An isobar-surfaces approach to multidimensional outlier-proneness.

Marie-Francoise Barme-Delcroix

Laboratoire de Statistique et Probabilités

Université des Sciences et Technologies de Lille

F - 59655 VILLENEUVE D'ASCQ CEDEX

Ursula Gather

Fachbereich Statistik

Universität Dortmund

D - 44221 DORTMUND

Abstract

The aim of this paper is to define and investigate outlier-proneness for multivariate distributions. This is done by using a concept of ordering multivariate data based on isobar-surfaces, which yields an utmost analogy of the results to the univariate case.

Key-words: isobars, order statistics, extreme values, weak stability, outliers, outlier-proneness.

A.M.S. Subject classification: 62 G 30, 60 D 05

1 Introduction

When modelling multivariate data, we often have some feeling about how outlier prone the underlying distribution should be. That is, there are phenomena where some outliers in the data will be a rule and are to be expected, rather than they are considered as very rare exceptions. In such situations we would not want of course to choose e.g. a multivariate normal distribution with very light tails as a model of the data-generating mechanism.

There are other multivariate distributions, different from the normal where even such a vague knowledge about their tendency to produce outliers is not known. Therefore, a classification of multivariate distributions w.r.t. their outlier-resistance and proneness as available in the univariate case ([Green 76], [Gather, Rauhut 90], [Schuster 84]) would be useful.

This paper gives a formal framework and definitions of the terms outlier-proneness and outlier-resistance of multivariate distributions based on an isobar - surfaces approach to multivariate extremes [Delcroix, Jacob 91].

More exactly, the limit behaviour of the difference of the two largest multivariate "extremes" $X_{n,n} - X_{n,n-1}$ is used -in the sense of [Green 76]- to define the outlier-proneness of the underlying distribution. Having to decide then, if some class of multivariate distribution functions is outlier-prone, we need a characterization, directly in terms of the distribution function, too. This paper gives such equivalence theorems which allow to check for outlier-proneness or outlier-resistance of a distribution in many different ways.

As mentioned, we choose a concept of ordering multivariate data based on the isobar-surfaces of the underlying distribution. Though this is a natural way of ordering multivariate data, in contexts with just a given data set, it cannot be applied, when the data generating distribution is not completely known. This is usually a deficiency but in our situation, where we want to check if a given distribution at hand is suitable for modelling a data structure, we are able to use this natural notion of ordering in terms of isobar-extremes, since the distribution is known then.

Also, as in the univariate case, this new notion of outlier-proneness via isobar-extremes is strongly related to weak stability of the extremes. Our approach differs from the one by [Mathar 85] who defines outlier-resistance via the limit behaviour of the distance of the upper extremes of the real valued norms of the sample points as ordering principle. His approach therefore yields a characterization of multivariate outlier-proneness via the minimum of the distribution function of the marginals, whereas our definition leads to characterizations depending on the behaviour of the conditional distribution functions given the angles. Hence, we take into account the complete shape of the multivariate distribution.

This paper is organized as follows. In section 2 we start with defining weak stability of multivariate extremes by the isobar surface ordering. In Section 3 we define outlier-resistance and outlier-proneness of multivariate distributions, we relate these properties to weak stability of the extremes and characterize outlier-resistance by the tail behaviour of the conditional radial distributions. Section 4 gives a generalization and examples.

2 Weak stability of multivariate extremes

We first recall the definition of the largest value of a multivariate sample, as given in [Delcroix, Jacob 91]. The motivation was to describe the asymptotic position of a multivariate sample, [Barme-Delcroix 93], without using classical convexity notions, [Geffroy 61].

We consider random variables with values in the Euclidean space $\mathbf{R}^k$.

For every x in $\mathbf{R}^k \setminus \{0\}$ we define a pair $(\|x\|, \frac{x}{\|x\|}) = (r, \theta)$ in $\mathbf{R}^{+*} \times \mathbf{S}^{k-1}$, where $\|\cdot\|$ is the Euclidean norm and $\mathbf{R}^{+*}$ is the set of the strictly positive real numbers. The unit sphere $\mathbf{S}^{k-1}$ in $\mathbf{R}^k$ is endowed with the induced topology of $\mathbf{R}^k$.

For each random variable (r.v. for short) $X = (R, \Theta)$ in $\mathbf{R}^k$ with radius R and angle Θ, we assume that the distribution of Θ, and for all θ in $\mathbf{S}^{k-1}$, the distribution of R given $\Theta = \theta$ respectively, has a continuous density. F_θ denotes the continuous and one-to-one conditional distribution function of R given $\Theta = \theta$. This means in particular that we suppose $F_\theta(r) < 1$ for all $r > 0$ and for all θ.

For each $0 < u < 1$, we call the mapping $\theta \rightsquigarrow F_\theta^{-1}(u)$ a u-level isobar of the distribution of R given $\Theta = \theta$. We suppose that this mapping is continuous and strictly positive. The surface given by $\rho_u(\theta) = F_\theta^{-1}(u)$, considered as a function of θ, is also called a u-isobar for all $0 < u < 1$.

Let $x_1 = (r_1, \theta_1), \dots, x_n = (r_n, \theta_n)$ be observations of an i.i.d. sample $X_1, \dots, X_n$ from the distribution of $X = (R, \Theta)$.
Let $u_j = F_{\theta_j}(r_j)$ for $1 \le j \le n$, $u_n^* = \max\limits_{1 \le j \le n} u_j$, and define $x_n^* = (r_n^*, \theta_n^*)$ by

$$F_{\theta_n^*}(r_n^*) = u_n^*.$$

Since F_θ is continuous and strictly increasing for all θ, like this we have defined almost surely unique r.v.'s $U_1, \dots, U_n$ as well as an almost surely unique r.v. $X_n^* = (R_n^*, \Theta_n^*)$ which is an element of $\{X_1, \dots, X_n\}$ for which

$$F_{\Theta_n^*}(R_n^*) = \max_{1 \le j \le n} U_j.$$

We call X_n^* the isobar-maximum of $X_1, \dots, X_n$.

Obviously, to find this isobar-maximum of a multivariate sample, the underlying distribution has to be known. However, this kind of extreme value, and more generally, the ordering of the sample according to the isobars, does not give up any information the sample carries, like the ordering by norms, e.g. It is possible to give an estimation of the isobars by regression methods for particular cases , [Jacob, Suquet 97] . One can also estimate the origin by using the barycentre of the sample points. However for many situations this origin is given in a natural way.

It has been shown in [Delcroix, Jacob 91] that the conditional distribution of R_n^* given Θ_n^* is F_θ^n , hence the distributions of (R_n^*, Θ_n^*) and (R, Θ) have the same set of isobars which led to the following definition.

Definition 1 *For a sequence $(E_n)_n$ of multivariate r.v.'s, the sequence $(X_n^*)_n = ((R_n^*, \Theta_n^*))_n$ of the isobar-maxima is called stable in probability if and only if there is a sequence $(g_n)_n$ of* *isobars satisfying*

$$R_n^* - g_n(\Theta_n^*) \xrightarrow{P} 0 \ . \tag{1}$$

Following [Geffroy 58] it is possible to choose $g_n(\theta) = F_\theta^{-1}(1 - \frac{1}{n})$.

It is convenient to fix a point $x_1 = (1, \theta_1)$, θ_1 in $\mathbf{S}^{k-1}$. For every point $x = (r, \theta_1)$, there is a unique surface $g(\theta, r)$, θ in $\mathbf{S}^{k-1}$, containing x, which has a level denoted by $u(r)$ and which is given by

$$g(\theta, r) = \rho_{u(r)}(\theta) = F_\theta{}^{-1}(F_{\theta_1}(r)). \tag{2}$$

Note that $g(\theta_1, r) = r$. Moreover the mapping $r \rightsquigarrow u(r)$ from $\mathbf{R}_+^*$ into $\mathbf{R}_+^*$ is increasing and one-to-one.

The following conditions (H) and (K) will be needed.

(H) There exist $0 < \alpha \leq \beta < \infty$ such that for all θ in $\mathbf{S}^{k-1}$ and for all $r > 0$:

$$\alpha \leq \frac{\partial g}{\partial r}(\theta, r) \leq \beta.$$

(K) For all $\varepsilon > 0$, there exists $\eta > 0$ such that for all $r > 0$:

$$\sup_\theta \{g(\theta, r + \eta) - g(\theta, r - \eta)\} < \varepsilon \ .$$

Clearly, (H) implies (K).

Remark 1 *Condition (H) entails a regularity property of the isobars following from the mean value theorem :*

For all $\beta_0 > 0$ there exists $\eta > 0$ and for all $r > 0$, there exist two isobars $h_{\beta_0}(\theta, r)$ and $\tilde{h}_{\beta_0}(\theta, r)$ such that for all θ,

$$g(\theta, r) - \beta_0 < \tilde{h}_{\beta_0}(\theta, r) < g(\theta, r) - \eta < g(\theta, r) + \eta < h_{\beta_0}(\theta, r) < g(\theta, r) + \beta_0.$$

Note that η does not depend on r.

Remark 2 *For a bivariate Gaussian sample with covariance matrix* $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix}$, *we have* $g(\theta, r) = r\phi(\theta)$ *with* $\phi(\theta) = \frac{1}{\sqrt{2}\sigma}(\frac{cos^2\theta}{2\sigma^2} + \frac{sin^2\theta}{2\tau^2})^{\frac{-1}{2}}$ *and the isobars are the density contours. Note that condition (H) is satisfied. For* $\sigma = \tau = 1$ *the distribution is spherically symmetric and the isobars are circles. Hence in this particular case the ordering of the sample is the ordering of the norms of the sample points.*

The next theorem gives conditions for stability similar to those of [Geffroy 58] in the univariate case (see [Delcroix, Jacob 91] for a proof). For this purpose we define W_n^* by $F_{\theta_1}(W_n^*) = F_{\Theta_n^*}(R_n^*)$ for $X_n^* = (R_n^*, \Theta_n^*)$, i.e. $W_n^* = F_{\theta_1}^{-1}(F_{\Theta_n^*}(R_n^*))$ is the intersection of the half axis containing $x_1 = (1, \theta_1)$ and the isobar containing X_n^*.

Theorem 1 *a) Under condition (K) the sequence* $(X_n^*)_n$ *is stable in probability if* $(W_n^*)_n$ *is stable in probability .*

b) Under condition (H) the sequence $(W_n^*)_n$ *is stable in probability if and only if* $(X_n^*)_n$ *is stable in probability .*

c) Consider for some fixed integer α *the sequence* $(X_{n,n-\alpha+1})_n$, *this being defined by ordering the sample according to increasing levels by*

$$X_{n,1}, \cdots, X_{n,n-\alpha+1}, \cdots, X_{n,n} = X_n^*.$$

Let (H) be satisfied. Then $(X_n^*)_n$ *is stable in probability if and only if* $(X_{n,n-\alpha+1})_n$ *is stable in probability.*

For the proof of a) and b) see [Delcroix, Jacob 91], c) follows immediately from a) and b) and the univariate result of [Geffroy 58].

Possibilities to check for the weak stability of $(X_n^*)_n$ on the basis of the distribution function $F_\theta(r)$, $r > 0$, θ in $\mathbf{S}^{k-1}$, will be provided by the characterization results in Theorem 2 below.

3 Multidimensional outlier-prone and outlier-resistant distributions

We give an application of stability in probability of multivariate samples as defined in the previous section to the notion of outlier-resistant and outlier-prone distributions. In [Neyman, Scott 71] we find a definition which has been improved by Green, [Green 76]. The goal is to distinguish between two classes of distributions : distributions for which as a rule there exist observations far apart from the main group of the data, and distributions for which this phenomenon occurs with very small probability. So, Green called a univariate distribution F absolutely outlier-resistant if for all $\varepsilon > 0$:

$$(3) \qquad \lim_{n\to+\infty} P(X_{n,n} - X_{n,n-1} > \varepsilon) = 0$$

where $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$ are the usual univariate order statistics of $X_1, \cdots, X_n$, distributed identically according to F.

On the other hand, a distribution F is called absolutely outlier-prone if there exist $\varepsilon > 0$, $\delta > 0$ and an integer n_0 such that for $n \geq n_0$:

$$(4) \qquad P(X_{n,n} - X_{n,n-1} > \varepsilon) \geq \delta.$$

Remark 3 *The AOR-and AOP-property depend only on the tail-behaviour of F, e.g. AOR is equivalent to $\lim_{x\to+\infty} \frac{1-F(x)}{1-F(x-h)} = 0$ for all $h > 0$. Moreover, the behaviour of many types of univariate distributions is investigated in [Gather 79] yielding the following special results: the Gumbel distribution with $F(x) = exp(-e^{-x})$ for $x \in \mathbf{R}$ is not AOR but is AOP.*
Also the Frechet distribution with $F(x) = exp(-x^{-\alpha})\mathbf{1}_{\{x>0\}}$ is not AOR but is AOP for all $\alpha > 0$.
Distribution functions of the type $F(x) = (1 - cexp(-bx^{\alpha}))\mathbf{1}_{\{x>0\}}$ with constants c, b, $\alpha > 0$, are AOR as long as $\alpha > 1$.

Other definitions of outlier proneness of univariate distributions have been given for example by [O'Hagan 79], and [Goldstein 82] in a Bayesian framework. [Gather, Rauhut 90] discuss these different notions for univariate data. Here, we will extend the notion of outlier-proneness given by Green to multidimensional samples by using isobars.

Recall that for all θ, F_θ denotes the distribution function of R given $\Theta = \theta$ and that $G_\theta = 1 - F_\theta$. Note also that for each sample point $X_i = (R_i, \Theta_i)$, $i = 1, \cdots, n$, there exists almost surely a unique isobar containing X_i. Let $g_{n,n}$ denote the isobar containing $X_n^* = X_{n,n}$ and $g_{n,n-1}$ the isobar containing $X_{n,n-1}$. Thus for all θ in $\mathbf{S}^{k-1}$, $g_{n,n}$ and $g_{n,n-1}$ are real valued r.v's. Since W_n^* was defined as the intersection of the half axis containing $(1, \theta_1)$ and the isobar containing X_n^*, we now have $W_n^* = g_{n,n}(\theta_1)$. If we define analogously for all $1 \leq i \leq n$, $W_i = {F_{\theta_1}}^{-1}(F_{\Theta_i}(R_i))$, we get $W_{n,1}, \cdots, W_{n,n} = W_n^*$ as the usual order statistics of the real valued sample $W_1, \cdots, W_n$ distributed identically according to F_{θ_1}.

3.1 Multivariate AOR distributions

Definition 2 *The distribution of the multivariate r.v. (R, Θ) is absolutely outlier-resistant (AOR), if and only if for all θ :*

$$g_{n,n}(\theta) - g_{n,n-1}(\theta) \xrightarrow{P} 0. \tag{5}$$

For a real sample it has been shown in [Geffroy 58] and [Gnedenko 43], that $(X_{n,n})_n$ is stable in probability if and only if $X_{n,n} - X_{n,n-1} \xrightarrow{\mathrm{P}} 0$. The following theorem gives an analogous result and as mentioned in Remark 3 a characterization of weak stability by the tail behaviour of the underlying distribution.

Let condition (H) always be satisfied in the following.

Theorem 2 *All the following statements are equivalent :*

(i) The distribution of (R, Θ) is AOR.

(ii) $(X_n^*)_n$ is stable in probability.

(iii) For all $1 \leq \alpha \leq n$, $(X_{n,n-\alpha+1})_n$ is stable in probability.

(iv) There exists θ_1 such that $\lim\limits_{x \to +\infty} \frac{G_{\theta_1}(x)}{G_{\theta_1}(x-h)} = 0$, for all $h > 0$.

(v) For all θ, $\lim\limits_{x \to +\infty} \frac{G_\theta(x)}{G_\theta(x-h)} = 0$, for all $h > 0$.

(vi) $W_n^* - W_{n,n-1} \xrightarrow{P} 0$.

(vii) $(W_n^*)_n$ *is stable in probability.*

(viii) For all θ*, the distribution* F_θ *is AOR.*

(ix) There exists θ_1 *such that the distribution* F_{θ_1} *is AOR..*

Proof : Theorem 1c) shows that (ii) and (iii) are equivalent. Gnedenko's Theorem and Theorem 1b) show that (ii), (iv), (v) and (vii) are equivalent. Moreover, from [Geffroy 58], (vii) and (vi) are equivalent. Now, (i) involves (ii) : if for all θ, $g_{n,n}(\theta) - g_{n,n-1}(\theta) \xrightarrow{\mathrm{P}} 0$ we get $g_{n,n}(\theta_1) - g_{n,n-1}(\theta_1) \xrightarrow{\mathrm{P}} 0$, that is $W_n^* - W_{n,n-1} \xrightarrow{\mathrm{P}} 0$; and since (vi) and (vii) are equivalent, $(W_n^*)_n$ is stable in probability and from Theorem 1b) $(X_n^*)_n$ is stable in probability.

Conversely, if $(X_n^*)_n$ is stable in probability, $(W_n^*)_n$ is also stable and $W_n^* - W_{n,n-1} \xrightarrow{\mathrm{P}} 0$. Then $g_{n,n}(\theta_1) - g_{n,n-1}(\theta_1) \xrightarrow{\mathrm{P}} 0$; but θ_1 being arbitrary, we obtain (i).

Clearly, these properties are equivalent to (viii) and (ix). ∎

For univariate samples, it is possible, following [Gather, Rauhut 90], to give other characterisations of AOR distributions based on the mean residual life function (mrlf), which is defined for a real r.v. X by

$$e(x) = E(X - x \mid X > x).$$

For $i = 1, \cdots, n-1$, and for $x > 0$, $y > 0$ and $n > 2$, let

$$M_{i,n}(x, y, \theta) = P\{g_{n,i+1}(\theta) - g_{n,i}(\theta) > y \mid g_{n,i}(\theta) = x)\} \tag{6}$$

For fixed, $\theta = \theta_1$, $M_{i,n}(x, y, \theta)$ can be written as

$$M_{i,n}(x, y, \theta_1) = P\{W_{n,i+1} - W_{n,i} > y \mid W_{n,i} = x\}. \tag{7}$$

From [Gather, Rauhut 90] and Theorem 2 we obtain the following result.

Theorem 3 *The distribution of* (R, Θ) *is AOR if and only if there exists* θ_1 *such that for* *all* $y > 0$ *:*

$$\lim_{x \to +\infty} M_{i,n}(x, y, \theta_1) = 0, \tag{8}$$

for some $1 \leq i \leq n$ *.*

Proof : From Theorem 2, the distribution of (R, Θ) is AOR if and only if there exists θ_1 such that F_{θ_1} is AOR. But, from [Gather, Rauhut 90], F_{θ_1} is AOR if and only if (8) is valid. To show this we observe that the order statistics $W_{n,1}, \cdots, W_{n,n}$ form a Markov chain, [Arnold, Becker, Gather, Zahedi 84], [David 81] and that

$$M_{i,n}(x, y, \theta_1) = P\{W_{n,i+1} > x + y \mid W_{n,i} = x\} = \left(\frac{1 - F_{\theta_1}(x + y)}{1 - F_{\theta_1}(x)}\right)^{n-i}.$$

Assertion (iv) of the previous theorem completes the proof. ∎

In Definition 2, the sample size increases ; but in Theorem 3, the sample size is fixed which makes it intuitively easier to relate the definition of outlier resistance of the distribution to the non-occurence of outliers in the sample : the larger $X_{n,i}$ gets, the smaller the probability for the difference $X_{n,i+1} - X_{n,i}$ to be larger than an arbitrary positive number. The next theorem describes this fact in average.

For all θ in $\mathbf{S}^{k-1}$ and for all $i = 1, \cdots, n$, consider

$$\tilde{M}_{i,n}(x, \theta) = E(g_{n,i+1}(\theta) - g_{n,i}(\theta) \mid g_{n,i}(\theta) = x). \tag{9}$$

For fixed $\theta = \theta_1$, $\tilde{M}_{i,n}(x, \theta)$ can be written as

$$\tilde{M}_{i,n}(x, \theta_1) = E(W_{n,i+1} - W_{n,i} \mid W_{n,i} = x)$$

Theorem 4 *Suppose that for all* θ*,* $\int x\, dF_\theta$ *exists. Then, the distribution of* (R, Θ) *is AOR if and only if there exists* θ_1 *such that for all* $n \geq 3$ *:*

$$\lim_{x \to +\infty} \tilde{M}_{n-1,n}(x, \theta_1) = 0.$$

The proof is again only an application of Theorem 1b) and of [Gather, Rauhut 90].

3.2 AOP distributions

Definition 3 *The distribution of (R, Θ) is called absolutely outlier-prone, (AOP), if and only if for all θ there exist $\varepsilon > 0$, $\delta > 0$ and an integer n_θ, such that for all θ and for all $n \geq n_\theta$:*

$$P(g_{n,n}(\theta) - g_{n,n-1}(\theta) > \varepsilon) > \delta. \tag{10}$$

That is, for all θ, the distribution F_θ is AOP.

Theorem 5 *All the following statements are equivalent :*

(i) The distribution of (R, Θ) is AOP.

(ii)For all θ, there exist $\alpha > 0$, $\beta > 0$ such that for all x

$$\frac{1 - F_\theta(x + \beta)}{1 - F_\theta(x)} \geq \alpha. \tag{11}$$

(iii) There exist θ_0, $\alpha_0 > 0$ and $\beta_0 > 0$ such that for all x

$$\frac{1 - F_{\theta_0}(x + \beta_0)}{1 - F_{\theta_0}(x)} \geq \alpha_0. \tag{12}$$

(iv) There exists θ_0 such that F_{θ_0} is AOP.

Proof : From [Green 76] Theorem 3.3, we have that for fixed θ, the univariate distribution F_θ is AOP if and only if (11) is fulfilled. This proves that (i) and (ii) are equivalent.

Clearly, (ii) implies (iii).

To show that (iii) implies (ii) we consider $\theta_1 \neq \theta_0$; for all $r > 0$. There exists an isobar $g(\theta, r)$, θ in $\mathbf{S}^{k-1}$, containing the point (r, θ_1) . Let $u(r)$ denote the level of this isobar. Since (H) is satisfied (see Remark 1) there exist $\eta > 0$ and an isobar $h_{\beta_0}(\theta, r)$ such that for all r and for all θ

$$g(\theta, r) + \eta < h_{\beta_0}(\theta, r) < g(\theta, r) + \beta_0.$$

Let $u_{\beta_0}(r)$ denote the level of $h_{\beta_0}(\theta, r)$. Since $G_{\theta_1} = 1 - F_{\theta_1}$ is decreasing,

$$G_{\theta_1}(r) = G_{\theta_1}(g(\theta_1, r)) = 1 - u(r) = G_{\theta_0}(g(\theta_0, r))\,,$$

and

$$G_{\theta_1}(r+\eta) = G_{\theta_1}(g(\theta_1, r)+\eta) > G_{\theta_1}(h_{\beta_0}(\theta_1, r)) = 1 - u_{\beta_0}(r).$$

Moreover,

$$1 - u_{\beta_0}(r) = G_{\theta_0}(h_{\beta_0}(\theta_0, r)) > G_{\theta_0}(g(\theta_0, r)+\beta_0),$$

and

$$\frac{G_{\theta_0}(g(\theta_0, r)+\beta_0)}{G_{\theta_0}(g(\theta_0, r))} < \frac{G_{\theta_1}(r+\eta)}{G_{\theta_1}(r)}.$$

Thus, if $\frac{G_{\theta_0}(r+\beta_0)}{G_{\theta_0}(r)} \geq \alpha_0$ for all real x, then for all $\theta_1 \neq \theta_0$, there exist $\beta_1 = \eta > 0$ and $\alpha_1 = \alpha_0 > 0$ such that for all r

$$\frac{G_{\theta_1}(r+\beta_1)}{G_{\theta_1}(r)} \geq \alpha_1\,,$$

and we obtain (ii) .

Clearly, (iv) is equivalent to the other statements. ∎

Examples :

a) For a bivariate Gaussian sample such as in Remark 2, we have $F_\theta(r) = 1 - exp(-r^2\phi(\theta))$ and following Theorem 2iv) we can conclude that this distribution is AOR.

b) Suppose that $F_\theta(r) = 1 - cexp(-br^{\alpha(\theta)})\mathbf{1}_{\{r>0\}}$ with α a strictly positive continuous function and b, $c > 0$ (Gumbel type distribution). It has been shown in [Delcroix, Jacob 91] that neither (H) nor the regularity property of isobars from Remark 1 is satisfied for this distribution. But if $\inf_\theta(\alpha(\theta)) > 1$, condition (K) is fulfilled for r large. Moreover, as in the univariate case, from Theorem 1 a), $(X_n^*)_n$ is stable and the distribution of (R, Θ) is AOR. If α is constant and equal to 1, the distribution is AOP. And if there exists θ_0 such that $\alpha(\theta_0) < 1$ then the distribution is neither AOP nor AOR.

c) For the bivariate Morgenstern distribution with density $f(x,y) = e^{-(x+y)}(1+\alpha(2e^{-x}-1)(2e^{-y}-1))$ with $-1 \leq \alpha \leq 1$ it is possible to write down the distribution function F_θ explicitely:

$$\begin{aligned}
F_\theta(r) &= \frac{1}{d(\theta)}\left\{(1+\alpha)\left[\frac{(1-e^{-(\cos\theta+\sin\theta)r})}{(\cos\theta+\sin\theta)^2} - \frac{re^{-(\cos\theta+\sin\theta)r)}}{(\cos\theta+\sin\theta)}\right]\right.\\
&+ 4\alpha\left[\frac{1-e^{-2(\cos\theta+\sin\theta)r}}{4(\cos\theta+\sin\theta)^2} - \frac{re^{-2(\cos\theta+\sin\theta)r}}{2(\cos\theta+\sin\theta)}\right]\\
&+ 2\alpha\left[\frac{1-e^{-(2\cos\theta+\sin\theta)r}}{(2\cos\theta+\sin\theta)^2} - \frac{re^{-(2\cos\theta+\sin\theta)r}}{(2\cos\theta+\sin\theta)}\right]\\
&+ \left.2\alpha\left[\frac{1-e^{-(\cos\theta+2\sin\theta)r}}{(\cos\theta+2\sin\theta)^2} - \frac{re^{-(\cos\theta+2\sin\theta)r}}{(\cos\theta+2\sin\theta)}\right]\right\}
\end{aligned}$$

where $d(\theta)$ is a function of θ. Hence $F_\theta(r)$ is of the type

$$1 - Aexp(-ar) - Bexp(-br) - Cexp(-cr) - Dexp(-dr)$$

$$-A'rexp(-ar) - B'rexp(-br) - C'rexp(-cr) - D'rexp(-dr)$$

with $a = cos(\theta) + sin(\theta)$, $b = 2a$, $c = cos(\theta) + 2sin(\theta)$, $d = 2cos(\theta) + sin(\theta)$ and A, B, C, D, A', B', C', D' all depending only on θ and α. We can then apply Theorem 5 iii) which yields after some manipulations that the bivariate Morgenstern distribution is AOP.

The following corollary is also obvious from Theorem 5 as well as from using (6) (7) and (9).

Corollary 1 *a) The distribution of (R,Θ) is AOP if and only if there exists θ_0 such that for all $y>0$, there exist α_0 and x_0 such that*

$$M_{i,n}(x,y,\theta_0) \geq \alpha_0\,,$$

for all $x \geq x_0$, for some $1 \leq i \leq n-1$.

b) Suppose that $\int xdF_\theta$ exists for all θ and that the distribution of (R,Θ) is AOP, then there exist θ_0, δ_0 and x_0, such that for $x \geq x_0$ and for all $n \geq 3$

$$\tilde{M}_{n-1,n}(x,\theta_0) \geq \delta_0.$$

4 Generalization and examples

Of course, a lot of multidimensional distributions do not have stability properties. However we can generalize the notion of weak stability, to φ-stability, see e.g. [Delcroix, Jacob 91], [Gather, Rauhut 90], [Geffroy 58], [Gnedenko 43], [Green 76], [Resnick 87], [Tomkins, Wang 92]. For a positive, increasing, concave, one-to-one C^1-function defined on $\mathbf{R}^+$, we consider the set of points $((\varphi(R_1), \Theta_1) \cdots (\varphi(R_n), \Theta_n))$ instead of the initial sample. Then, for a suitable function φ,we obtain stability properties for many usual multivariate distributions (exponential distributions, Cauchy distributions..). Having defined φ-stability [Delcroix, Jacob 91], we can also define multidimensional φ-outlier-resistant or φ-outlier-prone distributions. It suffices to consider the distribution of $(\varphi(R), \Theta)$ instead of the distribution of (R, Θ). For example, if the distribution of (R, Θ) is AOR and if φ is a positive, increasing, concave, one-to-one C^1-function defined over $\mathbf{R}^+$, then the distribution of $(\varphi(R), \Theta)$ is also AOR. When $\varphi(x) = max(0, Log\, x)$, we come to the notions of relatively outlier-resistant or relatively outlier-prone distributions. In this case, φ-outlier-resistant and φ-outlier-prone are denoted by ROR and ROP as they are given in [Green 76] for univariate distributions.

Examples :

a) Exponentional-type distributions with

$$F_\theta(r) = (1 - cexp(-b(\theta)r))\mathbf{1}_{\{r>0\}}\,,$$

$c > 0$ and b being a strictly positive and continuous function, are ROR and AOP (see example 1b in section 3).

b) Cauchy distributions with conditional density

$$f_\theta(r) = \frac{2}{\pi}\frac{\lambda(\theta)}{r^2 + \lambda(\theta)^2}\mathbf{1}_{\{r>0\}}\,,$$

λ being a strictly positive, and continuous function, are ROP, but are φ-OR for $\varphi(x) = LogLog\, x$.

c) If $0 < m \leq 1$ and

$$F_\theta(r) = (1 - exp(-\alpha(\theta) r^m))\mathbf{1}_{\{r>0\}} ,$$

the distribution of (R, Θ) is φ-OR, with $\varphi(x) = x^{\frac{1}{2m}}$.

(For each example the general form of φ is given in [Delcroix, Jacob 91].)

References

[Arnold, Becker, Gather, Zahedi 84] ARNOLD B.C., BECKER A., GATHER U., ZAHEDI H. (1984) - On the Markov property of order statistics. *Journal of Statistical Planning and Inference, 147–154.*

[David 81] DAVID H.A. (1981) - Orders Statistics. 2^{nd} *edition. Wiley, New-York.*

[De Haan 70] DE HAAN L. (1970) - On regular variation and its application to the weak convergence of sample extremes. *Mathematical centre tracts. n° 32. Amsterdam.*

[Delcroix, Jacob 91] DELCROIX M.F. and JACOB P. (1991) - Stability of extreme value for a multidimensional sample. *Statistique et Analyse des Données, Vol. 16, 1-21.*

[Barme-Delcroix 93] BARME-DELCROIX M.F. (1993) - Localisation asymptotique des échantillons multidimensionnels. *Bulletin de la societé mathématique de Belgique, vol. 45, fasc 2 série B.*

[Gather 79] GATHER U. (1979) - Über Ausreissertests und Ausreisseranfälligkeit von Wahrscheinlichkeitsverteilungen. *Dissertation, Aachen Technical University.*

[Gather, Rauhut 90] GATHER U. and RAUHUT B. (1990) - Outlier behaviour of probability distributions. *Journal of Statistical Planning and Inference 26, 237–252. North Holland.*

[Geffroy 58] GEFFROY J. (1958) - Contribution à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris, 7, 8 37–185.*

[Geffroy 61] GEFFROY J. (1961) - Localisation asymptotique du polyèdre d'appui d'un échantillon laplacien à k dimension, *Publ. Inst. Stat. Univ. Paris, 1961, n°59.*

[Gnedenko 43] GNEDENKO B.V. (1943) - Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math. 44, 423–453.*

[Goldstein 82] GOLDSTEIN M. (1982) - Contamination distribution. *Ann. Statist. 10, 174–183.*

[Green 76] GREEN R. F. (1976) - Outlier-prone and outlier-resistant distributions. *J. Amer. Statist. Assoc. 71, 502–505.*

[Jacob, Suquet 97] JACOB P. and SUQUET C. (1997) - Regression and asymptotical location of a multivariate sample. *Statistics and Probability Letters, 35, 173–179..*

[Mathar 81] MATHAR R. (1981) - Ausreisser bei ein- und mehrdimensionalen Wahrscheinlichkeitsverteilungen. *Dissertation, Aachen Technical University.*

[Mathar 85] MATHAR R. (1985) - Outlier-prone and outlier resistant multidimensional distributions. *Statistics 16, 451–456.*

[Neyman, Scott 71] NEYMAN J. and SCOTT E.L. (1971) - Outlier proneness of phenomena and of related distribution. *In : J. Rustagi, Ed. Optimizing Methods in Statistics. Academic Press, New-York, 413–430.*

[O'Hagan 79] O'HAGAN A. (1979) - On outlier rejection phenomena in Bayes inference. *J. Roy. Statist. Soc. Ser. B 41, 358–367.*

[Resnick 87] RESNICK S.I. (1987) - Extreme values, Regular Variation, and Point Processes. *Springer Verlag.*

[Schuster 84] SCHUSTER E.F. (1984) - Classification of probability laws by tail behaviour. *J. Amer. Statist. Assoc. 79, 936–939.*

[Tomkins, Wang 92] TOMKINS R.J. and WANG H. (1992) - Stability theorems for large order statistics with varying ranks. *Statistics and Probability letters 14, 91–95.*