

55376
1992
1

55376
1992
1

N° d'ordre : 840

THÈSE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES DE LILLE

pour obtenir

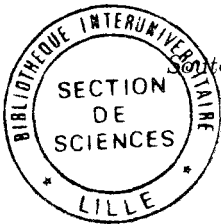
LE TITRE DE DOCTEUR DE L'UNIVERSITÉ

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

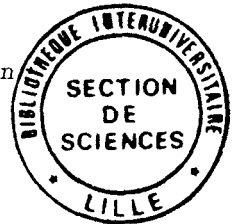
par

DELCROIX épouse BARME Marie-Françoise

PROBLÈMES DE VALEURS EXTRÊMES
POUR DES
ÉCHANTILLONS MULTIDIMENSIONNELS



Soutenue le 15 Janvier 1992 devant la Commission d'Examen



Président et Rapporteur : J. GEFFROY, Université de Paris VI

Rapporteur : G. HAIMAN, Université de Lille

Examineurs : U. GATHER, Université de Dortmund

P. JACOB, Université de Lille

A mon père,
A Laurent.

Je tiens à exprimer ma gratitude à :

Monsieur le Professeur J. Geffroy qui me fait l'honneur de présider le jury de cette thèse et d'en être le rapporteur.

Monsieur le Professeur G. Haiman, rapporteur de cette thèse, pour sa lecture approfondie de mon travail, pour ses critiques constructives et ses conseils.

Madame le Professeur U. Gather qui a eu l'extrême obligeance de participer au jury de cette thèse.

Monsieur le Professeur P. Jacob qui a guidé mon activité de recherche, me prodiguant sans cesse des conseils précieux et bienveillants avec la plus grande disponibilité.

Je remercie également toutes les personnes qui ont participé à la réalisation matérielle de ce travail, en particulier l'équipe du secrétariat scientifique ainsi que mes collègues Daniel Flipo et Charles Suquet qui avec beaucoup de patience m'ont aidée à utiliser les techniques du traitement de texte.

Enfin, la constante bonne humeur de mes collègues les plus proches a aussi contribué à l'aboutissement de cette thèse, je les en remercie chaleureusement.

Introduction

Le problème de l'estimation du support d'un processus ponctuel ou d'une loi de probabilité, étant donné un échantillon, a souvent été traité en supposant le support convexe. L'estimateur utilisé est alors l'enveloppe convexe du nuage des points de l'échantillon ; ce qui associe à ce problème d'estimation du support l'étude de cette enveloppe convexe.

Les premiers travaux relatifs à cette question sont ceux de Geffroy en 1961, [16], qui étudie le comportement asymptotique du polyèdre d'appui d'un échantillon laplacien de \mathbf{R}^k . Fisher, [13], en 1965, considère le cas d'une loi produit de lois de probabilité sur \mathbf{R} ; Renyi et Sulanke, [31], en 1964, Raynaud, [30], en 1965, Efron, [12], en 1965 étudient l'espérance mathématique du volume, du nombre de faces, d'arêtes ou de sommets de l'enveloppe convexe d'un échantillon de \mathbf{R}^2 ou de \mathbf{R}^3 . En 1973, Guilbart, [21], étudie la continuité de l'application qui à une probabilité de \mathbf{R}^3 associe l'enveloppe convexe fermée de son support.

Des travaux divers ont été réalisés sur le comportement asymptotique de l'enveloppe convexe d'un échantillon de \mathbf{R}^k : Ripley et Rasson, [32], en 1977 donnent un estimateur sans biais du support, supposé convexe, d'un processus de Poisson homogène ; Buchta en 1984, [4], Dwyer en 1988, [10], Barany et Larman en 1988, [1], Van Wel en 1989, [34], étudient le comportement asymptotique de l'enveloppe convexe d'un échantillon de la loi uniforme à support convexe ; en 1990, Massé, [27], en généralisant un résultat de Barany et Larman, [1], obtient deux majorations, indépendantes de la loi, de la vitesse de convergence vers 1 de la probabilité d'un dilaté de l'enveloppe convexe d'un échantillon vers l'enveloppe convexe du support ; Eddy et Gale en 1981, [11], dans le cas d'une loi à symétrie sphérique, puis Brozius et De Haan en 1987, [3], dans un cadre plus général étudient la loi limite de la fonction support de l'enveloppe convexe d'un échantillon de \mathbf{R}^k . En 1987, Davis, Mulrow et Resnick, [6], montrent que sous certaines conditions de régularité de la fonction de répartition commune de n variables de \mathbf{R}^2 indépendantes et de même loi, l'enveloppe convexe de l'échantillon converge en loi vers l'enveloppe convexe d'un processus de Poisson.

Le cas où le support est non convexe a été aussi envisagé. En 1964, Geffroy, [17], considère un support du type $\{(x, y) \in \mathbf{R}^2; 0 \leq x \leq 1; 0 \leq y \leq f(x)\}$ où f est une fonction continue ; Chevallier, [5], en 1976 considère le cas où f est de classe C^1 et Jacob, [24] en 1984 étudie le cas où f présente des discontinuités de première espèce. Gensbittel, [18], reprend en 1979 les études de Geffroy et de Chevallier en les adaptant au cas des répartitions ponctuelles

aléatoires. Enfin, en 1989 Jacob et Abbar, [25], proposent une estimation du support d'un processus ponctuel de Cox défini en coordonnées polaires dans \mathbf{R}^2 par : $\{(\rho, \theta) : 0 \leq \theta \leq 2\pi; 0 \leq \rho \leq \Phi(\theta)\}$.

Certains auteurs ont travaillé à partir d'ensembles « minimaux » contenant un échantillon : Silverman et Titterington en 1990, [33], ont développé un algorithme pour trouver l'ellipse de plus petite surface recouvrant un ensemble de points donné dans le plan ; Maller en 1990 [26], propose une définition des extrêmes d'un échantillon de vecteurs de \mathbf{R}^k , basée sur le plus petit convexe contenant l'échantillon.

L'étude qui est présentée ici se situe dans la ligne des travaux de Gefroy, [15], [16], sur la localisation asymptotique du polyèdre d'appui d'un échantillon laplacien et sur la stabilité des extrêmes d'un échantillon de \mathbf{R} . Il s'agit, étant donné un échantillon

$$\{X_1, \dots, X_n\}$$

de \mathbf{R}^k dont la loi de chaque variable est de support non borné, de rechercher une suite de boréliens, non nécessairement convexes, de \mathbf{R}^k , croissante avec n , qui contienne l'échantillon « au plus près ». Pour cela, nous avons étudié avec P. Jacob, [7], une nouvelle notion de valeur maximale d'un échantillon (X_1, \dots, X_n) :

$$X_n^* = (R_n^*, \Theta_n^*),$$

basée sur la représentation « polaire »

$$(R_1, \Theta_1), \dots, (R_n, \Theta_n)$$

des variables $\{X_i, i = 1 \dots n\}$.

L'étude de la stabilité de cette plus grande valeur X_n^* permet d'obtenir des résultats de localisation asymptotique des extrêmes de l'échantillon autour de courbes appelées isobares, définies à partir de la fonction de répartition conditionnelle de R étant donné $\Theta = \theta$ des variables $X_i = (R_i, \Theta_i)$.

Actuellement, la théorie des valeurs extrêmes concerne souvent des variables non identiquement distribuées, [23], [2], ou des variables dépendantes, [22], ou encore des variables multidimensionnelles, [26] ; cependant, de récents articles sur les outliers, [14], [20], [28], [29], donnent un nouvel intérêt à la notion déjà ancienne de stabilité, [15], [19].

Plus précisément, le plan de ce travail est le suivant : dans le chapitre 1, on reprend l'aspect géométrique de la question tel qu'il a été envisagé par Geffroy dans [16]. On considère :

$$\{X_1 = (R_1, \Theta_1), \dots, X_n = (R_n, \Theta_n)\}$$

un échantillon de la loi normale $N(O, Id_{\mathbf{R}^2})$. Etant donné un deuxième échantillon de variables positives (U_1, \dots, U_n) , indépendant du premier, on définit pour $1 \leq i \leq n$, la variable :

$$X'_i = U_i \cdot X_i.$$

On donne des exemples où cette perturbation de l'échantillon $\{X_1, \dots, X_n\}$ permet encore, comme pour la loi normale, [16], de localiser asymptotiquement l'enveloppe convexe de l'ensemble des points $\{X'_1, \dots, X'_n\}$ et d'obtenir des résultats semblables à ceux de Geffroy, [8]. On peut alors étudier le cas général où la déformation imposée à l'échantillon initial n'est pas isotrope. Aussi, l'objet des chapitres 2 et 3 est de trouver une méthode de localisation asymptotique d'un échantillon quelconque de \mathbf{R}^k .

Le chapitre 2 est consacré à l'étude de la stabilité de la valeur maximale X_n^* d'un échantillon $\{X_1 = (R_1, \Theta_1), \dots, X_n = (R_n, \Theta_n)\}$ de \mathbf{R}^k . La définition que l'on propose pour X_n^* a pour origine la remarque suivante : étant donné $\{Y_1, \dots, Y_n\}$ un échantillon de \mathbf{R} , dont chaque variable a pour fonction de répartition F , si Y_n^* désigne la valeur maximale de cet échantillon, $F(Y_n^*)$ est alors la valeur maximale de $\{F(Y_1), \dots, F(Y_n)\}$. Ainsi, si F_θ est la fonction de répartition conditionnelle de R étant donné $\Theta = \theta$, nous définissons la valeur maximale de l'échantillon comme l'observation X_n^* qui maximise $F_\theta(R)$. L'intérêt de cette définition réside dans le fait qu'elle n'est pas basée seulement sur l'éloignement des observations par rapport à l'origine O de \mathbf{R}^k , mais sur leur probabilité d'être à une certaine distance de l'origine. On peut alors définir une notion de stabilité pour X_n^* qui ne dépend que de la loi considérée. Bien sûr, toutes les distributions ne vérifient pas les propriétés de stabilité. Pour ces distributions nous avons recours à une notion plus faible. L'idée est de remplacer chaque observation $X = (R, \Theta)$ par une variable :

$$X_\varphi = (\varphi(R), \Theta),$$

où φ est une fonction convenable, afin d'obtenir les propriétés de stabilité pour la variable X_φ .

Dans le chapitre 3, nous considérons l'aspect géométrique de la stabilité et étudions la répartition de l'ensemble des points de l'échantillon dans un domaine non convexe limité par des courbes isobares. Ceci permet d'étendre à des ensembles non convexes les résultats de Geffroy, [15], sur le comportement asymptotique d'un échantillon de \mathbf{R}^k , [9].

Dans le chapitre 4, nous utilisons les résultats du chapitre 2 pour définir dans \mathbf{R}^k , et en donner quelques propriétés, les notions de distributions « outlier-resistant » ou « outlier-prone » afin de prolonger les définitions données par Green, [14], [20].

Références

- [1] BARANY I. and LARMAN D.G. (1988)- Convexes bodies, economics cap coverings random polytopes. *Mathematika* 35, 274–291.
- [2] BEIRLANT J. and TEUGELS J.L. (1991)- Asymptotics for compounded sums of extremes. *à paraître*.
- [3] BROZIUS H. and DE HAAN L. (1987)- On limiting laws for the convex hull of a sample. *J. Appl. Prob.* 24, 852–862.
- [4] BUCHTA C. (1984)- Stochastische approximation konvexer polygone. *Z. Wahrschein.* 67, 283–304.
- [5] CHEVALLIER J. (1976)- Estimation du support et du contenu du support d'une loi de probabilité. *Ann. Inst. Henri Poincaré*, section B, vol. 12, 4, 339–364.
- [6] DAVIS R., MULROW E., RESNICK S. (1987)- The convex hull of a random sample in R^2 . *Comm. statist. stochastic models*, 3(1), 1–27.
- [7] DELCROIX M.F et JACOB P. (1992)- Stability of extreme value for a multidimensional sample. *Statistique et Analyse des Données*, Juin 92, 1–21.
- [8] DELCROIX M.F (1992)- Un exemple de localisation asymptotique du polyèdre d'appui d'un échantillon de R^2 . *Accepté par la revue de l'I.S.U.P.*.
- [9] DELCROIX M.F (1992)- Localisation asymptotique des échantillons multidimensionnels. *A soumettre*.
- [10] DWYER R.A. (1988)- On the convex hull of random points in a polytope. *J. Appl. Prob.* 25, 688–699.
- [11] EDDY and GALE (1981)- The convex hull of a spherically symmetric sample. *Adv. Appl. Prob.* 13, 751–763.
- [12] EFRON B. (1965)- The convex hull of a random set of points. *Biometrika*, 52, 331–343.

- [13] FISHER L. (1966)- The convex hull of a sample. *Bull. Am. Math. Soc.* 72, 555-558.
- [14] GATHER U. and RAUHUT B. (1990)- The outlier behaviour of probability distributions. *Journal of Statistical Planning and Inference* 26, 237-256.
- [15] GEFFROY J. (1958-1959)- Contribution à la théorie des valeurs extrêmes. *Pub. Inst. Stat. Univ. Paris*, 1958 (fasc 4) et 1959 (fasc 1).
- [16] GEFFROY J. (1961)- Localisation asymptotique du polyèdre d'appui d'un échantillon laplacien à k dimensions. *Pub. Inst. Univ. Paris*, 1961, n° 59.
- [17] GEFFROY J. (1964)- Sur un problème d'estimation géométrique. *Pub. Inst. Univ. Paris*, 1964, 13, 191-200.
- [18] GENSBITTEL M.H. (1979)- Contribution à l'étude statistique de répartitions ponctuelles aléatoires. *Thèse, Université Pierre et Marie Curie, Paris.*
- [19] GNEDENKO B.V. (1943)- Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423-453.
- [20] GREEN R.F. (1976)- Outlier-prone and outlier-resistant distributions. *J. Amer. Statist. Assoc.* 71, 502-505.
- [21] GUILBART C. (1973)- Etude de la continuité de l'application H qui à toute mesure de probabilité définie sur R^n fait correspondre l'enveloppe convexe fermée du support de cette mesure. *C. R. Acad. Sci. Paris*, 277, A, 999-1002.
- [22] HAIMAN G. and PURI M. (1990)- A strong invariance principle concerning the J -upper order statistics for stationary m -dependent sequences. *Journal of Statistical Planning and Inference.*
- [23] HAIMAN G. and PURI M. (1990)- A strong invariance principle concerning the J -upper order statistics for stationary gaussian sequences. *A paraître dans Annals of Probability.*

- [24] JACOB P. (1984)- Estimation du contour discontinu d'un processus ponctuel sur le plan. *Pub. Inst. Univ. Paris*, 24, 3-4, 1-25.
- [25] JACOB P. and ABBAR H. (1988)- Estimating the edge of a Cox process area. *Cahiers du Centre d'Etudes de Recherche Opérationnelle*, vol. 31, n°3-4.
- [26] MALLER R.A. (1990)- Defining extremes and trimming by minimum covering sets. *Stoch. Proc. and their Appl.* 35, 169-180.
- [27] MASSE B. (1990)- Concentration et dispersion sur les convexes compacts d'une loi de probabilité multidimensionnelle. Problèmes statistiques associés. *Thèse USTL Flandres Artois*.
- [28] MATHAR R. (1989)- The outlier behaviour of distributions and the decay of power of optimal tests at decreasing level. *Statistics*, 20 (2), 247-254.
- [29] MUNOZ-GARCIA J., MORENO-REBOLLO J.L. and PASCUAL-ACOSTA A. (1990) - Outliers: a formal approach. *International Statistical Review* 58, 3.
- [30] RAYNAUD H. (1965)- Sur le comportement asymptotique de l'enveloppe convexe d'un nuage de points tirés au hasard dans R^n . *C. R. Acad. Sci. Paris* 262, A, 627-623.
- [31] RENYI A. und SULANKE R. (1964)- Uber die konvexe hülle von n zufalling gewählten punkten 2. *Z. Wahrschein.* 3, 138-147.
- [32] RIPLEY B.D. and RASSON J.P. (1977)- Finding the edge of Poisson forest. *J. Appl. Prob.* 14 ,483-491.
- [33] SILVERMAN B.W. and TITTERINGTON D.H. (1990)- Minimum covering ellipses. *SIAM. J. Sci. Stat. Comput.* vol.1, n° 4.
- [34] VAN WEL (1989)- The convex hull of an uniform sample from the interior of a simple d-polytope.

**Chapitre 1 : Exemples de localisation
asymptotique du polyèdre d'appui d'un
échantillon de \mathbf{R}^2**

I - INTRODUCTION

L'étude de l'enveloppe convexe d'un échantillon de \mathbf{R}^k a été envisagée sous de nombreux aspects. Geffroy (1958, 1961), Raynaud (1965 - [13]), Renyi et Sulanke (1964 - [14]), Efron (1965 - [6]), Fisher ([8]), semblent être les premiers à s'être intéressés à ce sujet. Plus récemment, Eddy et Gale (1981 - [5]), dans le cas d'une loi à symétrie sphérique, puis Brozius et de Haan (1987 - [1]) dans un cadre plus général, ont étudié la loi limite de la fonction de support de l'enveloppe convexe d'un échantillon de \mathbf{R}^k . Nous allons ici reprendre l'aspect géométrique de la question tel qu'il a été introduit par Geffroy. En 1961, Geffroy [10], a considéré une distribution normale dans \mathbf{R}^k : en notant Σ_n l'ellipsoïde lieu des points où la densité est $\frac{1}{n}$ et Δ_n l'écart maximum entre Σ_n et la frontière H_n de l'enveloppe convexe de l'échantillon ($\Delta_n = d^*(H_n, \Sigma_n) = \sup_{M \in H_n} \{d(M, \Sigma_n)\}$), il a obtenu $\Delta_n \xrightarrow{P.S.} 0$. En particulier en choisissant une loi normale de densité $f(M) = \pi^{-k/2} \exp(-M^2)$, il a montré que $d^*(H_n, S_n) \xrightarrow{P.S.} 0$ où S_n est la sphère de centre 0 de rayon $\sqrt{\text{Log } n}$. Nous nous proposons ici d'étendre ces résultats à d'autres échantillons, obtenus par perturbation d'un échantillon gaussien, avant d'envisager une étude plus générale dans les chapitres 2 et 3. Pour cela, quelques rappels sur la stabilité d'une suite de variables aléatoires réelles sont nécessaires. La référence pour ces résultats est Geffroy, [9].

II - STABILITÉ PRESQUE SURE DE LA VALEUR MAXIMALE D'UN ÉCHANTILLON DE VARIABLES ALÉATOIRES RÉELLES

Soit $(Z_n^*)_n$ une suite de variables aléatoires réelles. On dit que $(Z_n^*)_n$ est stable presque sûrement si et seulement s'il existe une suite déterministe $(a_n)_n$ telle que :

$$Z_n^* - a_n \xrightarrow{p.s.} 0,$$

ou encore si pour tout $\varepsilon > 0$,

$$Z_n^* \underset{ps}{\ll} a_n + \varepsilon \quad \text{et} \quad a_n - \varepsilon \underset{ps}{\ll} Z_n^*,$$

où selon les notations de [9], $Z_n^* \underset{ps}{\ll} a_n + \varepsilon$ traduit : $P\{\text{Lim inf}\{Z_n^* < a_n + \varepsilon\}\} = 1$.

Considérons le cas particulier où $(Z_n^*)_n$ est le maximum de n variables aléatoires réelles indépendantes et de même loi, Z_1, \dots, Z_n ; si f désigne la densité commune des Z_i et si F désigne la fonction de répartition des Z_i , une condition suffisante pour la stabilité presque sûre de $Z_n^* = \max\{Z_1, \dots, Z_n\}$ est donnée par le théorème 53 de [9] rappelé ci-dessous :

Si $G = 1 - F$ et f vérifient la condition $\lim_{x \rightarrow +\infty} \frac{f(x)}{G(x) \text{Log} |\text{Log} G(x)|} = +\infty$, $(Z_n^*)_n$ est stable presque sûrement.

Remarque 1.- On utilisera aussi dans les chapitres suivants la notion de stabilité en probabilité et en particulier le théorème de Gnedenko, [9].

Remarque 2.- La notion de stabilité est plus restrictive que les notions qui permettent de donner les différentes lois limites pour le maximum de n variables aléatoires réelles indépendantes. Cependant, en utilisant une notion affaiblie de la stabilité, on peut obtenir des propriétés de stabilité pour la plupart des lois usuelles. Ce travail est fait dans le chapitre 2 pour des variables multidimensionnelles.

On utilise aussi la définition suivante :

Soient $(b_n)_n$ une suite réelle, $(V_n)_n$ une suite de variables aléatoires réelles et F_{V_n} la fonction de répartition de V_n , pour tout n . Comme dans [9], on dit que $(b_n)_n$ minore presque complètement $(V_n)_n$ si :

$$\sum_n F_{V_n}(b_n) < +\infty.$$

On note $b_n \underset{p.co.}{\ll} V_n$.

a) **Lemme 1.**- Pour tout entier $s > 0$, et pour tout $\varepsilon > 0$, la suite $\alpha_n = [\text{Log } n]^s$ vérifie :

$$(15) \quad b_n - \varepsilon \underset{\text{p.co.}}{\ll} \bar{Z}_n^{\alpha_n}$$

où b_n est donné par (10).

■ On note $\bar{F}_{n,\alpha}(z)$ la fonction de répartition de \bar{Z}_n^α . Comme dans [10], on obtient :

$$(16) \quad \bar{F}_{n,\alpha}(z) = 1 - B(\alpha, n - \alpha + 1) \int_{+\infty}^{G(z)} (1 - \xi)^{n-\alpha} \xi^{\alpha-1} d\xi,$$

où

$$B(\alpha, n - \alpha + 1) = \frac{n!}{(n - \alpha)! (\alpha - 1)!}$$

Si bien que, pour tout $\varepsilon > 0$,

$$(17) \quad q_n = P\{\bar{Z}_n^{\alpha_n} \leq b_n - \varepsilon\} = B(\alpha_n, n - \alpha_n + 1) \int_{G_n}^1 (1 - \xi)^{n-\alpha_n} \xi^{\alpha_n-1} d\xi,$$

où $G_n = G(b_n - \varepsilon)$.

Sur $[0, 1]$, la fonction $h(\xi) = (1 - \xi)^{n-\alpha_n} \xi^{\alpha_n-1}$ a un maximum atteint en $\frac{\alpha_n-1}{n-1}$ tel que pour n assez grand :

$$(18) \quad \text{Log}\left(\frac{\alpha_n-1}{n-1}\right) = -\text{Log } n + s \text{Log } \text{Log } n + O(1),$$

puisque $\alpha_n = [\text{Log } n]^s$.

Par ailleurs, d'après (8), pour n suffisamment grand,

$$(19) \quad \text{Log } G_n = -\text{Log } n + \varepsilon \sqrt{2 \text{Log } n} - \text{Log } \text{Log } n + O(1).$$

La comparaison de (18) et (19) montre que pour n assez grand, $G_n > \frac{\alpha_n-1}{n-1}$; h étant décroissante sur $[G_n, 1]$, il vient :

$$(20) \quad q_n < B(\alpha_n, n - \alpha_n + 1) \cdot h(G_n) = q'_n.$$

Il suffit à présent de montrer que la série $\Sigma q'_n$ converge. On évalue pour cela $\text{Log } q'_n$:

$$(21) \quad \begin{aligned} \text{Log } q'_n &= \text{Log } B(\alpha_n, n - \alpha_n + 1) + (n - \alpha_n) \text{Log}(1 - G_n) + (\alpha_n - 1) \text{Log } G_n \\ &= A_n + A'_n + A''_n. \end{aligned}$$

D'après la formule de Stirling,

$$A_n \sim [\text{Log } n]^{s+1}.$$

Par ailleurs $A'_n \sim -nG_n$ et $A''_n \sim -[\text{Log } n]^{s+1}$, si bien que :

$$A_n \sim -A''_n.$$

De plus,

$$\text{Log } |A'_n| \sim \varepsilon \sqrt{2 \text{Log } n} \quad \text{et}$$

$$\text{Log } |A''_n| \sim (s+1) \text{Lcg}([\text{Log } n]).$$

Ainsi,

$$A''_n = o(A'_n).$$

Donc,

$$\text{Log } q'_n \sim A'_n = -nG_n,$$

et

$$\text{Log } |\text{Log } q'_n| \sim \varepsilon \sqrt{2 \text{Log } n}.$$

Ceci montre que q'_n est équivalent au terme général d'une série convergente, et que la série Σq_n converge. Ceci achève la preuve du lemme 1. ■

Le lemme 1 et (13) permettent d'écrire :

$$(22) \quad \sqrt{2 \text{Log } n} - \varepsilon \ll_{\text{ps}} \bar{Z}_n^{\alpha_n} < \bar{Z}_n^1 \ll_{\text{ps}} \sqrt{2 \text{Log } n} + \varepsilon.$$

De plus, si ν_n désigne le nombre des points de l'échantillon qui appartiennent à la couronne C'_n , les inégalités

$$\sqrt{2 \text{Log } n} - \varepsilon < \bar{Z}_n^{\alpha_n} < \bar{Z}_n^1 < \sqrt{2 \text{Log } n} + \varepsilon$$

impliquent $\nu_n \geq \alpha_n$. La relation (22) fournit alors une minoration presque sûre de ν_n . Le lemme suivant s'en déduit facilement :

Lemme 2.- *Le nombre ν_n des points de l'échantillon qui sont contenus dans la couronne C'_n est presque sûrement minoré, pour n assez grand, par $[\text{Log } n]^s$ où s est un entier positif arbitraire.*

b) Partageons le cercle S_n de centre O et de rayon $b_n - \varepsilon$ en μ_n arcs de cercles $\sigma_{n,m}$ ($m = 1, \dots, \mu_n$) de même longueur. Soit $\gamma_{n,m}$ la partie de la couronne C'_n découpée par les demi-droites d'origine O s'appuyant sur les extrémités de $\sigma_{n,m}$.

Lemme 3.- Presque sûrement, pour n suffisamment grand, chaque partie $\gamma_{n,m}$ ($m = 1, \dots, \mu_n$) contient au moins un point de l'échantillon si μ_n vérifie :

$$(23) \quad \text{Log } \mu_n = O(\text{Log Log } n).$$

■ Soit Q_n l'événement réalisé si chacun des domaines $\gamma_{n,m}$ ($m = 1, \dots, \mu_n$) contient au moins un point de l'échantillon. On pose :

$$(24) \quad \mathcal{E}_N = \bigcap_{n>N} Q_n.$$

On doit montrer que $\lim_{N \rightarrow +\infty} P(\mathcal{E}_N) = 1$, sous la condition (23).

Pour un entier $s > 0$ donné, on définit l'événement :

$$(25) \quad \mathcal{H}_N = \bigcap_{n>N} \{\nu_n > [\text{Log } n]^s\} \quad \text{pour } n > N.$$

Par l'inégalité de Boole,

$$(26) \quad P\{\mathcal{E}_N \mid \mathcal{H}_N\} > 1 - \sum_{n>N} P\{Q_n^c \mid \mathcal{H}_N\}.$$

Pour $n > N$, il est clair que la réalisation de Q_n ne dépend de ν_N, ν_{N+1}, \dots que par l'intermédiaire de ν_n . Ainsi,

$$P(Q_n^c \mid \nu_N, \nu_{N+1}, \dots) = P(Q_n^c \mid \nu_n) \leq \mu_n \left(1 - \frac{1}{\mu_n}\right)^{\nu_n}.$$

D'où :

$$P(Q_n^c \mid \mathcal{H}_N) \leq \mu_n \left(1 - \frac{1}{\mu_n}\right)^{[\text{Log } n]^s}.$$

On déduit alors de (26) :

$$P\{\mathcal{E}_N \mid \mathcal{H}_N\} > 1 - \sum_{n>N} \mu_n \left(1 - \frac{1}{\mu_n}\right)^{[\text{Log } n]^s}.$$

Or, $P\{\mathcal{E}_N\} > P\{\mathcal{E}_N \cap \mathcal{H}_N\} = P\{\mathcal{H}_N\} \cdot P\{\mathcal{E}_N \mid \mathcal{H}_N\}$, et d'après le lemme 2, $\lim_{N \rightarrow \infty} P\{\mathcal{H}_N\} = 1$. Ainsi, $\lim_{N \rightarrow \infty} P\{\mathcal{E}_N\} = 1$ si et seulement si la série de terme général $u_n = \mu_n \left(1 - \frac{1}{\mu_n}\right)^{[\text{Log } n]^s}$ converge.

On montrerait comme dans [10] que $\sum u_n$ converge si et seulement si (23) est réalisé, ce qui termine la démonstration. ■

Remarque 4.- On peut constater dans la démonstration précédente, que les hypothèses du lemme 3 sont vérifiées pour tout échantillon satisfaisant aux hypothèses du lemme 2.

c) **Lemme 4.-** Les hypothèses du lemme 3 sont vérifiées pour $\mu_n = [\text{Log } n]^2$. De plus pour cette valeur de μ_n , la longueur de chaque arc de cercle $\sigma_{n,m} (m = 1, \dots, \mu_n)$ tend vers 0 quand n tend vers l'infini.

■ On vérifie aisément que $[\text{Log } n]^2$ satisfait les conditions demandées. ■

d) Dans la dernière partie de ce paragraphe, il reste à montrer que le polyèdre d'appui H'_n est presque sûrement compris pour n suffisamment grand, entre deux cercles de rayons arbitrairement proches. Soit $\varepsilon > 0$, on peut, par les mêmes arguments géométriques que ceux de [10], montrer que H'_n est presque sûrement asymptotiquement compris entre S'_n , cercle de centre O de rayon $\sqrt{2 \text{Log } n} + \varepsilon$ et S''_n , cercle de centre O de rayon $\sqrt{2 \text{Log } n} - 2\varepsilon$. Si bien que si \tilde{S}_n désigne le cercle de centre O et de rayon $\sqrt{2 \text{Log } n}$,

$$\lim_{n \rightarrow +\infty} d^*(H'_n, \tilde{S}_n) = 0.$$

V - CAS OU $\Psi(t) = (1 - \exp(-t^\alpha))1_{\{t>0\}}$, ($\alpha > 0$).

1°) Stabilité presque-sûre de $(Z_n^*)_n$.

D'après (3) et (5) on obtient :

$$(27) \quad G(t) = 1 - F(t) = \int_0^{+\infty} \exp(-(t/r)^\alpha) r e^{-r^2/2} dr$$

et

$$(28) \quad f(t) = \alpha t^{\alpha-1} \int_0^{+\infty} r^{1-\alpha} \exp(-(t/r)^\alpha) e^{-r^2/2} dr.$$

Afin d'étudier la limite de $\frac{f(t)}{G(t) \text{Log} |\text{Log } G(t)|}$ et d'utiliser le théorème 53 de [9], on cherche des équivalents de f et de G . Pour obtenir un équivalent de f au voisinage de l'infini, on utilise la méthode de Laplace(1), ([4], [7]).

(1) cette méthode est rappelée dans l'appendice 1 de la thèse.

Dans $f(t)$, faisons le changement de variable $u(\alpha t^\alpha)^{\frac{1}{\alpha+2}}$. On obtient :

$$(29) \quad f(t) = t^{(3\alpha-2)/(\alpha+2)} \alpha^{4/(\alpha+2)} \cdot I(t)$$

où

$$(30) \quad I(t) = \int_0^{+\infty} u^{1-\alpha} \exp\left[-t^{2\alpha/(\alpha+2)}(\alpha^{-\alpha/(\alpha+2)} u^{-\alpha} + \alpha^{2/(\alpha+2)} \cdot u^2/2)\right] du.$$

Dans $I(t)$, on pose :

$$(31) \quad s = t^{2\alpha/(\alpha+2)}$$

$$(32) \quad g(u) = u^{1-\alpha}$$

$$(33) \quad H(u) = [-\alpha^{-\alpha/(\alpha+2)} u^{-\alpha} - \alpha^{2/(\alpha+2)} \cdot u^2/2].$$

Si bien que :

$$(34) \quad I(t) = \int_0^{+\infty} g(u) e^{sH(u)} du.$$

Les fonctions g et H sont continues, dérivables sur $]0, +\infty[$, H admet un maximum en $u_0 = 1$. D'après [4], chap. IV.2. ou [7](2), $I(t)$ est équivalent au voisinage de l'infini à :

$$(35) \quad g(u_0) e^{sH(u_0)} [-2\pi/sH''(u_0)]^{1/2}$$

avec :

$$(36) \quad g(u_0) = 1$$

$$(37) \quad H(u_0) = -\alpha^{-\alpha/(\alpha+2)} - \frac{\alpha^{2/(\alpha+2)}}{2} < 0$$

$$(38) \quad H''(u_0) = -\alpha^{2(\alpha+2)}(\alpha+2) < 0,$$

s étant donné par (31).

D'après (29), (30), et (35) $f(t)$ est donc au voisinage de l'infini équivalent à :

$$(39) \quad C_1(\alpha) t^{2(\alpha-1)/(\alpha+2)} \exp[-C_2(\alpha) t^{2\alpha/(\alpha+2)}]$$

avec

$$(40) \quad C_1(\alpha) = \alpha^{4/(\alpha+2)} (-2\pi/H''(u_0))^{1/2} > 0.$$

(2) voir aussi l'appendice 1, Lemme 1.

$$(41) \quad C_2(\alpha) = -H(u_0) > 0.$$

On supposera maintenant que $\alpha \geq 1$. Cherchons un équivalent de $G(t) = \int_t^{+\infty} f(u)du$, au voisinage de l'infini : f et (39) sont des fonctions de t positives continues sur $[0, +\infty]$, et l'intégrale sur $[0, +\infty]$ de (39) est convergente; on en déduit d'après le théorème III.10.2 de [4](3) qu'au voisinage de l'infini, $G(t)$ est équivalent à :

$$(42) \quad C_1(\alpha) \int_t^{+\infty} u^{2(\alpha-1)/(\alpha+2)} \exp[-C_2(\alpha)u^{2\alpha/(\alpha+2)}] du.$$

Posons maintenant $j(u) = u^{2(\alpha-1)/(\alpha+2)} \exp[-C_2(\alpha)u^{2\alpha/(\alpha+2)}]$. D'après le théorème III.10.7 de 4 $G(t)$ est alors équivalent au voisinage de l'infini à :

$$C_1(\alpha) \frac{(j(t))^2}{|j'(t)|}$$

soit à :

$$(43) \quad \{C_1(\alpha)(\alpha+2)/2\alpha \cdot C_2(\alpha)\} t^{\alpha/(\alpha+2)} \exp[-C_2(\alpha)t^{2\alpha/(\alpha+2)}].$$

Les expressions (39) et (43) montrent qu'au voisinage de l'infini $\frac{f(t)}{G(t) \text{Log} |\text{Log} G(t)|}$ est équivalent à :

$$(44) \quad C_3(\alpha) t^{(\alpha-2)/(\alpha+2)} [\text{Log} t^{2\alpha/(\alpha+2)}]^{-1},$$

C_3 étant une constante positive ne dépendant que de α . Si $\alpha \leq 2$, (44) tend vers 0 quand t tend vers l'infini et si $\alpha > 2$, cette même expression tend vers l'infini. On en déduit, par le théorème 53 ([9]), que si $\alpha > 2$, Z_n^* est stable presque sûrement et que $Z_n^* - G^{-1}(\frac{1}{n}) \xrightarrow{p.s.} 0$. Dans le cas où $\alpha \leq 2$ rien ne permet de conclure. D'où le théorème suivant:

Théorème 1.- *Si $\alpha > 2$, $(Z_n^*)_n$ est stable presque sûrement.*

On supposera par la suite que $\alpha > 2$. De même qu'au paragraphe précédent on peut donner une approximation de $a_n = G^{-1}(\frac{1}{n})$; comme $\text{Lim}_{n \rightarrow +\infty} a_n = +\infty$, on peut écrire d'après (43) :

$$\text{Log} G(a_n) = C_4(\alpha) + \alpha/(\alpha+2) \text{Log} a_n - C_2(\alpha) a_n^{2\alpha/(\alpha+2)} + O(1),$$

(3) voir l'appendice 1 de la thèse, Lemme 2.

(4) voir aussi l'appendice 1, Lemme 3.

où $C_4(\alpha)$ est une constante ne dépendant que de α . En posant

$$(45) \quad b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}$$

on obtient :

$$a_n^{2\alpha/(\alpha+2)} - b_n^{2\alpha/(\alpha+2)} = (C_4(\alpha) + \alpha/(\alpha+2) \text{Log } a_n + O(1)) / C_2(\alpha).$$

Si bien qu'en notant $r = 2\alpha/(\alpha+2)$, $a_n^r - b_n^r$ est équivalent au voisinage de l'infini à $C_5(\alpha) \text{Log } a_n$ (avec $C_5(\alpha) > 0$ et $1 < r < 2$ puisque $\alpha > 2$). De plus,

$$a_n^r - b_n^r = a_n^{r-1}(a_n - b_n^r a_n^{1-r}),$$

et, comme $0 < b_n < a_n$,

$$a_n^r - b_n^r > a_n^{r-1}(a_n - b_n).$$

D'où :

$$(46) \quad a_n^{1-r}(a_n^r - b_n^r) > a_n - b_n > 0.$$

Le premier membre de (46) étant équivalent à $C_5(\alpha)a_n^{1-r} \text{Log } a_n$, $a_n - b_n$ tend vers 0 ou encore :

$$(47) \quad a_n = b_n + o(1)$$

avec

$$(48) \quad b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}.$$

Ainsi, pour tout $\varepsilon > 0$,

$$(49) \quad b_n - \varepsilon \underset{\text{p.s.}}{\ll} Z_n^* \underset{\text{p.s.}}{\ll} b_n + \varepsilon$$

On utilise ensuite les mêmes méthodes qu'au paragraphe IV.

2°) **Lemme 5.**- Avec les mêmes notations qu'au paragraphe IV, on a encore avec les mêmes hypothèses : pour tout $\varepsilon > 0$,

$$b_n - \varepsilon \underset{\text{p.co.}}{\ll} \bar{Z}_n^{\alpha_n}$$

b_n étant donné par (48).

■ On peut encore écrire :

$$q_n = B(\alpha_n, n - \alpha_n + 1) \int_{G_n}^1 (1 - \xi)^{n - \alpha_n} \xi^{\alpha_n - 1} d\xi.$$

Ici, d'après (43) et (48),

$$\text{Log } G_n = -\text{Log } n + \frac{2\alpha\varepsilon}{\alpha + 2} C_2(\alpha) (\text{Log } n / C_2(\alpha))^{(\alpha-2)/2\alpha} + \frac{1}{2} \text{Log } \text{Log } n + O(1).$$

Donc, pour n assez grand, $\text{Log } G_n > \text{Log}(\frac{\alpha_n - 1}{n - 1})$. Si bien que :

$$q_n < B(\alpha_n, n - \alpha_n + 1)(1 - G_n)^{n - \alpha_n} G_n^{\alpha_n - 1} = q'_n.$$

Il reste à montrer que la série $\Sigma q'_n$ converge.

$\text{Log } q'_n = A_n + A'_n + A''_n$, ainsi qu'on l'a noté au paragraphe III. De la même façon qu'au paragraphe précédent,

$$\text{Log } q'_n \sim A'_n = -nG_n \quad \text{et}$$

$$\text{Log } |\text{Log } q'_n| \sim \frac{2\alpha\varepsilon C_2(\alpha)}{\alpha + 2} \left(\frac{\text{Log } n}{C_2(\alpha)} \right)^{(\alpha-2)/2\alpha}.$$

Ainsi, pour n assez grand, q'_n est majoré par le terme général d'une série convergente. La série Σq_n est donc convergente, ce qui achève la démonstration. ■

Comme on l'a remarqué plus haut (Remarque 4), le lemme 5 étant obtenu, la suite de la démonstration est analogue à celle faite au paragraphe III. Il reste à choisir μ_n pour que les conditions du lemme 4 soient vérifiées pour $b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}$. On constate que $\mu_n = [\text{Log } n]^2$ convient encore. La démonstration du résultat final est identique : on montre alors que

$$\lim d^*(H_n, \tilde{S}_n) = 0,$$

\tilde{S}_n étant le cercle de centre O et de rayon $b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}$.

$$(41) \quad C_2(\alpha) = -H(u_0) > 0.$$

On supposera maintenant que $\alpha \geq 1$. Cherchons un équivalent de $G(t) = \int_t^{+\infty} f(u)du$, au voisinage de l'infini : f et (39) sont des fonctions de t positives continues sur $[0, +\infty]$, et l'intégrale sur $[0, +\infty]$ de (39) est convergente ; on en déduit d'après le théorème III.10.2 de [4](3) qu'au voisinage de l'infini, $G(t)$ est équivalent à :

$$(42) \quad C_1(\alpha) \int_t^{+\infty} u^{2(\alpha-1)/(\alpha+2)} \exp[-C_2(\alpha)u^{2\alpha/(\alpha+2)}] du.$$

Posons maintenant $j(u) = u^{2(\alpha-1)/(\alpha+2)} \exp[-C_2(\alpha)u^{2\alpha/(\alpha+2)}]$. D'après le théorème III.10.7 de 4 $G(t)$ est alors équivalent au voisinage de l'infini à :

$$C_1(\alpha) \frac{(j(t))^2}{|j'(t)|}$$

soit à :

$$(43) \quad \{C_1(\alpha)(\alpha+2)/2\alpha \cdot C_2(\alpha)\} t^{\alpha/(\alpha+2)} \exp[-C_2(\alpha)t^{2\alpha/(\alpha+2)}].$$

Les expressions (39) et (43) montrent qu'au voisinage de l'infini $\frac{f(t)}{G(t) \text{Log} |\text{Log} G(t)|}$ est équivalent à :

$$(44) \quad C_3(\alpha) t^{(\alpha-2)/(\alpha+2)} [\text{Log} t^{2\alpha/(\alpha+2)}]^{-1},$$

C_3 étant une constante positive ne dépendant que de α . Si $\alpha \leq 2$, (44) tend vers 0 quand t tend vers l'infini et si $\alpha > 2$, cette même expression tend vers l'infini. On en déduit, par le théorème 53 ([9]), que si $\alpha > 2$, Z_n^* est stable presque sûrement et que $Z_n^* - G^{-1}(\frac{1}{n}) \xrightarrow{p.s.} 0$. Dans le cas où $\alpha \leq 2$ rien ne permet de conclure. D'où le théorème suivant :

Théorème 1.- *Si $\alpha > 2$, $(Z_n^*)_n$ est stable presque sûrement.*

On supposera par la suite que $\alpha > 2$. De même qu'au paragraphe précédent on peut donner une approximation de $a_n = G^{-1}(\frac{1}{n})$; comme $\text{Lim}_{n \rightarrow +\infty} a_n = +\infty$, on peut écrire d'après (43) :

$$\text{Log} G(a_n) = C_4(\alpha) + \alpha/(\alpha+2) \text{Log} a_n - C_2(\alpha) a_n^{2\alpha/(\alpha+2)} + O(1),$$

(3) voir l'appendice 1 de la thèse, Lemme 2.

(4) voir aussi l'appendice 1, Lemme 3.

où $C_4(\alpha)$ est une constante ne dépendant que de α . En posant

$$(45) \quad b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}$$

on obtient :

$$a_n^{2\alpha/(\alpha+2)} - b_n^{2\alpha/(\alpha+2)} = (C_4(\alpha) + \alpha/(\alpha+2) \text{Log } a_n + O(1)) / C_2(\alpha).$$

Si bien qu'en notant $r = 2\alpha/(\alpha+2)$, $a_n^r - b_n^r$ est équivalent au voisinage de l'infini à $C_5(\alpha) \text{Log } a_n$ (avec $C_5(\alpha) > 0$ et $1 < r < 2$ puisque $\alpha > 2$). De plus,

$$a_n^r - b_n^r = a_n^{r-1}(a_n - b_n a_n^{1-r}),$$

et, comme $0 < b_n < a_n$,

$$a_n^r - b_n^r > a_n^{r-1}(a_n - b_n).$$

D'où :

$$(46) \quad a_n^{1-r}(a_n^r - b_n^r) > a_n - b_n > 0.$$

Le premier membre de (46) étant équivalent à $C_5(\alpha) a_n^{1-r} \text{Log } a_n$, $a_n - b_n$ tend vers 0 ou encore :

$$(47) \quad a_n = b_n + o(1)$$

avec

$$(48) \quad b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}.$$

Ainsi, pour tout $\varepsilon > 0$,

$$(49) \quad b_n - \varepsilon \underset{\text{ps}}{\ll} Z_n^* \underset{\text{ps}}{\ll} b_n + \varepsilon$$

On utilise ensuite les mêmes méthodes qu'au paragraphe IV.

2°) **Lemme 5.-** Avec les mêmes notations qu'au paragraphe IV, on a encore avec les mêmes hypothèses : pour tout $\varepsilon > 0$,

$$b_n - \varepsilon \underset{\text{p.co.}}{\ll} \bar{Z}_n^{\alpha n}$$

b_n étant donné par (48).

■ On peut encore écrire :

$$q_n = B(\alpha_n, n - \alpha_n + 1) \int_{G_n}^1 (1 - \xi)^{n - \alpha_n} \xi^{\alpha_n - 1} d\xi.$$

Ici, d'après (43) et (48),

$$\text{Log } G_n = -\text{Log } n + \frac{2\alpha\varepsilon}{\alpha + 2} C_2(\alpha) (\text{Log } n / C_2(\alpha))^{(\alpha-2)/2\alpha} + \frac{1}{2} \text{Log } \text{Log } n + O(1).$$

Donc, pour n assez grand, $\text{Log } G_n > \text{Log}(\frac{\alpha_n - 1}{n - 1})$. Si bien que :

$$q_n < B(\alpha_n, n - \alpha_n + 1) (1 - G_n)^{n - \alpha_n} G_n^{\alpha_n - 1} = q'_n.$$

Il reste à montrer que la série $\Sigma q'_n$ converge.

$\text{Log } q'_n = A_n + A'_n + A''_n$, ainsi qu'on l'a noté au paragraphe III. De la même façon qu'au paragraphe précédent,

$$\text{Log } q'_n \sim A'_n = -nG_n \quad \text{et}$$

$$\text{Log } |\text{Log } q'_n| \sim \frac{2\alpha\varepsilon C_2(\alpha)}{\alpha + 2} \left(\frac{\text{Log } n}{C_2(\alpha)} \right)^{(\alpha-2)/2\alpha}.$$

Ainsi, pour n assez grand, q'_n est majoré par le terme général d'une série convergente. La série Σq_n est donc convergente, ce qui achève la démonstration. ■

Comme on l'a remarqué plus haut (Remarque 4), le lemme 5 étant obtenu, la suite de la démonstration est analogue à celle faite au paragraphe III. Il reste à choisir μ_n pour que les conditions du lemme 4 soient vérifiées pour $b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}$. On constate que $\mu_n = [\text{Log } n]^2$ convient encore. La démonstration du résultat final est identique : on montre alors que

$$\lim d^*(H_n, \tilde{S}_n) = 0,$$

\tilde{S}_n étant le cercle de centre O et de rayon $b_n = (\text{Log } n / C_2(\alpha))^{(\alpha+2)/2\alpha}$.

VI - CONCLUSION

Les calculs qui précèdent nous ont permis d'obtenir sur des exemples des résultats de localisation asymptotique du nuage de points défini par un échantillon de \mathbf{R}^2 dans le cas particulier où la déformation imposée à l'échantillon est isotrope. On peut dès à présent étendre ces résultats de la façon suivante : si A est une application linéaire de \mathbf{R}^2 , la frontière de l'enveloppe convexe de $A\{X'_1, \dots, X'_n\}$ n'est autre que la transformée par A de H'_n . Ainsi, si $\{X'_1, \dots, X'_n\}$ satisfait aux hypothèses des théorèmes cités plus haut, il en est de même pour $A\{X'_1, \dots, X'_n\}$. Plus généralement, on peut imaginer que les cercles \tilde{S}_n obtenus ici se déformeraient non uniformément dans toutes les directions si U était une fonction aléatoire de θ ; on serait alors amené à considérer des formes non convexes. Ce type de problème s'apparente à l'estimation du support d'une loi ([2], [11], [12]). Ceci nous conduit à étudier le problème suivant pour un échantillon de variables aléatoires à valeurs dans \mathbf{R}^k et de support non borné : trouver une enveloppe minimale non nécessairement convexe contenant le nuage de points défini par l'échantillon. Nous avons fait cette étude dans les chapitres 2 et 3, ([3], [3]').

BIBLIOGRAPHIE

- [1] BROZIUS, H. et DE HAAN, L. (1987) - *On limiting laws for the convex hull of a sample*. J. Appl. Prob. 24. p. 852-862.
- [2] CHEVALIER, J. (1976) - *Estimation du support et du contenu du support d'une loi de probabilité*. Ann. Inst. Henri Poincaré, section B, Vol. XII, 4, 1976, p. 339-364.
- [3] DELCROIX, M.F. et JACOB, P. (1992) - *Stability of extreme value for a multidimensional sample*. Statistique et Analyse des Données, Juin 92, 1-21.
- [3]' DELCROIX, M.F. - *Localisation asymptotique des échantillons multidimensionnels*. A soumettre.
- [4] DIEUDONNÉ, J. (1980) - *Calcul infinitésimal*. Hermann.
- [5] EDDY et GALE (1981) - *The convex hull of spherically symmetric sample*. Adv. Appl. Prob. 13. p. 751-763.
- [6] EFRON, B. (1965) - *The convex hull of a random set of points*. Biometrika, 52, p. 331-343.
- [7] ERDELYI, A. - *Asymptotic expansions*. Dovers.
- [8] FISHER, L.D. (1966) - *The convex hull of a sample*. Bull. Am. Math. Soc. 72, p. 555-558.
- [9] GEFFROY, J. (1958 - 1959) - *Contribution à la théorie des valeurs extrêmes*. Pub. Inst. Stat. Univ. Paris, 1958 (fasc. 4) et 1959 (fasc. 1).
- [10] GEFFROY, J. (1961) - *Localisation asymptotique du polyèdre d'appui d'un échantillon laplacien à k dimensions*. Pub. Inst. Stat. Univ. Paris, 1961 n° 59.
- [11] GENSBITTEL, M.H. (1979) - *Contribution à l'étude statistique des répartitions ponctuelles aléatoires*. Thèse, Université Pierre et Marie Curie, Paris.
- [12] JACOB, P. et ABBAR, H. (1988) - *Estimating the edge of a Cox process area*. Cahiers du centre d'Etudes de Recherche Opérationnelle, vol. 31, n°3-4.
- [13] RAYNAUD, H. (1970) - *Sur l'enveloppe convexe des nuages de points aléatoires dans \mathbf{R}^N* . J. Appl. Prob. 7 p. 35-48.
- [14] RENYI, A. et SULANKE, R. (1964) - *Über die konvexe hülle von n zufällig gewählten Punkten II*. Z. Wahrschein. 3, p. 138-147.

**Chapitre 2 : Stabilité de la valeur
maximale des échantillons
multidimensionnels**

I. INTRODUCTION

Actuellement, la théorie des valeurs extrêmes concerne souvent des variables non identiquement distribuées, des données non indépendantes, ([Haiman, Puri, 1990],[Haiman, Puri]) ou des données multidimensionnelles, ([Davis, Mulrow, Resnick, 1987]). Cependant, de récents articles sur les << outliers >> ([Gather, Rauhut, 1990], [Green, 1976], [Mathar, 1989], [Munoz-Garcia, Moreno-Rebollo, Pascual-Acosta, 1990]), donnent un nouvel intérêt à la notion déjà ancienne de stabilité ([Geffroy, 1958,1959], [Geffroy, 1961], [Gnedenko, 1943]). Nous proposons ici une nouvelle définition pour la valeur maximale d'un échantillon multidimensionnel et pour la stabilité de cette valeur maximale. Il est aussi possible de définir des lois << outlier-resistant >> ou << outlier-prone >> comme cela a été fait pour des variables aléatoires réelles, [Gather, Rauhut, 1990], [Green, 1976] ; dans ce chapitre nous étudions tout d'abord les propriétés de stabilité de la valeur maximale d'un échantillon multidimensionnel . Nous étudierons dans le chapitre 4 la notion de loi << outlier-resistant >> ou << outlier-prone >>.

Etant donné un échantillon (Y_1, \dots, Y_n) de \mathbb{R}^+ , où chaque variable a une fonction de répartition F continue, si Y_n^* est la valeur maximale de l'échantillon, alors $F(Y_n^*)$ est la valeur maximale de $(F(Y_1), \dots, F(Y_n))$; cette remarque est à l'origine de la définition donnée ci-dessous.

On considère dans ce chapitre des variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans l'espace euclidien \mathbb{R}^k .

Pour tout x dans $\mathbb{R}^k \setminus \{0\}$ on définit un couple $(\|x\|, \frac{x}{\|x\|}) = (r, \theta)$ dans $\mathbb{R}^{+*} \times S^{k-1}$, où $\| \cdot \|$ est la norme euclidienne. La sphère unité S^{k-1} de \mathbb{R}^k est munie de la topologie induite par celle de \mathbb{R}^k .

Pour chaque variable $X = (R, \Theta)$, on suppose que la loi de Θ , et pour tout θ , que la loi de R étant donné $\Theta = \theta$, sont définies par une densité continue. On désigne par F_θ la fonction de répartition conditionnelle de R , étant donné $\Theta = \theta$, et F_θ^{-1} son inverse généralisée.

Pour tout $0 < u < 1$, on appelle **isobare** de niveau u - de la loi de R étant donné $\Theta = \theta$ - l'application $\theta \rightsquigarrow F_{\theta}^{-1}(u)$. On suppose que cette application est continue et strictement positive ; la surface d'équation $\rho = F_{\theta}^{-1}(u)$ est aussi appelée isobare.

Soit $E_n = (X_1, \dots, X_n)$ un échantillon de variables indépendantes de même loi que X . Pour tout $1 \leq k \leq n$ il existe presque sûrement une unique isobare de niveau u_k , de la loi de R étant donné $\Theta = \theta$ qui contient (R_k, Θ_k) . On définit la **valeur maximale** de E_n comme le point $X_n^* = (R_n^*, \Theta_n^*)$ de l'échantillon qui appartient à l'isobare de plus haut niveau, c'est à dire l'isobare dont le niveau est $\max_{1 \leq k \leq n} u_k$. Bien

sûr on ne peut en pratique désigner la valeur maximale d'un échantillon de loi inconnue, comme il serait possible de le faire s'il s'agissait du point le plus éloigné de l'origine ou du point dont les coordonnées cartésiennes sont les maxima des coordonnées des points de l'échantillon. Cependant cette valeur extrême et plus généralement les valeurs extrêmes obtenues en ordonnant l'échantillon selon les niveaux croissants, recèlent beaucoup d'information sur la queue des distributions conditionnelles et permettent une étude statistique des isobares de ces distributions. De plus, l'intérêt de cette définition réside dans le fait qu'elle est non seulement basée sur une distance classique de \mathbb{R}^k , mais sur la probabilité d'être à une certaine distance de l'origine.

Dans le paragraphe II, nous donnons quelques propriétés de la loi du couple (R_n^*, Θ_n^*) . Dans les paragraphes III et IV, on définit la notion de stabilité en probabilité et de stabilité presque sûre de la valeur maximale. Il s'agit en fait de décrire la tendance de X_n^* à se situer au voisinage d'une surface donnée. Plus précisément, $(X_n^*)_n$ est dite **stable** en probabilité (où presque sûrement stable) s'il existe une suite (Γ_n) de surfaces, d'équation $\rho = g_n(\theta)$, telle que $R_n^* - g_n(\Theta_n^*) \rightarrow 0$ en probabilité (ou presque sûrement). Pour une classe de lois que nous allons préciser, ce phénomène se produit et (g_n) se révèle être une suite d'isobares de la loi de R étant donné $\Theta = \theta$. Cela généralise la notion de stabilité étudiée par J. Geffroy, [Geffroy, 1958, 1959], à qui nous faisons de nombreux emprunts. Quelques exemples sont donnés dans le paragraphe V. Dans le paragraphe VI, nous discutons les hypothèses que nous avons faites tout au long de ce chapitre, en particulier certaines hypothèses de régularité sur les isobares. Bien sûr, beaucoup de lois ne possèdent pas de propriétés de stabilité, aussi nous utilisons une notion plus faible. L'idée est de remplacer chaque observation $X = (R, \Theta)$ par la variable $X\phi = (\phi(R), \Theta)$ où ϕ est une fonction convenable, afin d'obtenir des propriétés de stabilité pour la variable $X\phi$. Cela consiste à considérer l'ensemble des

points $E_n^{\phi} = \{(\varphi(R_1), \Theta_1), \dots, (\varphi(R_n), \Theta_n)\}$ au lieu de l'échantillon initial. Nous donnons dans le paragraphe VII quelques propriétés de la φ -stabilité, ainsi que des exemples.

II - PRELIMINAIRES

Dans ce paragraphe nous donnons quelques propriétés de la loi conditionnelle de R_n^* étant donné $\Theta_n^* = \theta$ qui seront utiles par la suite.

Soit (X_1, \dots, X_n) un échantillon de représentation polaire $(R_1, \Theta_1), \dots, (R_n, \Theta_n)$. Pour tout $1 \leq i \leq n$, on pose :

$$(1) \quad E_i = \{F_{\Theta_i}(R_i) = \max_{j=1}^n F_{\Theta_j}(R_j)\}.$$

Pour tout θ et pour tout $0 \leq t \leq 1$, $P(F_{\Theta}(R) \leq t / \Theta = \theta) = F_{\theta}(F_{\theta}^{-1}(t)) = t$, donc $\{F_{\Theta_j}(R_j), j = 1, \dots, n\}$ est un échantillon de la loi uniforme sur $[0, 1]$. De plus, la valeur maximale de l'échantillon est presque sûrement définie comme le point X_n^* de représentation polaire :

$$(R_n^*, \Theta_n^*) = \sum_{i=1}^n (R_i, \Theta_i) \mathbf{1}_{E_i}.$$

Théorème 1. a) Θ_n^* et Θ sont de même loi

b) Toute isobare de niveau u de la loi de R étant donné Θ est aussi isobare de niveau u^n de la loi de R_n^* étant donné Θ_n^* .

— a) Comme $P(F_{\Theta}(R) \leq t / \Theta = \theta) = t$, pour tout $1 \leq j \leq n$ Θ_j et $F_{\Theta_j}(R_j)$ sont indépendantes. On en déduit que $\{\Theta_j; j = 1, \dots, n\}$ et $\{F_{\Theta_j}(R_j); j = 1, \dots, n\}$ sont indépendants.

Ainsi, pour tout $1 \leq j \leq n$, Θ_j et $\mathbf{1}_{E_j}$ sont indépendantes. Par conséquent, pour tout borélien C de S^{k-1} :

$$(2) \quad \begin{aligned} P(\Theta_n^* \in C) &= P\left(\sum_{i=1}^n \Theta_i \mathbf{1}_{E_i} \in C\right) = \sum_{i=1}^n P(\Theta_i \in C; E_i) \\ &= \sum_{i=1}^n P(\Theta_i \in C) P(E_i) = P(\Theta \in C). \end{aligned}$$

b) Soit $\rho = F_{\theta}^{-1}(u)$ une isobare de niveau u de la loi de R étant donné $\Theta = \theta$ et soit B l'événement $\{R_n^* \leq F_{\theta_n}^{-1}(u)\}$. Comme $B = \bigcap_{i=1}^n \{F_{\theta_i}(R_i) \leq u\}$, B est indépendant de $\{\Theta_j, j = 1, \dots, n\}$. Ainsi pour tout borélien C de S^{k-1} , (2) implique :

$$\begin{aligned} P(\Theta_n^* \in C; B) &= \sum_{i=1}^n P(\Theta_i \in C; E_i; B) = \sum_{i=1}^n P(\Theta_i \in C) P(E_i; B) \\ &= P(\Theta_n^* \in C) P(B) \end{aligned}$$

Ainsi Θ_n^* et $\mathbf{1}_B$ sont indépendantes ; et,

$$(3) \quad P(R_n^* \leq F_{\theta_n}^{-1}(u) / \Theta_n^* = \theta) = P(B) = \prod_{i=1}^n P(F_{\theta_i}(R_i) \leq u) = u^n . \quad -$$

Corollaire 1.- Soit $F_{n,\theta}^*$ la fonction de répartition conditionnelle de R_n^* étant donné $\Theta_n^* = \theta$. Pour tout θ , $F_{n,\theta}^* = F_{\theta}^n$.

— Soit S le support de la loi de X . Soit $x = (r, \theta)$ un point dont la distance à l'isobare la plus proche est strictement positive. D'après la continuité des isobares, il existe une boule ouverte $B(x, \varepsilon)$ dont la distance à l'isobare la plus proche est aussi strictement positive. Ainsi la loi de X affecte une masse nulle à $B(x, \varepsilon)$. Mais le support S est l'ensemble de tous les points z de R^k tels que $P(V) > 0$ pour chaque ensemble ouvert V contenant z ; donc x n'est pas un point de S . Ainsi la distance entre tout point de S et une isobare est nulle. Dans la suite de la démonstration, nous considérons comme isobare toute limite (uniforme) d'une suite décroissante d'isobares. S est alors l'union de toutes les isobares de la loi de R étant donné $\Theta = \theta$.

Pour tout ensemble ouvert $\mathbf{0}$ de S^{k-1} et pour tout couple (g, h) d'isobares tel que $g < h$, on définit :

$$D(\mathbf{0}, g, h) = \{x = (r, \theta) \in S : \theta \in \mathbf{0}, g(\theta) \leq r \leq h(\theta)\}.$$

La classe \mathcal{U} de ces ensembles est un π -système [Billingsley, 1968].

De plus, pour tout x dans S et pour tout $\varepsilon > 0$ il existe un ensemble D de \mathcal{U} , de diamètre inférieur à ε , tel que $x \in \overset{\circ}{D} \subset D$ ($\overset{\circ}{D}$ désigne l'intérieur de D pour la topologie induite de S). D'après [Billingsley, 1968] page 14, \mathcal{U} est une classe déterminante pour l'espace métrique séparable S .

Soit $\tilde{X} = (\tilde{R}, \tilde{\Theta})$ une variable aléatoire de \mathbb{R}^k telle que Θ et $\tilde{\Theta}$ soient de même loi et telle que la fonction de répartition de \tilde{R} étant donné $\tilde{\Theta} = \theta$ soit F_{θ}^n . Afin d'obtenir le corollaire 1 il suffit maintenant de montrer que \tilde{X} et X^* sont de même loi. Cela vient immédiatement de (3), de l'équation suivante :

$$(4) \quad P(\tilde{R} < g(\tilde{\Theta}) / \tilde{\Theta} = \theta) = F_{\theta}^n(g(\theta)) = u^n,$$

et du fait que \mathcal{U} est une classe déterminante. —

Les résultats précédents montrent que la loi de R étant donné Θ et la loi de R_n^* étant donné Θ_n^* définissent le même ensemble d'isobares. Dans la suite, nous utiliserons uniquement les isobares de la loi de R étant donné Θ et nous les appellerons simplement isobares.

Nous supposons dans la suite de ce chapitre que pour tout θ l'application F_{θ} est strictement croissante et donc bijective. Fixons un point θ_1 dans S^{k-1} et munissons la droite $(0\theta_1)$ d'un vecteur unité $\overset{\rightarrow}{0\theta_1}$. Pour tout point w (on désignera aussi par w l'abscisse de ce point) sur le demi-axe positif $0\theta_1^+$, il existe une unique isobare contenant w , dont le niveau est noté $u(w)$. Soit $\rho = g(\theta, w)$ l'équation de cette isobare (notons que $g(\theta_1, w) = w$). L'application $w \mapsto u(w)$ de \mathbb{R}^{+*} dans $]0, 1[$ est croissante et bijective. La condition (H) ci-dessous sera utilisée dans la plupart des théorèmes suivants.

$$(H) \quad \text{Il existe } 0 < \alpha_1 \leq \beta_1 < +\infty \text{ tel que pour tout } \theta \text{ dans } S^{k-1}, \\ \text{et pour tout } w > 0 : \alpha_1 \leq \frac{\partial g}{\partial w}(\theta, w) \leq \beta_1.$$

Une conséquence immédiate est donnée par le théorème suivant :

Propriété 1.- *Pour tout $\varepsilon > 0$, il existe $\eta > 0$, et pour tout $w > 0$, il existe deux isobares $h^{\varepsilon}(\theta, w)$ et $\tilde{h}^{\varepsilon}(\theta, w)$ tels que pour tout θ :*

$$(5) \quad g(\theta, w) - \varepsilon < \tilde{h}^{\varepsilon}(\theta, w) < g(\theta, w) - \eta < g(\theta, w) + \eta < h^{\varepsilon}(\theta, w) < g(\theta, w) + \varepsilon \\ (\text{Notons que } \eta \text{ ne dépend pas de } w).$$

— D'après le théorème des accroissements finis, on obtient :

$$\begin{array}{l} g(\theta, w) + \lambda \alpha_1 \leq g(\theta, w + \lambda) \leq g(\theta, w) + \lambda \beta_1 \quad \text{si } \lambda > 0 \\ \text{et} \quad g(\theta, w) + \lambda \beta_1 \leq g(\theta, w + \lambda) \leq g(\theta, w) + \lambda \alpha_1 \quad \text{si } \lambda < 0 \end{array}$$

Il suffit de choisir $\lambda = \frac{\varepsilon}{\beta_1}$ (resp. $\lambda = -\varepsilon/\beta_1$) et de poser :

$$(6) \quad \eta = \varepsilon \alpha_1 / \beta_1$$

$$(7) \quad h^\varepsilon(\theta, w) = g(\theta, w + \varepsilon/\beta_1)$$

$$(8) \quad \tilde{h}^\varepsilon(\theta, w) = g(\theta, w - \varepsilon/\beta_1).$$

—

Remarque 1.- Le niveau $u(w + \varepsilon/\beta_1)$ (resp. $u(w - \varepsilon/\beta_1)$) de $h^\varepsilon(\theta, w)$ (resp. de $\tilde{h}^\varepsilon(\theta, w)$) est une fonction croissante de w ou de $u(w)$.

Remarque 2.- L'inégalité (5) est une propriété clé mais (H) est plus facile à vérifier. (H) nous a été suggérée par un travail de Geffroy non publié dont certains détails peuvent être trouvés dans [Lecoutre, 1982]. Geffroy a considéré le cas d'une densité unimodale de \mathbb{R}^2 , décroissante dans toute direction. L'échantillon est alors ordonné selon les lignes de niveau de la densité bivariable et non pas selon les isobares. Par ailleurs, on montre dans le paragraphe VI que (H) n'est pas nécessaire à la stabilité de la valeur maximale.

Remarque 3.- Pour un échantillon gaussien de \mathbb{R}^2 de matrice de covariance $\begin{pmatrix} \sigma & 0 \\ 0 & \tau \end{pmatrix}$ l'hypothèse (H) est satisfaite. De plus, les isobares sont aussi dans ce cas les lignes de niveau de la densité bivariable. Leur équation polaire est $g(\theta, w) = w\varphi(\theta)$ où
$$\varphi(\theta) = \frac{1}{\sqrt{2}\sigma} \left(\frac{\cos^2\theta}{2\sigma^2} + \frac{\sin^2\theta}{2\tau^2} \right)^{-1/2}.$$

III - STABILITE EN PROBABILITE DE $X_n^* = (R_n^*, \Theta_n^*)$.

D'après le théorème 1, les lois de (R_n^*, Θ_n^*) et de (R, Θ) définissent le même ensemble d'isobares. Aussi nous pouvons proposer la définition suivante .

1. **Definition 1.-** $(X_n^*)_n$ est stable en probabilité si et seulement si il existe une suite $(g_n)_n$ d'isobares satisfaisant :

$$(9) \quad R_n^* - g_n(\Theta_n^*) \xrightarrow{P} 0 .$$

2. **Proposition 1.-** Supposons (H) satisfaite. Si $R_n^* - g_n(\Theta_n^*) \xrightarrow{P} 0$ (où $(g_n)_n$ est une suite d'isobares), alors pour tout $\varepsilon > 0$:

$$(10) \quad \min_{\theta} P(R_n^* < g_n(\theta) + \varepsilon / \Theta_n^* = \theta) \rightarrow 1 .$$

$$(11) \quad \max_{\theta} P(R_n^* < g_n(\theta) - \varepsilon / \Theta_n^* = \theta) \rightarrow 0 .$$

— Soit $\varepsilon > 0$; d'après la propriété 1, il existe $\eta > 0$ et deux suites d'isobares $(h_n^\varepsilon)_n$ et $(\tilde{h}_n^\varepsilon)_n$ tels que pour tout θ et pour tout n ,

$$(12) \quad g_n(\theta) - \varepsilon < \tilde{h}_n^\varepsilon(\theta) < g_n(\theta) - \eta < g_n(\theta) + \eta < h_n^\varepsilon(\theta) < g_n(\theta) + \varepsilon ,$$

Il vient alors pour tout θ fixé :

$$(R_n^* < g_n(\theta) + \eta) \subset (R_n^* < h_n^\varepsilon(\theta)) .$$

Comme $R_n^* - g_n(\Theta_n^*) \xrightarrow{P} 0$,

$$(13) \quad \text{Lim}_{n \rightarrow +\infty} P(R_n^* < h_n^\varepsilon(\Theta_n^*)) = 1 .$$

De plus si $v_{\varepsilon, n}$ est le niveau de h_n^ε :

$$(14) \quad \begin{aligned} P(R_n^* < h_n^\varepsilon(\Theta_n^*)) &= \int_{S^{k-1}} P(R_n^* < h_n^\varepsilon(\theta) / \Theta_n^* = \theta) P_{\Theta_n^*}(d\theta) \\ &= \int_{S^{k-1}} v_{\varepsilon, n} P_{\Theta_n^*}(d\theta) = v_{\varepsilon, n} . \end{aligned}$$

Et pour tout θ , $v_{\varepsilon,n} = P(R_n^* < h_n^\varepsilon(\theta)/\Theta_n^* = \theta) < P(R_n^* < g_n(\theta) + \varepsilon/\Theta_n^* = \theta)$.

On déduit alors (10) de (13) et de (14). La preuve de (11) se traite de la même façon. —

Cette proposition fournit un critère de stabilité en probabilité.

3. Théorème 2.-

i) Si (H) est satisfaite et si $(X_n^*)_n$ est stable en probabilité, alors pour tout

$0 < a < b < 1$:

$$(15) \quad \lim_{n \rightarrow \infty} \sup_{\theta \in S^{k-1}} [(F_{n,\theta}^*)^{-1}(b) - (F_{n,\theta}^*)^{-1}(a)] = 0$$

ii) Inversement, si (15) est réalisée, alors, $X_n^* = (R_n^*, \Theta_n^*)$ est stable en probabilité.

— Soit $\varepsilon > 0$. Par hypothèse il existe une suite d'isobares $(g_n)_n$ telle que $R_n^* - g_n(\Theta_n^*) \xrightarrow{P} 0$. D'après la proposition 1, pour n suffisamment grand, $b < P(R_n^* < g_n(\theta) + \varepsilon / \Theta_n^* = \theta)$ pour tout θ . Alors pour tout θ , $(F_{n,\theta}^*)^{-1}(b) < g_n(\theta) + \varepsilon$. De même $(F_{n,\theta}^*)^{-1}(a) > g_n(\theta) - \varepsilon$ pour tout θ ; d'où (15). Réciproquement, pour $\varepsilon > 0$, $(F_{n,\theta}^*)^{-1}(1-\varepsilon) - (F_{n,\theta}^*)^{-1}(\varepsilon)$ converge uniformément vers 0. On choisit par une méthode diagonale une suite (ε_n) qui décroît vers 0, telle que :

$$(F_{n,\theta}^*)^{-1}(1 - \varepsilon_n) - (F_{n,\theta}^*)^{-1}(\varepsilon_n)$$

converge vers 0 uniformément.

En posant $h_n(\theta) = (F_{n,\theta}^*)^{-1}(1 - \varepsilon_n)$ et $g_n(\theta) = (F_{n,\theta}^*)^{-1}(\varepsilon_n)$, il vient :

$$P(R_n^* < h_n(\theta) \mid \Theta_n^* = \theta) = 1 - \varepsilon_n \text{ et}$$

$$P(R_n^* < g_n(\theta) \mid \Theta_n^* = \theta) = \varepsilon_n \text{ pour tout } \theta ,$$

si bien que :

$$P(g_n(\Theta_n^*) < R_n^* < h_n(\Theta_n^*)) > 1 - 2\varepsilon_n .$$

Ce qui complète la démonstration du théorème 2. —

Le corollaire 1 montre que pour tout $0 < y < 1$: $(F_{n,\theta}^*)^{-1}(y) = F_{\theta}^{-1}(y^{1/n})$, ([Geffroy, 1958,1959], page 70). Ceci permet d'établir le théorème suivant. La démonstration est la même que celle du théorème 20 dans [Geffroy, 1958, 1959].

4. Théorème 3.-

i) Si (H) est satisfaite et si $(X_n^*)_n$ est stable en probabilité, alors pour tout $0 < \alpha < \beta$:

$$(16) \quad \lim_{n \rightarrow +\infty} \sup_{\theta \in S^{k-1}} |F_{\theta}^{-1}(1 - \alpha/n) - F_{\theta}^{-1}(1 - \beta/n)| = 0.$$

ii) Si (16) est satisfaite, alors $(X_n^*)_n$ est stable en probabilité.

Remarque 4.-

a) $F_{\theta}^{-1}(1 - \beta/n)$ et $F_{\theta}^{-1}(1 - \alpha/n)$ n'ont un sens que si $\beta/n < 1$ et $\alpha/n < 1$.

b) pour tout θ et pour tout $0 < t < 1$, on pose $G_{\theta}^{-1}(t) = F_{\theta}^{-1}(1-t)$ (où $G_{\theta} = 1 - F_{\theta}$).

On a donc montré que si $(X_n^*)_n$ est stable en probabilité, on peut choisir dans la définition, $g_n(\theta) = \gamma_n(\theta) = G_{\theta}^{-1}(\frac{1}{n})$, pour n suffisamment grand.

5. Le théorème de Gnedenko, [Geffroy, 1958,1959],[Gnedenko, 1943]) donne, dans notre contexte, un critère simple de stabilité en probabilité.

Théorème 4.- Supposons (H) satisfaite, alors $(X_n^*)_n$ est stable en probabilité si et seulement si il existe θ_1 tel que :

$$(17) \quad \lim_{x \rightarrow +\infty} \frac{G_{\theta_1}(x)}{G_{\theta_1}(x-h)} = 0, \text{ pour tout } h > 0$$

(Alors (17) est vraie pour tout θ).

— Posons W_n^* l'intersection de $\theta\theta_1^+$ et de l'isobare contenant X_n^* (cf le paragraphe II).

Pour tout $w > 0$,

$$P(W_n^* \leq w) = P(R_n^* \leq g(\theta_n^*, w)),$$

où g est l'isobare d'équation $g(\theta, w)$.

Comme g est une isobare, $P(R_n^* \leq g(\theta, w) / \Theta_n^* = \theta)$ ne dépend pas de θ ,

si bien que :

$$\begin{aligned}
P(W_n^* \leq w) &= P(R_n^* \leq g(\theta_1, w) / \Theta_n^* = \theta_1) \\
&= P(R_n^* \leq w / \Theta_n^* = \theta_1) \\
&= F_{\theta_1}^n(w).
\end{aligned}$$

Soit W_i l'intersection de $0\theta_1^+$ et de l'isobare contenant $X_i = (R_i, \Theta_i)$, ($i=1, \dots, n$). Pour tout $w > 0$,

$$(18) \quad P(W_i \leq w) = P(R_i \leq w / \Theta_i = \theta_1) = F_{\theta_1}(w) .$$

La variable aléatoire W_n^* est donc le maximum de n variables indépendantes de même loi, à valeurs dans \mathbb{R}^+ . De plus, les variables W_1, \dots, W_n ont pour fonction de répartition F_{θ_1} . D'après le théorème de Gnedenko, $(W_n^*)_n$ est stable en probabilité si et seulement si :

$$\lim_{x \rightarrow +\infty} \frac{G_{\theta_1}(x)}{G_{\theta_1}(x-h)} = 0 \text{ , pour tout } h > 0$$

Il suffit de montrer maintenant que $(W_n^*)_n$ est stable en probabilité si et seulement si $(X_n^*)_n$ est stable en probabilité.

Soit $\varepsilon > 0$, supposons que $(W_n^*)_n$ est stable en probabilité ; il existe une suite $(a_n)_n$ telle que $W_n^* - a_n \xrightarrow{P} 0$. Soit \mathcal{L}_n l'isobare contenant a_n et $\tilde{h}_n^\varepsilon, h_n^\varepsilon$ les isobares vérifiant (5) pour tout θ :

$$(19) \quad \mathcal{L}_n(\theta) - \varepsilon < \tilde{h}_n^\varepsilon(\theta) < \mathcal{L}_n(\theta) - \eta < \mathcal{L}_n(\theta) < \mathcal{L}_n(\theta) + \eta < h_n^\varepsilon(\theta) < \mathcal{L}_n(\theta) + \varepsilon$$

(η ne dépend pas de n).

Alors,

$$(20) \quad \{|W_n^* - a_n| \leq \eta\} \subset \{\tilde{h}_n^\varepsilon(\Theta_n^*) \leq R_n^* \leq h_n^\varepsilon(\Theta_n^*)\} \subset \{|R_n^* - \mathcal{L}_n(\Theta_n^*)| \leq \varepsilon\}$$

si bien que :

$$R_n^* - \mathcal{L}_n(\Theta_n^*) \xrightarrow{P} 0 .$$

Réciproquement, s'il existe une suite d'isobares $(g_n)_n$ telle que $R_n^* - g_n(\Theta_n^*) \xrightarrow{P} 0$, on note a_n l'intersection de $0\theta_1^+$ avec g_n . Pour $\varepsilon > 0$, il existe $\eta > 0$ et pour tout n , il existe \tilde{h}_n^ε et h_n^ε satisfaisant (19) pour tout θ ; alors,

$$(21) \quad \{|R_n^* - g_n(\Theta_n^*)| \leq \eta\} \subset \{\tilde{h}_n^\varepsilon(\Theta_n^*) \leq R_n^* \leq h_n^\varepsilon(\Theta_n^*)\} \subset \{|W_n^* - a_n| \leq \varepsilon\},$$

ce qui achève la démonstration. —

Remarque 5.- En fait, la démonstration ci dessus reste vraie si l'on considère un ensemble quelconque de courbes $\{g(\theta, w); w \in 0\theta_1^+\}$ tel que $w \leq w'$ implique $g(\theta, w) \leq g(\theta, w')$, pour tout θ . Si (H) est satisfaite, la stabilité de W_n^* et la stabilité de X_n^* sont équivalentes. Ceci permet d'envisager de nombreux cas, mais tous ne sont pas intéressants car il faut pouvoir calculer la loi de W_n^* : c'est ce qui est fait dans [Lecoutre, 1982].

IV - STABILITE PRESQUE-SURE DE $(X_n^*)_n$.

1.- Définitions.

a) **Définition 2.**- La suite $(X_n^*)_n$ est presque-sûrement stable si et seulement si il existe une suite d'isobares $(g_n)_n$ telle que :

$$(22) \quad R_n^* - g_n(\Theta_n^*) \xrightarrow{PS} 0.$$

D'après le théorème 3, si $(X_n^*)_n$ est presque sûrement stable, on peut choisir dans

$$(22) \quad g_n(\theta) = \gamma_n(\theta) = G_\theta^{-1}\left(\frac{1}{n}\right). \text{ Ainsi, dans la suite, } \gamma_n(\theta) \text{ désigne l'isobare } G_\theta^{-1}\left(\frac{1}{n}\right) \text{ et } \Gamma_n$$

l'ensemble des points $\{(\rho, \theta); \rho \leq \gamma_n(\theta)\}$, $(n \geq 2)$.

Pour $\varepsilon > 0$, posons :

$$(23) \quad \Gamma_n^\varepsilon = \{(\rho, \theta) : \rho \leq \gamma_n(\theta) + \varepsilon\}$$

$$(24) \quad \Gamma_n^{-\varepsilon} = \{(\rho, \theta) : \rho \leq \gamma_n(\theta) - \varepsilon\}$$

Alors, $(X_n^*)_n$ est presque sûrement stable si et seulement si pour tout $\varepsilon > 0$,

$$(25) \quad P\{\text{Lim inf } (X_n^* \in \Gamma_n^\varepsilon - \Gamma_n^{-\varepsilon})\} = 1$$

2.- Comme on l'a fait avec le théorème de Gnedenko, il est possible de prouver le théorème suivant, à l'aide des théorèmes 49 et 50 de [Geffroy, 1958, 1959].

Théorème 5.- Si (H) est satisfaite, alors $(X_n^*)_n$ est presque sûrement stable si et seulement si il existe θ_1 tel que :

$$(26) \quad \text{Lim}_{x \rightarrow +\infty} \frac{G_{\theta_1}(x-h)}{G_{\theta_1}(x) \text{ Log } G_{\theta_1}(x)} = +\infty, \text{ pour tout } h > 0.$$

(Si (26) est vraie pour θ_1 , alors (26) est vraie pour tout θ).

— Comme pour le théorème 4, on note W_n^* l'intersection de $0\theta_1^+$ avec l'isobare contenant X_n^* . Si $(W_n^*)_n$ est presque sûrement stable, alors il existe une suite $(a_n)_n$ telle que pour tout $\varepsilon > 0$:

$$P\{\text{Lim inf}(a_n - \varepsilon \leq W_n^* \leq a_n + \varepsilon)\} = 1.$$

Soit ℓ_n l'isobare contenant a_n . Pour tout $\varepsilon > 0$, il existe $(h_n^\varepsilon)_n$ et $(\bar{h}_n^\varepsilon)_n$ deux suites d'isobares et $\eta > 0$ tels que pour tout θ (19) soit satisfaite. Les inclusions (20) et (21) montrent que sous l'hypothèse (H), $(X_n^*)_n$ est presque sûrement stable si et seulement si $(W_n^*)_n$ est presque sûrement stable. De plus, d'après [Geffroy, 1958, 1959] (26) est une condition nécessaire et suffisante pour la stabilité de $(W_n^*)_n$, d'où le résultat.

Remarque 6.- Il est aussi possible de donner une preuve directe de ce résultat sans utiliser la suite $(W_n^*)_n$. Cette démonstration est donnée dans l'appendice 2 ; elle est plus longue, cependant elle met en évidence les conditions de majoration et de minoration presque sûre de la définition (25), [Geffroy, 1958, 1959].

Soit f_θ la densité de F_θ .

Corollaire 2.- Si (H) est satisfaite, s'il existe θ_1 tel que :

$$(27) \quad \lim_{x \rightarrow +\infty} \frac{f_{\theta_1}(x)}{G_{\theta_1}(x) \text{Log} |\text{Log} G_{\theta_1}(x)|} = +\infty,$$

alors $(X_n^*)_n$ est presque sûrement stable.

Pour la démonstration, voir [Geffroy, 1958,1959].

V - EXEMPLES

Dans ce paragraphe, on suppose que $k = 2$ et on utilise les coordonnées polaires dans \mathbb{R}^2 .

1. **Exemple 1.-** Dans ce premier exemple, $F_\theta(x) = (1 - e^{-\alpha(\theta)x^m}) \mathbb{1}_{\{x > 0\}}$ où $m > 0$ et α est une fonction continue strictement positive sur $[0, 2\pi]$ telle que $\alpha(0) = \alpha(2\pi)$. Alors $f_\theta(x) = m \alpha(\theta) x^{m-1} e^{-\alpha(\theta)x^m} \mathbb{1}_{\{x > 0\}}$.

Pour θ_1 fixé et pour tout $w > 0$, l'isobare $g(\theta, w)$ est définie par :

$$g(\theta, w) = \left(\frac{\alpha(\theta_1)}{\alpha(\theta)} \right)^{1/m} w.$$

Si bien que (H) est satisfaite.

Cependant (27) n'est vraie que pour $m > 1$ car :

$$\frac{f_\theta(x)}{G_\theta(x) \text{Log} |\text{Log} G_\theta(x)|} = \frac{m \alpha(\theta) x^{m-1}}{\text{Log} (\alpha(\theta) x^m)}.$$

Ainsi $(X_n^*)_n$ est presque sûrement stable pour $m > 1$.

2. **Exemple 2.** Supposons que $m = 1$, donc $F_\theta(x) = (1 - e^{-\alpha(\theta)x}) \mathbb{1}_{\{x > 0\}}$ (loi exponentielle). Les conditions du théorème 4 ne sont pas satisfaites : $(X_n^*)_n$ n'est pas stable en probabilité et donc n'est pas stable presque sûrement. Cet exemple suggère d'étudier une notion plus faible que la stabilité : la φ -stabilité. Avant de développer la φ -stabilité, nous examinons un peu plus précisément dans le paragraphe suivant la condition (H).

VI - QUELQUES REMARQUES A PROPOS DE (H).

Tout d'abord, on peut montrer que (H) n'est pas une condition nécessaire pour la stabilité de $(X_n^*)_n$; pour cela il suffit de considérer l'exemple suivant :

1. **Exemple 3.-** Supposons que (X_1, \dots, X_n) est un échantillon issu de la fonction de répartition multidimensionnelle :

$$(28) \quad F_\theta(x) = (1 - e^{-x^{\alpha(\theta)}}) 1_{\{x > 0\}},$$

où α est continue sur $[0, 2\pi]$, $\alpha(0) = \alpha(2\pi)$ et $\alpha(\theta) > 0$ pour tout θ .

Les isobares contenant $w > 0$ sont définies par :

$$g(\theta, w) = w^{\frac{\alpha(0)}{\alpha(\theta)}}.$$

a) Il est clair que (H) n'est pas vérifiée dans cet exemple. En fait, même (5) n'est pas satisfaite :

Choisissons $\alpha(\theta) = \pi + \theta$ sur $[0, \pi]$, alors pour $0 < \theta \leq \pi$,

$$(29) \quad \frac{1}{2} \leq \frac{\pi}{\pi + \theta} < 1$$

Supposons (5) vraie; alors pour tout $\varepsilon > 0$ il existe $\eta > 0$ et pour tout $w > 0$ il existe $v(w)$ tels que pour tout θ :

$$(30) \quad w^{\pi/\pi+\theta} < w^{\pi/\pi+\theta} + \eta < (v(w))^{\pi/\pi+\theta} < v^{\pi/\pi+\theta} + \varepsilon.$$

En particulier,

$$w < v(w) < w + \varepsilon.$$

Donc,

$$w^{\pi/\pi+\theta} - (v(w))^{\pi/\pi+\theta} < w^{\pi/\pi+\theta} \left[\left(1 + \frac{\varepsilon}{w}\right)^{\pi/\pi+\theta} - 1 \right],$$

et, pour w suffisamment grand, pour tout $\theta > 0$:

$$w^{\pi/\pi+\theta} - (v(w))^{\pi/\pi+\theta} < w^{\pi/\pi+\theta} \left(\frac{\varepsilon}{w} \frac{\pi}{\pi+\theta} + 0 \left(\frac{1}{w} \right) \right)$$

De (29) on déduit :

$$\lim_{n \rightarrow +\infty} (w^{\pi/\pi+\theta} - (v(w))^{\pi/\pi+\theta}) = 0,$$

ce qui contredit (30).

b) Pourtant $(X_n^*)_n$ est presque sûrement stable.

Soit $\beta(\theta) = 1/\alpha(\theta)$, supposons que β ait un maximum $\beta(0) = \beta < 1$. L'équation des isobares est :

$$(31) \quad g(\theta, w) = w^{\beta(\theta)/\beta};$$

comme pour tout θ , $\beta(\theta) \leq \beta < 1$,

$$(32) \quad \sup_{\theta} |w^{\beta(\theta)/\beta} - v^{\beta(\theta)/\beta}| = |w - v|$$

pour w et v suffisamment grands. Par conséquent,

$$(33) \quad \sup_{\theta} |g(\theta, w) - g(\theta, v)| = |g(0, w) - g(0, v)|$$

Notons maintenant W_n^* l'intersection de l'isobare contenant X_n^* et du demi-axe positif

$\theta = 0$; $(W_n^*)_n$ est presque sûrement stable puisque (27) est vérifiée.

Donc, pour tout $\varepsilon > 0$:

$$P \{ \liminf [a_n - \varepsilon < W_n^* < a_n + \varepsilon] \} = 1,$$

où $a_n = G_0^{-1} \left(\frac{1}{n} \right) = (\text{Log } n)^\beta$.

Soit $\varepsilon > 0$, on désigne par $\mathcal{L}_n^{-\varepsilon}(\theta)$ (resp. $\mathcal{L}_n(\theta), \mathcal{L}_n^{\varepsilon}(\theta)$) l'isobare contenant $a_n - \varepsilon$ (resp. $a_n, a_n + \varepsilon$). Comme $\mathcal{L}_n^{-\varepsilon}$ et $\mathcal{L}_n^{\varepsilon}$ sont des isobares,

$$\begin{aligned} 1 &= P \{ \text{Lim inf} [a_n - \varepsilon < W_n^* < a_n + \varepsilon] \} \\ &= P \{ \text{Lim inf} [\mathcal{L}_n^{-\varepsilon}(\Theta_n^*) < R_n^* < \mathcal{L}_n^{\varepsilon}(\Theta_n^*)] \} . \end{aligned}$$

De plus, d'après (33) ,

$$\lim_{n \rightarrow \infty} \sup_{\theta} | \mathcal{L}_n^{-\varepsilon}(\theta) - \mathcal{L}_n(\theta) | = \varepsilon ,$$

et

$$\lim_{n \rightarrow \infty} \sup_{\theta} | \mathcal{L}_n^{\varepsilon}(\theta) - \mathcal{L}_n(\theta) | = \varepsilon .$$

Donc,

$$R_n^* - \mathcal{L}_n(\Theta_n^*) \xrightarrow{\text{p.s.}} 0 ,$$

et $(X_n^*)_n$ est presque sûrement stable.

2.- A partir de cet exemple nous pouvons donner une autre condition suffisante de stabilité pour $(X_n^*)_n$. Supposons qu'il existe θ_1 tel que pour tout $w > 0$ et pour tout $\eta > 0$:

$$(K) \quad \sup_{\theta} (g(\theta, w + \eta) - g(\theta, w - \eta)) = g(\theta_1, w + \eta) - g(\theta_1, w - \eta) .$$

Soit W_n^* l'intersection de l'isobare contenant X_n^* et du demi-axe positif $0\theta_1^+$.

De même que dans l'exemple 3 on peut prouver le théorème suivant.

Théorème 6. *Si (K) est satisfaite et si $(W_n^*)_n$ est presque sûrement stable, alors $(X_n^*)_n$ est presque sûrement stable.*

3.- Les conditions (H) et (K) sont de nature différente : (H) est une condition d'uniformité tandis que (K) utilise l'existence d'une direction θ_1 , qui peut être interprétée comme la direction la moins favorable. Il est clair que (H) et (K) ne sont pas équivalentes (voir l'exemple 3). Bien sûr si $g(\theta, w) = k(\theta)w + l(\theta)$, alors (H) implique (K). On peut remarquer aussi que le théorème 4 est encore vrai sous la condition (H).

VII - φ -STABILITE

1.- Soit φ une fonction définie sur \mathbb{R}^+ , positive, de classe C^1 , croissante et bijective. Dans ce paragraphe on considère les variables :

$$X_1^\varphi = (\varphi(R_1), \Theta_1), \dots, X_n^\varphi = (\varphi(R_n), \Theta_n).$$

On note pour tout θ , f_θ^φ la densité conditionnelle de $\varphi(R)$ étant donné $\Theta = \theta$ et

$F_\theta^\varphi = 1 - G_\theta^\varphi$ la fonction de répartition conditionnelle de $\varphi(R)$ étant donné $\Theta = \theta$.

Pour tout $t > 0$,

$$(34) \quad F_\theta^\varphi(t) = F_\theta(\varphi^{-1}(t))$$

$$(35) \quad f_\theta^\varphi(t) = (\varphi^{-1}(t))' \cdot f_\theta(\varphi^{-1}(t)).$$

On désigne par $X_\theta^{\varphi*} = ((\varphi(R_n))^*, \Theta_n^*)$ la valeur maximale de $(X_1^\varphi, \dots, X_n^\varphi)$; alors,

$$X_n^{\varphi*} = (\varphi(R_n^*), \Theta_n^*).$$

Il est clair que si $(g_n)_n$ est une suite d'isobares de niveau u pour la loi de (R, Θ) , alors $(\varphi(g_n))_n$ est une suite d'isobares de niveau u pour la loi de $(\varphi(R), \Theta)$ et réciproquement.

Cette remarque suggère la définition suivante qui généralise la définition de la stabilité relative donnée par Geffroy dans [Geffroy, 1958, 1959] et par Gnedenko dans [Gnedenko, 1943].

Définition 3.- La suite $(X_n^*)_n$ est dite φ -stable en probabilité (presque sûrement) s'il existe une suite d'isobares $(g_n)_n$ (pour la loi de (R, Θ)) telle que :

$$(36) \quad R_n^* - g_n(\Theta_n^*) \xrightarrow[p.s.]{P} 0.$$

Remarque 7.- Quand $\varphi(x) = \widehat{\text{Log}}(x) = \text{Max}(0, \text{Log}x)$, la φ -stabilité n'est autre que la stabilité relative, [Geffroy, 1958,1959], [Gnedenko, 1943], [Green,1976] ; (36) peut alors s'écrire :

$$(37) \quad \frac{R_n^*}{g_n(\Theta_n^*)} \xrightarrow[p.s.]{P} 1 .$$

2.- Tous les résultats des paragraphes III et IV peuvent être utilisés pour l'ensemble des points $(X_1^\varphi, X_2^\varphi, \dots, X_n^\varphi)$:

a) L'équation de l'isobare contenant le point $v > 0$ du demi-axe $0\theta_1^+$ est :

$$(38) \quad g^\varphi(\theta, v) = \varphi(g(\theta, \varphi^{-1}(v))) .$$

b) La condition (H) devient $(H)^\varphi$; par exemple, si $\varphi = \widehat{\text{Log}}$, $(H)^\varphi$ s'écrit:

$$(\hat{H}) \quad \alpha_1 \leq \frac{w \frac{\partial g(\theta, w)}{\partial w}}{g(\theta, w)} \leq \beta_1 .$$

c) Les théorèmes 4 et 5 peuvent être utilisés en remplaçant $G_\theta, f_\theta, \dots$ par $G_\theta^\varphi, f_\theta^\varphi, \dots$

d) Bien sûr, si $(X_n^*)_n$ est stable, les propriétés de stabilité subsistent pour $(X_n^{\varphi*})_n$ si φ est une fonction concave, [Gather and Rauhut, 1990].

3.- Exemples

a) **Exemple 4 :** loi de Cauchy

Supposons que E_n est un échantillon tel que :

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

où λ est une fonction continue et positive .

Alors,

$$F_{\theta}(x) = \frac{2}{\pi} \operatorname{Arctg} \left(\frac{x}{\lambda(\theta)} \right) \mathbf{1} \{x > 0\} .$$

Comme la distribution de Cauchy ne possède pas de moments, $(X_n^*)_n$ n'est pas stable, [Geffroy, 1958, 1959].

Cependant, pour tout θ et pour x suffisamment grand,

$$F_{\theta}(x) \sim 1 - \frac{2}{\pi} \frac{\lambda(\theta)}{x} .$$

Ainsi les conditions du théorème 4 sont satisfaites par G_{θ}^{φ} si et seulement si :

$$(39) \quad \lim_{x \rightarrow +\infty} \frac{\varphi^{-1}(x-h)}{\varphi^{-1}(x)} = 0, \text{ pour tout } h > 0 .$$

Il suffit de choisir $\varphi(x) = \sqrt{\operatorname{Log} x}$ ou $\varphi(x) = \operatorname{Log} \operatorname{Log} x$ qui est une fonction plus concave que $\sqrt{\operatorname{Log} x}$. De plus, pour ces deux fonctions, (H^{φ}) est vraie pour v suffisamment grand. Ainsi $(X_n^*)_n$ est φ -stable en probabilité. Remarquons que pour $0 < \alpha < 1$, la fonction $\varphi(x) = (\operatorname{Log} x)^{\alpha}$ convient également ; cependant $(X_n^*)_n$ n'est pas relativement stable.

b) Exemple 5 : loi exponentielle

Supposons que E_n est un échantillon tel que :

$$F_{\theta}(x) = (1 - e^{-\alpha(\theta)x}) \mathbf{1} \{x > 0\} ;$$

α est une fonction continue et positive .

De même, pour une loi exponentielle, la condition sur φ est :

$$(40) \quad \lim_{x \rightarrow +\infty} \varphi^{-1}(x-h) - \varphi^{-1}(x) = -\infty, \text{ pour tout } h > 0 .$$

Les fonctions $\varphi(x) = \sqrt{x}$ ou $\varphi(x) = \widehat{\operatorname{Log}}(x)$ conviennent et $(H)^{\varphi}$ est aussi vérifiée.

c) Exemple 6.

On considère l'exemple 1 avec $0 < m \leq 1$. Choisissons $\varphi(x) = x^{1/2m}$, alors $F_{\theta}^{\varphi}(x) = (1 - e^{-\alpha(\theta)x^2}) \mathbb{1}_{\{x > 0\}}$ et $(X_n^{\varphi*})_n$ est presque sûrement stable, comme on l'a vu dans l'exemple 1.

d) Exemple 7.

Supposons que E_n est un échantillon tel que :

$$F_{\theta}(x) = \left(1 - \frac{\lambda(\theta)}{x^{\alpha(\theta)}}\right) \mathbb{1}_{\{x > 0\}} ;$$

λ et α sont des fonctions continues et positives. (Pour $\alpha(\theta) = 1$, on obtient asymptotiquement une loi de Cauchy). Choisissons $\varphi(x) = \sqrt{\text{Log } x}$; alors ,

$$F_{\theta}^{\varphi}(x) = (1 - \lambda(\theta) e^{-\alpha(\theta)x^2}) \mathbb{1}_{\{x > 0\}} ,$$

et l'exemple 1 montre que $(X_n^{\varphi*})_n$ est presque sûrement stable.

VIII - CONCLUSION

Les résultats de ce chapitre nous amènent à considérer l'aspect géométrique de la stabilité : dans le chapitre suivant, nous décrivons comment les points de l'échantillon se répartissent dans un domaine non convexe limité par les isobares γ_n . En utilisant la notion d'isobares, il est possible d'étudier la localisation asymptotique des points d'un échantillon, sans hypothèse de convexité.

BIBLIOGRAPHIE

- Billingsley, P.**, (1968), Convergence of Probability Measures. *Wiley*.
- Fisher, L.**, (1966), The convex hull of a sample, *Bull. Am. Math. Soc.* 72, 555-558 .
- Gather, U. and Rauhut, B.**, (1990), The outlier behaviour of probability distributions, *Journal of statistical planning and Inference* 26, 237-252.
- Geffroy, J.**, (1958,1959), Contribution à la théorie des valeurs extrêmes, *Pub. Inst. Stat. Univ. Paris*, 1958 (fasc. 4) et 1959 (fasc. 1).
- 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, B.V.**, (1943), Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423-453.
- Green, R.F.**, (1976), Outlier-prone and outlier-resistant distributions, *J. Amer. Statist. Assoc.* 71, 502-505.
- Haiman, G., Puri, M.**, (1990), A strong invariance principle concerning the J-upper order statistics for stationary m-dependent sequences, *Journal of Statistical Planning and inference*.
- Haiman, G., Puri, M.**, A strong invariance principle concerning the J-upper order of statistics for stationary gaussian sequences" (à paraître dans *Annals of Probability*).
- Mathar, R.**, (1989), The outlier-behaviour of distributions and the decay power of optimal tests at decreasing level. *Statistics* 20 (2), 247-254.
- Muñoz-Garcia, J., Moreno-Rebollo, J.L. and Pascual-Acosta, A.**, (1990), Outliers : A formal approach. *International Statistical Review*, 58, 3, 215-226.
- Lecoutre, J.P.**, (1982), Contribution à l'estimation non-paramétrique de la régression. *Thèse, Université Pierre et Marie Curie, Paris*.
- Davis, R., Mulrow, E., Resnick, S.**, (1987), The convex hull of a random sample in \mathbb{R}^2 . *Comm. statist. - stochastic models* ; 3(1), 1-27.

**Chapitre 3 : Localisation
asymptotique des échantillons
multidimensionnels**

I. INTRODUCTION

Dans ce nouveau chapitre, nous poursuivons notre travail en nous attachant à décrire avec une certaine précision le comportement asymptotique du nuage formé par un échantillon multidimensionnel. Geffroy ([6]) et Fisher ([3]) furent à notre connaissance les premiers à traiter ce genre de problème. Comme référence plus récente, nous pouvons citer Davis, Mulrow, Resnick ([1]), Van Wel ([11]).

En fait, tous ces auteurs ont travaillé essentiellement sur l'enveloppe convexe des échantillons. Pour ce qui nous concerne, nous avons tenté de définir et d'étudier des sortes d'enveloppes non nécessairement convexes. Les résultats obtenus dépendent de propriétés que toutes les lois de probabilité multidimensionnelles ne possèdent pas. Pour aller plus loin, nous exploitons alors la notion déjà ancienne de stabilité relative, en la généralisant ([7], [5],[8],[2]). Nous remplaçons chaque observation X par une variable X_φ de coordonnées polaires $(\varphi(R), \Theta)$, où φ est une fonction concave bien choisie, et nous voyons ce qu'il advient des propriétés de localisation asymptotique du nuage de points ainsi obtenu.

On rappelle ci-dessous les notations du chapitre 2. On considère des variables aléatoires, définies sur un espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans l'espace euclidien \mathbf{R}^k muni de la norme euclidienne $\|\cdot\|$. Soit X une telle variable aléatoire; si $\|X(\omega)\|$ est non nul, on pose :

$$(1) \quad R(\omega) = \|X(\omega)\|$$

$$(2) \quad \Theta(\omega) = \frac{X(\omega)}{\|X(\omega)\|}$$

La sphère unité S^{k-1} de \mathbf{R}^k est munie de la topologie induite par celle de \mathbf{R}^k .

Notre étude nécessite un certain nombre d'hypothèses sur la loi du couple (R, Θ) , que nous faisons une fois pour toutes: la loi de Θ est donnée par une densité continue $j(\theta)$. Pour tout θ , la loi conditionnelle de R étant donné $\Theta = \theta$ est donnée par une densité continue $f_\theta(\rho)$, ou par la fonction de répartition conditionnelle de R étant donné $\Theta = \theta$, $F_\theta(\rho) = P\{R \leq \rho | \Theta = \theta\} = 1 - G_\theta(\rho)$. On suppose que F_θ est bijective. Pour tout $0 < u < 1$, on appelle *isobare de niveau u* , de la loi de R étant donné $\Theta = \theta$, l'application $\theta \rightarrow F_\theta^{-1}(u)$. On suppose que cette application est continue et on appelle aussi isobare la surface de \mathbf{R}^k d'équation $\rho = F_\theta^{-1}(u)$.

Soit $E_n = (X_1, \dots, X_n)$ un échantillon de variables aléatoires indépendantes et de même loi que X . Pour tout $1 \leq k \leq n$, il existe presque sûrement une unique isobare de la loi de R étant donné $\Theta = \theta$ contenant $X_k = (R_k, \Theta_k)$. Comme dans le chapitre 2, nous définissons la valeur maximale de E_n comme le point $X_n^* = (R_n^*, \Theta_n^*)$ de l'échantillon contenu dans l'isobare de plus haut niveau.

Les lois de (R, Θ) et de (R_n^*, Θ_n^*) possèdent le même ensemble d'isobares, ce qui a justifié dans le chapitre précédent la définition suivante : la suite d'extrêmes $(X_n^*)_n$ est **stable** en probabilité (p.s.) s'il existe une suite (g_n) d'isobares telle que $(R_n^* - g_n(\Theta_n^*))$ converge en probabilité (p.s.) vers zéro. Cela traduit la tendance de $(X_n^*)_n$ à se rapprocher d'une certaine surface.

La notion de localisation asymptotique de E_n permet une description plus précise du comportement des extrêmes : nous dirons que E_n est **localisé asymptotiquement** si, pour tout $\varepsilon > 0$, il existe deux suites (h_n^ε) et $(\tilde{h}_n^\varepsilon)$ d'isobares délimitant des domaines

$$C_n^\varepsilon = \{(\rho, \theta) : h_n^\varepsilon(\theta) - \varepsilon < \tilde{h}_n^\varepsilon(\theta) < \rho < h_n^\varepsilon(\theta)\}$$

et si chacun de ces domaines peut être lui-même partagé en sous-ensembles $C_{n,i}^\varepsilon$ ($i = 1, \dots, \mu_n$) tels que :

- a) $\mu_n \rightarrow \infty$
- b) $\lim_{n \rightarrow \infty} \sup_i \lambda^k(C_{n,i}^\varepsilon) = 0$ (λ^k désigne la mesure de lebesgue sur \mathbf{R}^k .)
- c) $P(\cup_{r=1}^\infty \cap_{n=r}^\infty \{X_n^* \in C_n^\varepsilon\}) = 1$
- d) $P(\cup_{r=1}^\infty \cap_{n=r}^\infty \cap_{i=1}^{\mu_n} \{E_n \cap C_{n,i}^\varepsilon \neq \emptyset\}) = 1$

Cet ensemble de propriétés exprime le fait que, asymptotiquement, les points extrémaux de E_n remplissent régulièrement les couronnes C_n^ε . Il s'agit d'une notion plus forte que la stabilité presque sûre, exprimée ici par la seule condition c). Evidemment de nombreuses lois de probabilité ne conduisent pas à de telles propriétés. L'idée de φ -stabilité déjà utilisée dans le chapitre 2 permet alors, dans de nombreux cas, de donner quand même une réponse intéressante à ce type de question. Il s'agit de remplacer l'échantillon initial E_n par un nouvel ensemble de points $E_n^\varphi = ((\varphi(R_1), \Theta_1); \dots; (\varphi(R_n), \Theta_n))$, où φ est une fonction définie sur \mathbf{R}^+ , positive, de classe C_1 , strictement croissante, de façon que E_n^φ présente des propriétés de stabilité et de localisation asymptotique. Intuitivement, φ doit être "suffisamment concave" pour ramener les extrêmes de E_n à un tel

comportement. En fait, pour $\varphi(x) = \text{Max}(0, \text{Log } x) = \widehat{\text{Log}} x$, on retrouve la notion de stabilité relative introduite par Gnedenko [7], et Geffroy ([5],[6]) pour un échantillon unidimensionnel.

II. PRELIMINAIRES.

Nous avons utilisé à plusieurs reprises dans le chapitre 2, des hypothèses de régularité pour les isobares, afin de donner des conditions suffisantes de stabilité. Nous les reprenons dans ce chapitre.

a) Fixons θ_1 dans S^{k-1} et un vecteur unité $\overrightarrow{0\theta_1}$ sur la droite $0\theta_1$; pour tout point w , (on désignera aussi par w l'abscisse de ce point), sur le demi axe positif $0\theta_1^+$, il existe une unique isobare de niveau $u(w)$ contenant w . Soit $\rho = g(\theta, w)$ l'équation de cette isobare, (remarquons que $g(\theta_1, w) = w$). La condition (H) introduite dans le chapitre 2 sera parfois utile dans la suite :

(H) Il existe $0 < \alpha_1 \leq \beta_1 < +\infty$ tels que pour tout θ et pour tout $w > 0$:

$$\alpha_1 \leq \frac{\partial g}{\partial w}(\theta, w) \leq \beta_1.$$

Il est alors aisé de déduire de (H) la propriété suivante des isobares :

Pour tout $\varepsilon > 0$, il existe $\eta > 0$, et pour tout $w > 0$, il existe deux isobares $h^\varepsilon(\theta, w)$ et $\tilde{h}^\varepsilon(\theta, w)$ tels que pour tout θ :

$$(1) \quad g(\theta, w) - \varepsilon < \tilde{h}^\varepsilon(\theta, w) < g(\theta, w) - \eta < g(\theta, w) + \eta < h^\varepsilon(\theta, w) < g(\theta, w) + \varepsilon$$

b) Le théorème suivant, démontré dans le chapitre 2 ([2],[5]), donne une condition suffisante de stabilité presque sûre pour $(X_n^*)_n$.

Théorème A : Si (H) est satisfaite et s'il existe θ_1 tel que

$$(2) \quad \lim_{x \rightarrow +\infty} \frac{f_{\theta_1}(x)}{G_{\theta_1}(x) \text{Log} |\text{Log} G_{\theta_1}(x)|} = +\infty, \text{ alors } (X_n^*)_n$$

est stable presque sûrement.

c) Nous avons aussi montré dans le chapitre 2 que la condition (H) n'est pas nécessaire à la stabilité de $(X_n^*)_n$ (cf. exemple 4). Une autre hypothèse peut assurer la stabilité presque sûre de $(X_n^*)_n$: nous l'appellons la condition (K).

Il existe θ_1 tel que pour tout $w > 0$ et pour tout $\eta > 0$:

$$(K) \quad \sup_{\theta} (g(\theta, w + \eta) - g(\theta, w - \eta)) = g(\theta_1, w + \eta) - g(\theta_1, w - \eta).$$

Soit W_n^* l'intersection de l'isobare contenant X_n^* et du demi-axe positif $0\theta_1^+$.

Théorème B : *Supposons la condition (K) vérifiée, si de plus $(W_n^*)_n$ est presque sûrement stable, alors $(X_n^*)_n$ est presque sûrement stable.*

On a déjà remarqué que les conditions (K) et (H) étaient de nature différente ; (H) étant plutôt une hypothèse d'uniformité alors que (K) privilégie une certaine direction. Nous supposons dans la suite du chapitre que $(X_n^*)_n$ est stable presque sûrement.

III. LOCALISATION ASYMPTOTIQUE DE L'ECHANTILLON.

La suite des valeurs maximales $(X_n^*)_n$ étant stable presque sûrement, pour tout $\varepsilon > 0$,

$$(3) \quad P\{\liminf(X_n^* \in \Gamma_n^\varepsilon - \Gamma_n^{-\varepsilon})\} = 1$$

où $\Gamma_n^\varepsilon = \{(\rho, \theta) : \rho \leq \gamma_n(\theta) + \varepsilon\}$ et $\Gamma_n^{-\varepsilon} = \{(\rho, \theta) : \rho \leq \gamma_n(\theta) - \varepsilon\}$, où, comme on l'a montré dans le chapitre 2 :

$$(4) \quad \gamma_n(\theta) = G_\theta^{-1}\left(\frac{1}{n}\right) \quad \text{pour } n \geq 2.$$

Les résultats de localisation asymptotique sont basés sur l'étude de la répartition des points de l'échantillon qui appartiennent au domaine limité par des surfaces d'équation $\gamma_n(\theta) + \varepsilon$ et $\gamma_n(\theta) - \varepsilon$ (cf. exemples 1,2,3) ou par des isobares adéquates (cf. exemple 4).

Ordonnons l'échantillon selon les niveaux croissants :

$$X_1^{n*}, X_2^{n*}, \dots, X_n^{n*} = X_n^*.$$

Pour $1 \leq \alpha \leq n$, on note $\bar{X}_n^\alpha = (\bar{R}_n^\alpha, \bar{\Theta}_n^\alpha) = X_{n-\alpha+1}^{n*}$; en particulier : $\bar{X}_n^n = X_n^*$.

1) **Théorème 1.**

a) $\bar{\Theta}_n^\alpha$ et Θ ont même loi.

b) Soit g une isobare de niveau u pour la loi de (R, Θ) :

$$P(\bar{R}_n^\alpha < g(\bar{\Theta}_n^\alpha)) = \sum_{k=0}^{\alpha-1} C_n^k u^{n-k} (1-u)^k.$$

Les lois de $(\bar{R}_n^\alpha, \bar{\Theta}_n^\alpha)$ et de (R, Θ) définissent donc le même ensemble d'isobares.

c) La fonction de répartition conditionnelle de \bar{R}_n^α étant donné $\bar{\Theta}_n^\alpha = \theta$ vérifie :

$$(5) \quad d\bar{F}_{\theta, \alpha, n}(z) = \frac{n!}{(n-\alpha)!(\alpha-1)!} F_\theta^{n-\alpha}(z) (1-F_\theta(z))^{\alpha-1} dF_{\theta(z)}.$$

La démonstration est identique à celle du théorème 1 du chapitre 2.

Soit $(\gamma_n)_n$ la suite d'isobares de niveau $1 - \frac{1}{n}$, donnée par (4). On suppose que (H) est réalisée. Pour $\varepsilon > 0$, il existe $\eta > 0$ et deux suites d'isobares $(\tilde{h}_n^\varepsilon)_n$ et $(h_n^\varepsilon)_n$ tels que pour tout n et pour tout θ :

$$(6) \quad \gamma_n(\theta) - \varepsilon/2 < \tilde{h}_n^\varepsilon(\theta) < \gamma_n(\theta) - \eta < \gamma_n(\theta) < \gamma_n(\theta) + \eta < h_n^\varepsilon(\theta) < \gamma_n(\theta) + \varepsilon/2$$

Puisque $(h_n^\varepsilon)_n$ est une isobare, (5) montre que $\bar{F}_{\theta, \alpha, n}(\tilde{h}_n^\varepsilon(\theta))$ ne dépend pas de θ .

On note comme dans l'introduction C_n^ε le domaine compris entre \tilde{h}_n^ε et h_n^ε et ν_n^ε le nombre de points de l'échantillon situés dans C_n^ε .

Le lemme 1 donne une minoration asymptotique presque sûre de ν_n^ε ; pour cela on utilise la définition suivante :

2) Définition 1. Soit $(D_n)_n$ une suite croissante d'ensembles ordonnés par inclusion ; $\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\}$ est presque complètement asymptotiquement inclus dans D_n^c , ($\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\} \subset D_n^c$ p.co.a), si

$$(7) \quad \sum_{n=1}^{\infty} P\left[\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\} \not\subset D_n^c\right] < \infty$$

3) Lemme 1. Soit $\varepsilon > 0$. S'il existe une suite $(\alpha_n)_n$ telle que $\lim_{n \rightarrow +\infty} \alpha_n = +\infty$ et telle que $\sum_{n=2}^{\infty} \bar{F}_{\theta, \alpha, n}(\tilde{h}_n^\varepsilon(\theta))$ converge, alors $\nu_n^\varepsilon \geq \alpha_n$ presque sûrement asymptotiquement.

■ Comme on l'a déjà remarqué, $q_n = \bar{F}_{\theta, \alpha_n, n}(\tilde{h}_n^\varepsilon(\theta))$ ne dépend pas de θ , puisque \tilde{h}_n^ε est une isobare. Posons $\tilde{\Gamma}_n^\varepsilon = \{(\rho, \theta) : \rho \leq \tilde{h}_n^\varepsilon(\theta)\}$. Si $\sum_{n=2}^\infty q_n$ converge alors $\{X_{n-\alpha_n+1}^{n*}, \dots, X_n^*\} \subset (\tilde{\Gamma}_n^\varepsilon)^c$ p.co.a. Comme la suite $(X_n^*)_n$ est stable, asymptotiquement il n'y a presque sûrement pas de point de E_n en dehors du domaine limité par h_n^ε ; si bien que $\nu_n^\varepsilon \geq \alpha_n$ presque sûrement asymptotiquement. ■

On suppose de plus que la densité j de Θ vérifie :

$$(8) \quad m = \inf_{\theta \in S^{k-1}} j(\theta) > 0$$

On note : $M = \sup_{\theta \in S^{k-1}} j(\theta)$.

4) Théorème 2. *On suppose que les conditions du lemme 1 sont vérifiées. Soit $(\mu_n)_n$ une suite telle que $\lim_{n \rightarrow +\infty} \mu_n = +\infty$ et telle que $\sum \mu_n (1 - \frac{1}{\mu_n})^{\alpha_n}$ converge. Alors pour tout $\varepsilon > 0$, C_n^ε peut être partagé en μ_n parties telles que asymptotiquement chaque partie contienne presque sûrement, au moins un point de E_n .*

■ Soit $\varepsilon > 0$ et (P_1, \dots, P_{μ_n}) une partition de S^{k-1} en μ_n ensembles de même aire ([6]).

Posons :

$$C_{n,i}^\varepsilon = \{(\rho, \theta) \in C_n^\varepsilon; \theta \in P_i\}, \quad i = 1, \dots, \mu_n.$$

Alors

$$C_n^\varepsilon = \bigcup_{i=1}^{\mu_n} C_{n,i}^\varepsilon.$$

Pour tout $i = 1, \dots, \mu_n$,

$$\frac{m}{M\mu_n} \leq P(X \in C_{n,i}^\varepsilon | X \in C_n^\varepsilon) \leq \frac{M}{m\mu_n}.$$

Soit Q_n l'événement réalisé si chaque partie $C_{n,i}^\varepsilon (i = 1, \dots, \mu_n)$ contient au moins un point de E_n , et soit :

$$(9) \quad \mathcal{E}_N = \bigcap_{n > N} Q_n$$

$$(10) \quad \mathcal{H}_N = \bigcap_{n > N} \{\nu_n^\varepsilon > \alpha_n\}, \quad \text{pour } n > N.$$

L'inégalité de Boole fournit :

$$(11) \quad P\{\mathcal{E}_N | \mathcal{H}_N\} > 1 - \sum_{n>N} P\{Q_n^c | \mathcal{H}_N\}.$$

Il est clair que pour $n > N$, Q_n ne dépend de $\nu_N^\varepsilon, \nu_{N+1}^\varepsilon, \dots$ qu'à travers ν_n^ε .
Si bien que

$$P(Q_n^c | \nu_N^\varepsilon, \nu_{N+1}^\varepsilon, \dots) = P(Q_n^c | \nu_n^\varepsilon) \leq \mu_n \left(1 - \frac{m}{M\mu_n}\right)^{\nu_n^\varepsilon}.$$

Donc

$$P(Q_n^c | \mathcal{H}_N) \leq \mu_n \left(1 - \frac{m}{M\mu_n}\right)^{\alpha_n}.$$

Et, puisque les conditions du lemme 1 sont satisfaites,

$$\lim_{N \rightarrow +\infty} P\{\mathcal{H}_N\} = 1.$$

Ainsi, $\lim_{N \rightarrow +\infty} P(\mathcal{E}_N) = 1$ si la série $\sum \mu_n \left(1 - \frac{1}{\mu_n}\right)^{\alpha_n}$ converge. ■

Une condition suffisante pour que le volume des $C_{n,i}^\varepsilon$ ($i = 1, \dots, \mu_n$) tende vers zéro est donnée dans le théorème 3; λ^k désigne la mesure de Lebesgue dans \mathbf{R}^k .

5) Théorème 3. Si $\lim_{n \rightarrow +\infty} \sup_{\theta \in S^{k-1}} \left(\frac{(\gamma_n(\theta))^{k-1}}{\mu_n} \right) = 0$, alors pour ε fixé, $\lim_{n \rightarrow +\infty} \lambda^k(C_{n,i}^\varepsilon) = 0$ pour $i = 1, 2, \dots, \mu_n$.

■ Comme $C_{n,i}^\varepsilon = \{x = (\rho, \theta) : \theta \in P_i; \tilde{h}_n^\varepsilon(\theta) \leq \rho \leq h_n^\varepsilon(\theta)\}$, alors $\lambda^k(C_{n,i}^\varepsilon) \leq \lambda^k\{x = (\rho, \theta) : \theta \in P_i; \gamma_n(\theta) - \varepsilon \leq \rho \leq \gamma_n(\theta) + \varepsilon\}$. Si on note $\gamma_n^* = \sup_{\theta \in S^{k-1}} \{\gamma_n(\theta)\}$, il existe $c(k)$ tel que :

$$\lambda^k(C_{n,i}^\varepsilon) \leq \frac{\varepsilon c(k)}{\mu_n} (\gamma_n^* + \varepsilon)^{k-1}.$$

■

6) Remarque 1. La condition (K) permet aussi d'obtenir les résultats précédents. Supposons qu'il existe θ_1 satisfaisant (K). Posons $a_n = G_{\theta_1}^{-1}\left(\frac{1}{n}\right)$; ε étant fixé, on pose :

$$h_n^\varepsilon(\theta) = g(\theta, w_n) \quad \text{isobare passant par } w_n = a_n + \varepsilon/2$$

$$\tilde{h}_n^\varepsilon(\theta) = g(\theta, v_n) \quad \text{isobare passant par } v_n = a_n - \varepsilon/2$$

on obtient donc :

$$\sup_{\theta} |h_n^\varepsilon(\theta) - \tilde{h}_n^\varepsilon(\theta)| = g(\theta_1, w_n) - g(\theta_1, v_n) = \varepsilon.$$

Si bien que le domaine limité par \tilde{h}_n^ε et h_n^ε peut s'écrire encore :

$$C_n^\varepsilon = \{(\rho, \theta) : h_n^\varepsilon(\theta) - \varepsilon < \tilde{h}_n^\varepsilon(\theta) < \rho < h_n^\varepsilon(\theta)\}.$$

Si W_i , ($i = 1, \dots, n$), désigne l'intersection de $0\theta_1^+$ et de l'isobare contenant X_i , la variable aléatoire W_n^* , intersection de $0\theta_1^+$ et de l'isobare contenant X_n^* , peut être considérée comme le maximum de variables indépendantes W_1, \dots, W_n de fonction de répartition F_{θ_1} . Le lemme 1 montre alors que si le nombre de points W_i ($i = 1, \dots, n$) situés entre $a_n - \varepsilon/2$ et $a_n + \varepsilon/2$ sur l'axe $0\theta_1^+$ est presque sûrement asymptotiquement minoré par α_n , alors il en est de même pour le nombre de points de E_n situés dans C_n^ε .

On peut alors de la même façon que précédemment démontrer les théorèmes 2 et 3.

IV. EXEMPLES

1) Exemple 1.

On suppose que la loi de R étant donné $\Theta = \theta$ est définie dans \mathbf{R}^2 par :

$$(12) \quad F_\theta(x) = (1 - e^{-\alpha(\theta)x^m})\mathbf{1}_{\{x>0\}}$$

où α est une fonction continue, strictement positive telle que $\alpha(0) = \alpha(2\pi)$. On a montré dans le chapitre 2 que si $m > 1$ les conditions (H) et (2) sont satisfaites, si bien que la suite $(X_n^*)_n$ est stable presque sûrement ; on peut choisir :

$$\gamma_n(\theta) = G_\theta^{-1}\left(\frac{1}{n}\right) = \left(\frac{\text{Log } n}{\alpha(\theta)}\right)^{1/m}$$

dans la définition ([2],[5]).

Les conditions du lemme 1 sont satisfaites pour $\alpha_n = [\text{Log } n]^s$ où s est un entier positif arbitraire. La démonstration est semblable à celle du lemme 1 de [6].

Soit $\varepsilon > 0$. Comme $\bar{F}_{\theta, \alpha_n, n}(\tilde{h}_n^\varepsilon(\theta))$ ne dépend pas de θ , on peut supposer θ fixé. D'après (5), on obtient :

$$(13) \quad q_n = \bar{F}_{\theta, \alpha_n, n}(\tilde{h}_n^\varepsilon(\theta)) = B(\alpha_n, n - \alpha_n + 1) \int_{G_n}^1 (1 - \xi)^{n - \alpha_n} \xi^{\alpha_n - 1} d\xi$$

où

$$(14) \quad G_n = G_\theta(\tilde{h}_n^\varepsilon(\theta)) \quad \text{ne dépend pas de } \theta$$

et où

$$(15) \quad B(\alpha_n, n - \alpha_n + 1) = \frac{n!}{(n - \alpha_n)!(\alpha_n - 1)!}$$

La fonction $(1 - \zeta)^{n - \alpha_n} \zeta^{\alpha_n - 1}$ est continue sur $[0, 1]$ et atteint son maximum en $\frac{\alpha_n - 1}{n - 1}$. Pour comparer G_n et $\frac{\alpha_n - 1}{n - 1}$, on compare $\text{Log } G_n$ et $\text{Log}(\frac{\alpha_n - 1}{n - 1})$.

$$(16) \quad \text{Log} \frac{\alpha_n - 1}{n - 1} = -\text{Log } n + s \text{Log Log } n + o(1)$$

Par ailleurs, puisque (H) est satisfaite, d'après (1) il existe $\eta > 0$ tel que pour tout n et pour tout θ :

$$(17) \quad \tilde{h}_n^\varepsilon(\theta) = \gamma_n(\theta) - \varepsilon(n, \theta)$$

avec

$$(18) \quad 0 < \eta < \varepsilon(n, \theta) < \varepsilon/2, \quad \text{pour tout } n \quad \text{et pour tout } \theta.$$

Ainsi

$$\text{Log } G_n = \text{Log}[G_\theta \left(\left(\frac{\text{Log } n}{\alpha(\theta)} \right)^{1/m} - \varepsilon(n, \theta) \right)]$$

et

$$(19) \quad \text{Log } G_n = -\text{Log } n + m\varepsilon(n, \theta)(\alpha(\theta))^{1/m}(\text{Log } n)^{1 - \frac{1}{m}} + O\left((\text{Log } n)^{1 - 2/m}\right).$$

De (16) (18) et (19) on déduit que pour n assez grand ,

$$G_n > \frac{\alpha_n - 1}{n - 1}.$$

On obtient donc :

$$q_n < q'_n = B(\alpha_n, n - \alpha_n + 1)(1 - G_n)^{n - \alpha_n} G_n^{\alpha_n - 1}.$$

Il reste à prouver que $\Sigma q'_n$ converge. Or, quand n tend vers l'infini, on montre que $\text{Log} |\text{Log } q'_n|$ est équivalent à $m\varepsilon(n, \theta)(\alpha(\theta))^{1/m}(\text{Log } n)^{1 - 1/m}$. Comme α est bornée sur le compact $[0, 2\pi]$ par un nombre strictement positif, et d'après (18), q'_n est majorée par le terme général d'une série convergente.

De plus, il est clair que les conditions du théorème 2 sont satisfaites pour $\alpha_n = [\text{Log } n]^s$ et pour $\text{Log } \mu_n = O(\text{Log } \text{Log } n)$, ([6]). Il suffit alors de choisir $\mu_n = [\text{Log } n]^{\frac{1}{m} + \delta}$ ($\delta > 0$) pour que les conditions a) b) c) d) définies dans l'introduction soient réalisées.

2) Exemple 2. (Loi exponentielle)

On suppose ici que $F_\theta(x) = (1 - e^{-\alpha(\theta)x})\mathbf{1}_{\{x>0\}}$. On sait, d'après le chapitre 2, que pour cette loi, $(X_n^*)_n$ n'est pas stable en probabilité, cependant, si l'on choisit $\varphi(x) = \widehat{\text{Log}} x = \text{Max}(0, \text{Log } x)$ (ou $\varphi(x) = \sqrt{x}$), la suite de variables aléatoires $\{X_n^{\varphi*}\}_n = \{(\varphi(R_n^*), \Theta_n^*)\}_n$ est stable presque sûrement. On choisit par la suite $\varphi(x) = \widehat{\text{Log}} x$ ce qui correspond à la stabilité relative presque sûre. Il est possible d'obtenir pour $\hat{E}_n = ((\widehat{\text{Log}}(R_1), \Theta_1), \dots, (\widehat{\text{Log}}(R_n), \Theta_n))$ des propriétés de localisation asymptotique, en utilisant la même méthode que pour l'exemple précédent. La fonction de répartition conditionnelle de $\widehat{\text{Log}} R$ étant donné $\Theta = \theta$ est :

$$(20) \quad \hat{F}_\theta(x) = (1 - e^{-\alpha(\theta)e^x})\mathbf{1}_{\{x>0\}}$$

Puisque la suite $(X_n^*)_n$ est relativement stable presque sûrement,

$$\widehat{\text{Log}} R_n^* - \hat{\gamma}_n(\Theta_n^*) \xrightarrow{p.s.} 0$$

où :

$$\hat{\gamma}_n(\theta) = \hat{G}_\theta^{-1}(1/n) = \text{Log } \text{Log } n - \text{Log } \alpha(\theta) .$$

On vérifie facilement que les isobares définies à partir de (20) satisfont la condition (H); ainsi, pour tout $\varepsilon > 0$, il existe une suite d'isobares $(\hat{h}_n^\varepsilon)_n$ et $\eta > 0$ tels que pour tout n et pour tout θ .

$$\hat{h}_n(\theta) = \hat{\gamma}_n(\theta) - \varepsilon(n, \theta)$$

où

$$0 < \eta < \varepsilon(n, \theta) < \varepsilon/2 \quad \text{pour tout } n \quad \text{et tout } \theta .$$

On ordonne \hat{E}_n selon les niveaux croissants en :

$$\hat{X}_1^n, \dots, \hat{X}_{n-\alpha+1}^n = \hat{X}_n^\alpha, \dots, \hat{X}_n^n = \hat{X}_n^\alpha ;$$

$\hat{F}_{\theta, \alpha_n, n}$ désigne la fonction de répartition conditionnelle de \hat{R}_n^α étant donné $\bar{\Theta}_n^\alpha = \theta$. Le lemme 1 est encore satisfait pour $\alpha_n = [\text{Log } n]^s$ (s entier arbitraire). En effet, comme dans l'exemple 1,

$$q_n = \hat{F}_{\theta, \alpha_n, n}(\hat{h}_n^\varepsilon(\theta)) = B(\alpha_n, n - \alpha_n + 1) \int_{\hat{G}_n}^1 (1 - \xi)^{n - \alpha_n} \xi^{\alpha_n - 1} d\xi,$$

où $\hat{G}_n = \hat{G}_\theta(\hat{h}_n^\varepsilon(\theta)) = e^{-[\varepsilon(n, \theta) \cdot \text{Log } n]}$. Donc $\text{Log } \hat{G}_n = -e^{-\varepsilon(n, \theta)} \text{Log } n$ et pour n assez grand,

$$\hat{G}_n > \left(\frac{\alpha_n - 1}{n - 1} \right).$$

si bien que :

$$q_n \leq B(\alpha_n, n - \alpha_n + 1)(1 - \hat{G}_n)^{n - \alpha_n} \hat{G}_n^{\alpha_n - 1} = q'_n.$$

On verrait facilement que lorsque n tend vers l'infini, $\text{Log } | \text{Log } q'_n |$ est équivalent à $(1 - e^{-\varepsilon(n, \theta)}) \text{Log } n$, et que la série $\Sigma q'_n$ converge. En choisissant $\mu_n = [\text{Log } n]$, les conditions a) b) c) d) de localisation asymptotique sont satisfaites.

Remarque 2. On pourrait traiter de la même façon le cas où $F_\theta(x) = (1 - e^{-\alpha(\theta)x^m}) \mathbf{1}_{\{x > 0\}}$ où $m < 1$: il suffit de choisir $\varphi(x) = x^{\frac{1}{2m}}$, on obtient en effet :

$$F_\theta^\varphi(x) = P\{\varphi(R) < x | \Theta = \theta\} = (1 - e^{-\alpha(\theta)x^2}) \mathbf{1}_{\{x > 0\}},$$

loi pour laquelle les résultats de localisation asymptotique ont été déjà obtenus dans l'exemple 1.

3) Exemple 3. (Loi de Cauchy).

On choisit ici dans \mathbf{R}^2 des variables X_i , de fonction de répartition conditionnelle :

$$F_\theta(x) = 1 - \frac{\lambda(\theta)}{x^{\alpha(\theta)}} \mathbf{1}_{\{x > 1\}},$$

où λ et α sont continues, positives sur $[0, 2\pi]$ et vérifient $\lambda(0) = \lambda(2\pi)$ ainsi que $\alpha(0) = \alpha(2\pi)$.

En prenant en particulier $\alpha(\theta) \equiv 1$, on obtient asymptotiquement la distribution d'une loi de Cauchy.

La loi de Cauchy n'ayant pas de moments, $(X_n^*)_n$ n'est pas stable en probabilité ([5]), de plus, le théorème de Gnedenko ([2],[5],[7]) a montré dans le chapitre 2

que $(X_n^*)_n$ n'est pas relativement stable. Par contre, si l'on choisit φ telle que :

$$\lim_{x \rightarrow +\infty} \frac{\varphi^{-1}(x-h)}{\varphi^{-1}(x)} = 0, \quad \text{pour tout } h > 0,$$

alors $(X_n^*)_n$ est φ stable. (cf le chapitre 2).

Par exemple $\varphi(x) = \sqrt{\text{Log } x}$ convient ($\varphi(x) = \text{Log } \text{Log } x$ convient aussi). On obtient en effet comme nouvelle fonction de répartition conditionnelle de R étant donné $\Theta = \theta$:

$$F_\theta^\varphi(x) = P\{\varphi(R) < x | \Theta = \theta\} = (1 - \lambda(\theta) \cdot e^{-\alpha(\theta)x^2}) \mathbf{1}_{\{x>0\}}.$$

pour laquelle on a déjà obtenu dans le chapitre 2 des résultats de stabilité presque sûre ([2]) et dans l'exemple 1 des résultats de localisation asymptotique.

4) Exemple 4.

On suppose dans cet exemple que les X_i sont des variables aléatoires à valeurs dans \mathbf{R}^2 et que la fonction de répartition conditionnelle de R étant donné $\Theta = \theta$ est :

$$(24), \quad F_\theta(x) = (1 - \bar{e}^{x^{\alpha(\theta)}}) \mathbf{1}_{\{x>0\}}$$

où α est une fonction continue sur $[0, 2\pi]$, vérifiant $\alpha(0) = \alpha(2\pi)$ et $\alpha(\theta) > 0$ pour tout θ . L'étude faite dans le chapitre 2 a montré que la condition (H) n'est pas vérifiée. Cependant, on a constaté que $(X_n^*)_n$ est stable presque sûrement pour $\alpha(0) > 1$ car les conditions du théorème B rappelé dans l'introduction sont satisfaites pour $\theta_1 = 0$. Si l'on définit, comme dans la remarque 1, a_n par $G_{\theta_1}^{-1}(\frac{1}{n}) = (\text{Log } n)^{1/\alpha(0)}$ et si pour $\varepsilon > 0$ fixé, on note $h_n^\varepsilon(\theta)$ (resp. $\tilde{h}_n^\varepsilon(\theta)$) l'isobare passant par $w_n = a_n + \varepsilon/2$ (resp. $v_n = a_n - \varepsilon/2$), le domaine limité par h_n^ε et \tilde{h}_n^ε est de la forme :

$$C_n^\varepsilon = \{(\rho, \theta) : h_n^\varepsilon(\theta) - \varepsilon < \tilde{h}_n^\varepsilon(\theta) < \rho < h_n^\varepsilon(\theta)\}.$$

On vérifie facilement que les conditions du lemme 1 sont satisfaites pour $\alpha_n = [\text{Log } n]^s$ (s entier positif arbitraire.). Le terme $q_n = \bar{F}_{\theta, \alpha_n, n}(\tilde{h}_n^\varepsilon(\theta))$ ne dépend pas de θ , on peut fixer $\theta = \theta_1 = 0$.

$$\begin{aligned} \text{Log } G_n &= -[(\text{Log } n)^{1/\alpha(0)} - \varepsilon]^{\alpha(0)} \\ &= -\{\text{Log } n - \alpha(0)\varepsilon(\text{Log } n)^{1-1/\alpha(0)} + (\text{Log } n) o\left(\frac{1}{(\text{Log } n)^{1/\alpha(0)}}\right)\}. \end{aligned}$$

La comparaison avec (16) montre que $G_n > \frac{\alpha_n - 1}{n - 1}$, pour n suffisamment grand.

Comme dans l'exemple 1, q_n est majoré par un terme q'_n tel que $\text{Log} |\text{Log} q'_n|$ est équivalent à $\alpha(0)\varepsilon(\text{Log } n)^{1-1/\alpha(0)}$ quand n tend vers l'infini. On en déduit que la série $\sum_{n=1}^{\infty} q_n$ est convergente. Pour obtenir la localisation asymptotique de E_n , il suffit de poser $\mu_n = [\text{Log } n]^{1+\delta}$ où $\delta > 0$.

V. CONCLUSION.

Dans les exemples que nous avons traités, il a toujours été possible d'obtenir les propriétés de stabilité et de localisation asymptotique par le biais de la φ -stabilité. Si ces exemples montrent qu'une telle fonction n'est pas toujours unique, ils montrent qu'elle existe au moins pour la plupart des lois usuelles.

De plus, cette étude devrait permettre de définir la notion de fonction de répartition outlier-prone ou outlier-resistant dans le cas de distributions multidimensionnelles, comme cela a été fait pour les lois unidimensionnelles ([4],[8],[9],[10]). C'est l'objet du chapitre 4.

BIBLIOGRAPHIE

- [1] DAVIS,R. - MULROW, E. - RESNICK, S. (1987). The convex hull of a random sample in \mathbf{R}^2 . *Comm. statist. - Stochastic models*; 3(1), 1-27
- [2] DELCROIX, M.F. and JACOB,P. (1992). Stability of extreme value for a multidimensional sample. *Statistique et Analyse des Données*, Juin 92, 1-21.
- [3] FISCHER, L. (1966). The convex hull of a sample. *Bull. Am. Math. Soc.* 72, 555-558.
- [4] GATHER, U. and RAUHUT, B. (1990). The outlier behaviour of probability distributions. *Journal of Statistical Planning and Inference* 26, 237-252.
- [5] GEFFROY, J. (1958-1959). Contribution à la théorie des valeurs extrêmes. *Pub. Inst. Stat. Univ. Paris, 1958 (fasc.4.) et 1959 (fasc. 1).*
- [6] GEFFROY, J. (1961). Localisation asymptotique du polyèdre d'appui d'un échantillon Laplacien à k dimensions. *Pub. Inst. Stat. Univ. Paris, 1961, n° 59.*
- [7] GNEDENKO, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423-453.
- [8] GREEN, R.F. (1976). Outlier-prone and outlier-resistant distribution. *J. Amer. Statist. Assoc.* 71, 502-505.
- [9] MATHAR, R. (1989). The outlier behaviour of distributions and the decay power of optimal tests at decreasing level. *Statistics* 20 (2), 247-254.
- [10] MUNOZ-GARCIA,J MORENO-REBOLLO,JL and PASCUAL-ACOSTA, A. (1990). Outliers : a formal approach. *International Statistical review*, 58, 3, 215-220.
- [11] VAN WEL, (1989). The convex hull of an uniform sample from the interior of a simple d -polytope.

**Chapitre 4 : Distributions
multidimensionnelles
«outlier-resistant » et «outlier-prone »**

1 Introduction

Dans ce chapitre, on considère à nouveau les problèmes de stabilité pour s'intéresser aux distributions multidimensionnelles « outlier-resistant » et « outlier-prone » ainsi qu'aux liens existant entre ces notions.

En 1971, Neyman et Scott, [9], ont introduit une première définition, modifiée en 1976 par Green, [7]. Il s'agit de distinguer deux classes de lois : les lois pour lesquelles l'existence de valeurs éloignées du groupe principal des observations est normale, et les lois pour lesquelles cela est exceptionnel. Par exemple, l'idée intuitive de la notion de loi « outlier-resistant » est que la probabilité pour qu'il existe des observations isolées et situées loin de la plus grande partie des observations est très faible.

Aussi, en 1976, Green, [7], propose la définition suivante : une distribution F est dite absolument « outlier-resistant » si, pour tout $\varepsilon > 0$:

$$(1) \quad \lim_{n \rightarrow +\infty} P(X_{n,n} - X_{n,n-1} > \varepsilon) = 0.$$

(L'échantillon réel (X_1, \dots, X_n) est ordonné par : $X_{n,1} \leq X_{n,2} \leq \dots \leq X_{n,n}$.)

De même, F est dite relativement « outlier-resistant » s'il existe $k > 1$ tel que :

$$(2) \quad \lim_{n \rightarrow +\infty} P\left(\frac{X_{n,n}}{X_{n,n-1}} > k\right) = 0.$$

De plus, une distribution F est dite absolument « outlier-prone » s'il existe $\varepsilon > 0$, $\delta > 0$ et un entier n_0 , tels que pour $n \geq n_0$:

$$(3) \quad P(X_{n,n} - X_{n,n-1} > \varepsilon) \geq \delta.$$

De la même façon, F est dite relativement « outlier-prone » s'il existe $k > 1$, $\delta > 0$ et un entier n_0 , tels que pour $n \geq n_0$:

$$(4) \quad P\left(\frac{X_{n,n}}{X_{n,n-1}} > k\right) \geq \delta.$$

D'autres définitions ont été proposées : par exemple, O'Hagan en 1979, [10], et Goldstein en 1982, [6] ont introduit une notion bayésienne basée sur l'influence qu'une observation supplémentaire, suffisamment grande, peut avoir

sur la suite des observations. Gather et Rauhut, [3] donnent et comparent ces différentes notions en dimension 1, ainsi que leurs conséquences statistiques. Notre but, dans ce chapitre, est d'étendre les notions de valeurs aberrantes, telles qu'elles ont été définies par Green, aux lois multidimensionnelles, en utilisant la notion d'isobare que l'on a définie au chapitre 2.

Dans tout ce chapitre, nous faisons les mêmes hypothèses sur la loi du couple (R, Θ) que dans les chapitres 2 et 3 et conservons les mêmes notations.

2 Préliminaires

Avant de proposer des définitions, un résultat supplémentaire sur la stabilité d'une suite de variables aléatoires multidimensionnelles est nécessaire. Comme dans les deux chapitres précédents, on considère un échantillon E_n de \mathbf{R}^k , défini en coordonnées polaires par :

$$\{(R_1, \Theta_1), \dots, (R_n, \Theta_n)\}.$$

On rappelle que pour tout $1 \leq \alpha \leq n$, l'échantillon a été ordonné selon les niveaux croissants en :

$$X_1^{*n}, \dots, X_{n-\alpha+1}^{*n}, \dots, X_n^{*n} = X_n^*.$$

Pour plus de commodité, on notera :

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

ainsi l'échantillon E_n est ordonné selon les niveaux croissants par :

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

Nous allons montrer que la stabilité en probabilité de $(X_{n,n})_n$ est équivalente à la stabilité en probabilité de $(X_{n,n-\alpha+1})_n$, pour tout $1 \leq \alpha \leq n$. Les théorèmes 3, 4 et 5 de Geffroy, [4], permettent de montrer le théorème suivant, si l'on suppose de plus que la condition de régularité (H) du chapitre 2 est satisfaite.

Théorème 1 *Supposons (H) satisfaite. Pour $1 \leq \alpha \leq n$, la suite $(X_{n,n-\alpha+1})_n$ est stable en probabilité si et seulement si la suite $(X_n^*)_n$ est stable en probabilité.*

Démonstration : Soit θ_1 une direction arbitraire de S^{k-1} . On note W_n^* l'intersection de l'isobare passant par X_n^* et du demi-axe positif $0\theta_1^+$. Dans le chapitre 2, on a montré que sous la condition (H), $(X_n^*)_n$ est stable en probabilité si et seulement si $(W_n^*)_n$ est stable en probabilité; de plus, les théorèmes 3, 4, 5 de [4] montrent que la suite de variables aléatoires réelles $(W_n^*)_n$ est stable si et seulement si la suite $(W_{n,n-\alpha+1})_n$ est stable; mais ceci est aussi équivalent à la stabilité de $(X_{n,n-\alpha+1})_n$. D'où le résultat. ■

Les définitions de [7], pour les lois réelles « outlier-resistant » et « outlier-prone » se basent sur le comportement de $X_{n,n} - X_{n,n-1}$. On procède de même

dans \mathbf{R}^k . Rappelons qu'étant donné un point $X_i = (R_i, \Theta_i)$ de l'échantillon, il existe presque sûrement une unique isobare contenant ce point. On notera par la suite $g_{n,n}$ l'isobare passant par $X_{n,n}$ et $g_{n,n-1}$ l'isobare passant par $X_{n,n-1}$.

3 Lois A. O. R.

Définition 1 La loi de (R, Θ) est dite « *absolutely outlier-resistant* », (on note A. O. R.), si et seulement si pour tout θ :

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

Rappelons que pour tout θ , F_θ désigne la fonction de répartition, supposée bijective, de R étant donné $\Theta = \theta$, et que l'on note $G_\theta = 1 - F_\theta$. Reprenons les mêmes notations que dans le chapitre 2 : si θ_1 est une direction arbitraire, on sait que $W_n^* = g_{n,n}(\theta_1)$ peut être considéré comme le maximum de n variables aléatoires réelles indépendantes W_1, \dots, W_n de fonction de répartition commune F_{θ_1} . Plus précisément, pour tout $1 \leq i \leq n$, W_i est l'intersection de l'isobare passant par (R_i, Θ_i) et du demi-axe positif $0\theta_1^+$. On ordonne W_1, \dots, W_n comme on a ordonné X_1, \dots, X_n selon les niveaux croissants en :

$$W_{n,1}, \dots, W_{n,n} = W_n^*.$$

Ainsi, la relation (5) de la définition exprime que pour tout θ , la loi réelle F_θ est A. O. R. En dimension 1, on peut montrer, [4], [5], que la stabilité en probabilité de $(X_{n,n})_n$ équivaut à la convergence en probabilité vers 0 de $X_{n,n} - X_{n,n-1}$. Le théorème suivant établit des résultats analogues.

Théorème 2 Supposons (H) satisfaite. Les assertions suivantes sont équivalentes :

- (i) La loi de (R, Θ) est A. O. R.
- (ii) $(X_n^*)_n$ est stable en probabilité.
- (iii) Pour tout $1 \leq \alpha \leq n$, $(X_{n,n-\alpha+1})_n$ est stable en probabilité.
- (iv) Il existe θ_1 tel que $\lim_{x \rightarrow +\infty} \frac{G_{\theta_1}(x)}{G_{\theta_1}(x-h)} = 0$, pour tout $h > 0$.
- (v) Pour tout θ , $\lim_{x \rightarrow +\infty} \frac{G_\theta(x)}{G_\theta(x-h)} = 0$, pour tout $h > 0$.

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

(vii) $(W_n^*)_n$ est stable en probabilité.

(viii) Pour tout θ , la loi F_θ est A. O. R.

(ix) Il existe θ_1 tel que la loi F_{θ_1} soit A. O. R.

Démonstration : Le théorème 1 montre que (ii) et (iii) sont équivalents. Le théorème 4 du chapitre 2 montre l'équivalence de (ii), (iv), (v), (vii).

Par ailleurs, (vii) et (vi) sont équivalents : c'est un résultat qui est prouvé dans les théorèmes 26 et 27 de [4].

Montrons que (i) implique (ii) : par hypothèse, pour tout θ :

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

En particulier, $g_{n,n}(\theta_1) - g_{n,n-1}(\theta_1) \xrightarrow{P} 0$, c'est à dire :

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

Par conséquent, d'après l'équivalence de (vi) et de (vii), $(W_n^*)_n$ est stable en probabilité et d'après la démonstration du théorème 4 du chapitre 2, $(X_n^*)_n$ est stable en probabilité.

Réciproquement, si $(X_n^*)_n$ est stable en probabilité, $(W_n^*)_n$ l'est aussi, et donc, $W_n^* - W_{n,n-1} \xrightarrow{P} 0$. Par conséquent, $g_{n,n}(\theta_1) - g_{n,n-1}(\theta_1) \xrightarrow{P} 0$; mais, θ_1 étant arbitraire, on obtient (i).

L'équivalence de ces propriétés avec (viii) et (ix) est alors claire. ■

Remarque 1 Dans le cas où l'hypothèse (K), introduite au chapitre 2 est vérifiée, il existe θ_1 tel que pour tout $w > 0$ et pour tout $\eta > 0$:

$$(K) \quad \sup_{\theta} \{g(\theta, w + \eta) - g(\theta, w - \eta)\} = g(\theta_1, w + \eta) - g(\theta_1, w - \eta).$$

(On rappelle que $g(\theta, u)$ désigne l'isobare passant par le point d'abscisse $u > 0$ du demi-axe positif $0\theta_1^+$.) Dans ce cas, on a même l'équivalence entre les assertions (ii) et :

$$(i)' \quad \sup_{\theta} \{g_{n,n}(\theta) - g_{n,n-1}(\theta)\} \xrightarrow{P} 0.$$

Remarque 2 L'hypothèse (H) est inutile pour la démonstration de l'équivalence de (vi) et de (vii), de même que l'hypothèse (K).

En dimension 1, il est possible, [3], de donner d'autres caractérisations des lois A. O. R., en utilisant la fonction d'espérance de survie, (« mean residual life function » ou encore mrlf), définie pour une variable réelle X par :

$$e(x) = E(X - x / X > x).$$

Posons pour $i = 1, \dots, n - 1$, pour $x > 0, y > 0$ et $n > 2$:

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

Si θ est fixé, $\theta = \theta_1$, $M_{i,n}(x, y, \theta)$ s'écrit :

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

On peut alors en utilisant [3] et le théorème précédent démontrer le théorème suivant :

Théorème 3 *Supposons (H) satisfaite. La loi de (R, Θ) est A. O. R. si et seulement s'il existe θ_1 tel que pour tout $y > 0$:*

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

pour un indice $1 \leq i \leq n$.

Démonstration : D'après le théorème précédent, la loi de (R, Θ) est A. O. R. si et seulement s'il existe θ_1 tel que la loi réelle F_{θ_1} soit A. O. R.. Or, d'après [3], (8) est une condition nécessaire et suffisante pour que F_{θ_1} soit A. O. R.. En effet, la statistique d'ordre $W_{n,1}, \dots, W_{n,n}$ forme une chaîne de Markov, [1], ainsi :

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

D'où le résultat, en utilisant l'assertion (iv) du théorème précédent. ■

Remarque 3 On peut aussi démontrer ce théorème sans utiliser les variables $W_{n,i}$ en remarquant que pour tout i et pour tout θ , $g_{n,i}(\theta) = F_{\theta}^{-1}(U_{n,i})$ où

$(U_{n,i})_n$ est la statistique d'ordre des niveaux des isobares $g_{n,i}$, et utiliser la propriété de chaîne de Markov de la suite réelle $(U_{n,i})_n$, statistique d'ordre d'un échantillon de la loi uniforme sur $[0, 1]$; on a en effet remarqué dans le chapitre 2 que $\{F_{\Theta_i}(R_i) = U_i, i = 1, \dots, n\}$ est un échantillon de la loi uniforme sur $[0, 1]$.

Remarque 4 Dans la définition 1, la taille de l'échantillon devient de plus en plus grande; par contre, dans le théorème précédent, la taille de l'échantillon est fixée. L'énoncé de ce théorème se rapproche plus de l'idée intuitive que l'on peut avoir d'un « outlier »: plus $X_{n,i}$ est grand, plus la probabilité pour que la différence $X_{n,i+1} - X_{n,i}$ dépasse un nombre positif arbitraire est faible. Le théorème suivant rend compte de ce phénomène en moyenne.

Posons maintenant pour tout θ et pour tout $i = 1, \dots, n$:

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

Quand θ est fixé, $\theta = \theta_1$, $\tilde{M}_{i,n}(x, \theta)$ s'écrit:

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

Théorème 4 Supposons (H) satisfaite. Supposons que pour tout θ , $\int x dF_\theta < +\infty$. Alors, la loi de (R, Θ) est A. O. R. si et seulement s'il existe θ_1 tel que pour tout $n \geq 3$:

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

Démonstration : [3] On considère la rarf définie pour tout θ , par:

$$e(x, \theta) = E(g(\theta) - x/g(\theta) > x),$$

où g désigne une isobare de la loi de (R, Θ) . En particulier, pour $\theta = \theta_1$ fixé:

$$e(x, \theta_1) = E(W - x/W > x).$$

D'après le théorème 2, la loi de (R, Θ) est A. O. R. si et seulement si la loi réelle F_θ l'est. Or, d'après [3]:

$$e(x, \theta_1) = \int_x^{+\infty} \frac{1 - F_{\theta_1}(y)}{1 - F_{\theta_1}(x)} dy = \int_0^{+\infty} \frac{1 - F_{\theta_1}(x + y)}{1 - F_{\theta_1}(x)} dy,$$

par conséquent :

$$e(x, \theta_1) = \int_0^{+\infty} P\{W_{n,n} > x + y/W_{n,n-1} = x\} dy = \tilde{M}_{n-1,n}(x, \theta_1).$$

Par ailleurs, un résultat de De Haan, [2], montre que si (W_1, \dots, W_n) est un échantillon de variables réelles, $(W_{n,n})_n$ est stable en probabilité si et seulement si la mrlf $e(x) = E(W - x/W > x)$ vérifie :

$$\lim_{x \rightarrow +\infty} e(x) = 0.$$

Ainsi, s'il existe θ_1 , tel que pour tout $n \geq 3$, $\lim_{x \rightarrow +\infty} \tilde{M}_{n-1,n}(x, \theta_1) = 0$, la suite $(W_n^*)_n$, intersection de $g_{n,n}$ et de $0\theta_1^+$, est stable en probabilité et d'après le théorème 2, F_{θ_1} est A. O. R.. On montre de même la réciproque. ■

4 Lois A. O. P.

On reprend dans ce paragraphe les mêmes notations que dans le paragraphe précédent.

Définition 2 La loi de (R, θ) est dite « absolument outlier-prone » (on note A. O. P.), si et seulement si pour tout θ il existe $\varepsilon > 0$, $\delta > 0$, et n_θ entier naturel, tels que pour tout θ et pour tout $n \geq n_\theta$:

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

C'est à dire que pour tout θ , la loi de F_θ est A. O. P.

On utilise dans le théorème 6 un résultat de Green, [7] :

Théorème 5 Soit F une fonction de répartition réelle vérifiant $F(+\infty) = 1$ et $F(x) < 1$ pour tout x réel. La loi F est A. O. P. si et seulement s'il existe $\alpha > 0$ et $\beta > 0$ tels que pour tout x :

$$\frac{1 - F(x + \beta)}{1 - F(x)} \geq \alpha.$$

Ce théorème permet de montrer le théorème suivant :

Théorème 6 *Supposons (H) satisfaite. Les assertions suivantes sont équivalentes :*

(i) *La loi de (R, Θ) est A. O. P.*

(ii) *Pour tout θ , il existe $\alpha > 0$ et $\beta > 0$ tels que pour tout x réel :*

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

(iii) *Il existe $\alpha_0 > 0$ et $\beta_0 > 0$ et θ_0 tels que :*

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

(iv) *Il existe θ_0 tel que F_{θ_0} soit A. O. P..*

Démonstration : Montrons que (i) et (ii) sont équivalents. Pour θ fixé, (10) exprime que la distribution réelle F_{θ} est A. O. P.. Or, d'après le théorème de Green cité ci-dessus, ceci est équivalent, pour θ fixé, à (11).

Il est clair que (ii) implique (iii).

Montrons que (iii) implique (ii) sous la condition (H). Considérons θ_1 distinct de θ_0 . Pour tout $x > 0$, il existe presque sûrement une isobare $g(\theta, x)$ passant par le point d'abscisse x du demi-axe positif $0\theta_1^+$, telle que $g(\theta_1, x) = x$; notons $u(x)$ le niveau de cette isobare. Puisque (H) est satisfaite, il existe $\eta > 0$ et une isobare $h^{\beta_0}(\theta, x)$ tels que pour tout x et pour tout θ :

$$g(\theta, x) + \eta < h^{\beta_0}(\theta, x) < g(\theta, x) + \beta_0.$$

Notons $u_{\beta_0}(x)$ le niveau de $h^{\beta_0}(\theta, x)$. On peut écrire, puisque $G_{\theta_1} = 1 - F_{\theta_1}$ est décroissante :

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

et,

$$G_{\theta_1}(x + \eta) = G_{\theta_1}(g(\theta_1, x) + \eta) > G_{\theta_1}(h^{\beta_0}(\theta_1, x)) = 1 - u_{\beta_0}(x).$$

De plus :

$$1 - u_{\beta_0}(x) = G_{\theta_0}(h^{\beta_0}(\theta_0, x)) > G_{\theta_0}(g(\theta_0, x) + \beta_0).$$

D'où :

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

Par conséquent, si $\frac{G_{\theta_0}(x + \beta_0)}{G_{\theta_0}(x)} \geq \alpha_0$ pour tout x réel, alors pour tout θ_1 distinct de θ_0 , il existe $\beta_1 = \eta > 0$ et $\alpha_1 = \alpha_0 > 0$ tels que pour tout x :

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

d'où (ii).

De plus, (iv) est clairement équivalente aux autres assertions. ■

Remarque 5 De même que pour le théorème 2, si l'hypothèse (K) est satisfaite pour θ_1 , et si θ_1 vérifie (12), l'assertion (iii) est alors équivalente à (i') :

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

Il est possible aussi de donner un théorème faisant intervenir la probabilité conditionnelle $M_{i,n}(x, y, \theta)$ donnée par (7).

Théorème 7 *On suppose (H) satisfaite. La loi de (R, Θ) est A. O. P. si et seulement si pour tout $y > 0$, il existe θ_0, α_0 et x_0 , tels que pour tout $x \geq x_0$:*

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

pour un indice $1 \leq i \leq n - 1$.

Démonstration : La démonstration est immédiate, il suffit de reprendre le calcul fait lors de la démonstration du théorème 3 et d'utiliser le théorème 6. ■

On peut donner à l'aide de (9) une condition nécessaire pour que la loi de (R, Θ) soit A. O. P., mais elle n'est pas suffisante.

Théorème 8 *Supposons la loi de (R, Θ) A. O. P., alors il existe δ_0, x_0 , et θ_0 tels que pour $x \geq x_0$ et pour tout $n \geq 3$:*

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

Démonstration : Supposons que la loi de (R, Θ) soit A. O. P.. D'après le théorème 6, il existe α_0, β_0 et θ_0 tels que pour tout x :

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

Posons : $\delta_0 = \alpha_0\beta_0$. Le calcul fait dans la démonstration du théorème 4 montre que :

$$\tilde{M}_{n-1,n}(x, \theta_0) = \int_0^{+\infty} \frac{1 - F_{\theta_0}(x + y)}{1 - F_{\theta_0}(x)} dy \geq \int_0^{\beta_0} \frac{1 - F_{\theta_0}(x + y)}{1 - F_{\theta_0}(x)} dy$$

et :

$$\tilde{M}_{n-1,n}(x, \theta_0) \geq \frac{1 - F_{\theta_0}(x + \beta_0)}{1 - F_{\theta_0}(x)} y_0 \geq \alpha_0 \geq \delta_0.$$

■

Remarque 6 La réciproque du théorème 8 est fautive : il suffit de considérer comme Gather et Rauhut, [3], l'exemple suivant dans \mathbf{R} :

Soit $x_0 = 0, x_n = x_{n-1} + 2^{n(n-1)}$ et F une fonction de répartition telle que :
 $F(0) = 0$

$1 - F(x)$ est continue, décroissante,

$1 - F(x) = 2^{-n^2}$ pour $x_n + 2^{-n^2} \leq x \leq x_{n+1}$.

Alors, pour tout $y > 0, \delta > 0$ et pour tout $x > 0$, il existe $\hat{x} = x_n$, (avec $2^{-n^2} < y; x_n < x$ et $2^{-2n+1} < \delta$), tel que :

$$\frac{1 - F(\hat{x} + y)}{1 - F(\hat{x})} = \frac{1 - F(x_{n+1})}{1 - F(x_n)} = 2^{-n^2} 2^{(n-1)^2} < \delta.$$

Si bien que F n'est pas A. O. P.. Cependant, si l'on choisit $\delta_0 = \frac{1}{2}$ et $x^* = x_2$, pour $x \geq x^*$, pour tout $n \geq 3$ et pour $x_{n-1} \leq x \leq x_n$:

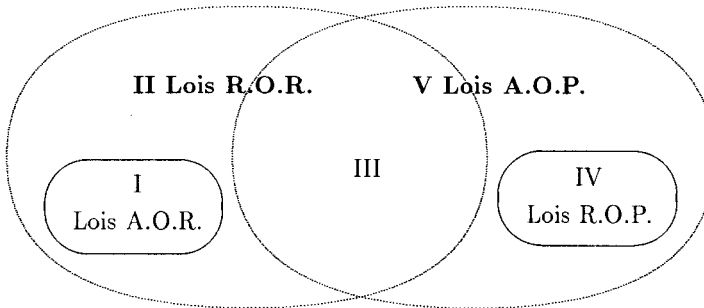
$$\tilde{M}_{n-1,n}(x) \geq \int_{x_n}^{+\infty} \frac{1 - F(y)}{1 - F(x_{n-1})} dy \geq \sum_{k=n}^{+\infty} 2^{(n-2)^2} 2^{-k^2} 2^{k^2-k} \geq \delta_0.$$

Les lois réelles pour lesquelles $\tilde{M}_{n-1,n}(x) > \delta$ pour $\delta > 0$ et pour $x \geq x_0$ sont appelées « mean absolutely outlier-prone » [11].

5 Généralisation

On pourrait de même que l'on a défini dans le chapitre 2 la φ -stabilité, définir les lois multidimensionnelles « φ -outlier-resistant » ou « φ -outlier-prone » : il suffit pour cela de considérer la loi de $(\varphi(R), \Theta)$ au lieu de la loi de (R, Θ) , φ étant une fonction telle qu'on l'a définie au chapitre 2. On peut alors exprimer autrement les propriétés de φ -stabilité que l'on a obtenues dans ce chapitre. Par exemple, si la loi de (R, Θ) est A. O. R. et si φ est une fonction définie sur \mathbf{R}^+ , positive de classe C^1 , croissante, bijective et concave, alors la loi de $(\varphi(R), \Theta)$ est encore A. O. R..

On peut résumer sur un schéma, comme cela a été fait dans [3], les relations entre ces diverses notions (les expressions « relatively outlier-resistant » et « relatively outlier-prone » sont abrégées par R. O. R. et R. O. P. ; il s'agit du cas où $\varphi(x) = \max(0, \text{Log } x)$).



Les exemples étudiés dans le chapitre 2 montrent que :

I Les lois pour lesquelles la fonction de répartition conditionnelle de R étant donné Θ est :

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

où $m > 1$ et α est une fonction strictement positive, sont A. O. R. , (exemple 1).

II On pourrait montrer que les lois de Poisson sont R. O. R. mais ni A. O. R. ni A. O. P..

III Les lois de type exponentiel, où

$$F_{\theta}(x) = (1 - e^{-\alpha(\theta)x})\mathbf{1}_{\{x>0\}},$$

sont R. O. R. et A. O. P., (exemple 5).

IV Les distributions de Cauchy, dont la densité conditionnelle de R étant donné Θ est :

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

sont R. O. P., mais sont φ -O. R. pour $\varphi(x) = \text{Log Log } x$, (exemple 4).

V Si l'on considère une loi de densité conditionnelle :

$$f(x) = \begin{cases} 2^{-(k-1)} & \text{si } m - \frac{1}{10} \leq x \leq m \\ 0 & \text{sinon} \end{cases}$$

avec $m = k + [10]^{\text{log}_2 k} - 2^{\text{log}_2 k}$ et k entier naturel, cette loi n'est ni R. O. R. ni R. O. P., [3].

VI Enfin, dans le cas où $0 < m \leq 1$, pour une loi de fonction de répartition conditionnelle donnée par :

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

la loi de (R, Θ) est φ -O. R., avec $\varphi(x) = x^{\frac{1}{2m}}$, (exemple 6).

6 Conclusion

Les définitions proposées par Green en dimension 1 se généralisent donc en dimension finie quelconque. Comme dans le chapitre 2, l'utilisation des isobares permet de ramener un problème multidimensionnel à un problème unidimensionnel. Certes, cela suppose que soient vérifiées les hypothèses (H) ou (K), mais, comme on a pu le constater sur les divers exemples étudiés, elles le sont pour la plupart des lois usuelles.

Références

- [1] DAVID H. A. (1981) - Orders Statistics. 2nd edition. Wiley, New-York.
- [2] DE HAAN L. (1970) - On regular variation and its application to the weak convergence of sample extremes. *Mathematical centre tracts. n°32. Amsterdam.*
- [3] GATHER U. and RAUHUT B. (1990) - Outlier behaviour of probability distributions. *Journal of Statistical Planning and Inference 26, 237-252. North Holland.*
- [4] GEFFROY J. (1958) - Contribution à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris, 7, 8 37-185.*
- [5] GNEDENKO B. V. (1943) - Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math. 44, 423-453.*
- [6] GOLDSTEIN M. (1976) - Contamination distribution. *Ann. Statist. 10, 174-183.*
- [7] GREEN R. F. (1976) - Outlier-prone and outlier-resistant distributions. *J. Amer. Statist. Assoc. 71, 502-505.*
- [8] KALE B.K. (1977) - A note on the outlier-resistant families and mixtures of distributions. *J. Indian. Statist. Assoc. 15, 119-130.*
- [9] 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.*
- [10] O'HAGAN A. (1979) - On outlier rejection phenomena in Bayes inference. *J. Roy. Statist. Soc. Ser. B 41, 358-367.*
- [11] RAUHUT B. (1987) - The modelling of outlier situation. In: *O. Opitz, B. Rauhut, Eds. Ökonomie und Mathematik. Springer, Berlin, 317-324.*

Conclusion

A ce point de notre recherche sur les problèmes de valeurs extrêmes multidimensionnelles, plusieurs perspectives de travail peuvent être envisagées.

Premièrement, dans une utilisation statistique des résultats des chapitres 3 et 4, il devrait être possible de donner, pour θ fixé, une estimation des isobares $g_n(\theta) = G_\theta^{-1}(\frac{1}{n})$, en combinant des méthodes de régression (BOSCQ D. et LECOUTRE J. P. (1987) - Théorie de l'estimation fonctionnelle, *Economica.*), et des méthodes d'approximation forte par les quantiles (CSÖRGÖ M. (1983) - Quantile processes with statistical applications, *S. I. A. M.* et, SHORACK G. R. and WELLNER J. A. (1986) - Empirical processes with applications to statistics, *Wiley.*).

La deuxième perspective qui s'ouvre aussi est la généralisation des résultats obtenus ici pour des échantillons multidimensionnels au nuage des points d'une répartition ponctuelle aléatoire de \mathbf{R}^k .

Appendice

Appendice 1

(voir le paragraphe V du chapitre 1)

Lemme 1 : méthode de Laplace (DIEUDONNE J. Calcul infinitésimal, chap. IV. 2. Hermann)

Supposons que dans l'intervalle ouvert (a, b) (borné ou non), les fonctions g et h soient deux fois continûment dérivables, que l'intégrale $\int_a^b |g(x)| e^{h(x)} dx$ soit définie et que h' ne change de signe qu'en un seul point c tel que $a < c < b$, où de plus h atteint un maximum et $g(c) \neq 0$, $h''(c) < 0$. Alors quand t tend vers $+\infty$:

$$\int_a^b |g(x)| e^{h(x)} dx \sim \sqrt{\pi} g(c) e^{th(c)} \sqrt{\frac{2}{-th''(c)}}.$$

Lemme 2 (DIEUDONNE J. Calcul infinitésimal, chap. III, 10. 2. Hermann)

Soient f et g deux fonctions complexes continues par morceaux dans $[a, +\infty)$ et supposons que $g(x) > 0$ dans cet intervalle. Supposons que l'intégrale $\int_a^{+\infty} g(t) dt$ soit convergente.

Alors, au voisinage de $+\infty$, la relation $f \sim g$ entraîne $\int_x^{+\infty} f(t) dt \sim \int_x^{+\infty} g(t) dt$.

Lemme 3 (DIEUDONNE J. Calcul infinitésimal, chap. III, 10. 7. Hermann)

Soit j une fonction positive définie et continûment dérivable sur $[0, +\infty)$.

Si au voisinage de $+\infty$:

$$\frac{1}{x} = o\left|\frac{j'(x)}{j(x)}\right|,$$

$\frac{j(x)}{j'(x)}$ est continûment dérivable,

$$\left(\frac{j}{j'}\right)'(x) = o(1),$$

$$j'(x) < 0,$$

l'intégrale $\int_a^{+\infty} j(u) du$ est convergente et au voisinage de $+\infty$, $\int_x^{+\infty} j(u) du$ est équivalente à $\frac{(j(x))^2}{|j'(x)|}$.

On utilise ce lemme pour donner un équivalent de $\int_t^{+\infty} j(u) du$ avec :

$$j(u) = u^{2\frac{(\alpha-1)}{(\alpha+2)}} \exp\{-C_2(\alpha)u^{\frac{2\alpha}{(\alpha+2)}}\}.$$

En effet :

$$j'(u) = \left\{ \frac{2(\alpha-1)}{\alpha+2} u^{\frac{\alpha-4}{\alpha+2}} - C_2(\alpha) \frac{2\alpha}{\alpha+2} u^{\frac{3\alpha-4}{\alpha+2}} \right\} \exp\{-C_2(\alpha)u^{\frac{2\alpha}{(\alpha+2)}}\},$$

j' est donc négative au voisinage de $+\infty$, puisque $C_2(\alpha) > 0$.

Des calculs précédents, on déduit que :

$$\left| \frac{j(u)}{uj'(u)} \right| \sim \left\{ C_2(\alpha) \frac{2\alpha}{\alpha+2} \right\}^{-1} u^{-\frac{2\alpha}{(\alpha+2)}}$$

ce qui montre que :

$$\frac{1}{u} = o\left(\left| \frac{j'(u)}{j(u)} \right|\right).$$

Par ailleurs, $\left(\frac{j}{j'}\right)'(u) = 1 - \frac{j(u)j''(u)}{j'^2(u)}$; et, comme au voisinage de $+\infty$:

$$j(u) \sim u^{2\frac{(\alpha-1)}{(\alpha+2)}} \exp\{-C_2(\alpha)u^{\frac{2\alpha}{(\alpha+2)}}\}$$

$$j'^2(u) \sim C_2^2(\alpha) \left(\frac{2\alpha}{\alpha+2}\right)^2 u^{\frac{6\alpha-8}{\alpha+2}} \exp\{-2C_2(\alpha)u^{\frac{2\alpha}{(\alpha+2)}}\},$$

$$j''(u) \sim C_2^2(\alpha) \left(\frac{2\alpha}{\alpha+2}\right)^2 u^{\frac{4\alpha-6}{\alpha+2}} \exp\{-C_2(\alpha)u^{\frac{2\alpha}{(\alpha+2)}}\},$$

on vérifie bien que :

$$\left(\frac{j}{j'}\right)'(x) = o(1).$$

Les conditions du Lemme 3 sont donc réalisées, on en déduit que $\int_t^{+\infty} j(u) du$ est équivalente à :

$$\frac{(\alpha+2)}{2\alpha C_2(\alpha)} t^{\frac{\alpha}{(\alpha+2)}} \exp\{-C_2(\alpha)t^{\frac{2\alpha}{(\alpha+2)}}\}.$$

Appendice 2

(voir la remarque 6 du chapitre 2 page 40)

- STABILITE PRESQUE SURE DE $(X_n^*)_n$.-

1 - Définitions.-

Soit $(D_n)_n$ une suite croissante d'ensembles (ordonnés pour l'inclusion) et $(Z_n)_n$ une suite de variables aléatoires (en fait, $Z_n = X_n$ ou X_n^*). On utilise les définitions suivantes .

Définition 1.- La variable Z_n appartient presque sûrement asymptotiquement à D_n , ($Z_n \in D_n$ p.s.a.), si $P\{\text{Lim inf}(Z_n \in D_n)\} = 1$.

Définition 2.- La suite $(Z_n)_n$ appartient presque sûrement asymptotiquement à D_n , ($(Z_n) \subset D_n$ p.s.a.), si $P\{\text{Lim inf}(\{Z_1, \dots, Z_n\} \subset D_n)\} = 1$.

On a vu dans le chapitre 2 que $(X_n^*)_n$ est p.s. stable si et seulement si les conditions suivantes C^+ et C^- sont satisfaites.

C^+ Pour tout $\varepsilon > 0$, $(X_n) \subset \Gamma_n^\varepsilon$ p.s.a..

C^- Pour tout $\varepsilon > 0$, $X_n^* \in (\Gamma_n^{-\varepsilon})^c$ p.s.a.,

où $\Gamma_n^\varepsilon = \{(\rho, \theta) : \rho \leq \gamma_n(\theta) + \varepsilon\}$, $\Gamma_n^{-\varepsilon} = \{(\rho, \theta) : \rho \leq \gamma_n(\theta) - \varepsilon\}$, et $\gamma_n(\theta) = G\theta^{-1}(1/n)$.

Nous utiliserons les lemmes suivants.

2 -La condition C^+ .

a) Lemme 1.- Soit $(D_n)_n$ une suite croissante d'ensembles telle que $R^k = \cup_n D_n$; $X_n \in D_n$ p.s.a. si et seulement si $(X_n) \subset D_n$ p.s.a..

■ Supposons que $X_n \in D_n$ p.s.a.. Soit $\varepsilon > 0$. Par hypothèse, il existe N_ε tel que :

$$P\left(\bigcap_{n=N_\varepsilon}^{\infty} (X_n \in D_n)\right) > 1 - \varepsilon/2,$$

et il existe m tel que :

$$P\left(\bigcap_{n=1}^{N_\varepsilon-1} (X_n \in D_m)\right) > 1 - \varepsilon/2.$$

On suppose que $m > N_\varepsilon$.

Si les deux événements sont réalisés, on a : $X_1 \in D_m ; \dots ; X_{N_\varepsilon-1} \in D_m ; X_{N_\varepsilon} \in D_{N_\varepsilon} ; \dots ; X_m \in D_m, X_{m+1} \in D_{m+1} ; \dots$. Ainsi $\{X_1, \dots, X_m\} \subset D_m ; \{X_1, \dots, X_{m+1}\} \subset D_{m+1} ; \dots$. Si bien que :

$$P\left(\bigcap_{n=m}^{\infty} \{X_1, \dots, X_n\} \subset D_n\right) > 1 - \varepsilon.$$

■

b) Pour $x > 1$, il existe une unique isobare de niveau $(1 - \frac{1}{x})$, $\gamma_x(\theta) = g(\theta, w_x)$. Soit $\varepsilon > 0$; (H) étant satisfaite, il existe $\eta > 0$ et une isobare de niveau $v(\varepsilon, x)$, $h_x^\varepsilon(\theta) = g(\theta, w_x + \varepsilon/\beta_1)$ tels que :

$$\gamma_x(\theta) + \eta < h_x^\varepsilon(\theta) < \gamma_x(\theta) + \varepsilon,$$

pour tout θ et tout $x > 1$.

Pour $x > 1$, on peut écrire :

$$h_x^\varepsilon(\theta) = \gamma_x(\theta) + \varepsilon(x, \theta)$$

où

$$(1) \quad 0 < \eta < \varepsilon(x, \theta) < \varepsilon.$$

On définit aussi $h_x^\varepsilon(\theta)$ pour $0 \leq x \leq 1$: comme $\gamma_1(\theta) = 0$, on pose $w_1 = 0$ et $h_1^\varepsilon(\theta) = g(\theta, \varepsilon/\beta_1)$. Pour $0 \leq x < 1$, on pose $h_x^\varepsilon(\theta) = g(\theta, x\varepsilon/\beta_1)$. Alors, pour $0 \leq x \leq 1$, $h_x^\varepsilon(\theta) = g(\theta, x\varepsilon/\beta_1)$. Ainsi, pour $x \geq 0$, h_x^ε est une isobare telle que $h_0^\varepsilon(\theta) = 0$ pour tout θ . Pour $x \geq 0$, le niveau $v(\varepsilon, x)$ est $v(\varepsilon, x) = F_\theta(h_x^\varepsilon(\theta))$. Posons $v_{\varepsilon, n} = v(\varepsilon, n) = F_\theta(h_n^\varepsilon(\theta))$ pour $n \geq 0$; $v_{\varepsilon, n}$ ne dépend pas de θ car h_n^ε est une isobare.

Lemme 2.- Pour $\varepsilon > 0$, si la série $\sum G_\theta(h_n^\varepsilon(\theta)) = \sum(1 - v_{\varepsilon,n})$ converge, alors $(X_n) \subset \Gamma_n^\varepsilon$ p.s.a..

■ Soit $\varepsilon > 0$. Posons $L_{\varepsilon,n} = \{(\rho, \theta) : \rho \leq h_n^\varepsilon(\theta)\}$. Comme les variables aléatoires $X_i = (R_i, \Theta_i)$ sont indépendantes,

$$P(X_n \notin L_{\varepsilon,n}) = \int_0^{2\pi} P(R > h_n^\varepsilon(\theta) / \Theta = \theta) P_\Theta(d\theta) = 1 - v_{\varepsilon,n}.$$

D'après le lemme de Borel-Cantelli, $X_n \in L_{\varepsilon,n}$ p.s.a. et d'après le lemme 1, $(X_n) \subset L_{\varepsilon,n}$ p.s.a.. ■

c) Le lemme 3 donne une condition nécessaire et suffisante pour la convergence de la série $\sum(1 - v_{\varepsilon,n})$. Soit $\varepsilon > 0$; pour θ dans S^{k-1} , et pour $x \geq 0$, posons :

$$(2) \quad B_\theta^\varepsilon(x) = h_x^\varepsilon(\theta) = F_\theta^{-1}(v(\varepsilon, x)).$$

B_θ^ε est une fonction strictement croissante de x , θ étant fixé. De plus B_θ^ε est positive, continue et $B_\theta^\varepsilon(0) = 0$.

Lemme 3.- Soit $\varepsilon > 0$ fixé.

i) $J_\varepsilon = \int_0^{+\infty} (B_\theta^\varepsilon)^{-1}(x) dF_\theta(x)$ ne dépend pas de θ .

ii) $\sum G_\theta(h_n^\varepsilon(\theta))$ converge si et seulement si J_ε converge.

■ Soit $\varepsilon > 0$. $\sum G_\theta(h_n^\varepsilon(\theta)) = \sum(1 - v_{\varepsilon,n})$ est une série de terme général positif et décroissant; par conséquent, elle converge si et seulement si l'intégrale $I_\varepsilon = \int_0^{+\infty} G_\theta(B_\theta^\varepsilon(x)) dx = \int_0^{+\infty} (1 - v(\varepsilon, x)) dx$ converge. (I_ε ne dépend pas de θ). On montre que pour tout θ , $\int_0^{+\infty} (B_\theta^\varepsilon)^{-1}(x) dF_\theta(x) = I_\varepsilon$, ([Geffroy (1958-1959)] page 52) et cette intégrale ne dépend pas de θ . D'où (i) et (ii). ■

d) Théorème 1.-

i) Pour $\varepsilon > 0$, si $J_\varepsilon = \int_0^{+\infty} (B_\theta^\varepsilon)^{-1}(x) dF_\theta(x)$ converge, alors $\tilde{J}(\theta, \varepsilon) = \int_\varepsilon^{+\infty} \frac{dF_\theta(x)}{G_\theta(x-\varepsilon)}$ converge pour tout θ .

ii) Si pour un θ_0 et pour tout $\eta > 0$, $\tilde{J}(\theta_0, \eta)$ converge, alors pour tout $\varepsilon > 0$, J_ε converge. De plus, $\tilde{J}(\theta, \varepsilon)$ converge pour tout (θ, ε) .

■ Soit $\varepsilon > 0$, $s > \sup_\theta(B_\theta^\varepsilon)(1)$ et $x_\theta = (B_\theta^\varepsilon)^{-1}(s) > 1$. Comme pour tout θ et pour tout $x > 1$, $0 < \eta < \varepsilon(x, \theta) < \varepsilon$, (1) et (2) donnent :

$$s - \varepsilon < G_\theta^{-1}\left(\frac{1}{x_\theta}\right) = B_\theta^\varepsilon(x_\theta) - \varepsilon(x_\theta, \theta) < s - \eta.$$

Si bien que pour tout θ et pour tout $s > \sup_{\theta}(B_{\theta}^{\varepsilon})(1)$,

$$(3) \quad \frac{1}{G_{\theta}(s-\varepsilon)} < x_{\theta} = (B_{\theta}^{\varepsilon})^{-1}(s) < \frac{1}{G_{\theta}(s-\eta)}.$$

Ainsi, si J_{ε} converge, alors pour tout θ , $\tilde{J}(\theta, \varepsilon)$ converge; d'où (i).

De même d'après (3), si, pour tout $\eta > 0$, $\tilde{J}(\theta_0, \eta)$ converge, alors, pour tout $\varepsilon > 0$, J_{ε} converge. ■

Corollaire 1.- *S'il existe θ_0 tel que, pour tout $\varepsilon > 0$, $\tilde{J}(\theta_0, \varepsilon)$ converge, alors pour tout $\varepsilon > 0$, $(X_n) \subset \Gamma_n^{\varepsilon}$ p.s.a. et C^+ est réalisée.*

e) Comme dans [Geffroy (1958-1959)], on introduit K l'ensemble des fonctions de répartition vérifiant pour tout $\varepsilon > 0$,

$$(4) \quad \lim_{x \rightarrow +\infty} \frac{G(x-\varepsilon)}{G(x) \text{Log } G(x)} = -\infty.$$

ou, en posant $W(x) = -\text{Log } G(x)$,

$$(5) \quad \lim_{x \rightarrow +\infty} (W(x) - W(x-\varepsilon) - \text{Log } W(x)) = +\infty.$$

Remarque 1.- Si (H) est satisfaite et si, pour un θ_0 , F_{θ_0} est dans K , alors F_{θ} est dans K pour tout θ .

Théorème 2.- *Si, pour un θ_0 , F_{θ_0} est dans K , alors C^+ est réalisée.*

■ Si F_{θ_0} est dans K , alors ([Geffroy (1958-1959)] page 112) pour tout $\varepsilon > 0$ et pour x_0 suffisamment grand,

$$(6) \quad \int_{x_0}^x \frac{dF_{\theta_0}(t)}{G_{\theta_0}(t-\varepsilon)} < \frac{1}{\text{Log } G_{\theta_0}(x)} - \frac{1}{\text{Log } G_{\theta_0}(x_0)}.$$

Le deuxième terme de (6) est borné, donc $\lim_{x \rightarrow +\infty} \int_{x_0}^x \frac{dF_{\theta_0}(t)}{G_{\theta_0}(t-\varepsilon)} < \infty$. Ainsi C^+ est réalisée d'après le corollaire 1. ■

3 -La condition C^- .

On utilise ici des lemmes de minoration presque complète.

a) Définition 3.- *Ordonnons l'échantillon E_n selon les niveaux croissants : $X_1^{n*}, X_2^{n*}, \dots, X_n^{n*} = X_n^*$. Soit $1 \leq \alpha \leq n$, et soit $(D_n)_n$ une suite croissante d'ensembles ordonnés par inclusion. $\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\}$ est presque complètement asymptotiquement dans D_n^c , ($\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\} \subset D_n^c$ p.co.a.), si*

$$(7) \quad \sum P[\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\} \not\subset D_n^c] < \infty.$$

b) Proposition 1.- Soit $(j_n)_n$ une suite d'isobares de niveau u_n . Pour tout $1 \leq \alpha \leq n$, $\{X_{n-\alpha+1}^{n*}, \dots, X_n^{n*}\} \subset \{(\rho, \theta) : \rho \leq j_n(\theta)\}^c$ p.co.a. si et seulement si

i) $\lim_{n \rightarrow +\infty} n(1 - u_n) = +\infty$.

ii) $\sum (u_n)^n (n(1 - u_n))^{\alpha-1}$ converge.

(En particulier, $X_n^* \in \{(\rho, \theta) : \rho \leq j_n(\theta)\}^c$ p.co.a. si et seulement si $\sum u_n^n$ converge).

Pour la démonstration, voir [Geffroy (1958-1959)] page 57.

Nous utilisons la même méthode que pour la condition C^+ . Pour $x > 1$ il existe une unique isobare de niveau $(1 - \frac{1}{x})$, $\gamma_x(\theta)$. Pour tout $\varepsilon > 0$, il existe $\eta > 0$ et une isobare de niveau $\tilde{v}(\varepsilon, x)$, $\tilde{h}_x^\varepsilon(\theta)$ tels que pour tout θ dans S^{k-1} et pour tout $x > 1$:

$$(8) \quad \gamma_x(\theta) - \varepsilon < \tilde{h}_x^\varepsilon(\theta) < \gamma_x(\theta) - \eta.$$

On peut écrire :

$$(9) \quad \tilde{h}_x^\varepsilon(\theta) = \gamma_x(\theta) - \tilde{\varepsilon}(x, \theta),$$

où $0 < \eta < \tilde{\varepsilon}(x, \theta) < \varepsilon$.

$F_\theta^n(\tilde{h}_x^\varepsilon(\theta))$ ne dépend pas de θ car \tilde{h}_x^ε est une isobare. La condition C^- est plus faible que la condition " $X_n^* \in (\Gamma_n^{-\varepsilon})^c$ p.co.a.". Il suffit donc d'obtenir une condition pour que (7) soit satisfaite avec $\alpha = 1$, et $D_n = \{(\rho, \theta) : \rho \leq \tilde{h}_n^\varepsilon(\theta)\}$.

c) Lemme 4.- Soit $\varepsilon > 0$; $X_n^* \in (\Gamma_n^{-\varepsilon})^c$ p.co.a. si $\sum_{n \geq 2} F_\theta^n(\tilde{h}_n^\varepsilon(\theta))$ converge.

La démonstration est immédiate en utilisant la proposition 1.

Soit $\varepsilon > 0$; pour θ dans S^{k-1} , et pour $x > 1$, posons

$$(10) \quad \tilde{B}_\theta^\varepsilon(x) = \tilde{h}_x^\varepsilon(\theta).$$

Pour θ fixé, $\tilde{B}_\theta^\varepsilon$ est une fonction croissante de x . Soit $c(\varepsilon) = \sup_{\theta \in S^{k-1}} \{\frac{1}{G_\theta(\varepsilon)}\}$, alors $F_\theta^x(\tilde{B}_\theta^\varepsilon(x))$ est définie pour $x \geq c(\varepsilon)$. Pour tout réel x , $[x]$ désigne la partie entière de x .

d) Lemme 5.- Soit $\varepsilon > 0$.

i) $K_\varepsilon = \int_{c(\varepsilon)}^{+\infty} F_\theta^x(\tilde{B}_\theta^\varepsilon(x)) dx$ ne dépend pas de θ .

ii) $\sum_{n \geq [c(\varepsilon)]+1} F_\theta^n(\tilde{h}_n^\varepsilon(\theta))$ converge si et seulement si K_ε converge.

■ Soit $\varepsilon > 0$ fixé. Posons $v_n = \int_n^{n+1} F_\theta^x(\tilde{B}_\theta^\varepsilon(x))dx$ pour $n \geq [c(\varepsilon)] + 1$. Pour $x \geq c(\varepsilon)$, $F_\theta(\tilde{B}_\theta^\varepsilon(x)) = \tilde{v}(\varepsilon, x)$, ainsi $\int_{c(\varepsilon)}^{+\infty} F_\theta^x(\tilde{B}_\theta^\varepsilon(x))dx$ ne dépend pas de θ . De plus $\sum_{n \geq [c(\varepsilon)]+1} v_n$ converge si et seulement si K_ε converge. La suite de la démonstration est dans [Geffroy (1958-1959)] page 96. ■

e) Théorème 3.-

i) Pour $\varepsilon > 0$, si $K_\varepsilon = \int_{c(\varepsilon)}^{+\infty} F_\theta^x(\tilde{B}_\theta^\varepsilon(x))dx$ converge, alors pour tout θ , $\tilde{K}(\theta, \varepsilon) = \int_{c(\varepsilon)}^{+\infty} F_\theta^x(G_\theta^{-1}(\frac{1}{x}) - \varepsilon)dx$ converge.

ii) Si pour un (θ_0, ε) $\tilde{K}(\theta_0, \varepsilon)$ converge, alors K_ε converge et $X_n^* \in (\Gamma_n^{-\varepsilon})^c$ p.co.a.; de plus $\tilde{K}(\theta, \varepsilon)$ converge pour tout θ .

■ Soit $\varepsilon > 0$; il existe $\eta > 0$ tel que (9) soit vraie. Pour $x > c(\varepsilon)$ et pour tout θ dans S^{k-1} , (9) et (10) montrent que :

$$G_\theta^{-1}(\frac{1}{x}) - \varepsilon < \tilde{B}_\theta^\varepsilon(x) = G_\theta^{-1}(\frac{1}{x}) - \tilde{\varepsilon}(x, \theta) < G_\theta^{-1}(\frac{1}{x}) - \eta.$$

f) Théorème 4.- Supposons qu'il existe θ_0 tel que pour $\varepsilon > 0$, l'intégrale $L(\theta_0, \varepsilon) = \int_\varepsilon^{+\infty} \frac{1}{G_{\theta_0}^2(x)} \exp(-\frac{G_{\theta_0}(x-\varepsilon)}{G_{\theta_0}(x)})dF_{\theta_0}(x)$ converge; alors, $X_n^* \in (\Gamma_n^{-\varepsilon})^c$ p.co.a..

■ Soit $\varepsilon > 0$. Posons $x = \frac{1}{G_{\theta_0}(t)}$ dans $\tilde{K}(\theta_0, \varepsilon)$. On obtient :

$$(11) \quad \tilde{K}(\theta_0, \varepsilon) = \int_{G_{\theta_0}^{-1}(\frac{1}{c(\varepsilon)})}^{+\infty} \exp(\frac{\text{Log } F_{\theta_0}(t-\varepsilon)}{G_{\theta_0}(t)}) \frac{dF_{\theta_0}(t)}{G_{\theta_0}^2(t)}.$$

De plus pour t suffisamment grand,

$$(12) \quad \text{Log } F_{\theta_0}(t-\varepsilon) = \text{Log}(1 - G_{\theta_0}(t-\varepsilon)) \leq -G_{\theta_0}(t-\varepsilon).$$

Donc, si $L(\theta_0, \varepsilon)$ converge, alors (11) et (12) montrent que $\tilde{K}(\theta_0, \varepsilon)$ converge. D'où le résultat en utilisant le théorème précédent. ■

Ces résultats permettent d'obtenir le théorème 5

g) Théorème 5.- Si, pour un θ_0 , F_{θ_0} appartient à K , alors C^- est satisfaite.

■ Si F_{θ_0} est dans K , alors pour tout $\varepsilon > 0$ et pour x suffisamment grand,

$$(13) \quad \frac{G_{\theta_0}(x-\varepsilon)}{G_{\theta_0}(x) \text{Log } G_{\theta_0}(x)} < -2.$$

D'où

$$\frac{-G_{\theta_0}(x - \varepsilon)}{G_{\theta_0}(x)} - 2 \text{Log } G_{\theta_0}(x) < 0,$$

et, pour x suffisamment grand,

$$(14) \quad \exp\left(\frac{-G_{\theta_0}(x - \varepsilon)}{G_{\theta_0}(x)}\right) \times \frac{1}{G_{\theta_0}^2(x)} < 1.$$

Ainsi $L(\theta_0, \varepsilon) < \infty$; et d'après le théorème précédent C^- est réalisée. ■

h) Finalement, d'après les théorèmes 2 et 5, si (H) est satisfaite et si F_{θ_0} est dans K pour un θ_0 , alors $(X_n^*)_n$ est presque sûrement stable. Ce qui démontre le théorème 5 du chapitre 2.

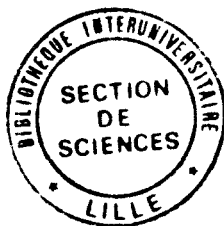
BIBLIOGRAPHIE

GEFFROY, J. (1958 - 1959) - *Contribution à la théorie des valeurs extrêmes*.
Pub. Inst. Stat. Univ. Paris, 1958 (fasc. 4) et 1959 (fasc. 1).

Table des matières

1	Introduction	1
2	Bibliographie	5
3	Chapitre 1 : Exemples de localisation asymptotique du polyèdre d'appui d'un échantillon de R^2	9
3.1	Introduction	11
3.2	Stabilité presque sûre de la valeur maximale d'un échantillon de variables aléatoires réelles	12
3.3	Exposition du problème	13
3.4	Cas où $\psi(t) = \mathbf{1}_{[0,1]}(t)$	14
3.5	Cas où $\Psi(t) = (1 - \exp(-t^\alpha))\mathbf{1}_{\{t>0\}}, (\alpha > 0)$	19
3.6	Conclusion	24
3.7	Bibliographie	25
4	Chapitre 2 : Stabilité de la valeur maximale des échantillons multidimensionnels	27
4.1	Introduction	29
4.2	Préliminaires	31
4.3	Stabilité en probabilité de $(X_n^*)_n$	35
4.4	Stabilité presque-sûre de $(X_n^*)_n$	39
4.5	Exemples	41
4.6	Quelques remarques à propos de (H)	42
4.7	φ -stabilité	45
4.8	Conclusion	48
4.9	Bibliographie	49

5	Chapitre 3 : Localisation asymptotique des échantillons multidimensionnels	51
5.1	Introduction	53
5.2	Préliminaires	55
5.3	Localisation asymptotique de l'échantillon	56
5.4	Exemples	60
5.5	Conclusion	65
5.6	Bibliographie	66
6	Chapter 4 : Distributions multidimensionnelles "outlier-resistant" et "outlier-prone"	67
6.1	Introduction	69
6.2	Préliminaires	71
6.3	Lois A. O. R.	72
6.4	Lois A. O. P.	76
6.5	Généralisation	80
6.6	Conclusion	81
6.7	Bibliographie	82
7	Conclusion	83
8	Appendice	87
8.1	Appendice 1	89
8.2	Appendice 2	91



RESUME

Notre travail concerne la localisation asymptotique presque-sûre d'un échantillon multidimensionnel. Nous étendons ainsi les travaux de Geffroy sur la stabilité de la valeur maximale d'un échantillon de variables réelles.

Dans le premier chapitre, nous considérons un échantillon de variables normales auquel nous imposons des déformations isotropes.

Dans le deuxième chapitre, nous introduisons une notion nouvelle de valeur maximale pour un échantillon multidimensionnel ; nous en donnons les propriétés de stabilité presque-sûre et en probabilité.

Dans le troisième chapitre, nous étudions la répartition asymptotique des extrêmes dans des domaines non convexes délimités par des isobares. Nous introduisons ensuite une notion affaiblie de stabilité dont la stabilité relative est un cas particulier.

Nous revenons dans le quatrième chapitre sur la notion de stabilité des valeurs extrêmes d'un échantillon multidimensionnel pour la comparer avec la notion de loi multidimensionnelle « outlier-resistant » ou « outlier-prone ».

Mots-clés : échantillon, valeur extrême, stabilité, stabilité relative, localisation asymptotique, statistique d'ordre, outlier.