Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2008/11

Marginal distributions of maximum-likelihood estimator when one or two components of the true parameter are on the boundary of the parameter space

Marco Barnabani

Università degli Studi di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Marginal Distributions of Maximum-likelihood estimator when one or two components of the true parameter are on the boundary of the parameter space

Marco Barnabani[1]
Dipartimento di Statistica “G. Parenti”
Università di Firenze
barnaban@ds.unifi.it

Abstract: When the true parameter lies on the boundary of the parameter space it is difficult to investigate the asymptotic distribution of maximum likelihood estimator. In some relatively simple cases it is a mixture of truncated normal distributions. In this paper we shall be concerned with the the marginal distributions of maximum likelihood estimator when one or two components of the true parameter are zero and can be on the boundary of the parameter space. We found that these distributions are (mixtures of) normal or truncated normal multiplied by ”skew functions” which distort the symmetry of the normality. Some of these are skew-normal

Keywords: non-regular problem; marginal density function; truncated multivariate normal; skew function; skew-normal

1 Introduction

To obtain the asymptotic distribution of maximum likelihood estimator, a standard assumption is that the true parameter is in the interior of the parameter space. This assumption allows one to make use of the fact that the first order conditions hold, at least asymptotically. When the true parameter lies on the boundary of the parameter space the asymptotic properties of maximum likelihood estimator are no more valid. In some relatively simple cases the asymptotic distribution is not normal but mixtures of truncated normal distributions while in more complicate cases it is much more difficult to investigate. This type of ”non-regularity” has been considered by several authors, Chernoff (1954), Moran (1971), Chant (1974), Shapiro (1985), Self and Liang (1987) whose paper reviewed all the earlier contributions and provided a uniform framework for the large sample distribution of maximum likelihood estimator. Following Self and Liang’s approach, recently Andrews (1999) established the asymptotic distribution of extremum estimators when the true parameter may be on the boundary providing general high level assumptions under which the results hold. In this paper we follow Self and Liang (1987) and we shall be concerned with the situation when one or two components of the true parameter

[1]Viale Morgagni, 59 - 50134 Firenze.

are zero and can be on the boundary of the parameter space. The asymptotic distribution of maximum likelihood estimator in this two cases is given in Moran (1971) and Chant (1974). In this paper we investigate the marginal distributions of the estimator. We found that these distributions are (mixtures of) normal or truncated normal multiplied by "skew functions" which distort the symmetry of the normality. Some of these are skew-normal as given by Azzalini and Dalla Valle (1996). We don't investigate the weights of the mixtures referring for this argument to the book of Sen and Silvapulle (2005).

2 Preliminaries

Let $X_1, \cdots X_n$ be *iid* observations from a population with density $f(x;\theta)$. Let $l_n(\theta)$ denote the log-likelihood with $\theta \in \Theta \subset \mathbb{R}^k$ where Θ is not necessarily open and B the Fisher information matrix in an observation. The true parameter θ_0 will be assumed to be a boundary point. Self and Liang (1987) assumed the classical Cramér conditions on the family of distributions - distinct values of θ corresponding to distinct probability distributions, existence and positive definiteness of B, existence of the first three derivatives of $l_n(\theta)$ with respect to θ, uniform boundedness of the third-order derivatives of the loglikelihood by a function of finite expectation. Moreover they assumed the convexity of the parameter space in a neighbourhood of θ_0.

Under the above conditions they showed

1. As sample size $n \to \infty$ there exists a sequence of points, $\widehat{\theta}_n \in \Theta$, which locally maximize $l_n(\theta)$ and that converges to θ_0 in probability.
2. $n^{1/2}(\widehat{\theta}_n - \theta_0) = O_p(1)$.
3. The loglikelihood function $l_n(\theta)$ can be approximated by

$$l_n(\theta) = l_n(\theta_0) + (1/2)Z_n' H_n(\theta_0) Z_n - (1/2) q_n(n^{1/2}(\theta - \theta_0)) + R_n(\theta)$$

where

$$\begin{aligned} &H_n(\theta_0) := -n^{-1} D^2 l_n(\theta_0) \qquad Z_n := H_n^{-1}(\theta_0) n^{-1/2} D l_n(\theta_0) \\ &q_n(\lambda) := (\lambda - Z_n)' H_n(\theta_0)(\lambda - Z_n) \qquad \lambda \in \mathbb{R}^k \\ &R_n(\theta) = n O_p(1) \|\theta - \theta_0\|^3 \end{aligned}$$

$D = [\partial / \partial\, \theta_i]\; i = 1, ..., k$ is the column vector of a differential operator; $D^2 = [\partial^2 / \partial\, \theta_i \partial\, \theta_j]$ $i, j = 1, ..., k$ is the matrix of second derivatives.

4. $n^{1/2}(\widehat{\theta}_n - \widetilde{\theta}_n) = o_p(1)$ where $\widetilde{\theta}_n = \textit{argmin}_{\theta \in \Theta}\, q_n(n^{1/2}(\theta - \theta_0))$.

Therefore, the asymptotic distribution of $\widehat{\theta}_n$ can be derived from that of $\widetilde{\theta}_n$. With respect to this, note that

$$\min_{\theta \in \Theta} q_n(n^{1/2}(\theta - \theta_0)) = \min_{T \in n^{1/2}(\Theta - \theta_0)} q_n(T)$$

where

$$n^{1/2}(\Theta - \theta_0) := \{T \in \mathbb{R}^k; T = n^{1/2}(\theta - \theta_0),\ for\ some\ \theta \in \Theta\}$$

and if the shifted and rescaled parameter space, $n^{1/2}(\Theta - \theta_0)$, can be approximated by a convex cone, Λ, it can be shown that

$$\min_{T \in n^{1/2}(\Theta - \theta_0)} q_n(T) = \min_{T \in \Lambda} q_n(T) + o_p(1) \quad and \quad n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \widehat{T}$$

with $\widehat{T}$ such that

$$q(\widehat{T}) = \inf_{T \in \Lambda} q(T) \quad where \quad q(T) := (T - Z)' B (T - Z) \quad and \quad Z \sim N_k(0, B^{-1})$$

In sum the asymptotic distribution of $\widehat{\theta}_n$ is given by the distribution of a random vector $\widehat{T}$ that minimizes a stochastic quadratic function over a convex cone Λ where the coefficients of the quadratic function have a multivariate normal distribution.

The vector $\widehat{T}$ is the projection of Z onto the convex cone Λ with respect to the metric B, and is denoted by $\Pi(Z, \Lambda)$; thus

$$\widehat{T} := \Pi(Z, \Lambda) = \arg\min_{T \in \Lambda} (T - Z)' B (T - Z)$$

therefore, the above results state that $n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \Pi(Z, \Lambda)$ which is a (non linear) function of a multivariate normal distribution defined on Λ.

Often in statistical applications we are interested on the asymptotic distribution of a subvector of θ that lies in a cone. With regard to this, partition θ, T and Z as follows, $\theta = [\alpha' \ \ \beta']'$, $T = [T'_\alpha \ \ T'_\beta]'$ and $Z = [Z'_\alpha \ \ Z'_\beta]'$ where $\alpha \in \mathbb{R}^p$, $\beta \in \mathbb{R}^q$, $p + q = k$ and assume Λ is given by a product set $\Lambda_\alpha \times \mathbb{R}^q$ where $\Lambda_\alpha \subset \mathbb{R}^p$ is a cone. This assumption on Λ requires that the true parameter β_0 is not on a boundary. Then it has been shown (Andrews, 1999) that

$$n^{1/2}(\widehat{\alpha}_n - \alpha_0) \xrightarrow{d} \widehat{T}_\alpha, \quad n^{1/2}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} \widehat{T}_\beta = B_{22}^{-1} G_\beta - B_{22}^{-1} B_{21} \widehat{T}_\alpha, \quad \widehat{T} = \begin{bmatrix} \widehat{T}_\alpha \\ \widehat{T}_\beta \end{bmatrix} \tag{1}$$

where

$$\begin{bmatrix} \underset{(p\times 1)}{G_\alpha} \\ \underset{(q\times 1)}{G_\beta} \end{bmatrix} = \begin{bmatrix} \underset{(p\times p)}{B_{11}} & \underset{(p\times q)}{B_{12}} \\ \underset{(q\times p)}{B_{21}} & \underset{(q\times q)}{B_{22}} \end{bmatrix} \begin{bmatrix} \underset{(p\times 1)}{Z_\alpha} \\ \underset{(q\times 1)}{Z_\beta} \end{bmatrix}, \qquad B^{-1} = \begin{bmatrix} \underset{(p\times p)}{B^{11}} & \underset{(p\times q)}{B^{12}} \\ \underset{(q\times p)}{B^{21}} & \underset{(q\times q)}{B^{22}} \end{bmatrix},$$

$$q_\alpha(\widehat{T}_\alpha) = \inf_{T_\alpha \in \Lambda_\alpha} q_\alpha(T_\alpha) \quad with \quad q_\alpha(T_\alpha) := (T_\alpha - Z_\alpha)'(B^{11})^{-1}(T_\alpha - Z_\alpha)$$

and

$$Z_\alpha \sim N_p(0, B^{11}) \quad with \quad B^{11} = (B_{11} - B_{12} B_{22}^{-1} B_{21})^{-1} \tag{2}$$

From (1) it emerges that the asymptotic distribution of $\widehat{\beta}_n$ depends on whether α_0 is on a boundary if and only if $B_{21} \neq 0$.

If $\Lambda_\alpha = \mathbb{R}^p$ which holds if α_0 is not on a boundary, then $\inf_{T_\alpha \in \Lambda_\alpha} q_\alpha(T_\alpha) = 0$, $\widehat{T}_\alpha = Z_\alpha$, $\widehat{T}_\beta = Z_\beta$ and $\widehat{T} = Z$ that corresponds to the standard case.

If Λ_α is a linear subspace of $\mathbb{R}^p$, which holds in the case of linear or nonlinear equality constraints then $\widehat{T}_\alpha = P_\alpha Z_\alpha$ where P_α is a B^{11}-orthogonal projector on Λ_α, a matrix that does not depend on Z. For example, if $\Lambda_\alpha = \{T_\alpha \in \mathbb{R}^p;\;\; QT_\alpha = 0\}$ where Q is a matrix of full row rank less or equal to p, then $P_\alpha = Id - B^{11}Q'(QB^{11}Q')^{-1}Q$ where Id is the identity matrix of appropriate order. The distribution of $\widehat{T}_\alpha$ is given by a linear transformation of a multivariate normal distribution, $\widehat{T}_\alpha \sim N_p(0, P_\alpha B^{11})$. By (1), $\widehat{T}_\beta = Z_\beta + HZ_\alpha$ with $H = -B^{21}Q'\left(QB^{11}\, Q'\right)^{-1} Q$ and $\widehat{T}_\beta \sim N_q(0, B^{22} + HB^{12})$. Moreover, $Cov\left(\widehat{T}_\alpha, \widehat{T}_\beta\right) = B^{12} + B^{11}H'$ and $Cov\left(\widehat{T}_\beta, \widehat{T}_\alpha\right) = B^{21} + HB^{11}$. This is the result given by Aitchison and Silvey (1958).

For other definitions of Λ_α the solution might be rather arduous. Later on we shall confine to the (polyhedral) cone given by equality/inequality,

$$\Lambda_\alpha = \{T_\alpha \in \mathbb{R}^p;\;\; QT_\alpha = 0,\;\; RT_\alpha \leq 0\} \tag{3}$$

with the matrix $[Q' \quad R']'$ of full row rank less or equal to p. This cone holds in many practical situations.

When the cone is given by (3), the distribution of the projection $\Pi(Z_\alpha, \Lambda_\alpha)$ could be investigated by simulating Z_α and computing $\widehat{T}_\alpha$ with a quadratic programming algorithm. This approach can be relatively simple but can not be of great help to know the analytic distribution of $\widehat{T}_\alpha$.

Alternatively we could proceed by describing the cone, to compute the projection of Z_α onto the appropriate edge and to investigate the distribution of the projection.

3 An analytic form of the projection

Let the constraint matrix R of the cone (3) be of dimension $r \times p$. Let $J = \{1, \cdots, m\}$ be a subset of $\{1, \cdots, r\}$; J may be empty. Let $I = \{1, \cdots, r\} \backslash J$. Let R_J and R_I denote the matrices with their rows given by the rows of R indexed by $j \in J$ and $i \in I$ respectively and denote with $F_J = \{T_\alpha \in \mathbb{R}^p;\;\; V_J T_\alpha = 0,\;\; R_I T_\alpha \leq 0\}$, $V_J = [Q'\; R_J']'$ a face of Λ_α. When $J = \{\emptyset\}$, $F_J = \Lambda_\alpha$, when $J = \{1, \cdots, r\}$, F_J is the vertex of the cone. Let $\mathbb{J}$ be the set of all subsets J which gives rise to a face. $\mathbb{J}$ has at most 2^r elements. We assume that F_J has no redundant columns. Let $ri(F_J) = \{T_\alpha \in \mathbb{R}^p;\;\; V_J T_\alpha = 0,\;\; R_I T_\alpha < 0\}$ be the relative interior of F_J. Then, there exists a collection of faces of Λ_α, say $\{F_J, J \in \mathbb{J}\}$ such that the collection of their relative interiors, $\{ri(F_J), J \in \mathbb{J}\}$, forms a partition of Λ_α (see Lemma 3.13.5 of Sen and Silvapulle (2005), p.128). Further,

$$\widehat{T}_\alpha = \sum_{J\in\mathbb{J}} (P_J Z_\alpha)\, I_{E_J}(Z_\alpha) := \sum_{J\in\mathbb{J}} Z_\alpha^{(J)} I_{E_J}(Z_\alpha), \quad I_{E_J}(Z_\alpha) = \begin{cases} 1 & \text{if } Z_\alpha \in E_J, \\ 0 & \text{if } Z_\alpha \notin E_J. \end{cases} \tag{4}$$

where $E_J = \{Z_\alpha \in \mathbb{R}^p; P_J Z_\alpha \in ri(F_J) \bigcap (Id - P_J) Z_\alpha \in F_J^\perp \cap \Lambda_\alpha^0\}$, Λ_α^0 is the polar cone, $P_J = Id - B^{11}V_J'(V_J B^{11} V_J')^{-1} V_J$ is the projection matrix onto the linear space spanned by F_J, Id is the identity matrix of appropriate order .

By (1), $\widehat{T}_\beta = Z_\beta - B_{22}^{-1} B_{21} \left(\widehat{T}_\alpha - Z_\alpha\right)$. Because $-B_{22}^{-1}B_{21} = B^{21}\left(B^{11}\right)^{-1}$, $Z_\alpha = \sum_{J\in\mathbb{J}} Z_\alpha I_{E_J}$ and $Z_\beta = \sum_{J\in\mathbb{J}} Z_\beta I_{E_J}$, by substitution, $\widehat{T}_\beta$ can be written as

$$\widehat{T}_\beta = \sum_{J\in\mathbb{J}} \left[Z_\beta - B^{21}\left(B^{11}\right)^{-1}(Z_\alpha - P_j Z_\alpha)\right] I_{E_J}(Z_\alpha) := \sum_{J\in\mathbb{J}} Z_\beta^{(J)} I_{E_J}(Z_\alpha) \tag{5}$$

Of course (4) and (5) are a possible representation of the estimator. Andrews (1999) proposed a similar formula for $\widehat{T}_\alpha$ defining a different indicator function but we found some problems with his results in some specific cases.

Stacking in $\widehat{T}$ the above two components, $\widehat{T}_\alpha$ and $\widehat{T}_\beta$, we get

$$\widehat{T} = \begin{bmatrix} \widehat{T}_\alpha \\ \widehat{T}_\beta \end{bmatrix} = \sum_{J\in\mathbb{J}} Z^{(J)} I_{E_J}(Z_\alpha) \quad with \quad Z^{(J)} = \begin{bmatrix} Z_\alpha^{(J)} \\ Z_\beta^{(J)} \end{bmatrix} \tag{6}$$

which is the form of the estimator we refer to.

As to the probability distributions of the events $\widehat{T} \le t$, $\widehat{T}_\alpha \le t_\alpha$ and $\widehat{T}_\beta \le t_\beta$ we observe that (Self and Liang, 1987)

$$\widehat{T} \le t = \bigcup_{J\in\mathbb{J}} \left(Z^{(J)} \le t \cap Z_\alpha \in E_J\right) \tag{7}$$

$$\widehat{T}_\alpha \le t_\alpha = \bigcup_{J\in\mathbb{J}} \left(Z_\alpha^{(J)} \le t_\alpha \cap Z_\alpha \in E_J\right) \tag{8}$$

$$\widehat{T}_\beta \le t_\beta = \bigcup_{J\in\mathbb{J}} \left(Z_\beta^{(J)} \le t_\beta \cap Z_\alpha \in E_J\right) \tag{9}$$

therefore,

$$Pr\left(\widehat{T} \le t\right) = \sum_{J\in\mathbb{J}} Pr\left(Z^{(J)} \le t / Z_\alpha \in E_J\right) w_J \tag{10}$$

$$Pr\left(\widehat{T}_\alpha \le t_\alpha\right) = \sum_{J\in\mathbb{J}} Pr\left(Z_\alpha^{(J)} \le t_\alpha / Z_\alpha \in E_J\right) w_J \tag{11}$$

$$Pr\left(\widehat{T}_\beta \le t_\beta\right) = \sum_{J\in\mathbb{J}} Pr\left(Z_\beta^{(J)} \le t_\alpha / Z_\alpha \in E_J\right) w_J \tag{12}$$

where

$$w_J = Pr\left(Z_\alpha \in E_J\right) = Pr\left(P_J Z_\alpha \in ri(F_J) \bigcap (Id - P_J) Z_\alpha \in F_J^\perp \cap \Lambda_\alpha^0\right)$$

Formulas (4)-(6) and (10)-(12) allow to investigate (at least in simple case) the probability distributions and the marginal distributions of the projector when the cone is given by (3).

4 Application I: $\Lambda = \Lambda_\alpha \times \mathbb{R}^q$ with $\Lambda_\alpha = \mathbb{R}^+ \times \mathbb{R}^{p-1}$

4.1 Analytic form of the estimator

Because Λ_α involves only an inequality constraint on the first component of the vector α, we can assume $\Lambda_\alpha = \mathbb{R}^+$ and $\Lambda = \mathbb{R}^+ \times \mathbb{R}^{k-1}$ considering α as a scalar and lumping in with β the other components of α. Then, the cone can be written as $\Lambda_\alpha = \{T_\alpha \in \mathbb{R}; -T_\alpha \leq 0\}$ and the polar cone as $\Lambda_\alpha^0 = \{y \in \mathbb{R}; y \leq 0\}$. There are two faces indexed by $J = \{\emptyset\}$, $F_{\{\emptyset\}} = \Lambda_\alpha$ and $J = \{1\}$ where $F_{\{1\}}$ is the vertex. Moreover, $F_{\{\emptyset\}}^\perp \cap \Lambda_\alpha^0 = \{y \in \mathbb{R}; y = 0\}$ and $F_{\{1\}}^\perp \cap \Lambda_\alpha^0 = \{y \in \mathbb{R}; y \leq 0\}$. The projectors are $P_{\{\emptyset\}} = 1$ and $P_{\{1\}} = 0$ with $E_{\{\emptyset\}} = \{Z_\alpha; Z_\alpha > 0\}$ and $E_{\{1\}} = \{Z_\alpha; Z_\alpha \leq 0\}$. Therefore, by (4) we get

$$\widehat{T}_\alpha = Z_\alpha^{(\{\emptyset\})} I_{E_{\{\emptyset\}}}(Z_\alpha) + Z_\alpha^{(\{1\})} I_{E_{\{1\}}}(Z_\alpha) \tag{13}$$

where $Z_\alpha^{(\{\emptyset\})} = Z_\alpha \sim N\left(0, b^{11}\right)$, $Z_\alpha^{(\{1\})}$ is a degenerate random variable with unit mass distribution at zero and $B^{-1} := \begin{bmatrix} b^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix}$

The component $\widehat{T}_\beta$. By (5) we have

$$\widehat{T}_\beta = Z_\beta^{(\{\emptyset\})} I_{E_{\{\emptyset\}}}(Z_\alpha) + Z_\beta^{(\{1\})} I_{E_{\{1\}}}(Z_\alpha) \tag{14}$$

with $Z_\beta^{(\{\emptyset\})} = Z_\beta$, $Z_\beta^{(\{1\})} = \left(Z_\beta - \frac{B^{21}}{b^{11}} Z_\alpha\right)$.

Stacking the above two components (formula (6)) we get

$$\widehat{T} = Z \; I_{E_{\{\emptyset\}}}(Z_\alpha) + Z^{(\{1\})} I_{E_{\{1\}}}(Z_\alpha) \tag{15}$$

which is the result of Andrews (1999), Self and Liang (1987).

4.2 Distributions

4.2.1 The distribution of $\widehat{T}$

By (10)

$$Pr\left(\widehat{T} \leq t\right) = Pr\left(Z \leq t / Z_\alpha > 0\right) Pr\left(Z_\alpha > 0\right) + Pr\left(Z^{(\{1\})} \leq t / Z_\alpha \leq 0\right) Pr\left(Z_\alpha \leq 0\right)$$

with $Pr\left(Z_\alpha > 0\right) = Pr\left(Z_\alpha \leq 0\right) = 1/2$.

The event $Z \leq t/Z_\alpha > 0$ has a k-variate truncated normal probability density function, denoted as $TN_k\left(0, B^{-1}, z_\alpha > 0\right)$. In the mathematical appendix we show that the denominator of the truncated density, $D = \int_0^{+\infty}\int_{I_{k-1}} \exp\left(-\frac{1}{2}\, z' B z\right) d\, z$ where $\int_{I_{k-1}}$ is a $(k-1)$-dimensional Riemann integral on $I_{k-1} = \{z_\beta; -\infty < z_\beta[i] < +\infty; i = 1, \cdots, k-1\}$ can be written as $D = (1/2)\,(2\pi)^{k/2}\,|B|^{-1/2}$. We denote with $z_\beta[i]$ the ith component of the vector z_β.

Therefore, $TN_k\left(0, B^{-1}, z_\alpha > 0\right)$ is $2\, N_k\left(0, B^{-1}\right) I_{E_{\{\emptyset\}}}(Z_\alpha)$ which is the result given by Moran (1971).

Consider the event $Z^{(\{1\})} \leq t/Z_\alpha \leq 0$. Because of the normality of the vector Z and the fact that $Cov\left[\left(Z_\beta - \frac{B^{21}}{b^{11}} Z_\alpha\right), Z_\alpha\right] = 0$ the variables $Z_\beta^{(\{1\})}$ and Z_α are independent. Then, $Pr\left(Z^{(\{1\})} \leq t/Z_\alpha \leq 0\right) = Pr\left(Z^{(\{1\})} \leq t\right)$. After simple algebra we can show that the variance-covariance matrix of the random vector $Z_\beta - \frac{B^{21}}{b^{11}} Z_\alpha$ is equal to B_{22}^{-1} then $\left(Z_\beta - \frac{B^{21}}{b^{11}} Z_\alpha\right) \sim N_{k-1}\left(0, B_{22}^{-1}\right)$. Therefore, $Z^{(\{1\})} \sim N_k(0, B^*)$ with $B^* = \begin{bmatrix} 0 & 0 \\ 0 & B_{22}^{-1} \end{bmatrix}$ (see Chant (1974), Moran (1971)).

4.2.2 The distribution of $\widehat{T}_\alpha$

The distribution of $\widehat{T}_\alpha$ has a continuous part and a discrete part. By (13) or by applying directly (11) we have

$$Pr\left(\widehat{T}_\alpha \leq t_\alpha\right) = Pr\left(Z_\alpha \leq t_\alpha / Z_\alpha > 0\right) Pr\left(Z_\alpha > 0\right) + \Phi(0) Pr\left(Z_\alpha \leq 0\right)$$

where the event $Z_\alpha \leq t_\alpha/Z_\alpha > 0$ has an half-normal probability density function, $TN\left(0, b^{11}, z_\alpha > 0\right)$ and $\Phi(0)$ is a degenerate distribution at 0 (Gourieroux and Monfort, 1989). For notational convenience we define $N(0,0)$ the normal density with mean and variance equal to zero to be the density that takes the value zero with probability one.

4.2.3 The distribution of $\widehat{T}_\beta$

We have,

$$\begin{aligned} Pr\left(\widehat{T}_\beta \leq t_\beta\right) &= Pr\left(Z_\beta \leq t_\beta / Z_\alpha > 0\right) Pr\left(Z_\alpha > 0\right) \\ &+ Pr\left[\left(Z_\beta - \frac{B^{21}}{b^{11}} Z_\alpha\right) \leq t_\beta / Z_\alpha \leq 0\right] Pr\left(Z_\alpha \leq 0\right) \end{aligned}$$

The variable $Z_\beta \leq t_\beta/Z_\alpha > 0$ has a $(k-1)$-dimensional skew-normal density function as given by Azzalini (1985), with location parameter zero, scale matrix B^{22} and shape parameter $\delta = \left(B^{22}\right)^{-1} B^{21} / \sqrt(b^{11} - B^{12}\left(B^{22}\right)^{-1} B^{21})$. Following the notation used by

Azzalini (1985), $Z_\beta/Z_\alpha > 0 \sim SN_{k-1}(0, B^{22}, \delta)$ and the marginal densities are skew-normal as well because of proposition 2 of Azzalini and Capitanio (1999).

As to the second component of $Pr\left(\widehat{T}_\beta \leq t_\beta\right)$, in section 4.2.1 we showed that $\left(Z_\beta - \frac{B^{21}}{b^{11}} Z_\alpha\right) \sim N_{k-1}\left(0, B_{22}^{-1}\right)$.

Summarizing the results of this section we can say that the probability distributions of the projector $\widehat{T}$ and of its components $\widehat{T}_\alpha$ and $\widehat{T}_\beta$, are mixtures of two distributions with weights $1/2$ and densities given in the following table.

Estimators	$Pr(Z_\alpha > 0) = 1/2$ Conditioning Variable $Z_\alpha > 0$	$Pr(Z_\alpha \leq 0) = 1/2$ Conditioning Variable $Z_\alpha \leq 0$
$\widehat{T}$	$TN_k\left(0, B^{-1}, z_\alpha > 0\right)$	$N_k\left(0, B^*\right)$
$\widehat{T}_\alpha$	$TN\left(0, b^{11}, z_\alpha > 0\right)$	$N\left(0, 0\right)$
$\widehat{T}_\beta$	$SN_{k-1}(0, B^{22}, \delta)$	$N_{k-1}\left(0, B_{22}^{-1}\right)$

Table 1: Table of probability densities

5 Application II: $\Lambda = \Lambda_\alpha \times \mathbb{R}^q$ with $\Lambda_\alpha = \left(\mathbb{R}^+\right)^2 \times \mathbb{R}^{p-2}$

5.1 Analytic form of the estimator

As in the previous case, we set $\Lambda_\alpha = \left(\mathbb{R}^+\right)^2$ lumping in with β the other components of α. Then, $\Lambda_\alpha = \{T_\alpha \in \mathbb{R}^2; -T_\alpha \leq 0\}$ and the polar cone is given by $\Lambda_\alpha^0 = \{y \in \mathbb{R}^2; By \leq 0\}$. There are four faces, $\mathbb{J} = \{\{\emptyset\}, \{1\}, \{2\}, \{1,2\}\}$. We introduce the notation $Z_\alpha[\,i\,]$, $i = 1, 2$ to denote the ith element of the vector Z_α. In this Section we refer to the matrix B^{-1} partitioned as follows,

$$B^{-1} = \left[\begin{array}{c|c} \underset{2\times 2}{B^{11}} & \underset{2\times(k-2)}{B^{12}} \\ \hline \underset{(k-2)\times 2}{B^{21}} & \underset{(k-2)\times(k-2)}{B^{22}} \end{array}\right] = \left[\begin{array}{cc|c} b^{11} & b^{12} & B^{13} \\ b^{21} & b^{22} & B^{23} \\ \hline B^{31} & B^{32} & B^{33} \end{array}\right]$$

We have the following regions

$$E_{\{\emptyset\}} = \left\{ Z_\alpha \in \mathbb{R}^2;\; Z_\alpha[\,1\,] > 0\,,\, Z_\alpha[\,2\,] > 0 \right\}$$
$$E_{\{1\}} = \left\{ Z_\alpha \in \mathbb{R}^2;\; Z_\alpha[\,1\,] \leq 0\,,\, Z_\alpha[\,2\,] - \frac{b^{21}}{b^{11}} Z_\alpha[\,1\,] > 0 \right\}$$
$$E_{\{2\}} = \left\{ Z_\alpha \in \mathbb{R}^2;\; Z_\alpha[\,1\,] - \frac{b^{12}}{b^{22}} Z_\alpha[\,2\,] > 0\,,\, Z_\alpha[\,2\,] \leq 0 \right\}$$
$$E_{\{1,2\}} = \left\{ Z_\alpha \in \mathbb{R}^2;\; C Z_\alpha \leq 0 \right\} \quad with \quad C = \begin{bmatrix} 1 & -\frac{b^{12}}{b^{22}} \\ -\frac{b^{21}}{b^{11}} & 1 \end{bmatrix}$$

The projectors are

$$P_{\{\emptyset\}} = Id, \quad P_{\{1\}} = \begin{bmatrix} 0 & 0 \\ -\frac{b^{21}}{b^{11}} & 1 \end{bmatrix}, \quad P_{\{2\}} = \begin{bmatrix} 1 & -\frac{b^{12}}{b^{22}} \\ 0 & 0 \end{bmatrix}, \quad P_{\{1,2\}} = 0$$

Then, by (4), $\widehat{T}_\alpha$ may be written as

$$\widehat{T}_\alpha = Z_\alpha\, I_{E_{\{\emptyset\}}}(Z_\alpha) + Z_\alpha^{(\{1\})} I_{E_{\{1\}}}(Z_\alpha) + Z_\alpha^{(\{2\})} I_{E_{\{2\}}}(Z_\alpha) + Z_\alpha^{(\{1,2\})}\, I_{E_{\{1,2\}}}(Z_\alpha) \qquad (16)$$

where $Z_\alpha^{(\{1,2\})}$ is a degenerate random vector at zero,

$$Z_\alpha^{(\{1\})} = \begin{bmatrix} 0 \\ Z_\alpha[\,2\,] - \frac{b^{21}}{b^{11}} Z_\alpha[\,1\,] \end{bmatrix} \quad and \quad Z_\alpha^{(\{2\})} = \begin{bmatrix} Z_\alpha[\,1\,] - \frac{b^{12}}{b^{22}} Z_\alpha[\,2\,] \\ 0 \end{bmatrix}$$

By (5),

$$\widehat{T}_\beta = Z_\beta\, I_{E_{\{\emptyset\}}}(Z_\alpha) + Z_\beta^{(\{1\})} I_{E_{\{1\}}}(Z_\alpha) + Z_\beta^{(\{2\})} I_{E_{\{2\}}}(Z_\alpha) + Z_\beta^{(\{1,2\})}\, I_{E_{\{1,2\}}}(Z_\alpha) \qquad (17)$$

where

$$Z_\beta^{(\{1\})} = \left(Z_\beta - \frac{B^{31}}{b^{11}} Z_\alpha[1] \right), \quad Z_\beta^{(\{2\})} = \left(Z_\beta - \frac{B^{32}}{b^{22}} Z_\alpha[2] \right),$$
$$Z_\beta^{(\{1,2\})} = \left[Z_\beta - B^{21} \left(B^{11} \right)^{-1} Z_\alpha \right]$$

Then, stacking the above two components we get $\widehat{T}$.

5.2 Distributions

5.2.1 The distribution of $\widehat{T}$

The probability distribution of the event $\widehat{T} \leq t$ is given by (10) with $w_J = Pr\left(Z_\alpha \in E_J\right)$. We analyze the four components of the estimator.

When $J = \{\emptyset\}$ the event to be analyzed is $Z \leq t / Z_1 > 0 \cap Z_2 > 0$. It has a k-variate truncated normal probability density function, $TN_k\left(0, B^{-1}, z_\alpha > 0\right)$ whose denominator is $D = \int_{0_2} \int_{I_q} \exp\left(-\frac{1}{2} z' B z\right) d\, z$ where $0_2 = \{z_\alpha; 0 < z_\alpha[i] < +\infty; i = 1, 2\}$, $I_q = \{z_\beta; -\infty < z_\beta[i] < +\infty; i = 1, \cdots, q\}$, $q + 2 = k$. In the mathematical appendix we show that $D = (2\pi)^{(k/2)} |B|^{-1/2} \frac{1}{2}\left(1 - \frac{\arccos r_{12}}{\pi}\right)$ where r_{12} is the correlation between $Z_\alpha[1]$ and $Z_\alpha[2]$. Therefore,

$$TN_k\left(0, B^{-1}, z_\alpha > 0\right) = \frac{2 N_k\left(0, B^{-1}\right)}{1 - \frac{\arccos r_{12}}{\pi}}\; I_{E_{\{\emptyset\}}}(Z_\alpha)$$

The density of $Z^{(\{1\})} / Z_\alpha[\,1\,] \leq 0 \cap Z_\alpha^{(\{1\})}[\,2\,] > 0$. We first observe that

$$\left[\begin{array}{c} Z_\alpha[1] \\ Z_\alpha^{(\{1\})}[2] \\ Z_\beta^{(\{1\})} \end{array}\right] = \left[\begin{array}{ccc} 1 & 0 & 0 \\ -b^{21}/b^{11} & 1 & 0 \\ -B^{31}/b^{11} & 0 & 1 \end{array}\right] \left[\begin{array}{c} Z_\alpha[1] \\ Z_\alpha[2] \\ Z_\beta \end{array}\right] \; with \; \left[\begin{array}{c} Z_\alpha[1] \\ Z_\alpha[2] \\ Z_\beta \end{array}\right] := Z \sim N_k\left(\,0\,,\; B^{-1}\,\right) \tag{18}$$

By a theorem on linear transformations of multivariate normal distributions, we have

$$\left[\begin{array}{c} Z_\alpha[1] \\ Z_\alpha^{(\{1\})}[2] \\ Z_\beta^{(\{1\})} \end{array}\right] \sim N_k\left(0\;,\; \left[\begin{array}{ccc} b^{11} & 0 & 0 \\ 0 & \sigma^2 & \gamma' \\ 0 & \gamma & \Omega \end{array}\right]\right) \tag{19}$$

where $\sigma^2 = b^{22} - b^{21}(b^{11})^{-1} b^{12}$, $\gamma = B^{32} - B^{31}(b^{11})^{-1} b^{12}$ and $\Omega = B^{33} - B^{31}(b^{11})^{-1} B^{13}$.

Given above results, it is immediate to observe that

$$Pr\left(Z^{(\{1\})} \leq t \;/\; Z_\alpha[1] \;\leq 0 \;\cap\; Z_\alpha^{(\{1\})}[2] > 0\right) = Pr\left(Z^{(\{1\})} \leq t \;/\; Z_\alpha^{(\{1\})}[2] > 0\right)$$

therefore, the density of $Z^{(\{1\})} /\, Z_\alpha^{(\{1\})}[2] > 0$ is a $k-1$ truncated normal density,

$$TN_{k-1}\left(0, \Sigma, z_\alpha^{(\{1\})}[2] > 0\right) = \xi\; \exp\left(-\frac{1}{2}\, z^{(\{1\})'} \Sigma^{-1} z^{(\{1\})}\right),\; Z_\alpha^{(\{1\})}[2] > 0$$

with

$$\Sigma = \left[\begin{array}{cc} \sigma^2 & \gamma' \\ \gamma & \Omega \end{array}\right] \quad and \quad \xi^{-1} = \frac{1}{2}\,(2\pi)^{(k-1)/2}\,(\det \Sigma)^{1/2}$$

Putting the results together, the density of $Z^{(\{1\})} /\, Z_\alpha^{(\{1\})}[2] > 0$ is $2N_{k-1}\left(0, \Sigma\right);\; Z_\alpha^{(\{1\})}[2] > 0$.

The density of the event $Z^{(\{2\})} \leq t / Z_\alpha[\,2\,] \leq 0 \cap Z_\alpha^{(\{2\})}[\,1\,] > 0$. As in the previous case we first observe that

$$\left[\begin{array}{c} Z_\alpha^{(\{2\})}[1] \\ Z_\alpha[2] \\ Z_\beta^{(\{2\})} \end{array}\right] = \left[\begin{array}{ccc} 1 & -b^{12}/b^{22} & 0 \\ 0 & 1 & 0 \\ 0 & -B^{32}/b^{22} & 1 \end{array}\right] \left[\begin{array}{c} Z_\alpha[1] \\ Z_\alpha[2] \\ Z_\beta \end{array}\right]$$

therefore

$$\left[\begin{array}{c} Z_{\alpha}^{(\{2\})}[1] \\ Z_{\alpha}[2] \\ Z_{\beta}^{(\{2\})} \end{array}\right] \sim N_k\left(0\;,\;\left[\begin{array}{ccc} \psi_{11} & 0 & \psi_{13} \\ 0 & b^{22} & 0 \\ \psi_{31} & 0 & \psi_{33} \end{array}\right]\right) \quad and \tag{20}$$

$$\left[\begin{array}{c} Z_{\alpha}^{(\{2\})}[1] \\ Z_{\beta}^{(\{2\})} \end{array}\right] \sim N_{k-1}\left(0\;,\;\Psi := \left[\begin{array}{cc} \psi_{11} & \psi_{13} \\ \psi_{31} & \psi_{33} \end{array}\right]\right) \tag{21}$$

with $\psi_{11} = b^{11} - b^{12}\left(b^{22}\right)^{-1} b^{21}$, $\psi_{13} = B^{13} - B^{23}\left(b^{22}\right)^{-1} b^{12}$, $\psi_{31} = B^{31} - B^{32}\left(b^{22}\right)^{-1} b^{21}$ and $\psi_{33} = B^{33} - B^{32}\left(b^{22}\right)^{-1} B^{23}$. Above result implies that

$$Pr\left(Z^{(\{2\})} \leq t / Z_{\alpha}[\,2\,] \leq 0 \cap Z_{\alpha}^{(\{2\})}[\,1\,] > 0\right) = Pr\left(Z^{(\{2\})} \leq t \,/\, Z_{\alpha}^{(\{2\})}[1] > 0\right)$$

therefore, the density of $Z^{(\{2\})} / \, Z_{\alpha}^{(\{2\})}[1] > 0$ is a $k-1$ truncated normal density with variance-covariance matrix equal to Ψ. As in the previous case, the denominator is equal to $\frac{1}{2}\left(2\pi\right)^{(k-1)/2}\left(\det\Psi\right)^{1/2}$ and

$$TN_{k-1}\left(0, \Psi, z_{\alpha}^{(\{2\})}[1] > 0\right) = 2N_{k-1}\left(0, \Psi\right); \;\; Z_{\alpha}^{(\{2\})}[1] > 0$$

Finally, we analyze the density of the event $Z^{(\{1,2\})} \leq t / CZ_{\alpha} \leq 0$ which occurs when the region E_J is indexed by $J = \{1, 2\}$. We recall that

$$Z^{(\{1,2\})} = \left[\begin{array}{c} Z_{\alpha}^{(\{1,2\})} \\ Z_{\beta}^{(\{1,2\})} \end{array}\right] \quad and \quad CZ_{\alpha} = \left[\begin{array}{c} Z_{\alpha}^{(\{2\})}[1] \\ Z_{\alpha}^{(\{1\})}[2] \end{array}\right]$$

with $Z_{\beta}^{(\{1,2\})} \sim N_{k-2}\left(0, B_{22}^{-1}\right)$, $B_{22}^{-1} = B^{22} - B^{21}\left(B^{11}\right)^{-1} B^{12}$. Simple algebra allows one to show that

$$\left[\begin{array}{cc} C & 0 \\ 0 & I \end{array}\right]\left[\begin{array}{c} Z_{\alpha} \\ Z_{\beta}^{(\{1,2\})} \end{array}\right] \sim N_k\left(0\;,\;\left[\begin{array}{cc} CB^{11}C' & 0 \\ 0 & B_{22}^{-1} \end{array}\right]\right)$$

Then, $Pr\left(Z^{(\{1,2\})} \leq t / CZ_{\alpha} \leq 0\right) = Pr\left(Z^{(\{1,2\})} \leq t\right)$ and $Z^{(\{1,2\})} \sim N_k\left(0, B^*\right)$ with $B^* = \begin{bmatrix} 0 & 0 \\ 0 & B_{22}^{-1} \end{bmatrix}$.

5.2.2 The distribution of $\widehat{T}_\alpha$

From (11) we have

$$\begin{aligned} Pr\left(\widehat{T}_\alpha \le t_\alpha\right) =& Pr\left(Z_\alpha \le t_\alpha / Z_\alpha \in E_{\{\emptyset\}}\right) Pr\left(Z_\alpha \in E_{\{\emptyset\}}\right) + \\ & Pr\left(Z_\alpha^{(\{1\})} \le t_\alpha / Z_\alpha \in E_{\{1\}}\right) Pr\left(Z_\alpha \in E_{\{1\}}\right) + \\ & Pr\left(Z_\alpha^{(\{2\})} \le t_\alpha / Z_\alpha \in E_{\{2\}}\right) Pr\left(Z_\alpha \in E_{\{2\}}\right) + \\ & Pr\left(Z_\alpha^{(\{1,2\})} \le t_\alpha / Z_\alpha \in E_{\{1,2\}}\right) Pr\left(Z_\alpha \in E_{\{1,2\}}\right) \end{aligned}$$

Then, the bivariate density of $Z_\alpha \le t_\alpha / Z_\alpha \in E_{\{\emptyset\}}$ is $TN_2\left(0, B^{11}, z_\alpha > 0\right)$. The marginal densities of the truncated normal, Z_α, are not truncated normal. In the mathematical appendix we show that

$$f_{Z_\alpha[1]} = \frac{2N\left(0, u^{-1}\right)\left(1 - F(a)\right)}{1 - \frac{\arccos r_{12}}{\pi}} \; = \; N\left(0, u^{-1}\right) \; g_1\left(z_\alpha[1]\right); \;\; z_\alpha[1] > 0$$

and

$$f_{Z_\alpha[2]} = \frac{2N\left(0, v^{-1}\right)\left(1 - F(b)\right)}{1 - \frac{\arccos r_{12}}{\pi}} \; = \; N\left(0, v^{-1}\right) \; g_2\left(z_\alpha[2]\right); \;\; z_\alpha[2] > 0$$

where $u = a_{11} - a_{12}(a_{22})^{-1}a_{21}$, $v = a_{22} - a_{21}(a_{11})^{-1}a_{12}$, $a = (a_{22})^{-1/2}a_{21}z_\alpha[1]$, $b = (a_{11})^{-1/2}a_{12}z_\alpha[2]$, r_{12} is the correlation between $Z_\alpha[1]$ and $Z_\alpha[2]$, $F(.)$ is the distribution function of a $N(0,1)$ and $(B^{11})^{-1} := A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$.

The functions $g_1\left(z_\alpha[1]\right)$ and $g_2\left(z_\alpha[2]\right)$ can be thought of as ”skew functions”. They distort the symmetry of the truncated normal density functions.

The event $Z_\alpha^{(\{1\})} \le t_\alpha / Z_\alpha \in E_{\{1\}}$. We first observe that

$$\begin{bmatrix} Z_\alpha[1] \\ Z_\alpha^{(\{1\})}[2] \end{bmatrix} \sim N_2 \begin{pmatrix} b^{11} & 0 \\ 0 & b_{22}^{-1} \end{pmatrix}$$

where $b_{22}^{-1} = b^{22} - \left(b^{21}\right)^2 / b^{11}$ and $Z_\alpha^{(\{1\})}[2] = Z_\alpha[\,2\,] - \frac{b^{21}}{b^{11}} Z_\alpha[\,1\,]$. Therefore,

$$Pr\left(Z_\alpha^{(\{1\})}[2] \le t_\alpha \;/\; Z_\alpha[1] \;\le 0 \;\cap\; Z_\alpha^{(\{1\})}[2] > 0\right) = \frac{Pr\left(0 < Z_\alpha^{(\{1\})}[2] \le t_\alpha\right)}{Pr\left(Z_\alpha^{(\{1\})}[2] > 0\right)}$$

and the density of $Z_\alpha[\,2\,] - \frac{b^{21}}{b^{11}} Z_\alpha[\,1\,]$ is $TN\left(0, b_{22}^{-1}, z_\alpha^{(\{1\})}[2] > 0\right)$.

We can apply the same line of reasoning to the event $Z_\alpha^{(\{2\})} \le t_\alpha / Z_\alpha \in E_{\{2\}}$ finding that $Z_\alpha[\,2\,]$ and $Z_\alpha^{(\{2\})}[1] = Z_\alpha[\,1\,] - \frac{b^{12}}{b^{22}} Z_\alpha[\,2\,]$ are independent and the density of $Z_\alpha^{(\{2\})}[1]$ is $TN\left(0, b_{11}^{-1}, z_\alpha^{(\{2\})}[1] > 0\right)$.

5.2.3 The distribution of $\widehat{T}_\beta$

By (12) the probability of the event $\widehat{T}_\beta \le t_\beta$ is given by,

$$
\begin{aligned}
Pr\left(\widehat{T}_\beta \le t_\beta\right) =& Pr\left(Z_\beta \le t_\beta / Z_\alpha \in E_{\{\emptyset\}}\right) Pr\left(Z_\alpha \in E_{\{\emptyset\}}\right) + \\
& Pr\left(Z_\beta^{(\{1\})} \le t_\beta / Z_\alpha \in E_{\{1\}}\right) Pr\left(Z_\alpha \in E_{\{1\}}\right) + \\
& Pr\left(Z_\beta^{(\{2\})} \le t_\beta / Z_\alpha \in E_{\{2\}}\right) Pr\left(Z_\alpha \in E_{\{2\}}\right) + \\
& Pr\left(Z_\beta^{(\{1,2\})} \le t_\beta / Z_\alpha \in E_{\{1,2\}}\right) Pr\left(Z_\alpha \in E_{\{1,2\}}\right)
\end{aligned}
$$

Then, the density of $Z_\beta \le t_\beta / Z_\alpha \in E_{\{\emptyset\}}$ is given by

$$
f_{Z_\beta} = \frac{\int_{0_2} \exp\left(-\frac{1}{2}\, z' B z\right)\, dz_\alpha}{D}; \quad z_\beta \in \mathbb{R}^q \tag{22}
$$

in the mathematical appendix we show that it can be written as

$$
f_{Z_\beta} = \frac{2\, N_q(0, W^{-1})\, F_2(c)}{1 - \frac{\arccos r_{12}}{\pi}} = N_q(0, W^{-1})\, h_2(z_\beta); \quad z_\beta \in \mathbb{R}^q \tag{23}
$$

where $c = -B_{11}^{-1} B_{12} z_\beta = [\, c_1 \ c_2\,]'$, $W = B_{22} - B_{21} B_{11}^{-1} B_{12}$ and $F_2(c) = \int_{-c_1}^{+\infty}\int_{-c_2}^{+\infty} N_2\left(y, 0, B_{11}^{-1}\right) d\,y$. Again, $h_2(z_\beta)$ can be thought of as a "skew function" that serves to distort the symmetry of the normal density.

The marginal density of a component of the vector $Z_\beta \le t_\beta / Z_\alpha \in E_{\{\emptyset\}}$. Without loss of generality let us derive the marginal density of the last component of Z_β, denoted Z_k, subject to the condition $Z_\alpha \in E_{\{\emptyset\}}$. Assume the following partitions of Z and B.

$$
Z = \begin{bmatrix} Z_\alpha \\ Z_\beta \end{bmatrix} := \begin{bmatrix} Z_\alpha \\ Z_\beta^* \\ Z_k \end{bmatrix} := \begin{bmatrix} \underset{(k-1)\times 1}{Z_1} \\ \underset{1\times 1}{Z_k} \end{bmatrix}
$$

lumping in with Z_1 any component different from Z_k, and

$$
B = \begin{bmatrix} \underset{(k-1)\times(k-1)}{B_{11}} & \underset{(k-1)\times 1}{b_{12}} \\ \underset{1\times(k-1)}{b_{21}} & \underset{1\times 1}{b_{22}} \end{bmatrix} \quad with \quad B_{11} = \begin{bmatrix} \underset{2\times 2}{C_{11}} & \underset{2\times(k-3)}{C_{12}} \\ \underset{(k-3)\times 2}{C_{21}} & \underset{(k-3)\times(k-3)}{C_{22}} \end{bmatrix}
$$

To derive the marginal density of Z_k, f_{Z_k}, we must integrate out the remaining variables of the numerator of f_{Z_β}, that is

$$
f_{Z_k} = \frac{\int_{0_2} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\, z' B z\right)\, dz_1}{D}; \quad z_k \in \mathbb{R} \tag{24}
$$

In mathematical appendix we show that f_{Z_k} is the same as (23) with $q=1$. That is,

$$f_{Z_k} = \frac{2\,N(0,W^{-1})\,F_2(c)}{1-\frac{\arccos r_{12}}{\pi}} := N(0,W^{-1})\,h_2(z_k); \quad z_k \in \mathbb{R} \tag{25}$$

where $c = -B_{11}^{-1} b_{12} z_k = [\, c_1 \ c_2 \,]'$, $W = b_{22} - b_{21} B_{11}^{-1} b_{12}$ and $F_2\,(c) = \int_{-c_1}^{+\infty}\int_{-c_2}^{+\infty} N_2\,(y, 0, V^{-1})\,d\,y$, $V = C_{11} - C_{12} C_{22}^{-1} C_{21}$.

The marginal density of the ith component of $Z_\beta \leq t_\beta / Z_\alpha \in E_{\{\emptyset\}}$ is again given by (25) once the matrix B has been modified changing the ith row with the kth row and the ith column with the kth column.

The density of the event $Z_\beta^{(\{1\})} \leq t_\beta / Z_\alpha \in E_{\{1\}}$. By (18) and (19) it is immediate to observe that

$$Pr\left(Z_\beta^{(\{1\})} \leq t_\beta \ / \ Z_\alpha[1] \ \leq 0 \ \cap \ Z_\alpha^{(\{1\})}[2] > 0 \right) = Pr\left(Z_\beta^{(\{1\})} \leq t_\beta \ / \ Z_\alpha^{(\{1\})}[2] > 0 \right)$$

therefore, the density of $Z_\beta^{(\{1\})}$ is skew-normal (Azzalini and Dalla Valle, 1996),

$$Z_\beta^{(\{1\})} \,/\, Z_\alpha^{(\{1\})}[2] > 0 \sim SN_{k-2}\,(0, \Omega, \alpha)$$

with $\alpha = \Omega^{-1}\gamma\,(\sigma^2 - \gamma'\Omega^{-1}\gamma)^{-1/2}$.

The marginal densities of $Z_\beta^{(\{1\})} \leq t_\beta / Z_\alpha \in E_{\{1\}}$ are skew-normal too.

The density of the event $Z_\beta^{(\{2\})} \leq t_\beta / Z_\alpha \in E_{\{2\}}$. As in the previous case, by (20) we observe that

$$Pr\left(Z_\beta^{(\{2\})} \leq t_\beta \ / \ Z_\alpha[2] \ \leq 0 \ \cap \ Z_\alpha^{(\{2\})}[1] > 0 \right) = Pr\left(Z_\beta^{(\{2\})} \leq t_\beta \ / \ Z_\alpha^{(\{2\})}[1] > 0 \right)$$

and the density of $Z_\beta^{(\{2\})}$ is skew-normal (Azzalini and Dalla Valle, 1996),

$$Z_\beta^{(\{2\})} \,/\, Z_\alpha^{(\{2\})}[1] > 0 \sim SN_{k-2}\,(0, \psi_{33}, \alpha)$$

with $\alpha = \psi_{33}^{-1}\psi_{13}\left(\psi_{11} - \psi_{13}\psi_{33}^{-1}\psi_{31}\right)^{-1/2}$.

As in the previous case, the marginal densities of $Z_\beta^{(\{2\})}$ are skew-normal.

Finally we investigate the density of the event $Z_\beta^{(\{1,2\})} \leq t_\beta / Z_\alpha \in E_{\{1,2\}}$. By the results of section 5.2.1 it is immediate to observe that $Z_\beta^{(\{1,2\})} \sim N_{k-2}\left(0, B_{22}^{-1}\right)$ with marginal normal densities.

Conclusions

In this paper we investigated the asymptotic distributions and marginal distributions of maximum likelihood estimator when the parameter space is $\Theta = \mathbb{R}^+ \times \mathbb{R}^{k-1}$ or $\Theta = (\mathbb{R}^+)^2 \times \mathbb{R}^{k-2}$ and the true parameter may be on the boundary. We found that these distributions are (mixtures of) normal or truncated normal multiplied by ”skew functions” which distort the symmetry of the normality. Some of these distributions are skew-normal.

References

Aitchison, J. and Silvey, S. D. (1958) Maximum-Likelihood Estimation of Parameters Subject to Restraints. *The Annals of Mathematical Statistics*, **29**, 813–828.

Andrews, D. W. K. (1999) Estimation when a Parameter is on a Boundary. *Econometrica*, **67**, 1341–1383.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171–178.

Azzalini, A. and Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution. *J. Royal Statist. Soc., series B*, **61**.

Azzalini, A. and Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83**, 715–726.

Chant, D. (1974) On Asymptotic tests of complete hypotheses in nonstandard conditions. *Biometrika*, **61**, 291–298.

Chernoff, H. (1954) On the distribution of the likelihood ratio. *Annals of Mathematical Statistics*, **25**, 573–578.

Gourieroux, C. and Monfort, A. (1989) *Statistics and Econometric Models. Volumes 1 and 2.* Cambridge: Cambridge University Press.

Graybill, F. A. (1983) *Matrices with Applications in Statistics*. Second Edition, Belmont: Wadsworth.

Moran, P. A. P. (1971) Maximum Likelihood Estimators in Nonstandard Conditions. In *Proceedings of The Cambridge Philosophical Society*, vol. 70, 441–450. Cambridge Philosophical Society: Cambridge UK.

Self, S. G. and Liang, K. Y. (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood ratio Tests under Nonstandard Conditions. *Journal of the American Statistical Association*, **82**, 605–610.

Sen, K. B. and Silvapulle, M. J. (2005) *Constrained Statistical Inference*. New York: Wiley.

Shapiro, A. (1985) Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, **72**, 133–144.

Mathematical Appendix

To compute the marginal densities we use repeatedly the Sheppard's result

$$P\left(X_1 > 0, X_2 > 0\right) = \frac{1}{2}\left(1 - \frac{\arccos r_{12}}{\pi}\right)$$

where r_{12} is the correlation between X_1 and X_2, and the solution of the multiple integral

$$\int_{I_k} \exp\left[-\left(x'Bx + x'b + b_0\right)\right] d\,x_1 \cdots d\,x_k$$

where B is a positive definite matrix, b an $n \times 1$ vector of constant, b_0 a scalar constant and $I_k = \{x; -\infty < x_i < +\infty; i = 1, \cdots, k\}$. The solution of the above multiple integral can be found in Graybill (1983) and is given by (Theorem 10.5.1, p. 342)

$$\exp\left(\frac{1}{4}\, b' B^{-1} b - b_0\right) \int_{I_k} \exp\left[-\frac{1}{2}(x-c)' R (x-c)\right] d\,x_1 \cdots d\,x_k$$

where $R = 2B$, $c = -(1/2)B^{-1}b$.

- Consider the denominator of the truncated normal distribution. D can be written as,
$$\int_{0_2} \left(\int_{I_q} \exp\left[-\left(z'_\beta \frac{B_{22}}{2} z_\beta + z'_\beta\, b + b_0 \right) \right] d\,z_\beta \right)\; d\,z_\alpha$$
where $b = B_{21} z_\alpha$, $b_0 = \frac{1}{2} z'_\alpha B_{11} z_\alpha$, $0_2 = \{z_\alpha; 0 < z_\alpha[i] < +\infty; i = 1, 2\}$ and $I_q = \{z_\beta; -\infty < z_\beta[i] < +\infty; i = 1, \cdots, q\}$.
We first apply Graybill's theorem to the multiple integral in round parentheses. We have the following result,
$$(2\pi)^{(k-2)/2}\, |B_{22}|^{-1/2}\, \exp\left(-\frac{1}{2} z'_\alpha U\, z_\alpha\right)\; d\,z_\alpha$$
with $U = B_{11} - B_{12}\,(B_{22})^{-1}\, B_{21}$. Then, we apply Sheppard's result getting the following expression for the denominator,
$$D = (2\pi)^{k/2}\, |B|^{-1/2}\, \frac{1}{2} \left(1 - \frac{\arccos\, r_{12}}{\pi}\right)$$
If the double integral from 0 to $+\infty$ were a simple integral then $D = (1/2)\,(2\pi)^{k/2}\,|B|^{-1/2}$.
- The marginal density $f_{Z_\alpha[2]}$. Because the density of $Z_\alpha \leq t_\alpha / Z_\alpha \in E_{\{\emptyset\}}$ is truncated normal, the marginal of $Z_\alpha[2]$ is given by,
$$f_{Z_\alpha[2]} = \frac{\int_0^{+\infty} \exp\left(-\frac{1}{2}\, z'_\alpha A z_\alpha\right) dz_\alpha[1]}{\int_{0_2} \exp\left(-\frac{1}{2}\, z'_\alpha A z_\alpha\right) dz_\alpha};\;\; z_\alpha[2] > 0$$
where $(B^{11})^{-1} := A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. By Sheppard's result we get the denominator,
$$\int_{0_2} \exp\left(-\frac{1}{2}\, z'_\alpha A z_\alpha\right) dz_\alpha = 2\pi\, |A|^{-1/2}\, \frac{1}{2}\, \left(1 - \frac{\arccos\, r_{12}}{\pi}\right)$$

The numerator is an application of Graybill's theorem,

$$
\begin{aligned}
& \int_0^{+\infty} \exp\,(-\frac{1}{2}\, z_\alpha' A z_\alpha)\, dz_\alpha[1] = \\
= \; & \int_0^{+\infty} \exp\,\left[-\left(\frac{z_\alpha[1]^2 a_{11}}{2} + z_\alpha[1] a_{12} z_\alpha[2] + \frac{z_\alpha[2]^2 a_{22}}{2}\right)\right]\, dz_\alpha[1] = \\
= \; & a_{11}^{-1/2}\,(2\pi)^{1/2}\; [1 - F(b)]\, \exp\left(-\frac{1}{2}\, z_\alpha^2[2]\; v\right)
\end{aligned}
$$

where $v = a_{22} - a_{21}(a_{11})^{-1} a_{12}$, $b = (a_{11})^{-1/2} a_{12} z_\alpha[2]$, r_{12} is the correlation between $Z_\alpha[1]$ and $Z_\alpha[2]$ and $F(.)$ is the distribution function of a $N(0, 1)$.
The ratio between the numerator and the denominator gives the marginal density. The same approach is used to obtain the marginal density $f_{Z_\alpha[1]}$.

- The density f_{Z_β}. Consider, first, the numerator of (22). By Graybill's theorem we have,

$$
\begin{aligned}
& \int_{0_2} \exp\,\left[-\left(z_\alpha' \frac{B_{11}}{2} z_\alpha + z_\alpha'\, b + b_0\right)\right]\, dz_\alpha \\
= \; & \exp\left(\frac{1}{4}\, b' \left(\frac{B_{11}}{2}\right)^{-1} b - b_0\right) \int_{0_2} \exp\,\left[-\frac{1}{2}(\,z_\alpha - c)' B_{11}(z_\alpha - c)\right]\, dz_\alpha
\end{aligned}
$$

with $b = B_{12} z_\beta$, $b_0 = \frac{1}{2} z_\beta' B_{22} z_\beta$, $c = -B_{11}^{-1} B_{12} z_\beta = [\,c_1\;\, c_2\,]'$. Then, the numerator is given by

$$
N = \exp\left(-\frac{1}{2} z_\beta'\, W\, z_\beta\right)\; F_2\left(B_{11}^{-1} B_{12} z_\beta\right)
$$

where $W = B_{22} - B_{21} B_{11}^{-1} B_{12}$ and $F_2\,(c) = \int_{-c_1}^{+\infty}\int_{-c_2}^{+\infty} N_2\left(y, 0, B_{11}^{-1}\right) d\,y$.

- The density f_{Z_k}. The main burden is to compute the numerator of (24). We have

$$
\begin{aligned}
& \int_{0_2}\int_{I_{q-1}} \exp\left(-\frac{1}{2}\, z' B z\right) d\, z_1 \\
= \; & \int_{0_2}\int_{I_{q-1}} \exp\left[-\left(z_1' \frac{B_{11}}{2} z_1 + z_1' b_{12} z_k + b_{22} z_k^2\right)\right] dz_1 \\
= \; & \exp\left[-\frac{1}{2}\left(b_{22} - b_{21} B_{11}^{-1} b_{12}\right) z_k^2\right] \int_{I_c}\int_{I_{q-1}} \exp\left(-\frac{1}{2}\, y' B_{11} y\right) dy
\end{aligned}
$$

with $c = -B_{11}^{-1} b_{12} z_k = \begin{bmatrix} c_\alpha \\ c_\beta^* \end{bmatrix}$, $y = z_1 - c = \begin{bmatrix} y_\alpha \\ y_\beta^* \end{bmatrix}$ according to the partition of Z_1 and $I_c = \{y_\alpha; -c_\alpha[i] < y_\alpha[i] < +\infty; i = 1, 2\}$, $c_\alpha[i]$ and $y_\alpha[i]$ are the ith components of the vectors c_α and y_α respectively. Moreover,

$$\begin{aligned}
&\int_{I_c}\int_{I_{q-1}} \exp\left(-\frac{1}{2}\,y' B_{11} y\right) d\,y \\
= &\int_{I_c}\int_{I_{q-1}} \exp\left[-\left(y_\beta^{*\prime}\,\frac{C_{22}}{2}\,y_\beta + y_\beta^{*\prime} C_{21} y_\alpha + y_\alpha' \frac{C_{11}}{2}\,y_\alpha\right)\right] d\,y_\beta^*\; d\,y_\alpha
\end{aligned}$$

Graybill's theorem applied to the integral with respect to y_β^* produces the following result

$$\exp\left(-\frac{1}{2}\,y_\alpha' V y_\alpha\right)(2\pi)^{\frac{k-3}{2}}\,|C_{22}|^{-\frac{1}{2}};\quad V = C_{11} - C_{12}C_{22}^{-1}C_{21}$$

that must be integrated in I_c. Therefore, the numerator of f_{Z_k} is given by

$$\exp\left[-\frac{1}{2}\left(b_{22} - b_{21}B_{11}^{-1}b_{12}\right) z_k^2\right](2\pi)^{\frac{k-3}{2}}\,|C_{22}|^{-\frac{1}{2}}\int_{I_c}\exp\left(-\frac{1}{2}\,y_\alpha' V y_\alpha\right) d\,y_\alpha$$

Some algebra applied to the ratio between this result and D gives the marginal density f_{Z_k}.