Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 1750–1755
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Limit laws for multidimensional extremes

Marie-Francoise Barme-Delcroix[a,*], Ursula Gather[b]

[a]*Laboratoire Painlevé, CNRS - UMR 8524, Université des Sciences et Technologies de Lille, F - 59655 Villeneuve d'Ascq cedex, France*
[b]*Fachbereich Statistik, Universitat Dortmund, D - 44221 Dortmund, Germany*

Received 27 June 2005; received in revised form 27 February 2007; accepted 10 April 2007
Available online 21 April 2007

Abstract

A definition of the ordering of a multivariate sample—based on the isobar surfaces—is used in order to obtain limit laws for the extreme values of a multidimensional sample.

MSC: 62G30; 60D05

Keywords: Isobars; Order statistics; Extreme values; Stability; Limit laws; Concomitant

1. Introduction

Today many papers on extreme values deal with multivariate extreme values (see Nadarajah et al., 1998; Schlatler and Tawn, 2003; Fougéres, 2004; Heffernan and Tawn, 2004; Falk and Reiss, 2005) see also for recent developments about max-stable processes and multivariate extremal index (Smith and Weissman, 1996; Zhang and Smith, 2004). Unfortunately, ordering multivariate data can be done in various ways and many definitions have been proposed (e.g. Barnett, 1976; Maller, 1990). Some papers of Abdous and Theodorescu (1992), Einmahl and Mason (1992), Massé and Theodorescu (1994), De Haan and Huang (1995), and more recently, of Berlinet et al. (2001), Serfling (2002) develop the notion of multivariate quantiles. In the classical scheme (cartesian coordinates) the multivariate variables are ordered coordinate by coordinate—see for example Galambos (1987) and the references therein. And in this way the maximum value thus obtained is not a sample point.

In this paper we will use the definition of the maximum value of a multidimensional sample, given in Delcroix and Jacob (1991). This definition is based on surfaces we have called isobars. Our approach is more geometric and as recalled just below, we use the level-surfaces of the conditional distribution function of the radius R given the angle Θ. So the maximum value is a sample point. The first motivation (Barme-Delcroix, 1993), was to find the asymptotic location of a multidimensional sample without using the convex hull of the sample as it is done classically in Geffroy (1958, 1961). By a unidimensional approach, stability, strong behaviour and outlier-proneness properties have been explored for this new notion of extreme value in

*Corresponding author.
E-mail address: marie-francoise.barme@math.univ-lille1.fr (M.-F. Barme-Delcroix).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2007.04.014

Barme-Delcroix and Brito (2001) and Barme-Delcroix and Gather (2002). In this paper, the goal is to find limit laws for such multidimensional extremes. Through the use of the isobar surfaces we will see that the classical results (Gnedenko, 1943) about limit laws for unidimensional extremes are still efficient in this multidimensional context.

2. Preliminaries

We consider random variables defined on a probability space $(\Omega, \mathscr{A}, \mathscr{P})$ and with values in the Euclidean space $\mathbf{R}^k$.

For every x in $\mathbf{R}^k \backslash \{0\}$ we define a pair $(\|x\|, x/\|x\|) = (r, \theta)$ in $\mathbf{R}^{+*} \times \mathbf{S}^{k-1}$, where $\| \cdot \|$ is the Euclidean norm and $\mathbf{R}^{+*}$ is the set of the 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 $X = (R, \Theta) = (\|X\|, X/\|X\|) \in \mathbf{R}^k$, we assume that the distribution of Θ, and for all θ, the distribution of R given $\Theta = \theta$, have a continuous density. F_θ denotes the continuous conditional distribution function of R, given $\Theta = \theta$ and F_θ^{-1} the generalized inverse of F_θ.

For each $0 < u < 1$, we call the mapping $\theta \rightsquigarrow F_\theta^{-1}(u)$ a u-level *isobar* from the distribution of R given $\Theta = \theta$. We suppose that this mapping is continuous and strictly positive; the surface given by $\rho = F_\theta^{-1}(u)$ is also called an isobar, see Fig. 1.

Let $E_n = (X_1, \dots, X_n)$ be a sample of independent random variables in $\mathbf{R}^k$ each with the same distribution as X. Clearly for each $1 \leqslant k \leqslant n$ there is almost surely a unique u_k-level isobar from the distribution of R given $\Theta = \theta$ which contains $(R_k, \Theta_k) = (\|X_k\|, X_k/\|X_k\|)$. We define *the maximum value* in E_n as the point of the sample $X_n^* = (R_n^*, \Theta_n^*)$ which belongs to the upper level isobar, i.e. the isobar with level $\max_{1 \leqslant k \leqslant n} u_k$, see Fig. 2. The multivariate sample is then ordered according to the increasing levels of the corresponding isobar surfaces, and the order statistics of the n-sample are denoted by

$$X_{n,1}, \dots, X_{n,i}, \dots, X_{n,n-1}, X_{n,n} = X_n^*.$$

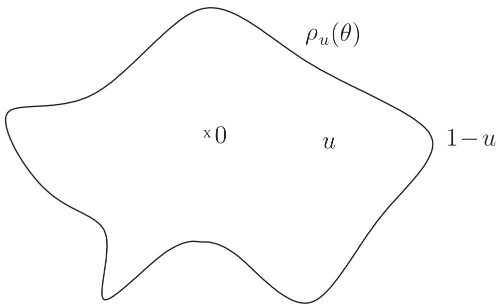

Fig. 1. u-Level isobar.

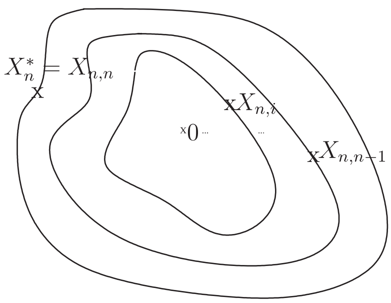

Fig. 2. Isobar-maximum.

Remark 1. We could imagine a more general way to order the sample. For example, by considering an increasing sequence of borelians, according to a criterion to define, and not necessarily related to the Euclidean norm. But it is not the purpose of this paper.

It has been shown in Delcroix and Jacob (1991) that the conditional distribution of R_n^* given Θ_n^* is F_θ^n. More exactly we have the following theorem.

Theorem 1. (*a*) *The variables Θ_n^* and Θ have the same distribution*. (*b*) *The conditional distribution function of R_n^* given $\Theta_n^* = \theta$ is given by $P(R_n^* \leqslant r/\Theta_n^* = \theta) = (F_\theta(r))^n$*.

Proof. See Delcroix and Jacob (1991). □

The second assertion of the previous theorem shows that this multidimensional problem can be reduced to a unidimensional one. That will be done in the next section to find limit laws for the pair (R_n^*, Θ_n^*). In a third section, we will focus on limit laws for the variables R_n^*.

From the previous theorem, we know that the distributions of (R_n^*, Θ_n^*) and (R, Θ) possess the same set of isobars. So it has been possible in Delcroix and Jacob (1991) to give the following definition.

Definition 1. $(X_n^*)_n$ is called stable in probability (resp. a.s.) if and only if there is a sequence $(g_n)_n$ of isobars satisfying

$$R_n^* - g_n(\Theta_n^*) \underset{\text{(a.s.)}}{\overset{\text{P}}{\longrightarrow}} 0. \tag{1}$$

We are well aware that the above definition of multivariate maxima depends on the underlying distribution. This may be crucial in practical situations when given data are to be ordered.[1] To find out more about the limit distribution of this maximum, however, it is not a severe additional obstacle. Note, that even to check the type of extreme value distribution in one dimension affords at least the knowledge of the tail of the underlying distribution.

3. Limit laws for some functionals of (R_n^*, Θ_n^*)

We know the three classical types of limit distributions for unidimensional extreme values and their domains of attraction (Gnedenko, 1943; Leadbetter et al., 1983; Galambos, 1987). Recall the three corresponding distribution functions with the Galambos notations:

(i) $G_{3,0}(x) = \exp(-e^{-x})$ for all x.
(ii) $G_{1,\alpha}(x) = \exp(-x^{-\alpha})$ for $x > 0$, with $\alpha > 0$.
(iii) $G_{2,\alpha}(x) = \exp(-(-x)^\alpha)\mathbf{1}_{(x \leqslant 0)} + \mathbf{1}_{(x > 0)}$, with $\alpha > 0$.

Similarly, in this section, the goal is to find a nondegenerate distribution function $\tilde{H}$ and sequences $a_n(\cdot) > 0$ and $b_n(\cdot)$ such that for all x

$$P\{a_n(\Theta_n^*)(R_n^* - b_n(\Theta_n^*)) \leqslant x\} \overset{\text{w}}{\to} \tilde{H}(x),$$

where $\overset{\text{w}}{\to}$ denotes convergence at continuity points of $\tilde{H}$.

Let $g_n(\theta)$ denote the $(1-(1/n))$ level isobar; that is $g_n(\theta) = F_\theta^{-1}(1-(1/n))$. Under given conditions (Delcroix and Jacob, 1991), we have proved that these isobars stabilize the sequence (R_n^*, Θ_n^*), according to the definition given in the first section.

Theorem 2. *Suppose that for all θ, $\sup_x\{F_\theta(x) < 1\} = +\infty$. If for all θ the distribution function F_θ satisfies*

$$\lim_{t\to\infty} \frac{1 - F_\theta(tx)}{1 - F_\theta(t)} = x^{-\alpha(\theta)} \quad \textit{for all } x > 0,$$

[1] Another work is in progress on the estimation of isobars.

where α *is a strictly positive function, then for all* $x>0$

$$P\{R_n^* \leqslant x g_n(\Theta_n^*)\} \to \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(\mathrm{d}\theta).$$

Proof. From Leadbetter et al. (1983, Theorem 1.6.2), for all θ there exist sequences a_n and b_n depending on θ such that

$$F_\theta^n\left(\frac{x}{a_n(\theta)} + b_n(\theta)\right) \to \exp(-x^{-\alpha(\theta)}).$$

Then from the Dominated Convergence Theorem we have

$$\int_{S^{k-1}} F_\theta^n\left(\frac{x}{a_n(\theta)} + b_n(\theta)\right) P_\Theta(\mathrm{d}\theta) \to \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(\mathrm{d}\theta).$$

But, we know that F_θ^n is the conditional distribution function of R_n^* given $\Theta_n^* = \theta$, so

$$\int_{S^{k-1}} P(R_n^* \leqslant \frac{x}{a_n(\theta)} + b_n(\theta) / \Theta_n^* = \theta) P_\Theta(\mathrm{d}\theta) \to \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(\mathrm{d}\theta).$$

And since Θ_n^* and Θ have the same distribution,

$$P\left(R_n^* \leqslant \frac{x}{a_n(\Theta_n^*)} + b_n(\Theta_n^*)\right) \to \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(\mathrm{d}\theta).$$

Hence

$$P(a_n(\Theta_n^*)(R_n^* - b_n(\Theta_n^*)) \leqslant x) \to \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(\mathrm{d}\theta).$$

Moreover from Leadbetter et al. (1983, Corollary 1.6.3) we know that for all θ $b_n = 0$ and $(a_n(\theta))^{-1} = F_\theta^{-1}(1 - (1/n)) = g_n(\theta)$. □

Example 1. If $F_\theta(x) = (1 - K(\theta) x^{-\alpha(\theta)}) \mathbf{1}_{(x > (K(\theta))^{1/\alpha(\theta)})}$, where α and K are strictly and continuous functions, then for all $x>0$

$$P\{R_n^* \leqslant x (nK(\Theta_n^*))^{1/\alpha(\Theta_n^*)}\} \to \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(\mathrm{d}\theta).$$

For the two others domains of attraction it is possible to give similar limit laws according to the results of Leadbetter et al. (1983). For all θ we use the notation $x_{F_\theta} = \sup(x : F_\theta(x) < 1)$.

Theorem 3. *If for all* θ $x_{F_\theta} < +\infty$ *and if for all* $x>0$ *and for all* θ *there exists a strictly positive function* α *such that*

$$\lim_{h\to 0} \frac{1 - F_\theta(x_{F_\theta} - xh)}{1 - F_\theta(x_{F_\theta} - h)} = x^{\alpha(\theta)}$$

then

$$P\{(x_{F_{\Theta_n^*}} - g_n(\Theta_n^*))^{-1}(R_n^* - x_{F_{\Theta_n^*}}) \leqslant x\} \xrightarrow{\mathrm{w}} \int_{S^{k-1}} (\exp(-(-x)^{\alpha(\theta)}) \mathbf{1}_{(x\leqslant 0)} + \mathbf{1}_{(x>0)}) P_\Theta(\mathrm{d}\theta).$$

Theorem 4. *If for all* θ *there exists a strictly positive function* l_θ *such that for all* θ *and for all* x

$$\lim_{t\to x_{F_\theta}} \frac{1 - F_\theta(t + x l_\theta(t))}{1 - F_\theta(t)} = \mathrm{e}^{-x}$$

then for all x

$$P\{(l_{\Theta_n^*}(g_n(\Theta_n^*)))^{-1}(R_n^* - g_n(\Theta_n^*)) \leqslant x\} \to \int_{S^{k-1}} \exp(-\mathrm{e}^{-x}) P_\Theta(\mathrm{d}\theta) = \exp(-\mathrm{e}^{-x}).$$

Example 2. If $F_\theta(x) = (1 - e^{-\alpha(\theta)x})\mathbf{1}_{(x>0)}$ where α is a strictly positive continuous function, then for all x

$$P\left\{\alpha(\Theta_n^*)\left(R_n^* - \frac{\mathrm{Log}(n)}{\alpha(\Theta_n^*)}\right) \leqslant x\right\} \to \exp(-e^{-x}).$$

Example 3. For a bivariate Gaussian sample with covariance matrix $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 \end{pmatrix}$, the density over $\mathbf{R}^2$ is $f(x,y) = (1/2\pi\sigma\tau)e^{-x^2/2\sigma^2 - y^2/2\tau^2}$; the conditional distribution of R given $\Theta = \theta$ is then given by $F_\theta(x) = (1 - e^{-x^2\Phi(\theta)})\mathbf{1}_{(x>0)}$, where $\Phi(\theta) = \cos^2(\theta)/2\sigma^2 + \sin^2(\theta)/2\tau^2$. Note that in this case, the isobars are also the level curves of the bivariate density. The condition of the previous theorem is satisfied and for all x,

$$P\left\{2\sqrt{\Phi(\Theta_n^*)\,\mathrm{Log}(n)}\left(R_n^* - \sqrt{\frac{\mathrm{Log}(n)}{\Phi(\Theta_n^*)}}\right) \leqslant x\right\} \to \exp(-e^{-x}).$$

4. Limit law of R_n^*

Now, we consider the pairs (R_k, U_k). In this part we use the following remark. By definition, R_n^* is the concomitant of $U_{n,n} = \mathrm{Max}(U_1, \ldots, U_n)$, where for all $1 \leqslant k \leqslant n$, $U_k = F_{\Theta_k}(R_k)$; see e.g. (Galambos, 1987; David, 1981). The two properties of the next theorem will be used in this section.

Theorem 5. *(a) $(U_1 = F_{\Theta_1}(R_1), \ldots, U_n = F_{\Theta_n}(R_n))$ is a sample from the uniform distribution on $[0,1]$. (b) U_k is independant of Θ_k for all $1 \leqslant k \leqslant n$.*

Proof. See Delcroix and Jacob (1991). $\square$

Note that the uniform distribution belongs to the type III—according to the notation of Leadbetter et al. (1983) of extreme values with $\alpha = 1$. It will be useful for the following. From Galambos results on concomitants it is possible to state this limit law for R_n^*.

Theorem 6. *If there exist A_n, $B_n > 0$ such that $\lim_{n\to\infty} P\{R \leqslant A_n + B_n u / U = 1 + z/n\} = T(u,z)$ (for $-n \leqslant z \leqslant 0$ and $u > 0$) is a nondegenerate distribution function, then $\lim_{n\to\infty} P\{R_n^* < A_n + B_n u\} = T(u)$, where*

$$T(u) = \int_{-\infty}^{0} T(u,z)e^z\,dz.$$

Proof. We have

$$P(R_n^* \leqslant r) = \int_{[0,1]} P(R_n^* \leqslant r / U_{n,n} = u) P_{U_{n,n}}(du).$$

But $(R_n^*, U_{n,n})$ is a point of the sample $((R_1, U_1), \ldots, (R_n, U_n))$ and the (R_k, U_k) are independent, so

$$P(R_n^* \leqslant r) = \int_{[0,1]} P(R \leqslant r / U = u) n F_1^{n-1}(u) f_1(u)\,du,$$

where F_1 and f_1 denote the uniform distribution function over $[0,1]$ and the uniform density function, respectively.

Now, put $u = 1 + z/n$. We obtain

$$P(R_n^* \leqslant r) = \int_{[-n,0]} P\left(R \leqslant r / U = 1 + \frac{z}{n}\right)\left(1 + \frac{z}{n}\right)^{n-1} dz.$$

And finally put $r = A_n + B_n u$, hence

$$P(R_n^* \leqslant r) = \int_{[-n,0]} P\left(R \leqslant A_n + B_n u / U = 1 + \frac{z}{n}\right)\left(1 + \frac{z}{n}\right)^{n-1} dz.$$

We conclude with the Dominated Convergence Theorem. $\square$

Remark 2. The condition of the previous theorem may be written more easily with the conditional law of R given $U = u$ if we suppose F_θ one-to-one. Since $U = F_\Theta(R)$, we have

$$P(R \leqslant r/U = u) = P(U \leqslant F_\Theta(r)/U = u).$$

And since the variables U and Θ are independent, we obtain

$$P(R \leqslant r/U = u) = P(u \leqslant F_\Theta(r)).$$

So the condition of the theorem is

$$P\left(1 + \frac{z}{n} \leqslant F_\Theta(A_n + B_n u)\right) \to T(u, z).$$

Of course this condition depends heavily on the form of the conditional distribution of R given $\Theta = \theta$.

Example 4. If $F_\theta(x) = (1 - \mathrm{e}^{\alpha(\theta)x})\mathbf{1}_{(x>0)}$ and $\alpha(\Theta)$ is uniformly distributed over [0, 1] then the previous theorem is true with $A_n = \operatorname{Log} n$, $B_n = \operatorname{Log} n$, the limit law is Pareto.

5. Concluding remarks

It would be possible to consider the pairs (U_k, Θ_k) as done in the last section for the variables U and R, but we obtain only the asymptotic independence between Θ and U and the fact that asymptotically the variables Θ_n^* and Θ have the same distribution, which is weaker than the results stated in Theorems 1 and 5.

References

Abdous, B., Theodorescu, R., 1992. Note on the spatial quantile of a random vector. Statist. Probab. letters 13, 333–336.

Barme-Delcroix, M.F., 1993. Localisation asymptotique des échantillons multidimensionnels. Bulletin de la societé mathématique de Belgique 45, fasc 2 serie B.

Barme-Delcroix, M.F., Brito, M., 2001. Multivariate stability and strong limiting behaviour of intermediate order statistics. J. Multivariate Anal. 79, 157–170.

Barme-Delcroix, M.F., Gather, U., 2002. An isobar surfaces approach to multidimensional outlier-proneness. Extremes 5, 131–144.

Barnett, V., 1976. The ordering of multivariate data. J. Roy. Statist. Soc. A 139 (Part 3), 318–355.

Berlinet, A., Gannoun, A., Matzner-Lober, E., 2001. Asymptotic normality of convergent estimates of conditional quantiles. Statistics 35, 139–169.

David, H.A., 1981. Orders Statistics, Second ed. Wiley, New York.

De Haan, L., Huang, X., 1995. Large quantile estimation in a multivariate setting. J. Multivariate Anal. 53, 247–263.

Delcroix, M.F., Jacob, P., 1991. Stability of extreme value for a multidimensional sample. Statistique et Analyse des Données 16, 1–21.

Einmahl, H.J., Mason, D.M., 1992. Generalized quantile processes. Ann. of Statist. 20 (2), 1062–1078.

Falk, M., Reiss, R.D., 2005. On Pickands coordinates in arbitrary dimensions. J. Multivariate Anal. 92, 426–453.

Fougéres, A.L., 2004. Multivariate extremes. In: Extreme Values in Finance Telecommunications and Environment. Chapman & Hall, CRC Press, London, Boca Raton, pp. 373–388.

Galambos, J., 1987. The Asymptotic Theory of Extreme Order Statistics. Robert E. Krieger publishing company, Malabar, Florida.

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

Geffroy, J., 1961. Localisation asymptotique du polyèdre d'appui d'un échantillon laplacien à k dimension, Publ. Inst. Stat. Univ. Paris 59.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. 44, 423–453.

Heffernan, J.E., Tawn, J.A., 2004. A conditional approach for multivariate extreme values. J. Roy. Statist. Soc. B 66 (Part 3), 497–546.

Leadbetter, M.R., Lindgren, G., Rootzen, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer, Berlin.

Maller, R.R., 1990. Defining extremes and trimming by minimum covering sets. Statist. Processes Appl. 35, 169–180.

Massé, J-C., Theodorescu, R., 1994. Halfplane trimming for bivariate distributions. J. Multivariate Anal. 48, 188–202.

Nadarajah, S., Anderson, C.W., Tawn, J.A., 1998. Ordered multivariate extremes. J. Roy. Statist. Soc. B 60 (Part 2), 473–496.

Schlatler, M., Tawn, J.A., 2003. A dependance measure for multivariate and spatial extreme values: properties and inference. Biometrika 90, 139–156.

Serfling, R., 2002. Quantile functions for multivariate analysis: approaches and applications. Statistica Neerlandica 56 (2), 214–232.

Smith, R.L., Weissman, I., 1996. Characterization and estimation of the multivariate extremal index. Technical Report, University of North Carolina.

Zhang, Z., Smith, R.L., 2004. The behavior of multivariate maxima of moving maxima processes. J. Appl. Probab. 41, 1113–1123.