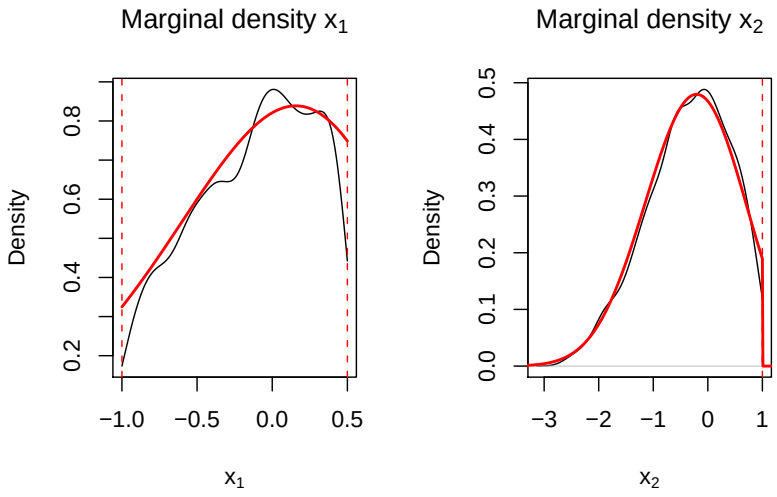



**Figure 2.** Marginal densities $F_k(x)$ $(k = 1, 2)$ for $x_1$ and $x_2$ obtained from Kernel density estimation of random samples and from direct evaluation of $F_k(x)$

which leads to the results $\boldsymbol{\mu}^* = (-0.152, -0.388)'$ and covariance matrix

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 0.163 & 0.161 \\ 0.161 & 0.606 \end{pmatrix}.$$

The trace plots in figures 3 and 4 show the evolution of a Monte Carlo estimate for the elements of the mean vector and the covariance matrix respectively for growing sample sizes. Furthermore, the 95% confidence interval obtained from Monte Carlo using the full sample of 10000 items is shown. All confidence intervals contain the true theoretical value, but Monte Carlo estimates still show substantial variation even with a sample size of 10000. Simulation from a truncated multivariate distribution and calculating the sample mean or the sample covariance also leads to consistent estimates of $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$. Since the rate of convergence of the MC estimator is $O(\sqrt{n})$, one has to ensure sufficient Monte Carlo iterations in order to have a good approximation or to choose variance reduction techniques.

### Example 2

Let $\boldsymbol{\mu} = (0, 0, 0)'$,the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.1 & 1.2 & 0 \\ 1.2 & 2 & -0.8 \\ 0 & -0.8 & 3 \end{pmatrix}$$

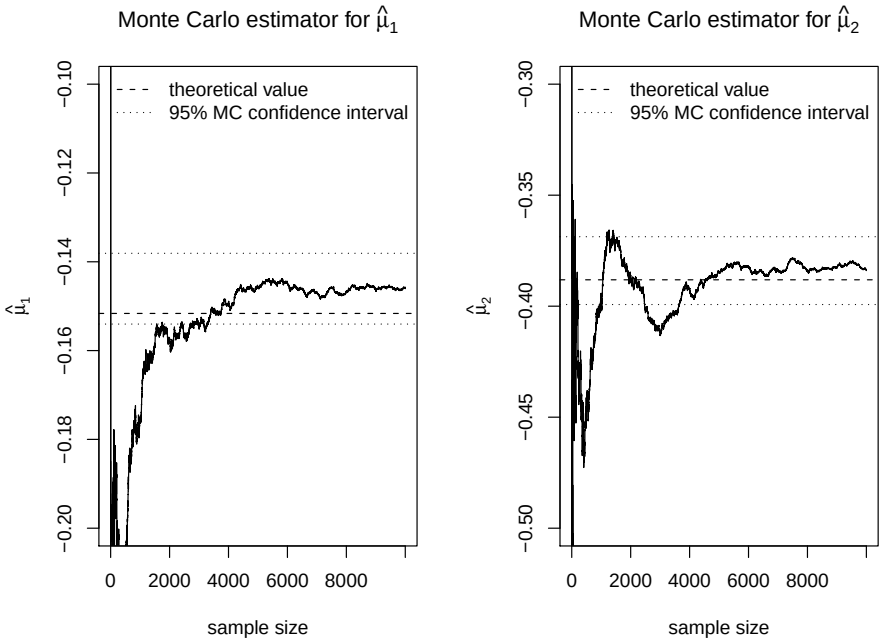



**Figure 3.** Trace plots of the Monte Carlo estimator for $\boldsymbol{\mu}^*$

and the lower and upper truncation points $\boldsymbol{a} = (-1, -\infty, -\infty)'$ and $\boldsymbol{b} = (0.5, \infty, \infty)'$. Then the only truncated variable is $x_1$, which is uncorrelated with $x_3$. Our formula results in $\boldsymbol{\mu}^* = c(-0.210, -0.229, 0)'$ and

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 0.174 & 0.190 & 0.0 \\ 0.190 & 0.898 & -0.8 \\ 0 & -0.8 & 3.0 \end{pmatrix}$$

For the special case of only $k < d$ truncated variables $(x_1, \ldots, x_k)$, the remaining $d - k$ variables $(x_{k+1}, \ldots, x_d)$ can be regressed on the truncated variables, and a simple formula for the mean and covariance matrix can be given (see Johnson & Kotz, 1971, p. 70).

Let the covariance matrix $\boldsymbol{\Sigma}$ of $(x_1, \ldots, x_d)$ be partitioned as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \tag{24}$$

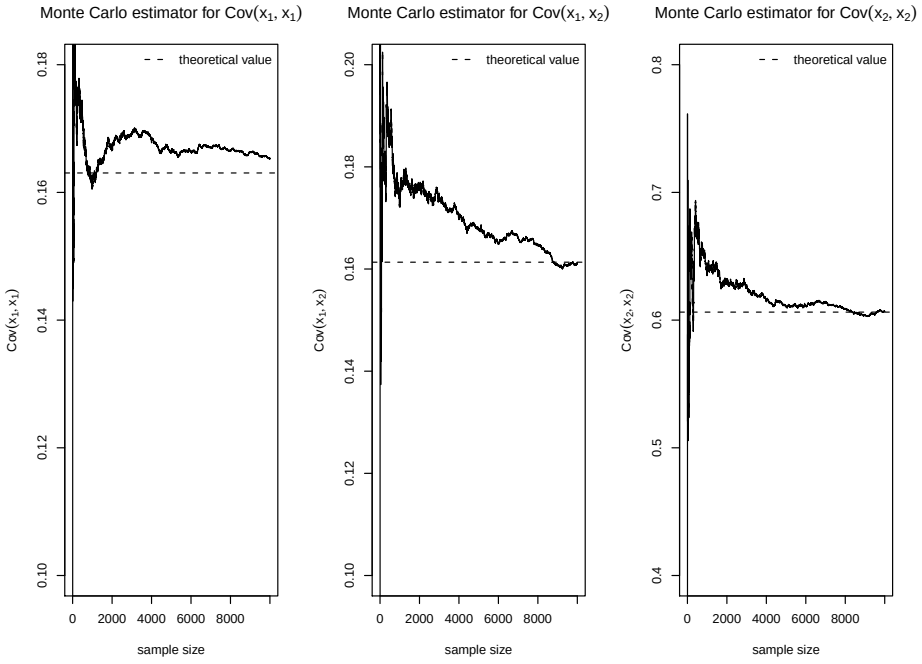



**Figure 4.** Trace plots of the Monte Carlo estimator for the 3 elements of $\boldsymbol{\Sigma}^*$ ($\sigma_{11}^*$, $\sigma_{12}^* = \sigma_{21}^*$ and $\sigma_{22}^*$)

where $\mathbf{V}_{11}$ denotes the $k \times k$ covariance matrix of $(x_1, \ldots, x_k)$. The mean vector[3] and the covariance matrix $\boldsymbol{\Sigma}^*$ of all $d$ variables can be computed as

$$(\boldsymbol{\xi}_\mathbf{1}', \boldsymbol{\xi}_\mathbf{1}' \mathbf{V}_{11}^{-1} \mathbf{V}_{12}) \tag{25}$$

and

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{11}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} \\ \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{U}_{11} & \mathbf{V}_{22} - \mathbf{V}_{21}(\mathbf{V}_{11}^{-1} - \mathbf{V}_{11}^{-1}\mathbf{U}_{11}\mathbf{V}_{11}^{-1})\mathbf{V}_{12} \end{pmatrix} \tag{26}$$

where $\boldsymbol{\xi}_\mathbf{1}'$ and $\mathbf{U}_{11}$ are the mean and covariance of the $(x_1, \ldots, x_k)$ after truncation.

The mean and standard deviation for the univariate truncated normal $x_1$ are

$$\xi_1 = \mu_1^* = \sigma_{11} \frac{\varphi_{\mu_1,\sigma_{11}}(a_1) - \varphi_{\mu_1,\sigma_{11}}(b_1)}{\Phi_{\mu_1,\sigma_{11}}(b_1) - \Phi_{\mu_1,\sigma_{11}}(a_1)},$$

$$\sigma_{11}^* = \sigma_{11} + \sigma_{11} \frac{a_1 \varphi_{\mu_1,\sigma_{11}}(a_1) - b_1 \varphi_{\mu_1,\sigma_{11}}(b_1)}{\Phi_{\mu_1,\sigma_{11}}(b_1) - \Phi_{\mu_1,\sigma_{11}}(a_1)}.$$

[3] The formula for the truncated mean given in Johnson and Kotz (1971, p. 70) is only valid for a zero-mean vector or after demeaning all variables appropriately. For non-zero means $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)'$ it will be $(\boldsymbol{\xi}_\mathbf{1}', \boldsymbol{\mu}_2 + (\boldsymbol{\xi}_\mathbf{1}' - \boldsymbol{\mu}_1)\mathbf{V}_{11}^{-1}\mathbf{V}_{12})$.

Letting $\mathbf{U}_{11} = \sigma_{11}^*$ and inserting $\xi_1$ and $\mathbf{U}_{11}$ into equations 25 and 26, one can verify our formula and the results for $\boldsymbol{\mu}^*$ and $\boldsymbol{\Sigma}^*$. However, the crux in using the Johnson/Kotz formula is the need to first compute the moments of the truncated variables $(x_1, \ldots, x_k)$ for $k \geq 2$. But this has been exactly the subject of our paper.

## 6 Moment Calculation and Conditional Independence

In this section we establish a link between our moment calculation and the theory of graphical models (Edwards, 1995; Lauritzen, 1996; Whittaker, 1990). We present some properties of the inverse covariance matrix and show how the dependence structure of variables is affected after selection.

Graphical modeling uses graphical representations of variables as nodes in a graph and dependencies among them as edges. A key concept in graphical modeling is the conditional independence property. Two variables $x$ and $y$ are conditionally independent given a variable or a set of variables $z$ (notation $x \perp\!\!\!\perp y|z$), when $x$ and $y$ are independent after partialling out the effect of $z$. For conditionally independent $x$ and $y$, the edge between them in the graph is omitted and the joint density factorizes as $f(x, y|z) = f(x|z)f(y|z)$.

Conditional independence is equivalent to having zero elements $\boldsymbol{\Omega}_{xy}$ in the inverse covariance matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ as well as having a zero partial covariance/correlation between $x$ and $y$ given the remaining variables:

$$x \perp\!\!\!\perp y|\text{Rest} \iff \boldsymbol{\Omega}_{xy} = 0 \iff \rho_{xy.Rest} = 0.$$

Both marginal independence and conditional independence between variables simplify the computations of the truncated covariance in equation 16. In the presence of conditional independence of $i$ and $j$ given $q$, the terms $\sigma_{ij} - \sigma_{iq}\sigma_{qq}^{-1}\sigma_{qj} = 0$ vanish as they reflect the partial covariance of $i$ and $j$ given $q$.

As has been shown by Marchetti and Stanghellini (2008), the conditional independence property is preserved after selection, i.e. the inverse covariance matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}^*$ before and after truncation share the same zero-elements. We prove that many elements of the precision matrix are invariant to truncation. For the case of $k < d$ truncated variables, we define the set of truncated variables with $T = \{x_1, \ldots, x_k\}$, and the remaining $d-k$ variables as $S = \{x_{k+1}, \ldots, x_d\}$. We can show that the off-diagonal elements $\boldsymbol{\Omega}_{i,j}$ are invariant after truncation for $i \in T \cup S$ and $j \in S$:

**Proposition 1.** *The off-diagonal elements $\boldsymbol{\Omega}_{i,j}$ and the diagonal elements $\boldsymbol{\Omega}_{j,j}$ are invariant after truncation for $i \in T \cup S$ and $j \in S$.*

*Proof.* The proof is a direct application of the Johnson/Kotz formula in equation 26 in the previous section. As a result of the formula for partitioned inverse

matrices (Greene, 2003, section A.5.3) , the corresponding inverse covariance matrix $\mathbf{\Omega}$ of the partitioned covariance matrix $\mathbf{\Sigma}$ is

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{V}_{11}^{-1}(\mathbf{I} + \mathbf{V}_{12}\mathbf{F}_2\mathbf{V}_{21}\mathbf{V}_{11}^{-1}) & -\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{F}_2 \end{pmatrix} \tag{27}$$

with $\mathbf{F}_2 = (\mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12})^{-1}$.

Inverting the truncated covariance matrix $\mathbf{\Sigma}^*$ in equation 26 using the formula for the partitioned inverse leads to the truncated precision matrix

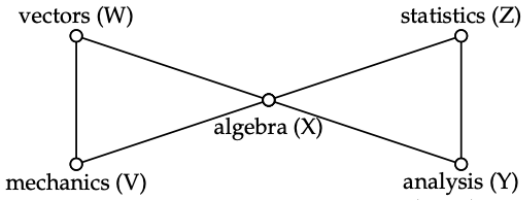
$$\mathbf{\Omega}^* = \begin{pmatrix} \mathbf{U}_{11}^{-1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{F}_2\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & -\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & \mathbf{F}_2 \end{pmatrix} \tag{28}$$

where the $\mathbf{\Omega}_{12}^*$ and $\mathbf{\Omega}_{21}^*$ elements are the same as $\mathbf{\Omega}_{12}$ and $\mathbf{\Omega}_{21}$ respectively. The same is true for the elements in $\mathbf{\Omega}_{22}^*$, especially the diagonal elements in $\mathbf{\Omega}_{22}^*$. □

Here, we prove this invariance property only for a subset of truncated variables. Based on our experiments we conjecture that the same is true also for the case of full truncation (i.e. all off-diagonal elements in $\mathbf{\Omega}_{11}^*$). However, we do not give a rigorous proof here and leave it to future research.

**Example 3**

We illustrate the invariance of the elements of the inverse covariance matrix with the famous mathematics marks example used in Whittaker (1990) and Edwards (1995, p. 49). The independence graph of the five variables $(W, V, X, Y, Z)$ in this example takes the form of a butterfly as shown in below.




Here, we have the conditional independence $(W, V) \perp\!\!\!\perp (Y, Z)|X$. A corresponding precision matrix might look like (sample data; zero-elements marked as "."):

$$\mathbf{\Omega} = \begin{pmatrix} 1 & 0.2 & 0.3 & . & . \\ 0.2 & 1 & -0.1 & . & . \\ 0.3 & -0.1 & 1 & 0.4 & 0.5 \\ . & . & 0.4 & 1 & 0.2 \\ . & . & 0.5 & 0.2 & 1 \end{pmatrix} \tag{29}$$

After truncation in some variables (for example $(W, V, X)$ as $-2 \leq W \leq 1$, $-1 \leq V \leq 1$, $0 \leq X \leq 1$), we apply equation 16 to compute the truncated second moment and the inverse covariance matrix as:

$$\boldsymbol{\Omega}^* = \begin{pmatrix} 1.88 & 0.2 & 0.3 & . & . \\ 0.2 & 3.45 & -0.1 & . & . \\ 0.3 & -0.1 & 12.67 & 0.4 & 0.5 \\ . & . & 0.4 & 1 & 0.2 \\ . & . & 0.5 & 0.2 & 1 \end{pmatrix} \tag{30}$$

The precision matrix $\boldsymbol{\Omega}^*$ after selection differs from $\boldsymbol{\Omega}$ only in the diagonal elements of $(W, V, X)$. From $\boldsymbol{\Omega}^*$, we can read how partial correlations between variables have changed due to the selection process.

Each diagonal element $\boldsymbol{\Omega}_{yy}$ of the precision matrix is the inverse of the partial variance after regressing on all other variables (Whittaker, 1990, p. 143). Since only those diagonal elements in the precision matrix for the $k \leq d$ of the truncated variables will change after selection, this leads to the idea to just compute these $k$ elements after selection rather than the full $k(k+1)/2$ symmetric elements in the truncated covariance matrix and applying the Johnson/Kotz formula for the remaining $d - k$ variables. However, the inverse partial variance of a scalar $y$ given the remaining variables $X = \{x_1, \ldots, x_d\} \setminus y$

$$\boldsymbol{\Omega}^*_{yy} = \left[\Sigma^*_{y.X}\right]^{-1} = \left[\Sigma^*_{yy} - \Sigma^*_{yX}\Sigma^{*-1}_{XX}\Sigma^*_{Xy}\right]^{-1}$$

still requires the truncated covariance results derived in Section 3.

## 7 Summary

In this paper, we derived a formula for the first and second moments of the doubly truncated multivariate normal distribution and for their bivariate marginal density. An implementation for both formulas has been made available in the R statistics software as part of the `tmvtnorm` package. We linked our results to the theory of graphical models and proved an invariance property for elements of the precision matrix. Further research can deal with other types of truncation (e.g. elliptical). Another line of research can look at the moments of the doubly truncated multivariate Student-t distribution, which contains the truncated multivariate normal distribution as a special case.

## References

Amemiya, T. (1974). Multivariate regression and simultaneous equations models when the dependent variables are truncated normal. *Econometrica*, *42*, 999–1012. doi: https://doi.org/10.2307/1914214

Arellano-Valle, R. B., & Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, *96*, 93–116. doi: https://doi.org/10.1016/j.jmva.2004.10.002

Arismendi, J. C. (2013). Multivariate truncated moments. *Journal of Multivariate Analysis*, *117*, 41–75. doi: https://doi.org/10.1016/j.jmva.2013.01.007

Cartinhour, J. (1990). One-dimensional marginal density functions of a truncated multivariate normal density function. *Communications in Statistics - Theory and Methods*, *19*, 197–203. doi: https://doi.org/10.1080/03610929008830197

Edwards, D. (1995). *Introduction to graphical modelling*. Springer.

Galarza C.E., R., Kan, & V.H., L. (2021). *MomTrunc: Moments of folded and doubly truncated multivariate distributions.* Retrieved from `https://cran.r-project.org/web/packages/MomTrunc` (R package version v5.97)

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, *1*, 141–149. doi: https://doi.org/10.2307/1390838

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2012). *mvtnorm: Multivariate normal and t distributions.* Retrieved from `http://CRAN.R-project.org/package=mvtnorm` (R package version 0.9-9992)

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Prentice-Hall.

Gupta, A. K., & Tracy, D. S. (1976). Recurrence relations for the moments of truncated multinormal distribution. *Communications in Statistics - Theory and Methods*, *5*(9), 855–865. doi: https://doi.org/10.1080/03610927608827402

Johnson, N. L., & Kotz, S. (1971). *Distributions in statistics: Continuous multivariate distributions*. John Wiley & Sons.

Kan, R., & Robotti, C. (2017). On moments of folded and truncated multivariate normal distributions. *Journal of Computational and Graphical Statistics*, *26(4)*, 930–934. doi: https://doi.org/10.1080/10618600.2017.1322092

Lauritzen, S. (1996). *Graphical models*. Oxford University Press.

Lee, L.-F. (1979). On the first and second moments of the truncated multinormal distribution and a simple estimator. *Economics Letters*, *3*, 165–169. doi: https://doi.org/10.1016/0165-1765(79)90111-3

Lee, L.-F. (1983). The determination of moments of the doubly truncated multivariate tobit model. *Economics Letters*, *11*, 245–250.

Leppard, P., & Tallis, G. M. (1989). Algorithm AS 249: Evaluation of the mean and covariance of the truncated multinormal distribution. *Applied Statistics*, *38*, 543–553. doi: https://doi.org/10.2307/2347752

Marchetti, G. M., & Stanghellini, E. (2008). A note on distortions induced by truncation with applications to linear regression systems. *Statistics & Probability Letters*, *78*, 824–829. doi: https://doi.org/10.1016/j.spl.2007.09.050

Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology*,

*43*, 131–143. doi: https://doi.org/10.1111/j.2044-8317.1990.tb00930.x

Regier, M. H., & Hamdan, M. A. (1971). Correlation in a bivariate normal distribution with truncation in both variables. *Australian Journal of Statistics*, *13*, 77–82. doi: https://doi.org/10.1111/j.1467-842x.1971.tb01245.x

Rosenbaum, S. (1961). Moments of a truncated bivariate normal distribution. *Journal of the Royal Statistical Society. Series B (Methodological)*, *23*, 405-408. doi: https://doi.org/10.1111/j.2517-6161.1961.tb00422.x

Shah, S. M., & Parikh, N. T. (1964). Moments of single and doubly truncated standard bivariate normal distribution. *Vidya (Gujarat University)*, *7*, 82–91.

Tallis, G. M. (1961). The moment generating function of the truncated multinormal distribution. *Journal of the Royal Statistical Society, Series B (Methodological)*, *23*(1), 223–229. doi: https://doi.org/10.1111/j.2517-6161.1961.tb00408.x

Tallis, G. M. (1963). Elliptical and radial truncation in normal populations. *The Annals of Mathematical Statistics*, *34*(3), 940–944. doi: https://doi.org/10.1214/aoms/1177704016

Tallis, G. M. (1965). Plane truncation in normal populations. *Journal of the Royal Statistical Society, Series B (Methodological)*, *27*(2), 301–307. doi: https://doi.org/10.1111/j.2517-6161.1965.tb01497.x

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. John Wiley & Sons.

Wilhelm, S., & Manjunath, B. G. (2010, June). tmvtnorm: A Package for the Truncated Multivariate Normal Distribution. *The R Journal*, *2*(1), 25–29. doi: https://doi.org/10.32614/rj-2010-005

Wilhelm, S., & Manjunath, B. G. (2012). *tmvtnorm: Truncated multivariate normal distribution and Student t distribution*. Retrieved from `http://CRAN.R-project.org/package=tmvtnorm` (R package version 1.4-5)