isid/ms/2011/03
July 1, 2011
http://www.isid.ac.in/~statmath/eprints

Some notes on extremal discriminant analysis

Manjunath B.G.
Melanie Frick
Rolf-Dieter Reiss

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Some notes on extremal discriminant analysis

Manjunath B.G., Melanie Frick[1]
and Rolf-Dieter Reiss[2]
All at Department of Mathematics, University Siegen, Siegen, Germany

Abstract

Classical discriminant analysis focusses on Gaussian and nonparametric models where in the second case the unknown densities are replaced by kernel densities based on the training sample. In the present article we assume that it suffices to base the classification on exceedances above higher thresholds, which can be interpreted as observations in a conditional framework. Therefore, the statistical modeling of truncated distributions is merely required. In this context, a nonparametric modeling is not adequate because the kernel method is inaccurate in the upper tail region. Yet one may deal with truncated parametric distributions like the Gaussian ones. Our primary aim is to replace truncated Gaussian distributions by appropriate generalized Pareto distributions and to explore properties and the relationship of discriminant functions in both models.

AMS 1991 subject classification: primary 62H30, 60G70, 62G32; secondary 62E20, 60G15

Key words: Discriminant analysis, Gaussian model, Truncation, Generalized Pareto distributions, Linear discriminant functions, Hüsler-Reiss triangular arrays.

0 Introduction

The basic idea of discriminant analysis is to classify an object of unknown origin to one of several given classes based on the measurement vector (also called discriminator) within a d-dimensional space. The available data sets to do this are samples of objects of which both their class memberships and their measurements are known. In the present article we confine ourselves to the case of two classes; the modifications required for dealing with more than two classes are straightforward. Consider a d-dimensional discriminator $\boldsymbol{x}$ from one of the two classes which are described by the densities $w(\boldsymbol{x}|1)$ and $w(\boldsymbol{x}|2)$. Let p_1 , p_2 and c_1 , c_2 be the corresponding prior probabilities and costs of misclassification to the first and the second

[1]Research enabled through a grant from the Deutsche Forschungsgemeinschaft

[2]Corresponding author. Department of Mathematics, University Siegen, Siegen, Germany

E–mail addresses: bgmanjunath@gmail.com (Manjunath B.G.), frick.melanie@googlemail.com (M. Frick), reiss@stat.math.uni-siegen.de (R.-D. Reiss)

population, respectively. The optimal discriminant decision is determined by the following rule: an observation vector $\boldsymbol{x}$ is classified to class 1 if the inequality

$$\frac{w(\boldsymbol{x}|1)}{w(\boldsymbol{x}|2)} \geq \frac{c_2 p_2}{c_1 p_1} \tag{1}$$

is fulfilled. The optimal common border or discriminant function is obtained by formulating (1) as an equation and solving it as a function in the discriminator $\boldsymbol{x}$.

Classical discriminant analysis focusses on the Gaussian model. In that case one gets an explicit representation of the discriminant function. Denote by

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

the d-dimensional Gaussian density with location parameter vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and non-singular covariance matrix Σ. The pertaining distribution function is denoted by $\Phi_{\boldsymbol{\mu},\Sigma}(\boldsymbol{x})$.

The corresponding discriminant function for classifying an observation $\boldsymbol{x}$ between $\varphi_{\boldsymbol{\mu}^{(1)},\Sigma^{(1)}}(\boldsymbol{x}|1)$ and $\varphi_{\boldsymbol{\mu}^{(2)},\Sigma^{(2)}}(\boldsymbol{x}|2)$ is

$$\begin{aligned} D_Q(\boldsymbol{x}) \quad = \quad & -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})^T{\Sigma^{(1)}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})^T \\ & {\Sigma^{(2)}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)}) - \log\frac{c_2 p_2 \left|\Sigma^{(1)}\right|^{1/2}}{c_1 p_1 \left|\Sigma^{(2)}\right|^{1/2}}. \end{aligned} \tag{3}$$

Therefore, the decision rule entails that an observation vector $\boldsymbol{x}$ is classified to $\varphi_{\boldsymbol{\mu}^{(1)},\Sigma^{(1)}}(\boldsymbol{x}|1)$ if $D_Q(\boldsymbol{x}) \geq 0$, cf. Lachenbruch [17], page 11, or Falk et al. [6], page 231. This function is quadratic in $\boldsymbol{x}$. In addition, in the case of identical covariance matrices $\Sigma^{(1)} = \Sigma^{(2)} = \Sigma$ the discriminant function

$$D_L(\boldsymbol{x}) \quad = \quad \left[\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})\right]^T \Sigma^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \log\frac{c_2 p_2}{c_1 p_1}. \tag{4}$$

is linear, and the common border constitutes a hyperplane. This result can be regarded as a benchmark in discriminant analysis.

We refer to Kocherlakota et al. [15] for discriminant analysis concerning truncated univariate Gaussian distributions. Discriminant analysis within univariate extreme value models and an application to life span classification was investigated by Abdalla [1]. Another important reference is Avery [2] dealing with discriminant analysis in the case of multivariate Gaussian distributions with linear truncation applied to credit scoring data. The present article may be regarded as a first systematic investigation with respect to the multivariate case; one can hope that it stimulates further theoretical research work and encourages practitioners to use such models.

The specification of multivariate extreme value models and exploring their properties is an active research area. The main goal of the present article is to introduce an extreme value model to discriminant analysis. We are primarily interested in modeling the upper tail of distributions, which can be done by using appropriate exceedance (truncated) distributions, and, later, explore properties and the relationship of discriminant functions in different models. We primarily confine ourselves to the classification among two classes.

It is well known that the asymptotic distribution of exceedances over high thresholds is that of a generalized Pareto (GP) random vector if, and only if, the corresponding maxima are asymptotically distributed according to an extreme value distribution (EVD). We mainly deal with the classical Gaussian model. According to Theorem 1 of Hüsler and Reiss [9] the asymptotic distribution of the maxima of a triangular scheme of Gaussian random vectors is the Hüsler–Reiss EVD. With the help of the characterization theorem we will deduce the GPD pertaining to the Hüsler–Reiss distribution. There are quite a few approaches how to construct GPDs from extreme value distributions and all these are closely related to each other. In the present article we confine to the procedure given by Tajvidi [24] and Rootzén and Tajvidi [22]. We finish by investigating properties and relationships between the classical Gaussian and the GPD discriminant functions.

1 Discriminant functions for the truncated Gaussian model

In this section we discuss the truncated Gaussian model, cf. Horrace [12], under rectangular truncations, and present the pertaining discriminant function. Truncation of distributions outside of the upper tail region is a crucial idea in extreme value theory.

Let $\boldsymbol{X} = (X_1, ..., X_d)^T$ be a d-dimensional Gaussian vector with non-singular covariance matrix Σ and location parameter $\boldsymbol{\mu} \in \mathbb{R}^d$. The rectangularly truncated version of $\boldsymbol{X}$ with truncation vector $\boldsymbol{c} = (c_1, ..., c_d)^T \in \mathbb{R}^d$ has the density

$$f_{RT}(\boldsymbol{x}) = \begin{cases} \frac{\varphi_{\boldsymbol{\mu},\Sigma}(\boldsymbol{x})}{P\{\mathbf{X} > \boldsymbol{c}\}}, & \text{for } \boldsymbol{x} > \boldsymbol{c}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The discriminant function for classifying an observation between two classes with the densities $f_{RT}(\boldsymbol{x}|1)$ and $f_{RT}(\boldsymbol{x}|2)$ which have different location parameters $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ and truncation vectors $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$, respectively, can be determined by using equation (1). We have

$$D_{RT}(\boldsymbol{x}) = \left[\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})\right]^T \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) - \log\frac{c_2 p_2}{c_1 p_1} + T_r(\boldsymbol{c}_1, \boldsymbol{c}_2),$$

where $T_r(\boldsymbol{c}_1, \boldsymbol{c}_2)$ is given by

$$T_r(\boldsymbol{c}_1, \boldsymbol{c}_2) = \log P\{\mathbf{X}_2 > \boldsymbol{c}_2\} - \log P\{\mathbf{X}_1 > \boldsymbol{c}_1\}$$

with $\mathbf{X}_1$ and $\mathbf{X}_2$ denoting the corresponding Gaussian random vectors. Note that $D_{RT}(\boldsymbol{x})$ is linear in $\boldsymbol{x}$. If one chooses different correlation matrices $\Sigma^{(1)}$ and $\Sigma^{(2)}$, one gets a function that is quadratic in $\boldsymbol{x}$ corresponding to D_Q in (3).

We add some remarks about elliptical truncation, cf. Tallis [25]. Let $\boldsymbol{X} = (X_1, ..., X_d)^T$ be again a d-dimensional Gaussian vector with non-singular covariance matrix Σ and location parameter $\boldsymbol{\mu} \in \mathbb{R}^d$. The elliptically truncated version of $\boldsymbol{X}$ has the density

$$f_{ET}(\boldsymbol{x}) = \begin{cases} \frac{\varphi_{\boldsymbol{\mu},\Sigma}(\boldsymbol{x})}{P\{\mathbf{X} \in E\}}, & \text{for } \boldsymbol{x} \in E, \\ 0, & \text{otherwise,} \end{cases}$$

where $E = \left\{\boldsymbol{x} : (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \geq u\right\}$ and u is non-negative real value.

The discriminant function for classifying an observation between two classes with the densities $f_{ET}(\boldsymbol{x}|1)$ and $f_{ET}(\boldsymbol{x}|2)$ which have different location parameters $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ and truncation regions E_1 and E_2, respectively, can again be determined by using equation (1). We have

$$D_{ET}(\boldsymbol{x}) = \left[\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})\right]^T \Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) - \log \frac{c_2 p_2}{c_1 p_1} + T_e(E_1, E_2) \tag{6}$$

where $T_e(E_1, E_2)$ is given by

$$T_e(E_1, E_2) = \log P\{\mathbf{X}_2 \in E_2\} - \log P\{\mathbf{X}_1 \in E_1\}$$

with $\mathbf{X}_1$ and $\mathbf{X}_2$ denoting the corresponding Gaussian random vectors. Obviously, the two truncation borders have the same shape but differ in the shift which depends on the type of truncation.

One could also think of certain convex truncations in this context. With a linear truncation one would not obtain an approximation by means of GP distributions. In the following section we will concentrate on the rectangular truncation. Finally, we also mention that some numerical work concerning the generation of random samples for the multivariate truncated Gaussian distribution using Gibbs sampling was done in Stefan and Manjunath [23].

2 Extreme value and generalized Pareto models

In this section we present the main results of Hüsler and Reiss [9] and the approach of Rootzén and Tajvidi [22] concerning the construction of a GPD and we deduce a simple form of the Hüsler–Reiss GP density.

Let (X_1, X_2) be a bivariate Gaussian vector with associated distribution function $F_{\rho_{1,2}}$, where X_1 and X_2 are standard Gaussian random variables and $\rho_{1,2}$ is the correlation coefficient. Subsequently, we consider n iid copies of (X_1, X_2) with the correlation coefficient depending on the sample size n. Then, according to Theorem 1 by Hüsler and Reiss [9] the following result holds. If

$$(1-\rho_{1,2}(n))\log n \to \lambda_{1,2}^2 \in [0,\infty], \quad n \to \infty, \tag{7}$$

then

$$\lim_{n\to\infty} F^n_{\rho_{1,2}(n)}(b_n + x_1/b_n, b_n + x_2/b_n) = H_{\lambda_{1,2}}(x_1, x_2)$$

for every x_1, $x_2 \in \mathbb{R}$, where $b_n = n\varphi(b_n)$, φ is the standard Gaussian density, and the limiting function is given by

$$\begin{aligned} H_{\lambda_{1,2}}(x_1,x_2) = \exp\Bigg[&-\Phi\left(\lambda_{1,2} + \frac{x_1 - x_2}{2\lambda_{1,2}}\right)e^{-x_2} \\ &-\Phi\left(\lambda_{1,2} + \frac{x_2 - x_1}{2\lambda_{1,2}}\right)e^{-x_1}\Bigg] \end{aligned} \tag{8}$$

with Φ being the standard Gaussian distribution function. For an explicit, approximate solution to the equation $b_n = n\varphi(b_n)$ we refer to Reiss [20], page 161. Moreover, independence and complete dependence are achieved at $\lambda_{1,2} = \infty$ and $\lambda_{1,2} = 0$, respectively, i.e.,

$$\begin{aligned} H_\infty(x_1,x_2) &= \exp(-e^{-x_1})\exp(-e^{-x_2}) \quad \text{and} \\ H_0(x_1,x_2) &= \exp\left(-e^{-\min(x_1,x_2)}\right). \end{aligned}$$

Now, following Section 3 by Hüsler and Reiss [9] and Section 12.1 by Reiss and Thomas [21], page 297, let $\mathbf{X} = (X_1, ..., X_d)^T$ be a d-dimensional standard Gaussian vector with df F_Σ, where $\Sigma = (\rho_{i,j})_{i,j\le d}$ is the correlation matrix. Again, we let the correlations depend on the sample size of n iid copies of $\mathbf{X}$, i.e. we consider a correlation matrix $\Sigma(n)$, $n \in \mathbb{N}$. Apparently, by imposing a certain rate of convergence on $\rho_{i,j}(n)$, i.e., for $1 \le i, j \le d$,

$$(1-\rho_{i,j}(n))\log n \to \lambda_{i,j}^2 \in [0,\infty], \quad n \to \infty,$$

the limit of the standardized Gaussian maxima distribution function $F^n_{\Sigma(n)}$, as $n \to \infty$, is the d-dimensional Hüsler-Reiss extreme value distribution

$$H_\Lambda(\boldsymbol{x}) = \exp\left(-\sum_{k=1}^d \int_{x_k}^\infty \Phi_{\Sigma(k)}\left(\left(\lambda_{i,k} + \frac{x_i - z}{2\lambda_{i,k}}\right)_{i=1}^{k-1}\right)e^{-z}dz\right) \tag{9}$$

(in the represention given by Joe [13]) where Λ is a symmetric $d \times d$-matrix $\Lambda = (\lambda_{i,j})$ with $\lambda_{i,j} > 0$ if $i \ne j$ and $\lambda_{i,i} = 0$, and $\Phi_{\Sigma(k)}$ is a $(k-1)$–variate Gaussian distribution function (with the convention $\Phi_{\Sigma(1)} = 1$). The mean vector of $\Phi_{\Sigma(k)}$ is zero and $\Sigma(k) = (\sigma_{i,j}(k))$ is the correlation matrix given by

$$\sigma_{i,j}(k) = \begin{cases} \frac{1}{2\lambda_{i,k}\lambda_{j,k}}\left(\lambda_{i,k}^2 + \lambda_{j,k}^2 - \lambda_{i,j}^2\right), & 1 \le i < j \le k-1, \\ 1, & i = j. \end{cases} \tag{10}$$

In some recent articles Hashorva [10], [11] shows that the multivariate Hüsler–Reiss distribution is as well the limiting distribution of multivariate maxima of elliptical triangular arrays if the random radius of the elliptical random vectors belongs to the max–domain of attraction of a Gumbel distribution. Concerning a corresponding result in the GPD case we refer to Manjunath [18].

Now we discuss the construction of a GPD belonging to an EVD. The derivation of univariate GPDs, which is presented in Section 1.3 of Reiss et al. [5], page 21, has to be modified in the multivariate case. In the framework of the construction of multivariate GPDs there are different approaches by different authors. One can be found in the dissertation of Tajvidi [24], another one in Kaufmann and Reiss [14] and in Section 5.1 by Reiss et al. [5], and still another one in Section 8.3 by Beirlant et al. [3]. In the present work we use the definition given by Tajvidi [24], which is investigated in detail in Rootzén and Tajvidi [22].

Now, let $H(\boldsymbol{x})$ be a d–variate EVD with $0 < H(\mathbf{0}) < \mathbf{1}$. Then the corresponding GPD has the representation

$$W(\boldsymbol{x}) := \begin{cases} 1 - \frac{\log H(\boldsymbol{x})}{\log H(\mathbf{0})}, & \text{if } \boldsymbol{x} \ge \mathbf{0} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The above definition has independently also been noted in Section 8.3.1 of Beirlant et al. [3], page 278. A similar definition can be found in Lemma 5.1.3 of Reiss et al. [5], where it is given for the entire negative quadrant, particularly for its upper region close to the origin.

Hence, the multivariate Hüsler–Reiss GPD has the form

$$\begin{aligned} W_\Lambda(\boldsymbol{x}) &= 1 - \log H_\Lambda(\boldsymbol{x}) / \log H_\Lambda(\mathbf{0}) \\ &= 1 - \frac{1}{C(\Lambda)}\left(\sum_{k=1}^{d}\int_{x_k}^{\infty} \Phi_{\Sigma(k)}\left(\left(\lambda_{i,k} + \frac{x_i - z}{2\lambda_{i,k}}\right)_{i=1}^{k-1}\right) e^{-z} dz\right) \end{aligned} \tag{12}$$

where

$$C(\Lambda) = \sum_{k=1}^{d}\int_{0}^{\infty} \Phi_{\Sigma(k)}\left(\left(\lambda_{i,k} - \frac{z}{2\lambda_{i,k}}\right)_{i=1}^{k-1}\right) e^{-z} dz.$$

Remark 2.1. *For $d = 2$ the constant $C(\Lambda)$ reduces to*

$$\begin{aligned} C(\lambda_{1,2}) &= \int_0^\infty e^{-z}dz + \int_0^\infty \Phi_{\Sigma(2)}\left(\lambda_{1,2} - \frac{z}{2\lambda_{1,2}}\right) e^{-z}dz \\ &= 2\Phi_{\Sigma(2)}(\lambda_{1,2}). \end{aligned}$$

Multivariate GPDs in the framework of extreme value theory are still under scrutiny. So, due to the limits in defining a multivariate GPD we use the above definition. One shortfall of it, as discussed by Tajvidi [24], is that there is some probability mass on each of the axes. i.e., the threshold line which consists of null sets with respect to the Lebesgue measure has a positive probability. This leads to one d-dimensional measure on $\mathbb{R}_+^d$ and d univariate measures on each axis. This point is also noted in Section 2 by Michel [19].

In the following theorem we present a simple form of the density of the multivariate Hüsler–Reiss GPD.

Theorem 2.2. *Let $W_\Lambda(\boldsymbol{x})$ be the Hüsler–Reiss GPD as defined in equation* (12)*. Then for each $0 < \lambda_{i,j} < \infty$, $i < j \leq d-1$, the multivariate Hüsler-Reiss GP conditional density given $\boldsymbol{x} > \boldsymbol{0}$ is of the form*

$$w_\Lambda(\boldsymbol{x}) = \frac{e^{-x_d}}{2^{d-1}\left(\prod_{i=1}^{d-1}\lambda_{i,d}\right)C^*(\Lambda)}\varphi_{\Sigma(d)}\left(\left(\lambda_{i,d} + \frac{x_i - x_d}{2\lambda_{i,d}}\right)_{i=1}^{d-1}\right), \tag{13}$$

where $C^*(\Lambda)$ is a scaling factor given by

$$C^*(\Lambda) = C(\Lambda)(1 - K(\Lambda)),$$

$K(\Lambda)$ being the mass on the axes, and $\varphi_{\Sigma(d)}$ is the $(d-1)$–variate Gaussian density. The mean vector of $\varphi_{\Sigma(d)}$ is zero and $\Sigma(d) = (\sigma_{i,j}(d))$ is the correlation matrix satisfying (10) *for $k = d$.*

Proof. We first prove the assertion for the bivariate case. Plugging equation (8) into (11) we obtain

$$W_{\lambda_{1,2}}(x_1, x_2) = 1 - \frac{\Phi(\lambda_{1,2} + \frac{x_1 - x_2}{2\lambda_{1,2}})e^{-x_2} + \Phi(\lambda_{1,2} + \frac{x_2 - x_1}{2\lambda_{1,2}})e^{-x_1}}{2\Phi(\lambda_{1,2})}.$$

If the continuous partial derivative of $W_{\lambda_{1,2}}$ exists on the open support, then according to Theorem A.2.2 in Bhattacharya and Rao [4], page 264, the density is given by

$$
\begin{aligned}
w^*_{\lambda_{1,2}}(x_1,x_2) &= \frac{\partial^2 W_{\lambda_{1,2}}(x_1,x_2)}{\partial x_2 \partial x_1} \\
&= \frac{1}{2\Phi(\lambda_{1,2})}\left[\frac{e^{-x_2}}{4\lambda_{1,2}^2}\varphi^{'}\left(\lambda_{1,2}+\frac{x_1-x_2}{2\lambda_{1,2}}\right)\right. \\
&+ \frac{e^{-x_1}}{4\lambda_{1,2}^2}\varphi^{'}\left(\lambda_{1,2}+\frac{x_2-x_1}{2\lambda_{1,2}}\right)+\frac{e^{-x_2}}{2\lambda_{1,2}}\varphi\left(\lambda_{1,2}+\frac{x_1-x_2}{2\lambda_{1,2}}\right) \\
&+ \left.\frac{e^{-x_1}}{2\lambda_{1,2}}\varphi\left(\lambda_{1,2}+\frac{x_2-x_1}{2\lambda_{1,2}}\right)\right]
\end{aligned}
$$

where $\varphi^{'}(a) = (-a)\varphi(a)$ is the derivate of φ . Note that

$$
e^{-x_2}\varphi\left(\lambda_{1,2}+\frac{x_1-x_2}{2\lambda_{1,2}}\right) = e^{-x_1}\varphi\left(\lambda_{1,2}+\frac{x_2-x_1}{2\lambda_{1,2}}\right)
$$

according to Reiss and Thomas [21], page 296. With this identity the function reduces to

$$
w^*_{\lambda_{1,2}}(x_1,x_2) = \frac{e^{-x_2}\varphi(\lambda_{1,2}+\frac{x_1-x_2}{2\lambda_{1,2}})}{4\lambda_{1,2}\Phi(\lambda_{1,2})}, \quad x_1,\, x_2 > 0. \tag{14}
$$

As discussed, if we integrate $w^*_{\lambda_{1,2}}(x_1,x_2)$ on the entire support, the total mass is less than one, namely, $(1-\Phi(\lambda_{1,2}))/\Phi(\lambda_{1,2})$. Of course, the total mass sums up to one if we additionally consider the mass on the axes. Now, the mass on the x_2–axis is equal to $W_{\lambda_{1,2}}(0,\infty) = (2\Phi(\lambda_{1,2})-1)/2\Phi(\lambda_{1,2})$. Since the bivariate Hüsler–Reiss distribution function is symmetric in x_1 and x_2, the same mass is obtained on the x_1–axis. One can easily see that the mass on the axes increases as $\lambda_{1,2}$ increases, i.e., the degree of independence between the two variables increases. In case of independence the entire mass lies on the axes. If $\lambda_{1,2}$ tends to zero, i.e., we move towards complete dependence, the mass on the axes converges to zero. Therefore it has been investigated that the mass on the axes is directly related to the strength of the tail dependence. The conditional bivariate density on $\mathbb{R}^2_+$ is obtained by truncating the mass on each axis. This implies that we are truncating the observations on each axis. Further, it means that we are modeling in an open rectangle of $\mathbb{R}^2_+$. So, by dividing the function in (14) by $(1-\Phi(\lambda_{1,2}))/\Phi(\lambda_{1,2})$ (which is calculated within the truncated model), we obtain the bivariate Hüsler–Reiss GP conditional density

$$
w_{\lambda_{1,2}}(x_1,x_2) = \frac{e^{-x_2}\varphi(\lambda_{1,2}+\frac{x_1-x_2}{2\lambda_{1,2}})}{4\lambda_{1,2}(1-\Phi(\lambda_{1,2}))}, \quad x_1,\, x_2 > 0. \tag{15}
$$

Now we generalize our proof to arbitrary dimensions. We use Theorem A.2.2 in Bhattacharya and Rao [4], page 264, again to deduce the multivariate density. The partial derivate of W_Λ with respect to $\boldsymbol{x}$ is given by

$$w^*_\Lambda(\boldsymbol{x}) = \frac{e^{-x_d}}{2^{d-1}\left(\prod_{i=1}^{d-1}\lambda_{i,d}\right)C(\Lambda)}\varphi_{\Sigma(d)}\left(\left(\lambda_{i,d}+\frac{x_i-x_d}{2\lambda_{i,d}}\right)_{i=1}^{d-1}\right). \tag{16}$$

Similarly as in the bivariate case the above function leads to positive mass on each axis. The mass on the i^{th} axis can easily be determined by calculating $W_\Lambda(0,...,0,\infty,0,...,0)$. The total mass on the d axes is denote by $K(\Lambda)$. We know that the sum of the mass on the axes and the mass on $\mathbb{R}^d_+$ will add up to one. Now, similar to the bivariate case we are interested in the density upon the open rectangle of $\mathbb{R}^d_+$. Therefore, the new scaling factor is given by

$$C^*(\Lambda) = C(\Lambda)(1-K(\Lambda)).$$

Replacing $C(\Lambda)$ by $C^*(\Lambda)$ in (16) completes the proof. □

Remark 2.3. *For $d=2$ the constant $C^*(\Lambda)$ is given by $C^*(\lambda_{1,2}) = 1-\Phi(\lambda_{1,2})$. Generally, if all $\lambda_{i,j}$ are close to* 0*, $C^*(\Lambda)$ is approximately equal to $C(\Lambda)$.*

3 Discriminant analysis for the GP model

In this section we construct the discriminant function within the multivariate Hüsler-Reiss GP model based on rectangular tructation. Therefore, we extend the multivariate Hüsler–Reiss GPD in (12) by a location parameter $\boldsymbol{\mu}\in\mathbb{R}^d$ and scale parameter $\boldsymbol{\sigma}>\boldsymbol{0}$ in the corresponding EVD. Then W_Λ becomes

$$\begin{aligned}
W_{\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma}}(\boldsymbol{x}) &= 1-\left(\log H_\Lambda\left(\frac{\boldsymbol{x}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)/\log H_\Lambda\left(\frac{\boldsymbol{0}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)\right)\\
&= 1-\frac{1}{C(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})}\\
&\times\left(\sum_{k=1}^{d}\int_{\left(\frac{x_k-\mu_k}{\sigma_k}\right)}^{\infty}\Phi_{\Sigma(k)}\left(\left(\lambda_{i,k}+\frac{\left(\frac{x_i-\mu_i}{\sigma_k}\right)-z}{2\lambda_{i,k}}\right)_{i=1}^{k-1}\right)e^{-z}dz\right)
\end{aligned} \tag{17}$$

where

$$C(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma}) \quad = \quad \sum_{k=1}^{d}\int_{\left(-\frac{\mu_k}{\sigma_k}\right)}^{\infty}\Phi_{\Sigma(k)}\left(\left(\lambda_{i,k}-\frac{z}{2\lambda_{i,k}}\right)_{i=1}^{k-1}\right)e^{-z}dz.$$

Theorem 3.1. *Let $w_{\Lambda,\boldsymbol{\mu}^{(1)},\boldsymbol{\sigma}}(\boldsymbol{x}|1)$ and $w_{\Lambda,\boldsymbol{\mu}^{(2)},\boldsymbol{\sigma}}(\boldsymbol{x}|2)$ be two multivariate Hüsler-Reiss GP densities which differ in the location parameters $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$. Then, using equation* (1) *we obtain the optimal* *common border*

$$\begin{aligned} D_{HR}(\boldsymbol{x}) &= (\Delta^{-1}\boldsymbol{x_\sigma})^T\Sigma(d)^{-1}(\Delta^{-1}(\Gamma^{(2)}-\Gamma^{(1)})) \\ &\quad +\frac{1}{2}(2\boldsymbol{L}+\Delta^{-1}(\Gamma^{(2)}+\Gamma^{(1)}))^T\Sigma(d)^{-1}(\Delta^{-1}(\Gamma^{(2)}-\Gamma^{(1)}))-C, \end{aligned}$$

where $\boldsymbol{L}=(\lambda_{1,d},...,\lambda_{d-1,d})^T$, $\boldsymbol{x_\sigma}=((\sigma_d x_1-\sigma_1 x_d),...,(\sigma_d x_{d-1}-\sigma_{d-1}x_d))^T$ *and* $\Gamma^{(i)}=((\sigma_1\mu_d^{(i)}-\sigma_d\mu_1^{(i)}),...,(\sigma_{d-1}\mu_d^{(i)}-\sigma_d\mu_{d-1}^{(i)}))^T$, $i=1,2$, *are* $(d-1)$*–dimensional vectors, and* $\Delta=diag(2\sigma_1\sigma_d\lambda_{1,d},...,2\sigma_{d-1}\sigma_d\lambda_{d-1,d})$ *is a* $(d-1)\times(d-1)$ *diagonal matrix. Obviously,* D_{HR} *is linear in* $\boldsymbol{x}$.

Proof. We rewrite the density $w_{\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma}}(\boldsymbol{x})$ in (13) using d–dimensional location and scale parameters, i.e.,

$$w_{\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma}}(\boldsymbol{x})=\frac{e^{-\left(\frac{x_d-\mu_d}{\sigma_d}\right)}\exp\left(-\frac{1}{2}\boldsymbol{z}^T\Sigma(d)^{-1}\boldsymbol{z}\right)}{2^{d-1}\left(\prod_{i=1}^d\sigma_i\right)\left(\prod_{i=1}^{d-1}\lambda_{i,d}\right)C^*(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})(2\pi)^{(d-1)/2}\left|\Sigma(d)\right|^{1/2}},\tag{18}$$

where $\boldsymbol{z}=\boldsymbol{L}+\Delta^{-1}(\boldsymbol{x_\sigma}+\Gamma)$ and $\boldsymbol{L}$, $\boldsymbol{x_\sigma}$, Γ and Δ are defined as above.

The scaling factor $C^*(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})$ is now given by

$$C^*(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})=C(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})(1-K(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})),$$

where $K(\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma})$ is the total mass on the d axes in the extended model.

Now, using (1), we obtain

$$-\frac{1}{2}\left(\boldsymbol{z}^{(1)}\right)^T\Sigma(d)^{-1}\boldsymbol{z}^{(1)}+\frac{1}{2}\left(\boldsymbol{z}^{(2)}\right)^T\Sigma(d)^{-1}\boldsymbol{z}^{(2)}=C,\tag{19}$$

where

$$\begin{aligned} C &= \log((c_2p_2)/(c_1p_1))+\frac{1}{\sigma_d}\left({\mu_d}^{(2)}-{\mu_d}^{(1)}\right) \\ &\quad +\log\left(C^*\left(\Lambda,\boldsymbol{\mu}^{(1)},\boldsymbol{\sigma}\right)/C^*\left(\Lambda,\boldsymbol{\mu}^{(2)},\boldsymbol{\sigma}\right)\right) \end{aligned}$$

is a constant and $\boldsymbol{z}^{(1)}=\boldsymbol{L}+\Delta^{-1}\left(\boldsymbol{x_\sigma}+\Gamma^{(1)}\right)$ and $\boldsymbol{z}^{(2)}=\boldsymbol{L}+\Delta^{-1}\left(\boldsymbol{x_\sigma}+\Gamma^{(2)}\right)$. By substituting $\boldsymbol{z}^{(1)}$ and $\boldsymbol{z}^{(2)}$ in equation (19) we obtain

$$\begin{aligned} -\frac{1}{2}\left(\boldsymbol{L}+\Delta^{-1}\left(\boldsymbol{x_\sigma}+\Gamma^{(1)}\right)\right)^T\Sigma(d)^{-1}\left(\boldsymbol{L}+\Delta^{-1}\left(\boldsymbol{x_\sigma}+\Gamma^{(1)}\right)\right)+ & \\ \frac{1}{2}\left(\boldsymbol{L}+\Delta^{-1}\left(\boldsymbol{x_\sigma}+\Gamma^{(2)}\right)\right)^T\Sigma(d)^{-1}\left(\boldsymbol{L}+\Delta^{-1}\left(\boldsymbol{x_\sigma}+\Gamma^{(2)}\right)\right) &= C, \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\left(\Delta^{-1}\boldsymbol{x_\sigma}\right)^T\Sigma(d)^{-1}\left(\boldsymbol{C}^{(2)}-\boldsymbol{C}^{(1)}\right) \\ &+\frac{1}{2}\left(\boldsymbol{C}^{(2)}+\boldsymbol{C}^{(1)}\right)^T\Sigma(d)^{-1}\left(\boldsymbol{C}^{(2)}-\boldsymbol{C}^{(1)}\right)=C \end{aligned}$$

where $C^{(1)} = L + \Delta^{-1}\Gamma^{(1)}$ and $C^{(2)} = L + \Delta^{-1}\Gamma^{(2)}$. Further simplification leads to the discriminant function $D_{HR}(x)$ as noted in the theorem. Notice that D_{HR} is linear in x. Hence the proof is complete. $\square$

Remark 3.2. *When the correlation matrices $\Sigma(d)$ and scale parameter $\boldsymbol{\sigma}$ are not identical between the two models, then the discriminant function will be a quadratic function in $\boldsymbol{x}$ which can be solved numerically. The pertaining discriminant function can be obtained by just plugging in* (19).

4 Convergence of the discriminant procedure

Having established the densities and discriminant functions within the truncated Gaussian and the Hüsler–Reiss GP model, we will now present a convergence theorem that relates both models to each other. By using the normalizing constants as Hüsler and Reiss [9] one can show that the density f_{RT} in (5) of the rectangularly truncated Gaussian model converges to the density $w_{\Lambda,\boldsymbol{\mu},\boldsymbol{\sigma}}$ of the Hüsler–Reiss GP model, cf. (18).

In the following theorem we restrict ourselves to the bivariate case. Nevertheless the proof can be generalized to arbitrary dimensions in a straightforward manner. Concerning density convergences in the univariate case we refer to Hüsler and Li [8].

Theorem 4.1. *Let $f_{RT,\boldsymbol{\mu},\Sigma}$ be the density of the bivariate rectangularly truncated Gaussian distribution with truncation vector $\boldsymbol{c}$ as in* (5)*, location parameter $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} \\ \sigma_1\sigma_2\rho_{1,2} & \sigma_2^2 \end{pmatrix}.$$

Now let $\boldsymbol{\mu}$ and Σ depend on n. We assume that the correlation coefficient $\rho_{1,2}(n)$ satisfies again

$$(1 - \rho_{1,2}(n)) \log n \to \lambda_{1,2}^2 \in [0, \infty], \quad n \to \infty,$$

cf. (7)*, and put*

$$\boldsymbol{\mu}(n) = (\mu_1 - \sigma_1 b_n^2, \mu_2 - \sigma_2 b_n^2)$$

and

$$\Sigma(n) = \begin{pmatrix} \sigma_1^2 b_n^2 & \sigma_1\sigma_2 b_n^2 \rho_{1,2}(n) \\ \sigma_1\sigma_2 b_n^2 \rho_{1,2}(n) & \sigma_2^2 b_n^2 \end{pmatrix},$$

with $b_n = n\varphi(b_n)$. Then we have

$$\lim_{n\to\infty} f_{RT,\boldsymbol{\mu}(n),\Sigma(n)}(x_1, x_2) = w_{\lambda_{1,2}}(x_1, x_2).$$

The limiting function is given by

$$
\begin{aligned}
&w_{\lambda_{1,2}}(x_1, x_2) \\
&= e^{-(x_2-\mu_2)/\sigma_2}\varphi\left(\lambda_{1,2} + \frac{(x_1-\mu_1)/\sigma_1 - (x_2-\mu_2)/\sigma_2}{2\lambda_{1,2}}\right) / \\
&\left\{2\lambda_{1,2}\sigma_1\sigma_2\left[\left(1-\Phi\left(\lambda_{1,2} + \frac{(c_2-\mu_2)/\sigma_2 - (c_1-\mu_1)/\sigma_1}{2\lambda_{1,2}}\right)\right)e^{-(c_1-\mu_1)/\sigma_1}\right.\right. \\
&\qquad\left.\left. + \left(1-\Phi\left(\lambda_{1,2} + \frac{(c_1-\mu_1)/\sigma_1 - (c_2-\mu_2)/\sigma_2}{2\lambda_{1,2}}\right)\right)e^{-(c_2-\mu_2)/\sigma_2}\right]\right\}.
\end{aligned} \tag{20}
$$

Proof. Using definition (5) for the truncation $\boldsymbol{x} > \mathbf{c}$ we obtain

$$
\begin{aligned}
&f_{RT,\boldsymbol{\mu}(n),\Sigma(n)}(x_1, x_2) \\
&= \frac{\exp\left(-\frac{1}{2}\frac{\left(\frac{x_1-\mu_1}{\sigma_1 b_n}+b_n\right)^2 - 2\rho_{1,2}(n)\left(\frac{x_1-\mu_1}{\sigma_1 b_n}+b_n\right)\left(\frac{x_2-\mu_2}{\sigma_2 b_n}+b_n\right) + \left(\frac{x_2-\mu_2}{\sigma_2 b_n}+b_n\right)^2}{(1-\rho_{1,2}(n)^2)}\right)}{2\pi\sigma_1\sigma_2 b_n^2\sqrt{1-\rho_{1,2}(n)^2}\,P\left\{X_1 > \frac{c_1-\mu_1}{\sigma_1 b_n} + b_n, X_2 > \frac{c_2-\mu_2}{\sigma_2 b_n} + b_n\right\}},
\end{aligned} \tag{21}
$$

where the distribution of (X_1, X_2) is the standard Gaussian distribution with correlation coefficient $\rho_{1,2}(n)$. Corresponding to the proof of Theorem 1 in Hüsler and Reiss [9] one gets

$$
\begin{aligned}
&nP\left\{X_1 > \frac{c_1-\mu_1}{\sigma_1 b_n} + b_n, X_2 > \frac{c_2-\mu_2}{\sigma_2 b_n} + b_n\right\} \\
&\to \left(1-\Phi\left(\lambda_{1,2} + \frac{(c_2-\mu_2)/\sigma_2 - (c_1-\mu_1)/\sigma_1}{2\lambda_{1,2}}\right)\right)e^{-(c_1-\mu_1)/\sigma_1} \\
&\qquad + \left(1-\Phi\left(\lambda_{1,2} + \frac{(c_1-\mu_1)/\sigma_1 - (c_2-\mu_2)/\sigma_2}{2\lambda_{1,2}}\right)\right)e^{-(c_2-\mu_2)/\sigma_2},
\end{aligned} \tag{22}
$$

as $n \to \infty$, and

$$b_n\sqrt{1-\rho_{1,2}(n)^2} \to 2\lambda_{1,2}, \tag{23}$$

as $n \to \infty$, which proves the convergence of the denominator.

From the proof in Hüsler and Reiss [9] we also deduce that

$$\frac{\frac{x_1-\mu_1}{\sigma_1 b_n} + b_n - \rho_{1,2}(n)\left(\frac{x_2-\mu_2}{\sigma_2 b_n} + b_n\right)}{\sqrt{1-\rho_{1,2}(n)^2}} \to \lambda_{1,2} + \frac{(x_1-\mu_1)/\sigma_1 - (x_2-\mu_2)/\sigma_2}{2\lambda_{1,2}}, \tag{24}$$

as $n \to \infty$, which we use to show the convergence of the numerator. We can write

$$\frac{n \exp\left(-\frac{1}{2}\frac{\left(\frac{x_1-\mu_1}{\sigma_1 b_n}+b_n\right)^2-2\rho_{1,2}(n)\left(\frac{x_1-\mu_1}{\sigma_1 b_n}+b_n\right)\left(\frac{x_2-\mu_2}{\sigma_2 b_n}+b_n\right)+\left(\frac{x_2-\mu_2}{\sigma_2 b_n}+b_n\right)^2}{(1-\rho_{1,2}(n)^2)}\right)}{2\pi b_n}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\frac{x_1-\mu_1}{\sigma_1 b_n} + b_n - \rho_{1,2}(n)\left(\frac{x_2-\mu_2}{\sigma_2 b_n} + b_n\right)}{\sqrt{1-\rho_{1,2}(n)^2}}\right)^2\right)$$

$$\times \frac{n}{b_n}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2 b_n} + b_n\right)^2\right), \tag{25}$$

where the first factor converges to

$$\varphi\left(\lambda_{1,2} + \frac{(x_1-\mu_1)/\sigma_1 - (x_2-\mu_2)/\sigma_2}{2\lambda_{1,2}}\right),$$

as $n \to \infty$, because of (25) and the second factor satisfies

$$\begin{aligned}
&\frac{n}{b_n}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2 b_n} + b_n\right)^2\right) \\
&= \frac{n}{b_n}\varphi(b_n) \exp\left(-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2 b_n}\right)^2\right) e^{-(x_2-\mu_2)/\sigma_2} \\
&= \exp\left(-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2 b_n}\right)^2\right) e^{-(x_2-\mu_2)/\sigma_2} \\
&\to e^{-(x_2-\mu_2)/\sigma_2}, \quad n \to \infty. \qquad (26)
\end{aligned}$$

Combining the above convergences completes the proof. □

Because the discriminant functions are obtained by using the inequality (1) which contains a ratio of densities, Theorem 4.1 directly implies the convergence of the discriminant functions. More precisely, the discriminant function D_{RT} (appropriately normalized) of the rectangularly truncated Gaussian model converges to the discriminant function D_{HR} of the Hüsler-Reiss GP model.

An analogous result still holds if different covariance matrices $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are chosen. In this case the quadratic discriminant function of the truncated Gaussian model converges to the quadratic one in the Hüsler–Reiss GP model, cf. Remark 3.2.

5 Concluding Remarks

We shortly discuss the question why one should carry out the discriminant analysis within the limiting GP model in place of the truncated Gaussian model. As already mentioned in Section

2 the limit results hold in the elliptical case under a certain condition on the random radius. We also refer to a recent article by Frick and Reiss [7] where it is verified that multivariate EVDs—including the Hüsler-Reiss EVD—occur as limiting dfs of maxima under a certain technical condition. Consequently, according to the insight gained from the limit results we know that the truncated Gaussian model as well as the GP model lead to statistical procedures which are approximately valid for a broader class of distributions.

Arguments from extreme value theory speak in favour of the GP model. First of all the GP models satisfy a certain pot-stability property, that is, truncations of such distributions are of the same type. Also, we may extend the GP model above by replacing the univariate margins. Further extensions of the initial model may be explored in the same manner; if the model is extended to the family of all GP distributions then we are in a nonparametric setup.

In practice the discriminant procedure can be executed in the following manner (formulated for $d = 2$):

1. Fix sufficiently high truncation vectors so that one is in the realm of extreme value theory. As a rule of thumb, take truncation vectors so that for each sub-sample about 15% of the original observations are in the truncated sample. In the extreme value literature one may find automatic procedures for the selection of truncation points.

2. Select a model of GP densities as given in (20).

3. Specify the discriminant function.

4. Estimate the ratio p_2/p_1 of expected frequencies by the ratio of corresponding sample frequencies, and estimate the GP parameters by MLEs. It would be desirable to explore other estimators of the GP parameters, e.g., in order to get initial estimators for the MLE procedure.

References

[1] E. M. O. Abdalla, Structural modeling of aggregated financial ratios, PhD Thesis, Dept. Math., University of Siegen, 2007.

[2] R.B. Avery, Robert B., Credit scoring models with discriminant analysis and truncated samples, No 54, (1981), Research Papers in Banking and Financial Economics, Board of Governors of the Federal Reserve System (U.S.).

[3] J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of Extremes, Theory and Applications, Wiley, Chichester, 2005.

[4] R. Bhattacharya, R. Rao, Normal Approximation and Asymptotic Expansions, Wiley, New York, 1976.

[5] M. Falk, J. Hüsler, R.-D. Reiss, Laws of Small Numbers: Extremes and Rare Events, 2nd ed., Birkhäuser, Basel, 2004 (1st ed., Birkhäuser, Basel, 1994).

[6] M. Falk, F. Marohn, B. Tewes, Foundations of Statistical Analyses and Applications with SAS, Birkhäuser, Basel, 2002.

[7] M. Frick, R.-D. Reiss, Limiting distributions of maxima under triangular schemes, J. Mult. Analysis. 101 (2010), 2346-2357.

[8] J. Hüsler, D. Li, Tail Approximations to the Density Function in EVT, Extremes 9 (2006), 131-149.

[9] J. Hüsler, R.-D. Reiss, Maxima of normal random vectors: between independence and complete dependence, Statist. Probab. Letters 7 (1989), 283-286.

[10] E. Hashorva, Elliptical triangular arrays in the max-domain of attraction of Hüsler-Reiss distribution, Statist. Probab. Letters 72 (2005), 125-135.

[11] E. Hashorva, On the multivariate Hüsler-Reiss distribution attracting the maxima of elliptical triangular arrays, Statist. Probab. Letters 76 (2006), 2027–2035.

[12] W. C. Horrace, Some results on the multivariate truncated normal distribution, J. Mult. Analysis 94 (2005), 209-221.

[13] H. Joe, Multivariate extreme-value distributions with applications to environmental data, Canad. J. Statist. 22 (1994), 47-64.

[14] E. Kaufmann, R.-D. Reiss, Approximation rates for multivariate exceedances, J. Statist. Plann. Inference 45 (1995), 235-245.

[15] S. Kocherlakota, N. Balakrishnan, K. Kocherlakota, The linear discriminant function: sampling from the truncated normal distribution, Biomet. J. 29 (1987), 131-139.

[16] P. A. Lachenbruch, M. Goldstein, Discriminant analysis, Biometrics 35 (1979), 69-85.

[17] P.A. Lachenbruch, Discriminant Analysis, Hafner Press, New York, 1975.

[18] B.G. Manjunath, Extremal discriminant analysis, PhD Thesis, Dept. Math., University of Siegen, 2010.

[19] R. Michel, Simulation of certain multivariate generalized Pareto, Extremes 10 (2007), 83-107.

[20] R.-D. Reiss, Approximate Distributions of Order Statistics, Springer, New York, 1989.

[21] R.-D. Reiss, M. Thomas, Statistical Analysis of Extreme Values, 3rd ed., Birkhäuser, Basel, 2007 (1st ed. 1997, 2nd ed. 2001, both in Birkhäuser, Basel).

[22] H. Rootzén, N. Tajvidi, Multivariate generalized Pareto distributions, Bernoulli 12 (2006), 917–930.

[23] Stefan Wilhelm and B. G. Manjunath, tmvtnorm: A Package for the Truncated Multivariate Normal Distribution, The R Journal 2(1) (2010), 25-29.

[24] N. Tajvidi, Characterization and some statistical aspects of univariate and multivariate generalised Pareto distributions, PhD Thesis, Dept. Math., University of Göteborg, 1996.

[25] G. M. Tallis, Elliptical and radial truncation in normal populations, Ann. Math. Statist. 34 (1963), 940-944.