



Limit laws for multidimensional extremes

Marie-Francoise Barne-Delcroix, Ursula Gather

► To cite this version:

Marie-Francoise Barne-Delcroix, Ursula Gather. Limit laws for multidimensional extremes. *Statistics and Probability Letters*, 2009, 77 (18), pp.1750. 10.1016/j.spl.2007.04.014 . hal-00580092

HAL Id: hal-00580092

<https://hal.science/hal-00580092>

Submitted on 26 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Limit laws for multidimensional extremes

Marie-Francoise Barne-Delcroix, Ursula Gather

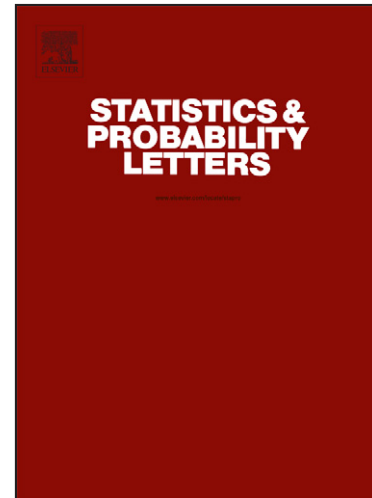
PII: S0167-7152(07)00161-7
DOI: doi:10.1016/j.spl.2007.04.014
Reference: STAPRO 4664

To appear in: *Statistics & Probability Letters*

Received date: 27 June 2005
Revised date: 27 February 2007
Accepted date: 10 April 2007

Cite this article as: Marie-Francoise Barne-Delcroix and Ursula Gather, Limit laws for multidimensional extremes, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.04.014

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

Limit laws for multidimensional extremes.

MARIE-FRANCOISE BARME-DELCROIX

Laboratoire Painlevé, CNRS - UMR 8524

Université des Sciences et Technologies de Lille

F - 59655 Villeneuve d'Ascq cedex

URSULA GATHER

Fachbereich Statistik

Universitat Dortmund

D - 44221 Dortmund

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.

Key-words : isobars , order statistics, extreme values, stability, limit laws, concomitant.

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

1 Introduction

Today many papers on extreme values deal with multivariate extreme values, see [Nadarajah, Anderson, Tawn 98], [Schlatler, Tawn 03], [Fougères 04], [Heffernan, Tawn 04], [Falk, Reiss 05]; see also for recent developments about max-stable processes and multivariate extremal index, [Smith, Weissman 96], [Zhang, Smith 04]. Unfortunately, ordering multivariate data can be done in various ways and many definitions have been proposed e.g. [Barnett 76], [Maller 90]. Some papers of [Abdous, Theodorescu 92], [Einmahl, Mason 92], [Massé, Theodorescu 94], [De Haan, Huang 95], and more recently, of [Berlinet, Gannoun, Matzner-Lober 01], [Serfling 02] develop the notion of multivariate quantiles. In the classical scheme (cartesian coordinates) the multivariate variables are ordered coordinate by coordinate - see for example [Galambos 87] 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, Jacob 91]. 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 93], 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 61], [Geffroy 58]. By a unidimensional approach, stability, strong behaviour and outlier-proneness properties have been explored for this new notion of extreme value in [Barme, Brito 01] and [Barme, Gather 02]. In this paper, the goal is to find limit laws for such mul-

tidimensional extremes. Through the use of the isobar surfaces we will see that the classical results -[Gnedenko 43] - about limit laws for unidimensional extremes are still efficient in this multidimensional context.

2 Preliminaries

We consider random variables defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and 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 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 $X = (R, \Theta) = (\|X\|, \frac{X}{\|X\|}) \in \mathbf{R}^k$, 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 Figure 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 $1 \leq k \leq n$ there is almost surely a unique u_k -level isobar from the distribution of R given $\Theta = \theta$ which contains $(R_k, \Theta_k) = (\|X_k\|, \frac{X_k}{\|X_k\|})$. 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 $\max_{1 \leq k \leq n} u_k$. See Figure 2. The

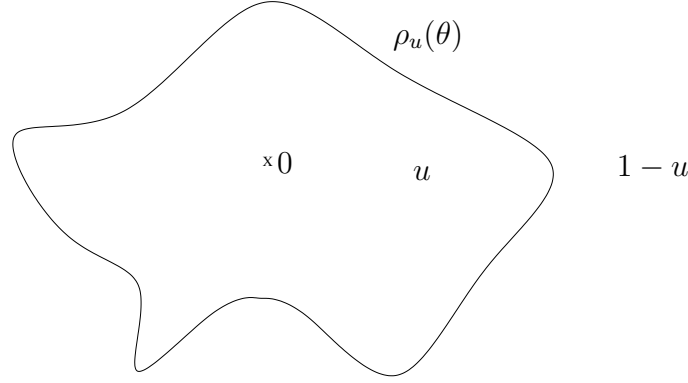


Figure 1: u -level isobar

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

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

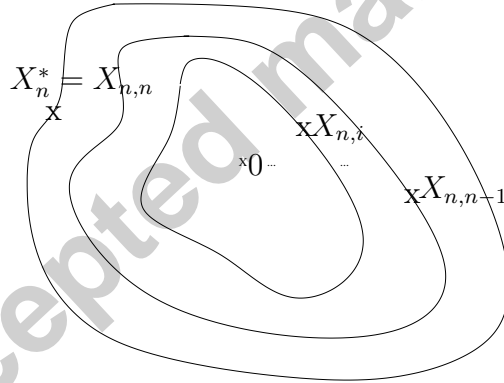


Figure 2: isobar-maximum

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, Jacob 91] 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 $P(R_n^* \leq r / \Theta_n^* = \theta) = (F_\theta(r))^n$.*

Proof : see [Delcroix, Jacob 91]. ■

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, Jacob 91] 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

$$(1) \quad R_n^* - g_n(\Theta_n^*) \xrightarrow[(a.s.)]{P} 0 .$$

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 ordered¹. 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.

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

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 43], [Leadbetter 83], [Galambos 87]. 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) $G_{2,\alpha}(x) = \exp(-(-x)^\alpha) \mathbf{1}_{(x \leq 0)} + \mathbf{1}_{(x > 0)}$, 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

$$P\{a_n(\Theta_n^*)(R_n^* - b_n(\Theta_n^*)) \leq x\} \xrightarrow{w} \tilde{H}(x)$$

where \xrightarrow{w} denotes convergence at continuity points of \tilde{H} .

Let $g_n(\theta)$ denote the $(1 - \frac{1}{n})$ -level isobar; that is $g_n(\theta) = F_\theta^{-1}(1 - \frac{1}{n})$. Under given conditions, [Delcroix, Jacob 91], 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*

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

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

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

Proof : From Theorem 1.6.2 of [Leadbetter 83], for all θ there exist sequences a_n and b_n depending on θ such that

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

Then from the Dominated Convergence Theorem we have

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

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

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

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

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

Hence

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

Moreover from Corollary 1.6.3. of [Leadbetter 83] we know that for all θ $b_n = 0$ and $(a_n(\theta))^{-1} = F_\theta^{-1}(1 - \frac{1}{n}) = g_n(\theta)$. ■

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

$$P\{R_n^* \leq x(nK(\Theta_{n,n}))^{1/\alpha(\Theta_n^*)}\} \rightarrow \int_{S^{k-1}} \exp(-x^{-\alpha(\theta)}) P_\Theta(d\theta).$$

For the two others domains of attraction it is possible to give similar limit laws according to the results of [Leadbetter 83]. 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*

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

then

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

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

$$\lim_{t \rightarrow x_{F_\theta}} \frac{1 - F_\theta(t + xl_\theta(t))}{1 - F_\theta(t)} = e^{-x}$$

then for all x

$$P\left\{\left(l_{\Theta_n^*}(g_n(\Theta_n^*))\right)^{-1} (R_n^* - g_n(\Theta_n^*)) \leq x\right\} \rightarrow \int_{S^{k-1}} \exp(-e^{-x}) P_\Theta(d\theta) = \exp(-e^{-x}).$$

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*

$$P\left\{\alpha(\Theta_n^*) \left(R_n^* - \frac{\text{Log}(n)}{\alpha(\Theta_n^*)}\right) \leq x\right\} \rightarrow \exp(-e^{-x}).$$

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 $f(x, y) = \frac{1}{2\pi\sigma\tau} e^{-\frac{x^2}{2\sigma^2} - \frac{y^2}{2\tau^2}}$; 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 $\Phi(\theta) = \frac{\cos^2(\theta)}{2\sigma^2} + \frac{\sin^2(\theta)}{2\tau^2}$. 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 ,

$$P\left\{2\sqrt{\Phi(\Theta_n^*)\text{Log}(n)} \left(R_n^* - \sqrt{\frac{\text{Log}(n)}{\Phi(\Theta_n^*)}}\right) \leq x\right\} \rightarrow \exp(-e^{-x}).$$

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 $U_{n,n} = \text{Max}(U_1, \dots, U_n)$ where for all $1 \leq k \leq n$, $U_k = F_{\Theta_k}(R_k)$; see e.g. [Galambos 87], [David 81]. The two properties of the next theorem will be used in this section.

Theorem 5 *a) $(U_1 = F_{\Theta_1}(R_1), \dots, U_n = F_{\Theta_n}(R_n))$ is a sample from the uniform distribution on $[0, 1]$.*

b) U_k is independant of Θ_k for all $1 \leq k \leq n$.

Proof : see [Delcroix, Jacob 91]. ■

Note that the uniform distribution belongs to the type III -according to the notation of [Leadbetter 83]- 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 $\lim_{n \rightarrow \infty} P\{R \leq A_n + B_n u / U = 1 + z/n\} = T(u, z)$ (for $-n \leq z \leq 0$ and $u > 0$) is a non degenerate distribution function, then $\lim_{n \rightarrow \infty} P\{R_n^* < A_n + B_n u\} = T(u)$ where*

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

Proof : We have

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

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

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

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

Now, put $u = 1 + \frac{z}{n}$. We obtain

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

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

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

We conclude with the Dominated Convergence Theorem. ■

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*

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

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

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

So the condition of the theorem is

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

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

Example 4 *If $F_\theta(x) = (1 - e^{\alpha(\theta)x})\mathbf{1}_{(x>0)}$ and $\alpha(\Theta)$ is uniformly distributed over $[0, 1]$ then the previous theorem is true with $A_n = \text{Log } n$, $B_n = \text{Log } n$, 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, Theodorescu 92] ABDIOUS B. and THEODORESCU R. (1992) - Note on the spatial quantile of a random vector. *Statistics and Probability letters* 13, 333–336.
- [Barne 93] BARME-DELCROIX M.F. (1993) - Localisation asymptotique des échantillons multidimensionnels. *Bulletin de la société mathématique de Belgique, vol. 45, fasc 2 serie B.*
- [Barne, Brito 01] BARME-DELCROIX M.F. and BRITO M. (2001) - Multivariate stability and strong limiting behaviour of intermediate order statistics. *Journal of Multivariate Analysis* 79, 157–170.

- [Barne, Gather 02] BARME-DELCROIX M.F. and GATHER U. (2002) - An isobar surfaces approach to multidimensional outlier-proneness. *Extremes*, 5, 131–144.
- [Barnett 76] BARNETT V. (1976) - The Ordering of Multivariate Data. *J. R. Statist. Soc. A*, 139 Part 3, 318–355.
- [Berlinet, Gannoun, Matzner-Lober 01] BERLINET A., GANNOUN A., MATZNER-LOBER E. (2001) - Asymptotic normality of convergent estimates of conditional quantiles. *Statistics* 35, 139–169.
- [David 81] DAVID H.A. (1981) - Orders Statistics. 2nd edition. Wiley, New-York.
- [De Haan, Huang 95] De Haan L. and Huang X. (1995) - Large Quantile Estimation in a Multivariate Setting. *Journal of Multivariate Analysis* 53, 247–263.
- [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.

- [Einmahl, Mason 92] EINMAHL H.J. and MASON D.M. (1992) - Generalized quantile processes. *The annals of Statistics*, 20, 2, 1062–1078.
- [Fougères 04] FOUGERES A.L. (2004) - Multivariate extremes, *In Extreme values in Finance, Telecommunications and environment*, 373–388, Chapman and Hall, CRC Press, London.
- [Galambos 87] GALAMBOS J. (1987) - The asymptotic Theory of Extreme Order Statistics, *Robert E. Krieger publishing company, Malabar, Florida*.
- [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, no59.
- [Gnedenko 43] GNEDENKO B.V. (1943) - Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423–453.

- [Heffernan, Tawn 04] HEFFERNAN J.E., and TAWN J.A. (2004)- A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66, Part 3, 497–546.
- [Leadbetter 83] LEADBETTER M.R., LINDGREN G.,ROOTZEN H. (1983) - Extremes and related properties of random sequences and processes, *Springer Verlag*.
- [Maller 90] MALLER R.R. (1990) -Defining Extremes and Trimming by Minimum Covering Sets. *Statistic Processes and their Applications* 35, 169–180.
- [Massé, Theodorescu 94] Massé J-C. and Theodorescu R. (1994) - Halfplane Trimming for bivariate distributions. *Journal of Multivariate Analysis* 48, 188–202.
- [Nadarajah, Anderson, Tawn 98] NADARAJAH S., ANDERSON C.W., TAWN J.A. (1998) - Ordered multivariate extremes. *J. R. Statist. Soc B*, 60, Part 2, 473–496.
- [Falk, Reiss 05] FALK M. and REISS R.D. (2005) - On Pickands coordinates in arbitrary dimensions. *Journal of Multivariate Analysis*, 92, 426–453.

- [Schlatler, Tawn 03] Schlatler M. and TAWN J.A. (2003) -A dependence measure for multivariate and spatial extreme values : properties and inference. *Biometrika*, 90, 139–156.
- [Serfling 02] SERFLING R. (2002) - Quantile functions for multivariate analysis : approaches and applications. *Statistica Neerlandica*, 56, nr 2, 214–232.
- [Smith, Weissman 96] SMITH R.L. and WEISSMAN I. (1996) -Characterization and Estimation of the Multivariate Extremal Index. *Tech. Rep., University of North Carolina*.
- [Zhang, Smith 04] ZHANG Z. and SMITH R.L. (2004) - The behavior of multivariate maxima of moving maxima processes. *J. Appl. Probab.* 41, 1113–1123.