UNIVERSITE MOHAMMED V-AGDAL
FACULTE DES SCIENCES-RABAT

THESE DE DOTORAT

Pr ésentée Par
MY HANAFI AZZAT

Caractérisation de lois limites en statistiques d'ordre

Soutenue le 01 JUIN 2013

Devant le jury composé de :
Président :
Abdelaziz CHAOUBI PES à l'INSEA de Rabat

Examinateurs :
Abdelhak ZOGLAT PES à la Faculté des Sciences de Rabat
Khalid EL HIMDI PES à la Faculté des Sciences de Rabat
Samir HAKEM PES à la Faculté des Sciences de Rabat
Abelouahid IMLAHI PES à la Faculté des Sciences et techniques de Tanger
Mohammadine BELBACHIR PES à la Faculté des Sciences de Rabat

Remerciements

Les travaux présentés dans notre thèse ont été effectués au laboratoire de Mathématiques Appliquées à la Faculté des Sciences de Rabat sous la direction de Professeur BELBACHIR Mohammadine.

Je tiens, à exprimer ma gratitude à Monsieur le Professeur BELBACHIR Mohammadine pour son soutien, ses conseils avisés et sa disponibilité qui ont contribué à la réalisation de cette thèse. Son encadrement a favorisé le développement de mes compétences et de mon goût pour la recherche. Je salue également sa souplesse et son ouverture d'esprit qui ont su me laisser une large marge de liberté pour mener à bien ce travail de recherche.

Je tiens, également à exprimer toute ma reconnaissance au Professeur A. CHAOUBI, qui m'a fait l'honneur d'accepter de présider mon jury de thèse.

Je tiens à adresser mes vifs remerciements aux Professeurs Mr Abdelhak ZOGLAT et Mr Abdelouahid IMLAHI pour avoir accepté d'être rapporteurs de cette thèse.

Comme je tiens à remercier vivement les Professeurs Mr Khalid EL HIMDI et Mr Samir HAKAM pour l'intérêt qu'ils ont porté à mon travail et pour avoir accepté de faire partie du jury.

Un grand remerciement à mes amis et compagnons doctorants et plus particulièrement à ceux qui m'ont soutenu pendant les moments difficiles en me faisant profiter de leur bonne humeur.

Enfin, je dédie cette thèse à ma famille qui, depuis de longues années, a su m'encourager dans mes choix et, sans qui, je ne serais pas devenu ce que je suis. Que ceux que j'ai oubliés de citer ici, me pardonnent, je leur dois sans doute beaucoup aussi.

Table des matières

Introduction Générale

La théorie classique des valeurs extrêmes est apparue dès les années vingt et a contribué à la solution de nombreux problèmes statistiques. La modélisation des événements extrêmes(tels les inondations, les crus, les tempêts, les pics de pollution etc.) est aujourd'hui un champ de recherche particulièrement actif. Ces événements extrêmes peuvent causer des dégâts humains et matériels considérables. De telles catastrophes ne peuvent pas toujours être évitées. Cependant, la société peut prendre des actions préventives pour minimiser leurs effets. Pour cela, les spécialistes ont à leur disposition la théorie statistique des valeurs extrêmes. Elle donne un résultat très intéressant parce qu'il est d'une portée générale. Quelle que soit la loi de la variable parente, la loi limite des extrêmes a toujours la même forme. IL est à noter avec intérêt que depuis quelques années, la théorie des valeurs extrêmes(TVE) est largement utilisée pour la modélisation de tels événements. Pour une présentation assez complète du sujet, nous renvoyons à l'ouvrage de référence de Embrechts, Kluppelberg et Mikosch [1997] rappelant les principaux résultats théoriques sur la TVE. De même, Reiss et Thomas [2001] ont proposé un certain nombre d'exemples pratiques. Les domaines d'application utilisant les modèles de la TVE n'ont cessé de se développer ces dernières années en touchant des domaines variés : en hydrologie, la prévision des crues, par exemple, est particulièrement importante (Davison et Smith [1990]; Katz [2002]), en assurance, l'une des préoccupations, est la prise en compte des grands sinistres (McNeil et al., [1997]; Rootzen et Tajvidi [1997]). Leur introduction en finance (Embrechts et al. [1997]; Danielsson et de Vries [1997]; McNeil [1998]; Longin [1998][2000]; Embrechts [1999]; Gen et Sel, [2004]) est une réponse immédiate à la remise en cause de l'hypothèse de normalité surtout avec les observations en hautes fréquences. En météorologie (Coles et Walshaw [1994]; Smith [2001]; Klajnmic [2003]) où l'étude de la vitesse du vent, par exemple, permet d'évaluer le degré de résistance des matériaux face à la pression exercée par le vent (au cours d'une tempête par exemple) sur les bâtiments ou les structures de génie civile. Dans les domaines des sciences humaines et sociales, la théorie des valeurs extrêmes pourra également contribuer à la compréhension de nombreux phénomènes sociologiques. Un grand débat sur "la durée extrême de la vie humaine" a été initié par Gumbel [1937], et dans lequel Fréchet[1927] puis plus tard Thatcher [1999] ont pris une part active.

Lorsqu' on modélise le maximum d'un ensemble de variables aléatoires, alors, sous certaines conditions que nous préciserons plus loin, la distribution ne peut appartenir qu' à l'une des trois lois suivantes : Weibull (à support borné), Gumbel et Fréchet (à support non borné). L'introduction de la TVE est devenue une exigence dans de nombreux domaines. En effet, les outils probabilistes traditionnels (développés dans un univers gaussien) sont inadaptés à l'appréhension de ces comportements extrêmes : la moyenne n'existe plus dans le nouvel univers des risques. L'outil de mesure des risques de marché communément reconnu, surtout en finance, est la Value at Risk (VaR). Elle a été introduite pour répondre à une question assez simple mais extrêmement précise :"Pour une valeur de probabilité donnée, à combien peut s'élever la perte potentielle associée à un portefeuille donné, au cours d'un horizon T fixé (1 jour, 10 jours,...)?". L'estimation de la VaR devra donc passer par l'utilisation des techniques issues de la théorie des extrêmes qui s'est particulièrement développée depuis le milieu des années 70 (voir les travaux de Pickands [1975]; Hill [1975]; Smith [1987]; ou encore Dekkers et al. [1989]). Récemment, cette théorie a été appliquée surtout à la finance et à l' assurance (voir McNeil [1997]; McNeil et Frey [2000]). La TVE permet d'évaluer les événements rares et les pertes associées à leur apparition.

Dans ce travail, le premier chapitre est un aperçu bibliographique sur la théorie des valeurs extrêmes. Nous rappelons que celle-ci consiste en l'étude du comportement asymptotique du maximum de n variables aléatoires(v.a.) indépendantes et identiquement distribuées(i.i.d.). Aussi nous nous intéressons à la caractérisation des lois limites. Nous y évoquons aussi le lien entre le comportement asymptotique de la distribution d'une somme de v.a.i.i.d. et celui de la distribution des maxima.

Au Chapitre 2, nous allons établir une loi limite du maximum de la somme de k variables aléatoires. Nous nous intéressons essentiellement à une famille particulière de lois, qui ont une fonction de survie qui décroit vers zéro à la vitesse exponentielle. A la fin nous présenterons quelques exemples afin d'illustrer ces différents résultats.

Dans le chapitre 3, nous étendons l'étude des lois limites extrêmes aux variables aléatoires indépendantes non identiquement distribuées dans le cas d'une normalisation linéaire.

Enfin dans le dernier chapitre, nous étudions la caractérisation des lois limites du maximum de n v.a. i.i.d. dans le cas d'une normalisation non linéaire. Nous donnons des conditions nécessaires et suffisantes d'appartenance aux domaines d'attraction de ces lois limites et établissons des expressions générales des constantes de normalisation.

CHAPITRE 1 : INTRODUCTION A LA THEORIE DES VALEURS EXTREMES

1.1 Introduction

Ce chapitre introduit la notion des valeurs extrêmes pour une suite de variables aléatoires et énonce les principaux résultats probabilistes les concernant. Nous y exposerons uniquement la théorie des valeurs extrêmes pour des variables aléatoires indépendantes. Cette théorie est en quelque sorte le parallèle de la théorie des lois stables caractérisant la distribution limite de la somme de variables aléatoires indépendantes. Nous donnerons ensuite une définition de telles fonctions et nous en concluons quelques propriétés afin d'étudier le domaine d'attraction de chaque loi limite possible.

1.2 Lois limites du $k^{ème}$ maximum

La théorie des valeurs extrêmes a pour but d'étudier la loi du maximum d'une suite de variables aléatoires réelles, même si la loi du phénomène n'est pas connue. Formellement, soit $X_1, X_2, ..., X_n$ une suite de n variables aléatoires indépendantes identiquement distribuées (i.i.d.) de fonction de répartition commune $F(x) = P(X_1 \leq x)$.

Soient $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ les statistiques d'ordre correspondantes. Nous rappelons que $X_{n-k+1,n}$ est le $k^{ème}$ maximum supérieur.

Sa loi est donnée par :

$$P(X_{n-k+1} \leq x) = \sum_{i=0}^{k-1} \binom{n}{i} (1 - F(x))^i F^{n-i}(x), \tag{1.1}$$

$$\text{où} \quad \binom{n}{i} = \frac{n!}{i!(n-i)!}.$$

Pour l'obtention de l'expression (1.1) voir Galambos [1978].

Le cas $k = 1$, correspond au maximum

$$X_{n,n} = \max(X_1, ..., X_n).$$

Comme les variables aléatoires sont i.i.d., la fonction de répartition de $X_{n,n}$ est $P(X_{n,n} \leq x) = F^n(x)$. Puisque le fait que $\lim_{n \to +\infty} F^n(x) = 0$ pour $0 \leq F(x) < 1$ ne présente aucun intérêt, il est intéressant de considérer le comportement asymptotique du maximum convenablement normalisé. Par analogie avec le théorème de la limite centrale, qui décrit la distribution limite d'une somme de variables aléatoires i.i.d., l'objectif de la théorie des valeurs extrêmes est de trouver un résultat analogue pour $X_{n,n}$. De nombreux auteurs ont étudié le comportement asymptotique du maximum $X_{n,n}$ normalisé lorsque $n \to +\infty$. Fisher-Tippet [1928] et Gnedenko[1943] ont démontré le résultat suivant :

Théorème 1.1
Soient $a_n > 0$ et b_n deux suites numériques.
Alors $$\lim_{n\to+\infty} P(X_{n,n} \leq a_n x + b_n) = G(x) \qquad (1.2)$$
où G est une fonction de répartition non dégénérée si et seulement si G est une fonction de répartition appartenant à l'un des trois types suivants :
- Loi de Fréchet :
$$\Phi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\}, & \text{si} \quad x > 0 \\ 0, & \text{si} \quad x \leq 0 \end{cases}$$
- Loi de Weibull :
$$\Psi_\alpha(x) = \begin{cases} 1, & \text{si} \quad x \geq 0 \\ \exp\{-(-x)^\alpha\}, & \text{si} \quad x < 0 \end{cases}$$

où $\alpha > 0$ est une constante.
- Loi de Gumbel :
$\Lambda(x) = \exp\{-e^{-x}\}, -\infty < x < +\infty,$

Ce théorème énonce un résultat très intéressant : quelle que soit la loi de la variable parente, la loi limite des extrêmes prend toujours l'une des trois formes ci-dessus. De plus il fournit, dans le cas d'événements extrêmes un résultat analogue au théorème central limite (TCL). Cependant, contrairement au TCL, où la loi normale est la seule loi limite possible, dans le cas des extrêmes, trois types de loi limites sont possibles. La preuve de ce théorème a été formulée plusieurs fois. Sa première version complète est donnée par Gnedenko[1943].

Remarquons que (1.2) équivaut à $\lim_{n\to+\infty} F^n(a_n x + b_n) = G(x)$ pour tout x point de continuité de G.

Notons $\underline{G} = \inf\{x : G(x) > 0\}$ et $\overline{G} = \sup\{x : G(x) < 1\}$,

Une des conséquences du théorème 2.1 est le résultat suivant qui décrit le comportement asymptotique du $k^{ème}$ maximum (voir par exemple Galambos [1978]).

Théorème 1.2
Soient $a_n > 0$ et b_n deux suites de constantes réelles.
Si $$\lim_{n\to+\infty} F^n(a_n x + b_n) = G(x) \quad \text{pour} \quad \underline{G} < x < \overline{G}\ , \qquad (1.3)$$
alors $$\lim_{n\to+\infty} P(X_{n-k+1,n} \leq a_n x + b_n) = G(x) \sum_{j=0}^{k-1} \frac{(-\log G(x))^j}{j!}, \qquad (1.4)$$
pour tout $k > 1$ fixé.

Définition 1.1
On dit que deux fonctions de répartition F et H sont de même type, si l'on peut trouver deux constantes $c_1 > 0$ et $c_2 \in IR$ telles que pour tout $x \in IR$, $H(x) = F(c_1 x + c_2)$.

Autrement dit, les variables aléatoires de même type ont la même loi, à un facteur de localisation et d'échelle près.

Définition 1.2

On dit qu'une fonction de répartition F appartient au domaine d'attraction de G, et on note $F \in DA(G)$ si la distribution du maximum normalisé converge vers G.

Remarque 1.1
Il est possible de regrouper ces trois lois limites sous une seule forme en utilisant la paramétrisation de **Von Mises**, en posant $\gamma = \frac{1}{\alpha}$ de telle façon que pour $\gamma \to 0$ on obtienne la loi de Gumbel Λ. La forme unifiée est la suivante :

$$G_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & 1+\gamma x > 0, \quad \gamma \neq 0 \\ \exp(-e^{-x}), & \text{si} \qquad \gamma = 0 \end{cases}$$

Ayant déterminé la forme des lois limites, il convient à présent de déterminer le domaine d'attraction pour une fonction de répartition F donnée. Pour cela supposons que $F \in DA(G)$, et rappelons que pour $n \to \infty$

$$(1 - \frac{x}{n})^n \to \exp(-x).$$

Nous obtenons pour n grand

$$F^n(a_n x + b_n) = (1 - \frac{n\bar{F}(a_n x + b_n)}{n})^n \sim \exp(-n\bar{F}(a_n x + b_n)),$$

en notant $\bar{F} = 1 - F$. Donc, si (1.2) est vérifiée, alors

$$n\bar{F}(a_n x + b_n) \to -\log G(x). \tag{1.5}$$

Ainsi, la loi limite $G(x)$ dépend-elle exclusivement de la queue de distribution à droite de F. En particulier, deux fonctions de répartition F et H sont dites à queue de distribution équivalente si elles possèdent la même borne supérieure pour le support ($x_F = x_H := x_0$), avec $x_F = \sup\{x : F(x) < 1\}$, et si pour une certaine constante positive $A > 0$, $\lim\limits_{x \uparrow x_0} \frac{1-F(x)}{1-H(x)} = A$.

On peut alors montrer **(Resnick, 1987)[prop. 1.19]** que dans ce cas, si F(resp. H) appartient au domaine d'attraction d'une loi limite G_1 (resp G_2), alors G_1 et G_2 sont de même type.

Ce résultat implique qu'il suffit de trouver un prototype de queue de distribution $1 - H(x)$ pour H appartenant à un domaine d'attraction de l'un des trois types de loi extrêmes, pour en déduire une caractérisation de toutes les fonctions de répartition F appartenant au même domaine, sous la forme

$$1 - F(x) = c(x)(1 - H(x)), \qquad x > z_0$$

où $z_0 < x_F$ est une constante et $\lim_{x \uparrow x_0} c(x) = c > 0$.

Bien sûr, le théorème 1.2 n'est valable que si les suites $a_n > 0$ et $b_n \in IR$ existent.

Par exemple dans le cas de la loi exponentielle de paramètre 1, nous avons, pour $a_n = 1$ et $b_n = \log n$,

$$\begin{aligned}
\lim_{n \to +\infty} F^n(a_n x + b_n) &= \lim_{n \to +\infty} (1 - \exp(-x - \log n))^n \\
&= \lim_{n \to +\infty} (1 - \frac{\exp(-x)}{n})^n \\
&= \exp(-\exp(-x)).
\end{aligned}$$

Remarque 1.2

Les distributions Λ, Φ_α et Ψ_α sont appelées distributions des valeurs extrêmes et les variables aléatoires correspondantes sont les variables aléatoires extrêmales. Nous analysons maintenant la différence entre ces trois distributions.

1.3 Caractérisation des domaines d'attraction

Les résultats de cette section sont présentés dans Resnick[1987] et Embrechts et al.[1997]. Dans la section précédente, nous avons vu que la distribution exponentielle appartient au domaine d'attraction de Λ. Nous constatons que les trois distributions des valeurs extrêmes sont très différentes en terme de domaine d'attraction :

- Dans le domaine d'attraction de la loi de Gumbel ($DA(\Lambda)$), où la loi de X présente une décroissance de type exponentiel pour la queue de la loi (c'est le cas par exemple des lois normale, exponentielle, gamma ou lognormale), nous trouvons des distributions qui n'ont pas de queue épaisse.

- Dans le domaine d'attraction de la loi de Fréchet ($DA(\Phi_\alpha)$), correspondant à une loi de X non bornée (c'est le cas par exemple de la loi de Cauchy ou Pareto), nous trouvons des distributions qui ont des queues épaisses.

- Dans le domaine d'attraction de la loi Weibull ($DA(\Psi_\alpha)$), où la loi de X est bornée (c'est le cas par exemple de la loi uniforme ou béta), nous trouvons des distributions à support fini, ce qui implique que le support du maximum est borné à droite.

Etant donné une fonction de répartition F, nous voudrions connaître à quel domaine d'attraction elle appartient et quelles sont les constantes de normalisation correspondantes. La réponse est relativement complexe et n'est pas unique. Nous indiquons ici les critères les plus utilisés en faisant appel à la notion de fonctions à variation régulière.

1.3.1 Fonctions à variation régulière

Les fonctions à variation régulière et leurs propriétés sont souvent rencontrées dans cette thèse. Pour plus d'informations sur ces notions, on peut consulter l'excellente référence de Bingham et al[1987].

Définition 1.3

Une fonction mesurable positive f est dite à variation régulière d'indice α à l'infini et on note $f \in VR_\alpha$ si pour tout $x > 0$, nous avons

$$\lim_{t \to +\infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

On dira que f est à variation régulière au voisinage de 0 si $f(x^{-1})$ est à variation régulière à l'infini. Si $\alpha = 0$, on dit que f est à variation lente, on la note souvent l(ou parfois L).

Remarque 1.3

Les exemples typiques de fonctions à variation lente sont les constantes positives ou les fonctions convergeant vers une constante positive. Citons, $\ln(1+x)$, $\ln\ln(e+x)$, etc.

Il est facile de montrer qu'une fonction f à variation régulière d'indice α peut toujours s'écrire sous la forme

$$f(x) = x^\alpha l(x) \quad \text{où} \quad l \in VR_0.$$

1.3.2 Domaine d'attraction de Fréchet

Considérons une fonction de répartition dans le domaine d'attraction de la loi de Fréchet Φ_α. Tout d'abord, le point terminal $x_F = \sup\{x : F(x) < 1\}$ est nécessairement infini.

Gnedenko(1943) a obtenu le résultat suivant

Théorème A

$$F \in DA(\Phi_\alpha) \Leftrightarrow 1 - F \in VR_{-\alpha}.$$

En plus

$$F^n(a_n x) \to \Phi_\alpha(x), \text{ lorsque } n \to \infty$$

avec

$$a_n = \left(\frac{1}{1-F}\right)^{\leftarrow}(n) \text{ et } b_n = 0$$

où $H^{\leftarrow}(y) = \inf\{x : H(x) \geq y\}$ est la fonction inverse généralisée de H.

Ce résultat permet d'obtenir la forme de F en utilisant la représentation de Karamata des fonctions L à variation lente.

Proposition 1.4 (Représentation de Karamata)

L est à variation lente ssi L peut être représentée sous la forme

$$L(x) = c(x)\exp\{\int_1^x t^{-1}\epsilon(t)dt\},$$

pour $x > 1$, où $c : IR^+ \to IR^+$, $\epsilon : IR^+ \to IR^+$ et

$$\begin{aligned} \lim_{x\to+\infty} c(x) &= c > 0 \\ \lim_{x\to+\infty} \epsilon(x) &= 0. \end{aligned}$$

Théorème B

$F \in DA(\Phi_\alpha)$ si et seulement si il existe des fonctions $c(.)$ et $\alpha(.)$ définies sur $]1, \infty[$ qui vérifient

$$\begin{aligned} \lim_{x\to+\infty} c(x) &= c > 0 \\ \lim_{x\to+\infty} \alpha(x) &= \alpha > 0, \end{aligned}$$

et telles que

$$1 - F(x) = c(x)\exp\{-\int_1^x t^{-1}\alpha(t)dt\}.$$

1.3.3 Domaine d'attraction de Weibull

Comme dans le cas précédent, il existe un lien entre les fonctions de répartition appartenant au domaine d'attraction de Weibull et la variation régulière de la queue de distribution supérieure. Cependant il apparait une différence majeure avec le cas précédent, à savoir que si $F \in DA(\Psi_\alpha)$, alors nécessairement $x_F < \infty$(cf. (Resnick, 1987)[prop. 1.13]).

Notons que pour tout $y > 0$,

$$\Psi_\alpha(-1/y) = \Phi_{-\alpha}(y),$$

et introduisons la fonction

$$F_\star(x) = \begin{cases} 0, & \text{si} \quad x \leq 0 \\ F(x_F - x^{-1}), & \text{si} \quad x > 0. \end{cases}$$

Supposons qu'il existe $a_n > 0$ et b_n tels que, lorsque $n \to \infty$,

$$\begin{aligned} F^n(a_n x + b_n) &\to \Psi_\alpha(x),\ x < 0 \\ F^n(b_n - a_n y^{-1}) &\to \Phi_\alpha(y),\ y > 0 \\ F_\star^n\left(\frac{y}{a_n - y(x_F - b_n)}\right) &\to \Phi_\alpha(y),\ y > 0 \end{aligned}$$

En prenant $b_n = x_F$, nous obtenons que $F_\star$ appartient au domaine d'attraction de Φ_α, d'où d'après le résultat de Gnedenko[1943]

$1 - F_\star \in VR_{-\alpha}$ et $a_n^{-1} = (1/(1 - F_\star))^{\leftarrow}(n)$, et par conséquent $a_n = x_F - (1/(1 - F))^{\leftarrow}(n)$.

Une caractérisation de F s'en déduit comme précédemment

Théorème C

$F \in DA(\Psi_\alpha)$ ssi $x_F < \infty$ et s' il existe des fonctions positives $c(x)$ et $\delta(x)$ définies sur IR^+ et une constante $c_0 > 0$ telles que

$$\lim_{t \uparrow x_F} \delta(t) = \alpha, \qquad \lim_{t \uparrow x_F} c(t) = c_0,$$

et pour tout $x < x_F$,

$$1 - F(x) = c(x) \exp\{- \int_{x_F^{-1}}^{x} \frac{\delta(t)}{x_F - t} dt\}.$$

1.3.4 Domaine d'attraction de Gumbel

La caractérisation du domaine d'attraction de la loi de Gumbel $DA(\Lambda)$ est basée sur la notion de fonction de Von Mises F de point terminal $x_0 = x_F$, dont la représentation est la suivante : il existe $c > 0$ et $z_0 < x_0$ tels que pour tout $z_0 < x < x_0$,

$$1 - F(x) = c \exp\{- \int_{z_0}^{x} \frac{1}{f(u)} du\} \tag{1.6}$$

où $f(u) > 0$ pour $z_0 < u < x_0$, et f est absolument continue sur $]z_0, x_0[$, de dérivée $f^{'}(u)$ telle que $\lim_{u \uparrow x_0} f^{'}(u) = 0$.

Un choix possible de f

$$f(x) = \int_{x}^{x_F} \frac{1 - F(t)}{1 - F(x)} dt, \qquad x < x_F$$

Resnick (1987) montre le résultat suivant :

Théorème D
Si $f(u)$ est définie comme ci-dessus, alors nécessairement
(a) si $x_0 = \infty$, $\lim\limits_{t\to\infty} t^{-1}f(t) = 0$
(b) si $x_0 < \infty$, $f(x_0) = \lim\limits_{t\to x_0} f(t) = 0$ et $\lim\limits_{t\uparrow x_0}(x_0 - t)^{-1}f(t) = 0$

et dans tous les cas

$$\lim_{t\to x_0}(t + xf(t)) = x_0.$$

D'autre part,

$$\lim_{t\to x_0} \frac{f(t+xf(t))}{f(t)} = 1, \quad \text{localement uniformément en } x$$

De cette dernière relation on déduit immédiatement que pour F de la forme (1.6)

$$\begin{aligned}
\lim_{t\to x_0} \frac{1-F(t+xf(t))}{1-F(t)} &= \lim_{t\to x_0} \exp\{-\int_t^{t+xf(t)} \frac{1}{f(u)}du\} \\
&= \lim_{t\to x_0} \exp\{-\int_0^x \frac{f(t)}{f(t+sf(t))}ds\} \quad \text{où} \quad s = \frac{u-t}{f(t)} \\
&= e^{-x},
\end{aligned} \tag{1.7}$$

Si (1.7) est vérifiée on dit que $\frac{1}{1-F}$ est à variation de type Γ et on note $\frac{1}{1-F} \in \Gamma$ d'où $F \in DA(\Lambda)$.
Ainsi les fonctions de Von Mises appartiennent au domaine d'attraction de la loi de Gumbel et on peut choisir pour constantes de normalisation

$$b_n = (\frac{1}{1-F})^{\leftarrow}(n), \qquad a_n = f(b_n).$$

Un résultat dû à Balkema et de Haan (1972)((Resnick, 1987)[prop. 1.4]) permet d'obtenir la caractérisation du domaine d'attraction de la loi de Gumbel sous la forme suivante :

Théorème E

$F \in DA(\Lambda)$ ssi il existe une fonction de Von Mises F telle que pour $x \in]z_0, x_0[$,

$$1 - F(x) = c(x)(1 - F(x)),$$

avec

$$\lim_{t \to x_0} c(x) = c > 0.$$

1.3.5 Conditions de Von Mises

Les conditions nécessaires et suffisantes pour qu'une fonction de répartition F appartienne à un domaine d'attraction d'une loi extrême donnée ci-dessus sont souvent difficiles à vérifier. Pour cette raison, nous présentons ci-après des conditions suffisantes, dues à Von Mises (1936), plus simples à vérifier et sans grandes restrictions. Néanmoins, elles ne seront applicables que pour les fonctions de répartition ayant une densité. Les conditions de Von Mises sont les suivantes (Resnick, 1987) :

Théorème F

Soit F une fonction de répartition de densité $F^{'}(x)$.

(a) Si pour un certain $\alpha > 0$,

$$\lim_{x \to \infty} \frac{xF^{'}(x)}{1 - F(x)} = \alpha,$$

alors $F \in DA(\Phi_\alpha)$.

(b) Si $x_F < \infty$, et pour un certain $\alpha < 0$,

$$\lim_{x \to x_F} \frac{(x_F - x)F^{'}(x)}{1 - F(x)} = -\alpha,$$

alors $F \in DA(\Psi_\alpha)$

(c) Si

$$\lim_{x \to x_F} F^{'}(x) \int_x^{x_F} \frac{1 - F(t)}{(1 - F(x))^2} dt = 1,$$

alors $F \in DA(\Lambda)$.

(d) Si F admet une dérivée seconde $F^{''}$, et si

$$\lim_{x \to x_F} \frac{F^{''}(x)(1 - F(x))}{(F^{'}(x))^2} = -1,$$

alors $F \in DA(\Lambda)$.

Remarque 1.4

Pour une v.a. $X > 0$, la fonction auxiliaire $f(t)$ s'exprime par

$$f(t) = E(X - t/X > t).$$

1.3.6 Lois de Pareto généralisées

Nous avons vu, dans le paragraphe précédent, la paramétrisation de Von Mises pour les lois extrêmes

$$G_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & 1+\gamma x > 0, \quad \gamma \neq 0 \\ \exp(-e^{-x}), & \text{si} \qquad \gamma = 0 \end{cases}$$

Comment caractériser le domaine d'attraction de G_γ? D'après (Embrechts et al., 1997)[theorem 3.4.5], il y a équivalence entre les trois assertions :

Théorème G

(a) $F \in DA(G_\gamma)$

(b) Il existe une fonction $a(.)$ positive et mesurable telle que pour tout x vérifiant, $1+\gamma x > 0$,

$$\lim_{x \to x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1+\gamma x)^{-1/\gamma}, & \text{si } \gamma \neq 0 \\ e^{-x}, & \text{si } \gamma = 0. \end{cases}$$

(c) Pour x, $y > 0$, $y \neq 1$,

$$\lim_{x \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\gamma - 1}{y^\gamma - 1}, & \text{si } \gamma \neq 0 \\ \frac{\log x}{\log y}, & \text{si } \gamma = 0, \end{cases}$$

avec $U = (\frac{1}{1-F})^{\leftarrow}$.

La forme des limites figurant dans (b) motive la définition des lois de Pareto généralisées de fonction de répartition $W\gamma$ définie par

$$W_\gamma(x) = \begin{cases} 1 - (1+\gamma x)^{-1/\gamma}, & \text{si } \gamma \neq 0 \\ 1 - e^{-x}, & \text{si } \gamma = 0 \end{cases}$$

où

$$\begin{array}{lcl} x & \geq & 0 \qquad\qquad\qquad \text{si} \quad \gamma \geq 0 \\ 0 & \leq & x \leq -1/\gamma \qquad \text{si} \quad \gamma < 0. \end{array}$$

Ainsi il y a équivalence entre l'appartenance de F au domaine d'attraction de G_γ et

$$\lim_{x \to x_F} F_u(xa(u)) = W_\gamma(x). \tag{1.8}$$

où $F_u(x) = P(X - u \leq x / X > u)$

Soit

$$W_{\gamma,\beta}(x) := W_\gamma(x/\beta),$$

alors (1.8) signifie que $F \in DA(G_\gamma)$ ssi

$$\lim_{x \to x_F} |F_u(x) - W_{\gamma, a(u)}(x)| = 0.$$

Ce résultat est à la base d'une méthode statistique d'étude des évènements extrêmes appelée Peaks Over Threshold (POT) basée sur la comparaison entre la queue de distribution et une loi de Pareto généralisée puisque

$$\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y) \approx \bar{F}(u)W_{\gamma, a(u)}(y).$$

1.4 La loi de la somme

Dans ce paragraphe, nous nous intéressons à l'étude des sommes de v.a.i.i.d. pour passer ensuite à celle des maxima de ces variables. En effet, ce passage est relativement facile et commode. Dans un premier temps, la théorie des valeurs extrêmes a concerné le comportement d'événements rares apparaissant dans des suites de variables aléatoires indépendantes. L'un des principaux problèmes abordés est relatif à la somme de v.a.i.i.d. Ceci correspond au comportement d'une marche aléatoire qui apparait comme un cas particulier du mouvement brownien. Une abondante littérature est consacrée à ce thème. Elle inclut des résultats classiques concernant le théorème central limite, la loi faible des grands nombres, la loi des logarithmes itérés et d'autres extensions. On peut citer comme exemple de travaux fondateurs ceux de **Breiman [1968], Feller [1971]**.

Nous allons rappeler un résultat fondamental qui permet de faire le lien entre la théorie fondée sur le comportement asymptotique de la distribution d'une somme de v.a. i.i.d. et celui de la distribution des maxima. Il s'agit du théorème de **Heyde [1967]** sur les grandes déviations. La théorie des grandes déviations (ou des grands écarts) concerne la loi faible des grands nombres.

Soient $X_1, ..., X_n$ une suite de v.a. i.i.d. de fonction de répartition commune F et $S_n = \sum_{i=1}^{n} X_i$. Selon la loi faible des grands nombres, $\frac{S_n}{n}$ converge en probabilité vers $E(X)$. On montre alors que les grandes déviations sont rares et que

$$P(S_n > na) \simeq e^{-nh(a)},$$

où $h(.)$ est une fonction continûment dérivable.

On définit le domaine d'attraction d'une loi $\alpha-$stable comme suit :

Définition 1.5

On dit que F appartient au domaine d'attraction d'une loi G_α $\alpha-$stable ($\alpha \in]0, 2[$) et on note $F \in D(G_\alpha)$ s'il existe des constantes a$_n \in IR$ et $b_n > 0$ telles que

$$\frac{S_n - a_n}{b_n} \overset{L}{\to} G_\alpha, \quad \text{lorsque } n \to +\infty$$

Théorème 1.4

Soient X une variable aléatoire appartenant au domaine d'attraction $D(G_\alpha)$ avec $\alpha \in]0,2[$ et (b_n) une suite croissante de réels telle que : $b_n \uparrow^{+\infty}$ et $P(X > b_n) \sim \frac{1}{n}$.

Si on note $M_1 = X_1$ et $M_n = \max(X_1, ..., X_n)$, $n > 2$

Alors

$$\lim_{n\to\infty} \frac{P(S_n > b_n x_n)}{nP(X > b_n x_n)} = \lim_{n\to\infty} \frac{P(S_n > b_n x_n)}{P(M_n > b_n x_n)} = 1,$$

pour toute suite $x_n \to \infty$, $n \to +\infty$.

Des résultats analogues à celui-ci peuvent être établis pour des variables aléatoires dont la distribution affiche des comportements de queue réguliers que l'on peut caractériser de la manière suivante :

$1 - F(x) \sim x^{-\alpha} L(x)$, pour $\alpha > 2$, quand $x \to \infty$, où L est une fonction à variation lente.

Le théorème précédent peut être compris et interprété comme le résultat suivant :

$$\lim_{n\to\infty} \frac{P(S_n > x)}{P(M_n > x)} = 1.$$

Autrement dit, pour un nombre entier n grand, la probabilité pour qu'une somme de réalisations d'une suite de n variables aléatoires dépasse un seuil x(grand) est proche de la probabilité pour que la plus grande des réalisations de cette suite dépasse x. Ainsi, pour qu'une somme de variables aléatoires dépasse un seuil x, il suffit qu'une des ses composantes prenne une valeur extrême très supérieure aux valeurs prises par chacun des autres éléments de la suite. Cette réalisation est l'évènement rare dont on cherche à identifier et à estimer la distribution.

Chapitre 2 : LOIS LIMITES POUR LE MAXIMUM DES SOMMES DANS LE DOMAINE D'ATTRACTION DE LA LOI DE GUMBEL

2.1 Introduction

Soient $X_1, ..., X_n$, n variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition F continue. Soient $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ les statistiques d'ordre correspondantes. D'après Fisher et Tippett, [1928], il existe deux suites de constantes $a_n > 0$ et $b_n \in IR$ telles que

$$\lim_{n \to +\infty} F^n(a_n x + b_n) = H_1(x) , \qquad x \in IR \tag{2.1}$$

pour tout réel x, point de continuité de $H_1 \in \{\Lambda, \Phi_\alpha, \Psi_\alpha\}$, ensemble de trois limites extrêmes définies d'une façon précise dans le chapitre 1. Lorsqu' une fonction de répartition F vérifie (2.1), on dit que F appartient au domaine d'attraction de H_1 et on note $F \in DA(H_1)$.

On peut trouver des applications à ces modèles dans des domaines très différents tels que les finances, les assurances, la climatologie et l'hydrologie. Le domaine d'attraction de Gumbel contient un faisceau très large de fonctions de répartition F dont les queues de distribution peuvent être très différentes. Les queues peuvent varier des plus légères (exemple de la loi normale) à celles modérément lourdes (exemple de la loi lognormale). Pour cette raison ce modèle d'attraction est peut être le plus intéressant des trois modèles d'attraction cidessus. Cependant, bien que ces trois modèles soient si différents, d'un point de vue mathématique ils sont étroitement liés. En effet, on peut vérifier facilement les propriétés suivantes :

Supposons la variable aléatoire $X > 0$, alors

$$X \text{ a pour } f.r.\ \Phi_\alpha \Longleftrightarrow -X^{-1} \text{a pour } f.r.\ \Psi_\alpha \Longleftrightarrow \log X^\alpha \text{a pour } f.r.\ \Lambda.$$

Par le passé, de nombreux articles ont étudié le comportement asymptotique des sommes partielles $S_n = \sum_{i=1}^n X_i$, pour une suite de v.a. $(X_n)_{n\geq 1}$. Plus récemment Dingcheny et Qihe [2004] se sont penché sur le comportement limite des maxima dans le cas de variables aléatoires négatives associées, à queues lourdes. Par la suite Barry, Kang et Yongcheng[2007] se sont intéressé à la loi limite conjointe de la somme du maximum de variables aléatoires gaussiennes multivariées. Pour plus de résultats sur ce sujet, on peut se référer au livre d'Embrecht et al. [1997] et les références qu'il contient.

Notre but dans ce chapitre est d'établir une loi limite de la variable aléatoire

$$\max(X_{i_1} + ... + X_{i_k}),$$

dûment normalisée d'abord dans le cas particulier simple $k = 2$, puis ensuite nous traiterons le cas général $k \geq 2$. A la fin nous présenterons quelques exemples afin d'illustrer ces différents résultats.

2.2 Loi jointe des k premiers maxima

Une des conséquences du Théorème 1.1 dans le chapitre 1 est le résultat suivant qui décrit le comportement asymptotique du $k^{ème}$ maximum supérieur (voir par exemple Galambos [1987]).

Théorème A

Soient $a_n > 0$ et b_n deux suites réelles non aléatoires. Si $\lim_{n\to+\infty} F^n(a_n x + b_n) = G(x)$ pour tout point de continuité de G, alors

$$\lim_{n\to+\infty} P(X_{n-k+1,n} \leq a_n x + b_n) = G(x) \sum_{j=0}^{k-1} \frac{(-\log G(x))^j}{j!}, \tag{2.2}$$

pour tout $k > 1$ fixé.

Pour k fixé et $n \to +\infty$, considérons maintenant la loi limite jointe de $(X_{n,n}, X_{n-1,n}, ..., X_{n-k+1,n})$ lorsque $n \to +\infty$.

Notons $I_n(A) = \#\{\frac{X_i - b_n}{a_n} \in A,\ 1 \leq i \leq n\}$

où $a_n > 0$ et b_n sont deux suites réelles et $\#B$ désigne le nombre d'éléments appartenant à l'ensemble B.

Lorsque $y \to 0$, nous avons $\log(1-y) \simeq -y$, (2.3)

et par conséquent

$\lim_{n\to+\infty} F^n(a_n x + b_n) = G(x)$

$\Leftrightarrow \lim_{n\to+\infty} n \log[1 - (1 - F(a_n x + b_n))] = \log G(x)$

$\Leftrightarrow \lim_{n\to+\infty} n[1 - F(a_n x + b_n)] = -\log G(x).$ (2.4)

Ici (2.4) est obtenue à partir de (2.3) avec $y = 1 - F(a_n x + b_n) \to 0$, lorsque $n \to +\infty$.

Weissman (1975) aboutit au résultat suivant.

Théorème B

Supposons que la condition (2.4) soit satisfaite. Soient $A_1, A_2, ..., A_k$ k ensembles boréliens (avec k fixé). Alors il existe sur IR une mesure aléatoire de Poisson I avec $E\{I(x, +\infty]\} = -\log G(x)$ et telle que :

$$(I_n(A_1), I_n(A_2), ..., I_n(A_k)) \xrightarrow{L} (I(A_1), I(A_2), ..., I(A_k))$$

où $^L \to$ désigne la convergence en loi.

En plus $I(A_1), I(A_2), ..., I(A_k)$ sont des variables aléatoires indépendantes lorsque $A_1, A_2, ..., A_k$ sont disjoints.

Notons

$$\begin{aligned} m_{i,n} &= \frac{X_{n-i+1,n} - b_n}{a_n}, 1 \leq i \leq k, \\ I_n(x) &= I_n\{(x, +\infty]\} \end{aligned}$$

et définissons

$$\begin{aligned} m_i &= \inf\{x : I(x) \leq i-1\} \\ M_{k,n} &= (m_{1,n}, m_{2,n}, ..., m_{k,n}) \end{aligned}$$

et

$$M_k = (m_1, m_2, ..., m_k).$$

Nous avons alors :

$$\{I_n(x) \leq k-1\} = \{m_{k,n} \leq x\} \tag{2.5}$$

et

$$\{I(x) \leq k-1\} = \{m_k \leq x\}. \tag{2.6}$$

Prenons $\{x_1, x_2, ..., x_k\}$ k points de continuité de G, nous avons :

$P\{\bigcap\limits_{i=1}^{k} [m_{i,n} \leq x_i]\} = P\{\bigcap\limits_{i=1}^{k} [I_n(x_i) \leq i-1]\}$ par (2.5).

Mais $P\{\bigcap\limits_{i=1}^{k} [I_n(x) \leq i-1]\} \to P\{\bigcap\limits_{i=1}^{k} [I(x_i) \leq i-1]\}$, $n \to +\infty$

lorsque $n \to +\infty$ en prenant $A_i = (x_i, +\infty]$, $1 \leq i \leq k$, dans le théorème B, et enfin

$P\{\bigcap\limits_{i=1}^{k} [I_n(x) \leq i-1]\} \to P\{\bigcap\limits_{i=1}^{k} [m_i \leq x_i]\}$, $n \to +\infty$ par (2.6)

soit $\lim\limits_{n \to +\infty} P\{\bigcap\limits_{i=1}^{k} [m_{i,n} \leq x_i]\} = P\{\bigcap\limits_{i=1}^{k} [m_i \leq x_i]\}$.

Ceci nous permet d'énoncer le résultat suivant :

Théorème C

Sous les hypothèse du Théorème B,

$$M_{k,n} \xrightarrow{L} M_k.$$

Le même auteur Weissman (1978) traite le cas où $G(x) = \Lambda(x)$.

Théorème D

Si $G(x) = \exp(-e^{-x})$, $-\infty \leq x \leq +\infty$, alors
lorsque $n \to +\infty$,

$$\{\frac{X_{n,n} - X_{n-1,n}}{a_n}, \frac{X_{n-1,n} - X_{n-2,n}}{a_n}, ..., \frac{X_{n-k+1,n} - X_{n-k,n}}{a_n}\} \xrightarrow{L} \{w_1, \frac{w_2}{2}, ..., \frac{w_k}{k}\}$$

où $a_n > 0$ est une suite de réels et $w_1, w_2, ..., w_k$ sont des variables aléatoires indépendantes et identiquement distribuées, exponentielles de moyenne 1.

A cette étape de la démonstration nous rappelons le résultat suivant de Billingsley [1968] (Corollaire 1, page 31).

Lemme 2.1
Soit S et $S^{'}$ deux espaces métriques. Supposons que $h : S \to S^{'}$ soit une fonction continue, alors

$$X_n \to X \Rightarrow h(X_n) \to h(X), \text{ lorsque } n \to +\infty.$$

Appliquons ce lemme avec $S = IR^{k+1}$, $S^{'} = IR^k$,
$h : IR^{k+1} \to IR^k$
$(x_1, x_2, ..., x_{k+1}) \longmapsto (x_1 - x_2, ..., x_k - x_{k+1})$ est une fonction continue,
$X_n = (m_{1,n}, ..., m_{k+1,n})$, $X = (m_1, ..., m_{k+1})$. Nous obtenons alors

$$(m_{1,n} - m_{2,n}, m_{2,n} - m_{3,n}, ..., m_{k,n} - m_{k+1,n}) \xrightarrow{L} (m_1 - m_2, m_2 - m_3, ..., m_k - m_{k+1}) \text{ lorsque } n \to +\infty.$$

Weisman [1978] montre que $m_1 - m_2$, $m_2 - m_3, ...,$ $m_k - m_{k+1}$ sont des variables aléatoires indépendantes de loi jointe.

$$\varphi_k(y_1, y_2, ..., y_k) = \prod_{i=1}^{k} (i \exp(-iy_i)),\ y_i \geq 0 \text{ pour } 1 \leq i \leq k.$$

Par conséquent, pour tout $1 \leq i \leq k$, $m_i - m_{i+1}$ est une variable aléatoire exponentielle de moyen $\frac{1}{i}$; $i(m_i - m_{i+1})$ est une variable exponentielle de moyen 1.

Si nous posons $w_i = i(m_i - m_{i+1})$ nous avons alors

$$(m_{1,n} - m_{2,n}, m_{2,n} - m_{3,n}, ..., m_{k,n} - m_{k+1,n}) \xrightarrow{L} (w_1, \tfrac{w_2}{2}, ..., \tfrac{w_k}{k}).$$

Ce qui met fin à la démonstration du théorème D.

2.3 Résultats limites

Soient $X_1, ..., X_n$ n variables aléatoires i.i.d. de fonction de répartition commune $F \in DA(\Lambda)$. Supposons qu'il existe deux suites de constantes $a_n > 0$ et $b_n \in IR$ telles que la fonction de répartition de la variable aléatoire

$$\max_{1 \leq i \neq j \leq n} \frac{(X_i + X_j) - 2b_n}{a_n}$$

converge vers une fonction de répartition non dégénérée H_2.

Notre but dans cette section est de caractériser cette loi limite. A cette fin, nous rappelons deux résultats généraux qui seront nécessaires pour la démonstration des résultats principaux des sections 2.3 et 2.4 ci-dessous.

Lemme 2.2 (Galambos [1978], Lemme 2.2.1, p :58)
Soient $\mathbf{U}_n$ et V_n deux suites de variables aléatoires. Supposons qu'il existe une fonction de répartition $G(x)$ telle que, $\lim_{n\to\infty} P(U_n \leq x) = G(x)$ pour tout x

point de continuité de $G(.)$. Supposons qu' en plus, $\forall \varepsilon > 0 \ \lim_{n\to\infty} P(|V_n| > \varepsilon) = 0$. (i.e. $V_n \xrightarrow{P} 0$, $U_n \xrightarrow{L} Y \Rightarrow U_n + V_n \xrightarrow{L} Y$ où Y v.a. de f.r. G)

Alors,

$$\lim_{n\to\infty} P(U_n + V_n \leq x) = \lim_{n\to\infty} P(U_n \leq x) = G(x),$$

pour tout x point de continuité de $G(.)$.

Lemme 2.3 (Galambos [1978], Lemme 2.9.1 p : 107)

Supposons que la fonction de répartition conjointe du couple aléatoire (Y_n, U_n) converge vers $T(y)E(u)$, où T et E sont des fonctions de répartition continues. Alors

$$\lim_{n\to\infty} P(U_n + Y_n \leq x) = \int_{-\infty}^{+\infty} E(x-y)dT(y).$$

2.3.1 Cas particulier où k=2

Ici le résultat principal est donné par le théorème suivant.

Théorème 2.4

Si $F \in DA(\Lambda)$, alors il existe deux suites de constantes $a_n > 0$ et $b_n \in IR$ telles que

$$\lim_{n\to\infty} P\left(\frac{\max\limits_{1\leq i\neq j\leq n}(X_i + X_j) - 2b_n}{a_n} \leq x\right) = H_2(x) \tag{2.7}$$

où

$$H_2(x) = \exp\{-e^{-\frac{x}{2}}(1 + e^{-\frac{x}{2}})\} - e^{-x}\int_{e^{-\frac{x}{2}}}^{+\infty} \frac{e^{-y}}{y}dy.$$

Remarque 2.5

Dans cette remarque nous présentons un cas particulier, celui où j est fixé dans le théorème 2.4 et la variable aléatoire X suit une loi normale standard. Nous obtenons

$$\max_{1\leq i\neq j\leq n}(X_i + X_j) = X_j + \max_{1\leq i\neq j\leq n}(X_i).$$

Si nous montrons que $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\max_{1\leq i\neq j\leq n}(X_i) - b_n|) > \varepsilon) = 0,$$

nous pouvons alors conclure, en utilisant le lemme 2.2, que

$$\lim_{n\to\infty} P(|\max_{1\leq i\neq j\leq n}(X_i) + X_j - b_n) \leq x) = P(X_j \leq x) = \Phi(x),$$

où $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$ est la fonction de répartition de la loi normale standard.

En effet, pour $\varepsilon > 0$ et puisque $a_n > 0$, nous obtenons

$$\begin{aligned}
\lim_{n\to\infty} P(|\max_{1\leq i\neq j\leq n}(X_i) - b_n|) &> \varepsilon) = \lim_{n\to\infty} P(|\frac{\max_{1\leq i\neq j\leq n}(X_i) - b_n}{a_n}| > \varepsilon a_n^{-1}) \\
&= 1 - \lim_{n\to\infty} P(|\frac{\max_{1\leq i\neq j\leq n}(X_i) - b_n}{a_n}| \leq \varepsilon a_n^{-1}) \\
&= 1 - \lim_{n\to\infty} \{\Lambda(\varepsilon a_n^{-1}) - \Lambda(-\varepsilon a_n^{-1})\}
\end{aligned}$$

$$= 1 - \lim_{n\to\infty} \{\exp(-e^{-\varepsilon a_n^{-1}}) - \exp(-e^{\varepsilon a_n^{-1}})\}$$

$$\begin{aligned}
&= 1 - 1 \\
&= 0
\end{aligned}$$

puisque pour la loi normale, nous pouvons choisir $a_n = \frac{1}{(\log n)^{\frac{1}{2}}}$ c-à-d $\varepsilon a_n^{-1} = \varepsilon(2\log n)^{1/2} \to +\infty$, lorsque $n \to +\infty$. D'où le résultat.

2.3.2 Démonstration du théorème 2.4

Dans notre cas, où $k = 2$, par le Théorème C il vient que

$$(m_{1,n}, m_{2,n}) \xrightarrow{L} (m_1, m_2) \text{ lorsque } n \to +\infty,$$

puis le Théorème D nous assure que les variables aléatoires $D_1 = m_1 - m_2$ et m_2 sont indépendantes et que D_1 suit une loi exponentielle de moyenne 1.

Maintenant considérons la fonction $h : IR^2 \to IR$ définie par $h(x,y) = x+y$. Puisque h est continue, nous obtenons

$$h(m_{1,n}, m_{2,n}) \xrightarrow{L} h(m_1, m_2) \text{ lorsque } n \to +\infty.$$

Ainsi

$\lim_{n\to\infty} P(m_{1,n} + m_{2,n} \leq x) = P(m_1 + m_2 \leq x) = P(D_1 + 2m_2 \leq x)$

$$= \frac{1}{2} \int_{-\infty}^{x} (1 - e^{-(x-y)}) \exp[-(y + e^{-\frac{y}{2}})] dy$$
$$= \frac{1}{2}(J_1 - J_2)$$

où

$$\begin{aligned}
J_1 &= \frac{1}{2} \int_{-\infty}^{x} \exp(-(y + e^{-\frac{y}{2}})) dy = \frac{1}{2} \int_{-\infty}^{x} \exp(-y) \exp(-e^{-\frac{y}{2}}) dy \\
&= 2 \int_{-\infty}^{x} (\exp(-e^{-\frac{y}{2}})) \frac{1}{2} e^{-\frac{y}{2}} e^{-\frac{y}{2}} dy \\
&= 2(\exp(-(\frac{x}{2} + e^{-\frac{x}{2}})) + \frac{1}{2} \int_{-\infty}^{x} \exp(-(\frac{y}{2} + e^{-\frac{y}{2}})) dy)
\end{aligned}$$

$$= 2(\exp(-(\frac{x}{2} + e^{-\frac{x}{2}})) + \left[\exp -e^{-\frac{y}{2}}\right]_{-\infty}^{x})$$

$$= 2(\exp(-(\frac{x}{2} + e^{-\frac{x}{2}})) + \exp(-e^{-\frac{x}{2}}))$$

$$= 2 \exp\{-e^{-\frac{x}{2}}(1 + e^{-\frac{x}{2}})\}$$

et

$$J_2 = \frac{1}{2} \int_{-\infty}^{x} \exp(-(x + e^{-\frac{y}{2}})) dy = e^{-x} \int_{-\infty}^{x} \exp -e^{-\frac{y}{2}} dy$$
$$= 2e^{-x} \int_{e^{-\frac{x}{2}}}^{+\infty} \frac{e^{-y}}{y} dy.$$

Ceci complète la démonstration de (2.7).

2.3.3 Cas général

Dans le cas général, nous obtenons le résultat suivant.

Théorème 2.6

Si $F \in DA(\Lambda)$, alors il existe deux suites de constantes $a_n > 0$ et $b_n \in IR$ telles que

$$\lim_{n\to\infty} P(\frac{\max_{1\leq i_1 \ldots \leq i_k \leq n}(X_{i_1} + X_{i_2} + \ldots + X_{i_k}) - kb_n}{a_n} \leq x) = H_k(x)$$

où

$$H_k(x) = e^{-e^{-\frac{x}{k}}} \sum_{j=0}^{k-1} \frac{e^{-j\frac{x}{k}}}{j!} - \frac{e^{-x}}{k!} \sum_{j=0}^{k-2} \int_{-\infty}^{x} e^{-e^{-\frac{y}{k}}} \frac{(x-y)^j}{j!} dy.$$

2.3.4 Démonstration du Théorème 2.6

Par le théorème C, nous avons

$$(m_{1,n}, ..., m_{k,n}) \xrightarrow{L} M_k, \quad \text{lorsque} \quad n \to +\infty.$$

En définissant la fonction $h : IR^k \to IR$ par $h(x_1, ..., x_k) = x_1 + x_2 + ... + x_k$, qui est continue, nous obtenons

$$h(m_{1,n}, ..., m_{k,n}) \xrightarrow{L} h(m_1, ..., m_k), \qquad \text{lorsque} \quad n \to \infty.$$

Il vient que

$$\lim_{n\to\infty} P(m_{1,n} + ... + m_{k,n} \leq x) = P(m_1 + ... + m_k \leq x)$$

$$= P(\sum_{j=1}^{k-1} jD_j + km_k \leq x)$$

$$= F_{D_1} * F_{2D_2} * ... * F_{(k-1)D_{k-1}} * G(x),$$

où

$D_i = m_i - m_{i+1} \quad 1 \leq i \leq k-1$,
F_{iD_i} est la fonction de répartition de iD_i
G est la fonction de répartition de km_k et
F est la fonction de répartition de la loi exponentielle de paramètre 1.

En utilisant le lemme 2.3, nous obtenons
$F_{D_1} * F_{2D_2} * ... * F_{(k-1)D_{k-1}} * G(x)$
$= \int\limits_{-\infty}^{+\infty} F^{\circledast(k-1)}(x-y) * dG(y)$
$= \int\limits_{-\infty}^{+\infty} [\int_0^{x-y} \frac{1}{(k-2)!} z^{k-2} e^{-z} dz] \frac{1}{k!} \exp(-e^{-\frac{y}{k}} - y) dy.$
Soit
$J = \int_0^{x-y} \frac{1}{(k-2)!} z^{k-2} e^{-z} dz = 1 - e^{-(x-y)} \sum_{j=0}^{k-1} \frac{(x-y)^j}{j!}$
$= \frac{1}{k!} \int_{-\infty}^{x} \exp(-e^{-\frac{y}{k}} - y) dy - \frac{1}{k!} \exp(-e^{-\frac{y}{k}} - y) dy$
$\quad - \int_{-\infty}^{x} \frac{1}{k!} \sum_{j=0}^{k-2} \exp(-e^{-\frac{y}{k}} - x) \frac{(x-y)^j}{j!} dy,$

si nous posons $v = e^{-\frac{y}{k}}$, il vient que

$$\int_{-\infty}^{x} \exp(-e^{-\frac{y}{k}} - y)dy = k \int_{e^{\frac{x}{k}}}^{+\infty} v^{k-1} e^{-v} dv$$
$$= k! \int_{e^{-\frac{x}{k}}}^{+\infty} \frac{v^{k-1}}{\Gamma(k)} e^{-v} dv$$
$$= k! e^{-e^{-\frac{x}{k}}} \sum_{j=0}^{k-1} \frac{e^{-j\frac{x}{k}}}{j!}.$$

Ainsi nous obtenons le résultat du théorème 2.6.

2.4 Exemples

Les exemples ci-dessous illustrent les résultats théoriques de la section précédente lorsque la fonction de répartition appartient au domaine d'attraction de Gumbel.

Exemple 1

Considérons n variables aléatoires i.i.d. $X_1, ..., X_n$ de la loi $N(0,1)$. Puisque $\Phi \in MDA(\Lambda)$, par le théorème 2.4, nous obtenons l'existence de deux suites de constantes $a_n = (\frac{2}{\ln n})^{-\frac{1}{2}}$ et $b_n = \frac{1}{a_n} - \frac{1}{2} a_n (\log \log n + \log 4\pi)$ telle que

$$\begin{aligned} \lim_{n\to\infty} P(\frac{\max\limits_{1\leq i\neq j\leq n}(X_i + X_j) - 2b_n}{a_n} &\leq x) \\ &= \exp\{-e^{-\frac{x}{2}}(1 + e^{-\frac{x}{2}})\} - e^{-x} \int_{e^{-\frac{x}{2}}}^{+\infty} \frac{e^{-y}}{y} dy. \end{aligned}$$

Exemple 2

Soient $X_1, ..., X_n$ n variables aléatoires i.i.d. de loi exponentielle de paramètre $\lambda > 0$, et de fonction de répartition : $F(x) = (1 - e^{-\lambda x}) \,\mathrm{II}_{[0,+\infty[}\,(x)$.

Lorsque $x \to x_F = +\infty$,

$$F_n(x) := P(\lambda M_n - \log n \leq x) = F^n(\frac{x + \log n}{\lambda}) = (1 - \frac{e^{-x}}{n})^n.$$

Il s'en suit que

$$\lim_{n\to\infty} F_n(x) = \exp(-e^{-x}), x \in IR.$$

En utilisant le théorème 2.4 et en prenant pour constantes de normalisation $a_n = \frac{1}{\lambda}$ et $b_n = \frac{\log n}{\lambda}$, il vient que

$$\lim_{n\to\infty} P(\lambda[\max_{1\leq i\neq j\leq n} X_i + X_j - 2\frac{\log n}{\lambda}] \leq x) = \exp\{-e^{-\frac{x}{2}}(1 + e^{-\frac{x}{2}})\} - e^{-x} \int_{e^{-\frac{x}{2}}}^{+\infty} \frac{e^{-y}}{y} dy.$$

CHAPITRE 3 : LOIS LIMITES POUR LES STATISTIQUES D'ORDRE DANS LE CAS DE V.A. INDEPENDANTES NON IDENTIQUEMENT DISTRIBUEES

3.1 Loi du $k^{ème}$ maximum

Soit $X_1, ..., X_n$ une suite de variables aléatoires indépendantes telles que $P(X_i \leq x) = F_i(x)$, $1 \leq i \leq n$.

Soit $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ les statistiques d'ordre correspondantes.

$X_{n-k+1,n}$ désigne le $k^{ème}$ maximum supérieur lorsque $1 \leq k \leq n$.

La loi simple du $k^{ème}$ maximum est donnée par Mejzler et Weissman (1969).

$$P(X_{n-k+1,n} \leq x) = \prod_{i=1}^{n} F_i(x) + \sum_{s=1}^{k-1} \sum_{(n,s)} \{ \prod_{j \in (n,s)^c} F_j(x) \prod_{i \in (n,s)} (1 - F_i(x)) \}. \quad (3.1)$$

où (n, s) désigne un sous-ensemble de $\{1, 2, ..., n\}$ de s éléments et $(n, s)^c$ son complémentaire.

Une expression plus commode de (3.1) est la suivante.

$$P(X_{n-k+1,n} \leq x) = \prod_{i=1}^{n} F_i(x) \{1 + \sum_{s=1}^{k-1} \sum_{(n,s)} \prod_{i \in (n,s)} (\frac{1}{F_i(x)} - 1)\}.$$

La loi limite du maximum (cas $k = 1$) a été étudiée par Mejzler (1949, 1950, 1953, 1956).

Galambos (1987)(Chapitre 3, p.181) énonce cette loi limite comme suit.

Théorème 3.1.1 Supposons que

(i) $\lim\limits_{n \to \infty} \max\limits_{1 \leq i \leq n} \{1 - F_i(a_n x + b_n)\} = 0$;

(ii) $\lim\limits_{n \to \infty} \sum\limits_{i=1}^{[nt]} \{1 - F_i(a_n x + b_n)\} = w(t, x)$ existe et finie pour tout $t \in (0, 1]$ lorsqu'elle est finie pour $t = 1$, $a_n > 0$ et b_n sont deux suites réelles non aléatoires,

où $[u]$ désigne la partie entière de u.

Alors une fonction de répartition non dégénérée $H(x)$ est telle que $\lim\limits_{n \to \infty} P(X_{n,n} \leq a_n x + b_n) = H(x)$ si et seulement si

• $\log H(x)$ est concave;

• ou $\log H(\overline{H} - e^{-x})$ est concave, $x > 0$, $\overline{H} < +\infty$;

• ou $\log H(\underline{H} + e^{+x})$ est concave, $x > 0$, $\underline{H} < +\infty$;

$\overline{H} = \sup\{x : H(x) < 1\}$ et $\underline{H} = \inf\{x : H(x) > 0\}$.

La loi limite du $k^{ème}$ maximum est obtenue par Mejzler et Weissman(1969) de la façon suivante.

Théorème 3.1.2

Supposons que

(a) $\lim\limits_{n \to \infty} \max\limits_{1 \leq i \leq n} \{1 - F_i(a_n x + b_n)\} = 0$. et

(b) $\lim_{n\to\infty} P(X_{n,n} \le a_n x + b_n) = H(x)$ où $H(x)$ vérifie les conditions du Théorème 3.1.1.

Alors $\lim_{n\to\infty} P(X_{n-k+1,n} \le a_n x + b_n) = H(x) \sum_{i=0}^{k-1} \frac{(-\log H(x))^i}{i!}$

où k est fixé et $a_n > 0$ et b_n sont deux suites numériques.

3.2 Loi jointe des k premiers maxima

Nous allons procéder de la même façon que pour le cas des variables aléatoires i.i.d.

Notons

$$I_n(A) = \#\{\frac{X_i - b_n}{a_n} \in A,\ 1 \le i \le n\}.$$

Maintenant, puisque

$$\lim_{n\to\infty} \min_{1\le i\le n} F_i(a_n x + b_n) = 1$$

et

$$\begin{aligned} \prod_{i=1}^{n} F_i(a_n x + b_n) &= P(X_{n,n} \le a_n x + b_n) \\ &\to H(x),\ \forall x > \underline{H},\ n \to +\infty, \end{aligned}$$

nous avons

$$\sum_{i=1}^{n} \{1 - F_i(a_n x + b_n)\} \to -\log H(x), \forall x > \underline{H},\ \text{lorsque } n \to +\infty.$$

Le résultat suivant a été obtenu par Weissman(1975).

Théorème 3.2.1

Supposons que :

(i) $\lim_{n\to\infty} \max_{1\le i\le n} \{1 - F_i(a_n x + b_n)\} = 0;$

(ii) $\lim_{n\to\infty} P(X_{n,n} \le a_n x + b_n) = H(x),\ \forall x > \underline{H}.$

Soient $A_1, A_2, ..., A_k$ k ensembles boréliens (k fixé).

Alors il existe sur $\mathbb{R}$ une mesure aléatoire de Poisson I, avec $E\{I(x, +\infty)\} = -\log H(x)$ et telle que

$$(I_n(A_1), ..., I_n(A_k)) \overset{L}{\to} (I(A_1), ..., I(A_k)).$$

En plus $I(A_1), ..., I(A_k)$ sont des variables aléatoires indépendantes lorsque $A_1, ..., A_k$ sont disjoints.

Notons $m_{i,n} = \frac{X_{n-i+1,n} - b_n}{a_n}$ pour $1 \le i \le k$ et $I_n(x) = I_n\{(x, +\infty]\}$.

Définissons $m_i = \inf\{x : I(x) \le i - 1\}$, $M_{k,n} = (m_{1,n}, m_{2,n}, ..., m_{k,n})$ et $M_k = (m_1, m_2, ..., m_k)$.

Nous avons

$$\{I_n(x) \le k - 1\} = \{m_{k,n} \le x\}$$

et $\{I(x) \leq k-1\} = \{m_k \leq x\}$.
Nous pouvons, maintenant énoncer le résultat suivant :
Théorème 3.2.2
Sous les hypothèses du Théorème 3.2.1, pour $k \geq 1$ fixé, nous avons

$$M_{k,n} \overset{L}{\rightarrow} M_k, \text{ lorsque } n \rightarrow +\infty.$$

Démonstration
En considérant $x_1, x_2, ..., x_k$, k points de continuité de H, tels que $\overline{H} > x_1 > x_2 > ... > x_k > \underline{H}$ et en utilisant le Théorème 3.2.2, nous avons

$$\lim_{n\rightarrow+\infty} P\{\bigcap_{i=1}^{k} [m_{i,n} \leq x_i]\} = P\{\bigcap_{i=1}^{k} [m_i \leq x_i]\}$$

soit

$$\lim_{n\rightarrow+\infty} P(\frac{X_{n,n} - b_n}{a_n} \leq x_1, ..., \frac{X_{n-k+1,n} - b_n}{a_n} \leq x_k) = H(x_1, ..., x_k)$$

dont nous allons chercher la densité $h(x_1, ..., x_k)$.

Weissman(1975), en généralisant un résultat de Dwass (1966), avait obtenu l'expression suivante :

$$\begin{aligned} P\{v_1 &< m_1 \leq u_1, ..., v_k < m_k \leq u_k\} \\ &= \{H(u_k) - H(v_k)\} \prod_{i=1}^{k-1} \{-\log \frac{H(v_i)}{H(u_i)}\} \end{aligned}$$

où $\underline{H} < v_k < u_k < v_{k-1} < ... < v_1 < u_1 < \overline{H}$.
Si nous prenons $u_i = x_i$ et $v_i = x_i + dx_i, 1 \leq i \leq k$, nous obtenons
$P\{x_1 < m_1 \leq x_1 + dx_1, ..., x_k < m_k \leq x_k + dx_k\}$

$$\begin{aligned} &= \{H(x_k + dx_k) - H(x_k)\} \prod_{i=1}^{k-1} \{-\log \tfrac{H(x_i)}{H(x_i+dx_i)}\} \\ &= h(x_k)dx_k \prod_{i=1}^{k-1} \{\log H(x_i + dx_i) - \log H(x_i)\} \\ &= h(x_k)dx_k \prod_{i=1}^{k-1} \tfrac{h(x_i)}{H(x_i)} dx_i \\ &= h(x_k) \prod_{i=1}^{k-1} \tfrac{h(x_i)}{H(x_i)} \prod_{i=1}^{k} dx_i. \end{aligned}$$

D'où

$$h(x_1, ..., x_k) = h(x_k) \prod_{i=1}^{k-1} \frac{h(x_i)}{H(x_i)}$$

où $\underline{H} < x_k < x_{k-1} < ... < x_1 < \overline{H}$ et $h(x) = H'(x)$.
3.3. Résultats limites des statistiques d'ordre centrales

Dans cette section, nous généralisons les résultats de Smirnov (1962) et Wu (1966) au cas général non i.i.d. Le cas $X_{n,n}$ a déjà été traité par Leadbetter (1978).

3.3.1 Lois limites de type Smirnov

Rappelons la définition d'une statistique d'ordre centrale.

Définition 3.3.1

Nous dirons que $X_{k(n),n}$ est une statistique d'ordre centrale lorsque $1 \leq k = k(n) < n$, $k \to +\infty$

$\lim_{n \to +\infty} \sqrt{n}(\frac{k}{n} - \lambda)$ où $\lambda \in (0,1)$

Notons

$$F_{k(n)}(x) = P(X_{k,n} \leq a_n x + b_n) = F_{k(n)}(a_n x + b_n)$$

et

$$u_n(x) = \sqrt{n} \frac{\sum_{i=1}^{n} F_i(x) - k(n)}{\{k(n)(n-k(n))\}^{1/2}}.$$

Nous montrons un résultat qui nous sera très utile par la suite. Ce résultat a été obtenu par Mejzler (1984) sous d'autres hypothèses.

Théorème 3.3.1

Soit H une fonction de répartition propre (c'est-à-dire pour laquelle $\underline{H} < \overline{H}$). Supposons que $k = k(n)$ vérifie les conditions $k \to +\infty$, $n - k \to +\infty$ et $\frac{k}{n} \to \lambda \in (0,1)$ (3.2)

Supposons aussi que les suites a_n et b_n vérifient

$$\begin{aligned} b_n &\to b \ , \ (b_n - b)/a_n \ \to 0 \\ a_n &\to 0 \text{ et } a_{n+1}/a_n \to 1 \end{aligned} \tag{3.3}$$

alors, lorsque $n \to \infty$ et pour tout $x \in (\underline{H}, \overline{H})$,

$$F_{k(n)}(a_n x + b_n) \to H(x) \Leftrightarrow u_n(a_n x + b_n) \to u(x)$$

où la fonction u est déterminée par l'équation fonctionnelle $H(x) = \Phi\{u(x)\}$ où Φ est la f.r. de la loi normale standard.

Remarque. Les conditions pour la version de Mejzler du Théorème 3.3.1 sont les mêmes sauf pour la condition (3.3). La condition de Mejzler est : pour tout $x \in (\underline{H}, \overline{H})$

$$\sup\{|F_i(a_n x + b_n) - \lambda| \ : \ 1 \leq i \leq n\} \to 0.$$

On peut montrer facilement que cette condition de Mejzler entraine la condition (3.3).

Théorème 3.3.2

Supposons que $k = k(n)$ vérifie les conditions $1 \leq k \leq n$ et $\sqrt{n}(\frac{k}{n} - \lambda) \to 0$ pour une valeur de $\lambda \in (0, 1)$.

Supposons en outre que pour $a_n > 0$ et b_n vérifiant (3.3), la suite $\{F_{k(n)}: n \geq 1\}$ vérifie,

$$F_{k(n)}(a_n x + b_n) \to H(x),\ x \in (\underline{H}, \overline{H})$$

où H est une fonction de répartition non dégénérée.

Alors $H(x)$ ne peut prendre que l'une des quatre formes suivantes :

$$H_{1,a}(x) = \begin{cases} \Phi(cx^a) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

$$H_{2,a}(x) = \begin{cases} 1 & \text{si } x \geq 0; \\ \Phi(-c_1 |x|^a) & \text{si } x < 0. \end{cases}$$

$$H_{3,a}(x) = \begin{cases} \Phi(-c_2 |x|^a) & \text{si } x < 0; \\ \Phi(c_3 x^a) & \text{si } x \geq 0. \end{cases}$$

$$H_{4,a}(x) = \begin{cases} 0 & \text{si } x < -1; \\ 1/2 & \text{si } -1 \leq x \leq 1; \\ 1 & \text{si } x > 1. \end{cases}$$

où c, c_1, c_2, c_3 et a sont des réels strictement positifs et $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$.

Démonstration du Théorème 3.3.1

Considérons le système de variables aléatoires

$$\{Y_{in} : 1 \leq i \leq n, n = 1, 2, ...\}$$

avec $P(Y_{in} = 1) = F_i(a_n x + b_n) = 1 - P(Y_{in} = 0)$.

Soit $S_n = \sum_{i=1}^n Y_{in}$, nous avons

$E(S_n) = \sum_{i=1}^n F_i(a_n x + b_n),\ V(S_n) = \sum_{i=1}^n F_i(a_n x + b_n)\{1 - F_i(a_n x + b_n)\}$ et

$$F_{kn}(a_n x + b_n) = P(S_n \geq k) \tag{3.4}$$

Mais $V(S_n) \geq \min\limits_{1 \leq i \leq n} F_i(a_n x + b_n) \sum_{i=1}^n \{1 - F_i(a_n x + b_n)\}$ et puisque

$0 < \lim\limits_{n \to \infty} \min\limits_{1 \leq i \leq n} F_i(a_n x + b_n) < 1$ à cause de (3.3), il suffit de montrer que $\sum_{i=1}^n \{1 - F_i(a_n x + b_n)\} \to \infty$ pour que

$$V(S_n) \to \infty, \text{ lorsqu en} \to \infty. \tag{3.5}$$

L'expression $u_n(x)$ peut s'écrire comme

$$u_n(x) = \sqrt{\frac{n}{k}} [\frac{\sum_{i=1}^n \{F_i(x) - 1\}}{\sqrt{n-k}} + \sqrt{n-k}]$$

d'où

$$\sum_{i=1}^{n}\{1-F_i(a_nx+b_n)\} = \sqrt{n(1-\frac{k}{n})}\{\sqrt{n(1-\frac{k}{n})}-\sqrt{\frac{k}{n}}u_n(a_nx+b_n)\}$$

puisque $u_n(a_nx+b_n) \to u(x)$ et $\frac{k}{n} \to \lambda$ lorsque $n \to \infty$.

Ainsi $\sum_{i=1}^{n}\{1-F_i(a_nx+b_n)\} \sim n(1-\lambda) \to \infty$, ce qui montre (3.5). Nous avons aussi

$$\frac{F_n(a_nx+b_n)\{1-F_n(a_nx+b_n)\}}{V(S_n)} \to 0 \tag{3.6}$$

lorsque $n \to \infty$, en utilisant (3.5) et le fait que, à partir de la condition (3.3),

$$\lim_{n\to\infty} F_n(a_nx+b_n)\{1-F_n(a_nx+b_n)\} = F_\infty(b)\{1-F_\infty(b)\}$$

est finie avec $F_\infty(x) = \lim\limits_{n\to\infty} F_n(x)$. Nous avons en outre la condition de Lindeberg qui est satisfaite pour le système de variables aléatoires

$$\{\overline{Y_{in}} \quad = \quad \frac{Y_{in}-F_i(a_nx+b_n)}{\sqrt{V(S_n)}} : 1 \leq i \leq n,\ n=1,2,...\}.$$

En notant

$$h_n(x) = \frac{\sum_{i=1}^{n} F_i(x) - k}{[\sum_{i=1}^{n} F_i(x)\{1-F_i(x)\}]^{1/2}}$$

la condition de Lindeberg et les résultats (3.5) et (3.6) nous permettent de conclure que

$$F_{k(n)}(a_nx+b_n) - \Phi\{h_n(a_nx+b_n)\} \to 0 \tag{3.7}$$

lorsque $n \to \infty$, en utilisant (3.4) et le théorème central limite pour le cas indépendant non identiquement distribué.

Par conséquent nous aurons

$F_{kn}(a_nx+b_n) \to H(x)$ si et seulement si

$$h_n(a_nx+b_n) \to u(x) \tag{3.8}$$

où la fonction u satisfait l'équation $H = \Phi(u)$. Nous pouvons écrire $h_n(a_nx+b_n)$ comme suit

$$h_n(a_nx+b_n) = n^{-1/2}\frac{\{k(n-k)\}^{1/2}}{[\sum_{i=1}^{n} F_i(a_nx+b_n)\{1-F_i(a_nx+b_n)\}]^{1/2}}u_n(a_nx+b_n).$$

Puisque $u_n(a_nx+b_n) \to u(x)$, pour avoir (3.8) il suffit de montrer que

$$n^{-1/2}\frac{\{k(n-k)\}^{1/2}}{[\sum_{i=1}^{n} F_i(a_nx+b_n)\{1-F_i(a_nx+b_n)\}]^{1/2}} \to 1$$

lorsque $n \to \infty$, soit

$$n^{-1} \sum_{i=1}^{n} F_i(a_n x + b_n)\{1 - F_i(a_n x + b_n)\} \to \lambda(1-\lambda) \tag{3.9}$$

sachant que $\frac{k}{n} \to \lambda$, lorsque $n \to \infty$. En remarquant que

$$\lim_{n\to\infty} F_i(a_n x + b_n)\{1 - F_i(a_n x + b_n)\} = F_\infty(b)\{1 - F_\infty(b)\} = \lambda(1-\lambda)$$

où $F_\infty(b) = \lambda \in (0,1)$, nous obtenons (3.9) en utilisant le résultat de Cesaro sur les suites.

Nous avons ainsi montré que $F_{k(n)}(a_n x + b_n) \to H(x) = \Phi\{u(x)\}$ et par conséquent le théorème 3.3.1.

Démonstration du Théorème 3.3.2

La démonstration du théorème 3.3.2 nécessite les deux propositions suivantes :

Proposition 3.3.1 Supposons que $k(n)$ et $k^{'}(n)$ vérifient les conditions suivantes :

$k(n) \to +\infty$, $k'(n) \to +\infty$, $n - k(n) \to \infty$, $n - k'(n) \to \infty$, $\sqrt{n}(\frac{k(n)}{n} - \lambda) \to 0$ et $\sqrt{n}(\frac{k'(n)}{n} - \lambda) \to 0$, pour une valeur de $\lambda \in (0,1)$. Supposons aussi que pour $a_n > 0$, $a_n^{'} > 0$, b_n et b_n' satisfaisant la condition (3.3), les suites $\{F_{k(n)} : n \geq 1\}$ et $\{F_{k'(n)} : n \geq 1\}$ vérifient, pout tout $x \in (\underline{H}, \overline{H})$,

1) $F_{k(n)}(a_n x + b_n) \to H_1(x)$,

2) $F_{k'(n)}(a_n' x + b_n') \to H_2(x)$.

Alors H_1 et H_2 sont deux fonctions de répartition de même type c'est-à-dire qu'il existe $a > 0$ et b, deux constantes réelles telles que $H_1(x) = H_2(ax + b)$.

Proposition 3.3.2 Supposons que $k = k(n)$ vérifie les conditions $k \to \infty, n - k \to \infty$, $\frac{k(n)}{n} \to \lambda$ et $\sqrt{n}(\frac{k}{n} - \lambda) \to 0$ pour une valeur de $\lambda \in (0,1)$. Supposons aussi que, pour $a_n > 0$ et b_n satisfaisant la condition (3.3), la suite $\{F_{k(n)}\colon n \geq 1\}$ vérifie, pour tout $x \in (\underline{H}, \overline{H})$,

$$F_{k(n)}(a_n x + b_n) \to H(x) = \Phi\{u(x)\} \tag{3.10}$$

Alors, pour tout $\upsilon > 1$, $u(x) = \sqrt{\upsilon}u(a_\upsilon x + b_\upsilon)$ avec $a_\upsilon > 0$ et b_υ deux constantes.

Pour la démonstration de ces deux propositions nous aurons besoin du résultat classique de Khintchine sur la convergence des types.

Lemme 3.3.1 (Théorème des convergences des types)

Soit $F_n(x)$ une fonction de répartition, soit $c_n > 0$, d_n et $\tau_n > 0$ et ρ_n des suites de réels et G et T deux fonctions de répartition non dégénérées telles que $\lim_{n\to\infty} F_n(d_n x + c_n) = G(x)$ et $\lim_{n\to\infty} F_n(\tau_n x + \rho_n) = T(x)$ pour tout x point de continuité des limites.

Alors, il existe deux réels $A > 0$ et B tels que $T(x) = G(Ax + B)$ avec $A = \lim_{n\to\infty} \frac{\tau_n}{d_n}$ et $B = \lim_{n\to\infty} \frac{\rho_n - c_n}{d_n}$.

Preuve : (voir Galambos 1978 p :61 L 2.2.3).

Démonstration de la proposition 3.3.1

Puisque $k(n) \to +\infty$, $n - k(n) \to +\infty$, $\frac{k(n)}{n} \to \lambda \in]0, 1[$ et $\sqrt{n}(\frac{k(n)}{n} - \lambda) \to 0$, et $a_n > 0$ et b_n satisfont la condition (3.3), les hypothèses du théorème 3.3.1 sont satisfaites.

Nous avons alors, lorsque $n \to +\infty$,

$$\begin{aligned} F_{k(n)}(a_n x + b_n) &\to H_1(x) \\ \Leftrightarrow &\quad \sqrt{n}\frac{\sum_{i=1}^n F_i(a_n x + b_n) - k(n)}{\{k(n)(n-k(n))\}^{1/2}} \to u_1(x) \end{aligned} \tag{3.11}$$

où $u_1(x)$ est solution de l'équation fonctionnelle $H_1(x) = \Phi\{u_1(x)\}$.

Puisque $\frac{k(n)}{n} - \lambda = \circ(\frac{1}{\sqrt{n}})$ l'expression (3.11) équivaut à

$$\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n F_i(a_n x + b_n) - \lambda}{\sqrt{\lambda(1-\lambda)}} \to u_1(x),\ n \to +\infty.$$

De la même façon nous obtenons, lorsque $n \to +\infty$,

$$\begin{aligned} F_{k'(n)}(a'_n x + b'_n) &\to H_2(x) \\ \Leftrightarrow &\quad \sqrt{n}\frac{\sum_{i=1}^n F_i(a'_n x + b'_n) - \lambda}{\sqrt{\lambda(1-\lambda)}} \to u_2(x). \end{aligned}$$

Nous avons donc $F_{kn}(a_n x + b_n) \to H_1(x)$ et $F_{kn}(a'_n x + b'_n) \to H_2(x)$. D'après le lemme 3.3.1, il existe deux réels $a > 0$ et b où $a = \lim_{n\to\infty} \frac{a'_n}{a_n}$ et $b = \lim_{n\to\infty} \frac{b'_n - b_n}{a_n}$ tels que $\Phi\{u_2(x)\} = \Phi\{u_1(ax + b)\}$ soit $H_2(x) = H_1(ax + b)$. Ce qui met fin à la démonstration de la proposition 3.3.1.

Démonstration de la proposition 3.3.2

Avec les conditions sur $k(n)$ et la condition (3.3), les hypothèses du théorème 3.3.1 sont satisfaites et nous avons alors

$$\begin{aligned} u_n(x) &= \sqrt{n}\{\frac{1}{n}\sum_{i=1}^n F_i(a_n x + b_n) - \lambda\}\frac{1}{\sqrt{\lambda(1-\lambda)}} \\ &\to u(x), n \to +\infty. \end{aligned} \tag{3.12}$$

Pour $\upsilon > 1$ un entier naturel fixé, choisissons $(u_{s\upsilon})_{s>0}$ une sous-suite de u_n.Nous obtenons, en utilisant (3.11)

$$u_{m\upsilon}(x) = \sqrt{m\upsilon}\{\frac{1}{m\upsilon}\sum_{i=1}^{[m\upsilon]} F_i(a_{m\upsilon}x + b_{m\upsilon}) - \lambda\}\frac{1}{\sqrt{\lambda(1-\lambda)}} \to u(x)$$

qui implique

$$\begin{aligned} \sqrt{m}\{\frac{1}{m\upsilon}\sum_{i=1}^{[m\upsilon]} F_i(a_{m\upsilon}x + b_{m\upsilon}) - \lambda\}\frac{1}{\sqrt{\lambda(1-\lambda)}} &\to \frac{u(x)}{\sqrt{\upsilon}} \\ &= u_1(x) \end{aligned} \tag{3.13}$$

lorsque $m \to +\infty$.

Posons $k'(m) = [\frac{k(m\upsilon)}{\upsilon}]$ où $[x]$ est la partie entière de x, nous avons

$$k'(m) \to +\infty, \text{ lorsque } m \to +\infty. \tag{3.14}$$

Puisque $m - k'(m) \geq \{m\upsilon - k(m\upsilon)\}/\upsilon \to \infty$, alors

$$m - k'(m) \to \infty \tag{3.15}$$

Nous avons en plus $k\prime(m)$ qui vérifie

$$\frac{k'(m)}{m} - \lambda = \frac{k(m\upsilon)}{m\upsilon} - \lambda = o(\frac{1}{\sqrt{m}}) \tag{3.16}$$

Les conditions (3.3), (3.14), (3.15) et (3.16) satisfont les hypothèses du théorème 3.3.1 et nous avons alors

$$(3.13) \Leftrightarrow F_{k'(m)m}(a_{m\upsilon}x + b_{m\upsilon}) \to \Phi\{u_1(x)\} \tag{3.17}$$

En résumé (3.14), (3.15), (3.16), (3.17) et les hypothèses de la proposition 3.3.2, par conséquent $\Phi(u(x))$ et $\Phi(u_1(x))$ sont de même type et il existe deux constantes $a_\upsilon > 0$ et b_υ(avec $\upsilon > 1$) telles que

$$\begin{aligned} \Phi(u_1(x)) &= \Phi(u(a_\upsilon x + b_\upsilon)), \text{ soit} \\ u_1(x) &= u(a_\upsilon x + b_\upsilon). \end{aligned}$$

Puisque $u_1 = \frac{u(x)}{\sqrt{\upsilon}}$, nous avons donc

$$u(x) = \sqrt{\upsilon}u(a_\upsilon x + b_\upsilon).$$

Ce qui met fin à la démonstration de la proposition 3.3.2.

Démonstration du Théorème 3.3.2

La résolution de l'équation $u(x) = \sqrt{\upsilon}u(a_\upsilon x+b_\upsilon)$(voir Smirnov (1962, p.101-106) donne les solutions suivantes

$$\begin{aligned}
u_{1,\alpha}(x) &= \begin{cases} -\infty & \text{si} \quad x < 0 \\ cx^\alpha & \text{si} \quad x \geq 0; \end{cases} \\
u_{2,\alpha}(x) &= \begin{cases} -c_1 |x|^\alpha & \text{si} \quad x < 0 \\ \infty & \text{si} \quad x \geq 0; \end{cases} \\
u_{3,\alpha}(x) &= \begin{cases} -c_2 |x|^\alpha & \text{si} \quad x < 0 \\ c_3 x^\alpha & \text{si} \quad x \geq 0; \end{cases} \\
u_4(x) &= \begin{cases} -\infty & \text{si} \quad x < -1 \\ 0 & \text{si} \quad -1 \leq x \leq 1 \\ -\infty & \text{si} \quad x > 1; \end{cases}
\end{aligned}$$

où c, c_1, c_2, c_3 et α sont des réels strictement positifs.

D'où les quatre expressions différentes de $H(x) = \Phi\{u(x)\}$ du Théorème 3.3.1.

3.3.2 Les lois limites de type Wu

Nous rappelons que, pour les v.a. i.i.d., Wu(1966) avait obtenu les différents types de lois limites possibles pour $H(x)$ lorsque $F_{k(n)}(a_nx + b_n) \to H(x)$ avec $\frac{k(n)}{n} \to \lambda \in]0, 1[$ et $\sqrt{n}(\frac{k(n)}{n} - \lambda) \to +\infty$ $(n \to +\infty)$.

Ces lois sont de la forme $\Phi(u(x))$,

$u_1(x) = \begin{cases} \beta \log x & \text{si} \quad x > 0 \\ -\infty & \text{si} \quad x \leq 0 \end{cases}$, $u_2(x) = \begin{cases} +\infty & \text{si} \quad x \geq 0 \\ -\beta \log |x| & \text{si} \quad x < 0 \end{cases}$,

$u_3(x) = x$ si $x \in \mathbb{R}$ avec $\beta > 0$ une constante et $\Phi(.)$ une fonction de répartition de la loi $N(0, 1)$

Dans le théorème suivant, nous allons étendre ce résultat au cas des variables indépendantes mais non identiquement distribuées.

Théorème 3.3.3

Supposons que $k(n)$ soit une suite croissante en n et telle que $k(n) \to \infty$, $\frac{k(n)}{n} \to \lambda \in]0, 1[$ et $\sqrt{n}\{\frac{k(n)}{n} - \lambda\} \to \infty$ lorsque $n \to \infty$. Supposons que, pour $a_n > 0$ et b_n satisfaisant la condition (3.3) la suite$\{F_{k(n)} : n \geq 1\}$ vérifie, pour tout $x \in (\underline{H}, \overline{H})$, $F_{k(n)}(a_nx + b_n) \to H(x) = \Phi(u(x))$.

Alors $H(x)$ ne peut appartenir qu'à l'un des trois types suivants

$$H_{1,a}(x) = \begin{cases} \Phi(a \log x) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0; \end{cases}$$

$$H_{2,a}(x) = \begin{cases} \Phi(-a \log |x|) & \text{si } x < 0 \\ 0 & \text{si } x \geq 0; \end{cases}$$

$$H_3(x) = \Phi(x), \text{ si } -\infty < x < +\infty.$$

où $\Phi(x)$ est la fonction de répartition d'une loi normale centrée réduite.

Démonstration du théorème 3.3.3 La démonstration du théorème 3.3.3 nécessite la proposition suivante

Proposition 3.3.3

Supposons que, lorsque $n \to +\infty$

1) $k(n) \to +\infty$, $n - k(n) \to +\infty$, $k(n)$ croissante et $\frac{k(n)}{n} \to \lambda \in]0, 1[$,

2) $F_{k(n)}(a_nx+b_n) \to H(x) = \Phi(u(x))$,et $a_n > 0$ et b_n satisfaisant la condition (3.3).

Supposons aussi l'existence de deux suites $(m_s)_s$ et $(n_s)_s$ telles que $m_s > n_s$ pour tout s, $\frac{m_s}{n_s} \to \mu^2$ et $\frac{n_s}{\sqrt{k(n_s)}}(\frac{k(n_s)}{n_s} - \frac{k(m_s)}{m_s}) \to \upsilon\sqrt{1 - \lambda}$, lorsque $s \to +\infty$ où $-\infty < \upsilon < +\infty$ et $1 \leq \mu < +\infty$.

Alors il existe deux réels $a > 0$ et b tels que $u(x) = \mu\{u(ax + b) + \upsilon\}$.

Démonstration de la proposition 3.3.3

Nous remarquons d'abord que, puisque $\frac{k(n)}{n} \to \lambda \in]0, 1[$, lorsque $n \to +\infty$, nous avons

$$u_n(a_nx + b_n) = \sqrt{n}\frac{\sum_{i=1}^{n} F_i(a_nx + b_n) - k(n)}{\{k(n)(n - k(n))\}^{1/2}} \to u(x)$$

qui est équivalente à

$$\frac{n}{\sqrt{k(n)}}\frac{1}{\sqrt{1-\lambda}}\{n^{-1}\sum_{i=1}^{n}F_i(a_nx+b_n)-\frac{k(n)}{n}\}\to u(x).$$

Nous remarquons ensuite que

$$\lim_{n\to+\infty}\frac{m_s}{n_s}=\lim_{n\to+\infty}\frac{k(m_s)}{k(n_s)}=\mu^2. \tag{3.18}$$

En effet

$$\frac{n_s}{\sqrt{k(n_s)}}(\frac{k(n_s)}{n_s}-\frac{k(m_s)}{m_s})=\frac{n_s}{m_s}\sqrt{k(n_s)}(\frac{m_s}{n_s}-\frac{k(m_s)}{k(n_s)})\to\upsilon\sqrt{1-\lambda},$$

par hypothèse. Mais $\frac{m_s}{n_s}\to\mu^2$, lorsque $s\to+\infty$ et nous avons alors

$$\frac{m_s}{n_s}-\frac{k(m_s)}{k(n_s)}\simeq\frac{\mu^2\upsilon\sqrt{1-\lambda}}{\sqrt{k(n_s)}}\to 0, s\to+\infty \text{ car } k(n_s)\to+\infty.$$

Donc

$$\lim_{n\to+\infty}\frac{m_s}{n_s}=\lim_{n\to+\infty}\frac{k(m_s)}{k(n_s)}=\mu^2.$$

Maintenant nous avons, lorsque $s\to\infty$,

$$\begin{aligned}&\frac{n_s}{\sqrt{k(n_s)}}\frac{1}{\sqrt{1-\lambda}}\{\frac{1}{n_s}\sum_{i=1}^{n_s}F_i(a_{n_s}x+b_{n_s})-\frac{k(n_s)}{n_s}\}\\ \to\quad& u(x),\ s\to+\infty\end{aligned} \tag{3.19}$$

et

$$\begin{aligned}&\frac{m_s}{\sqrt{k(m_s)}}\frac{1}{\sqrt{1-\lambda}}\{\frac{1}{n_s}\sum_{i=1}^{n_s}F_i(a_{n_s}x+b_{n_s})-\frac{k(n_s)}{n_s}\}\\ =\quad&\frac{m_s}{n_s}\sqrt{\frac{k(n_s)}{k(m_s)}}\{\frac{n_s}{\sqrt{k(n_s)}}\frac{1}{\sqrt{1-\lambda}}[\frac{1}{n_s}\sum_{i=1}^{n_s}F_i(a_{n_s}x+b_{n_s})-\frac{k(n_s)}{n_s}]\}\\ \to\quad&\mu u(x),\ \text{lorsque } s\to+\infty\end{aligned}$$

puisque $\frac{m_s}{n_s}\to\mu^2$ et $\sqrt{\frac{k(n_s)}{k(m_s)}}\to\frac{1}{\mu},\ \ s\to+\infty.$
En plus

$$\begin{aligned}&\frac{m_s}{\sqrt{k(m_s)}}\frac{1}{\sqrt{1-\lambda}}\{\frac{1}{n_s}\sum_{i=1}^{n_s}F_i(a_{n_s}x+b_{n_s})-\frac{k(m_s)}{m_s}\}\\ =\quad&\frac{1}{\sqrt{1-\lambda}}\{\frac{m_s}{\sqrt{k(m_s)}}[\frac{1}{n_s}\sum_{i=1}^{n_s}F_i(a_{n_s}x+b_{n_s})-\frac{k(n_s)}{n_s}]+\frac{m_s}{\sqrt{k(m_s)}}[\frac{k(n_s)}{n_s}-\frac{k(m_s)}{m_s}]\}\\ \to\quad&\mu u(x)+\mu\upsilon, s\to+\infty.\end{aligned} \tag{3.20}$$

Mais, d'après le théorème 3.3.1,

$$(3.19) \Leftrightarrow F_{k(m_s)m_s}(a_{m_s}x + b_{m_s}) \to \Phi(u(x)) \tag{3.21}$$

et

$$(3.20) \quad \Leftrightarrow \quad F_{k(m_s)m_s}(a_{n_s}x + b_{n_s}) \to \Phi\{\mu(u(x) + \upsilon)\}. \tag{3.22}$$

En appliquant le lemme 3.3.1 à (3.21) et (3.22), nous obtenons l'existence de deux réelles $a > 0$ et b où $a = \lim_{n\to+\infty} \frac{a_{m_s}}{a_{n_s}}$, $b = \lim_{n\to+\infty} \frac{b_{m_s} - b_{n_s}}{a_{n_s}}$ telles que

$$\Phi(u(x)) = \Phi\{\mu(u(ax + b) + \upsilon)\}$$

soit

$$u(x) = \mu(u(ax + b) + \upsilon). \tag{3.23}$$

Ce qui achève la démonstration de la proposition 3.3.3.

Démonstration du théorème 3.3.3

La résolution de l'équation fonctionnelle (3.23) (résolution faite par Wu (1966)) donne les solutions suivantes:

$$\begin{aligned}
u_1(x) &= \begin{cases} a \log x & \text{si} \quad x \geq 0, \\ -\infty & \text{si} \quad x < 0. \end{cases} \\
u_2(x) &= \begin{cases} +\infty & \text{si} \quad x \geq 0, \\ -a \log |x| & \text{si} \quad x < 0. \end{cases} \\
u_3(x) &= x \quad \text{si} \quad -\infty < x < \infty
\end{aligned}$$

où $a > 0$ est un réel. D'où les expressions différentes de $H(x) = \Phi\{u(x)\}$ dans le théorème 3.3.3.

CHAPITRE 4 : LOIS LIMITES DES STATISTIQUES D'ORDRE DANS LE CAS D'UNE NORMALISATION NON LINEAIRE

4.1 Introduction

Soit $X_1, ..., X_n$ un échantillon de variables aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition (f.r.) commune $F(x) = P(X \leq x)$. Soient $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ les statistiques d'ordre associées à cet échantillon et $F_{k(n)}(x) = P(X_{k,n} \leq x)$ la f.r. de $X_{k,n}$. Comme nous avons vu dans le chapitre 1, la théorie des valeurs extrêmes concerne, en partie, la détermination de la loi du maximum $X_{n,n}$ lorsque n est assez grand. Le cas d'une normalisation linéaire $a_n x + b_n$ avec $a_n > 0$ et b_n des constantes réelles a fait l'objet de nombreux travaux initiés par Fisher et Tipett [1928]. Gnedenko [1943] a montré qu'il y a exactement trois types de lois limites pour $\frac{X_{n,n} - b_n}{a_n}$.

Le cas non linéaire a été rarement étudié. Une brève étude de ce cas a été entreprise par Panchéva [1984]. Récemment Ravi et Praveena[2011] ont travaillé sur la normalisation non linéaire $b_n |x|^{a_n} sign(x)$ avec $a_n > 0$ et b_n des constantes réelles. Notre but est d'abord la caractérisation des lois limites du maximum $X_{n,n}$ dans le cas d'une normalisation non linéaire. Il s'agit de trouver les expressions différentes de la fonction de répartition limite $G(x)$ telle que $\lim_{n\to\infty} P(X_{n,n} \leq b_n(x)^{a_n}) = G(x)$, de donner les conditions nécessaires et suffisantes (C.N.S.) d'appartenance aux domaines d'attraction de ces différentes lois limites et d'expliciter les expressions de a_n et b_n. Notre but aussi est d'étudier la caractérisation des fonctions de répartition non dégénérées $L(x)$ d'une statistique centrale telles que $\lim_{n\to\infty} F_{k(n)}(b_n x^{a_n}) = L(x)$. Dans le cas d'une normalisation linéaire, ce problème a fait l'objet de nombreuses études. Smirnov [1962] a étudié ce problème sous l'hypothèse $\frac{k(n)}{n} = \lambda + \frac{t}{\sqrt{n}} + o(\frac{1}{\sqrt{n}})$ quand $n \to \infty$, pour une valeur de t réelle. Wu [1966] a considéré le cas où $t = \infty$. Cette étude a été développée par Balkema et de Haan [1978a] sous la forme d'une théorie plus générale sur les limites des statistique d'ordre. Dans la section 2 nous établissons les différents types de lois limites dans le cas non linéaire et nous donnons des conditions nécessaires et suffisantes d'appartenance aux domaines d'attraction de ces lois limites et définissons des expressions générales des constantes de normalisation $a_n > 0$ et $b_n > 0$. Dans la section 3, nous exposons les démonstrations des différents résultats de la section 2 et donnons des exemples. Dans la section 4 nous évoquons les différentes lois limites pour une statistique d'ordre centrale dans le cas d'une normalisation non linéaire en précisant les C.N.S d'appartenance aux domaines d'attraction de ces lois. Dans la section 5, nous donnons les démonstrations des résultats de la section 4.

4.2 Caractérisation des lois limites du maximum.

L'un des buts de cette section est d'établir le théorème 4.2.1 ci-dessous et la caractérisation des lois limites.

Théorèm 4.2.1

Soient $a_n > 0$ et $b_n > 0$ deux suites numériques, alors $\lim_{n\to\infty} P(X_{n,n} \leq b_n x^{a_n}) = G(x)$, $x > 0$, si et seulement si G appartient à l'un des trois types suivants :

$$\Phi_1(x) = \begin{cases} 0 & \text{si } x < 0 \\ \exp(-x^{-1}) & \text{si} \quad x \geq 0 \end{cases} ;$$

$$H_{1,\alpha}(x) = \begin{cases} \exp(-(\log x)^{-\alpha}) & \text{si } x > 1 \\ 0 & \text{si } x \leq 1 \end{cases} ;$$

$$H_{2,\alpha}(x) = \begin{cases} \exp(-(-\log x)^{\alpha}) & \text{si } 0 < x \leq 1 \\ 1 & \text{si } x > 1 \end{cases} ;$$

où ■ est un réel strictement positif.

L'autre but de ce paragraphe est de donner des conditions nécessaires et suffisantes d'appartenance aux domaines d'attraction d'une loi limite.

Définition 4.2.1

Soit F une f.r. non dégénérée, nous dirons que F est dans le domaine d'attraction de H , et nous notons $F \in DA(H)$ s'il existe deux suites de constantes réelles strictement positives $a_n > 0$ et $b_n > 0$ telles que

$$\lim_{n \to \infty} P(X_{n,n} \leq b_n x^{a_n}) = H(x), \qquad\qquad x > 0.$$

Les conditions d'appartenance aux domaines d'attraction des lois limites Φ_1, $H_{1,\alpha}$ et $H_{2,\alpha}$ tiennent en compte le comportement de queue de la fonction de répartition F. En notant $x_F = \sup\{x \in IR : F(x) < 1\}$, nous avons :

Théorème 4.2.2

(i) $F \in DA(\blacksquare_1)$ ssi $x_F = +\infty$ et $1 - F(x) = x^{-1} L_1(x)$.

(ii) $F \in DA(H_{1,\alpha})$ ssi $x_F = +\infty$ et $1 - F(x) = (\log x)^{-\alpha} L_2(x)$.

(iii) $F \in DA(H_{2,\alpha})$ ssi $x_F < \infty$ et $1 - F(x_F e^{-\frac{1}{x}}) = (\log x_F - 1/x)^{-\alpha} L_3(x)$,

où L_1, L_2 et L_3 sont des fonctions à variation lente au voisinage de $+\infty$ (c-à-d $\lim_{x\to\infty} \frac{L_i(tx)}{L_i(x)} = 1$, $\forall t > 0$ et $i = 1, 2, 3$).

Il reste enfin à déterminer les constantes de normalisation $a_n > 0$ et $b_n > 0$ évoquées dans le théorème 4.2.1. En définissant la fonction inverse généralisée de F par

$$F^{\leftarrow}(y) = \inf\{x \in IR,\ F(x) > y\}, \quad y \in]0, 1[,$$

nous obtenons le résultat suivant

Théorème 4.2.3

(i) Lorsque $F \in DA(\Phi_1)$, un choix possible des constantes de normalisation est

$$b_n = 1 \text{ et } a_n = F^{\leftarrow}(1 - n^{-1}).$$

(ii) Lorsque $F \in DA(H_{1,\alpha})$, un choix possible des constantes de normalisation est

$$a_n = \log F^{\leftarrow}(1 - n^{-1}) \text{ et } b_n = 1.$$

(iii) Lorsque $F \in DA(H_{2,\alpha})$, un choix possible des constantes de normalisation est

$$a_n = \log x_F - \log F^{\leftarrow}(1 - n^{-1}) \text{ et } b_n = x_F.$$

4.3 Démonstrations.

Pour la démonstration du théorème 4.2.1, nous avons besoin du résultat suivant :

Lemme 4.3.1

Soient $X_1, ..., X_n$ n variables aléatoires i.i.d. positives de f.r. commune F.

Notons $Y_i = \log X_i$, $1 \leq i \leq n$ de fonction de répartition commune F_1 nous avons

$$F \in DA\{\Phi_1,\ H_{1,\alpha},\ H_{2,\alpha}\} \Leftrightarrow F_1 \in DA\{\Lambda,\ \Phi_\alpha,\ \Psi_\alpha\} \text{ respectivement.}$$

Remarque

Nous avons $x_{F_1} = \log x_F$.

En effet, soient $Y_i = \log X_i$, $1 \leq i \leq n$, (Y_i) sont des variables aléatoires i.i.d. de f.r. F_1 et puisque la fonction log est croissante sur IR^{*+}, $Y_{i,n} = \log X_{i,n}$ sont les statistiques d'ordre correspondantes à $Y_1, ..., Y_n$.

Soient $a_n > 0$ et $b_n > 0$ deux suites numériques, nous avons

$P(X_{n,n} \leq b_n x^{a_n}) = P(\log X_{n,n} \leq \log b_n x^{a_n})$, $x > 0$

$= P(Y_{n,n} \leq a_n \log x + \log b_n)$

$= F_1^n(a_n y + \log b_n)$, $y \in IR$.

D'après le résultat de Gnedenko [1943], nous avons $\lim_{n\to\infty} F_1^n(a_n y + \log b_n) = G(y)$ avec $G \in \{\Phi_\alpha, \Psi_\alpha, \Lambda\}$.

Nous allons discuter trois cas possibles :

(i) Si $G = \Phi_\alpha$, nous avons :

$\lim_{n\to\infty} F_1^n(a_n y + \log b_n) = \Phi_\alpha(y)$, $y > 0$

et comme $F_1(x) = F(\exp(x))$, nous obtenons

$\lim_{n\to\infty} F^n(\exp(a_n y + \log b_n)) = \Phi_\alpha(y)$

$\Leftrightarrow$

$\lim_{n\to\infty} F^n(b_n x^{a_n}) = \Phi_\alpha(\log x) = H_{1,\alpha}(x)$, $x > 1$.

(ii) Si $G = \Psi_\alpha$, nous avons :

$\lim_{n\to\infty} F_1^n(a_n y + \log b_n) = \Psi_\alpha(y)$, $y < 0$

$\Leftrightarrow$

$\lim_{n\to\infty} F^n(\exp(a_n y + \log b_n)) = \Psi_\alpha(y)$, $y < 0$

$\Leftrightarrow$

$\lim_{n\to\infty} F^n(b_n x^{a_n}) = \Psi_\alpha(\log x) = H_{2,\alpha}(x)$, $0 < x \leq 1$.

(iii) Si $G = \Lambda$, nous obtenons :

$\lim_{n\to\infty} F_1^n(a_n y + \log b_n) = \Lambda(y)$, $y \in IR$

$\Leftrightarrow$

$\lim_{n\to\infty} F^n(\exp(a_n y + \log b_n)) = \Lambda(y)$, $y \in IR$

$\Leftrightarrow$

$\lim_{n\to\infty} F^n(b_n x^{a_n}) = \Lambda(\log x) = \Phi_1(x)$, $x > 0$.

D'après (i), (ii) et (iii), nous obtenons le résultat du lemme 4.3.1.

4.3.1 Démonstration du théorème 4.2.1

Nous avons :

$P(X_{n,n} \leq b_n x^{a_n}) = P(\log X_{n,n} \leq \log b_n x^{a_n})$, $x > 0$

$= P(Y_{n,n} \leq a_n y + \log b_n)$, $y \in IR$

$= F_1^n(a_n y + \log b_n)$

D'après le résultat de Gnedenko [1943], nous avons $\lim_{n\to\infty} F_1^n(a_n y+\log b_n) = G(y)$ avec $G \in \{\Phi_\alpha, \Psi_\alpha, \Lambda\}$, en utilisant la démonstration du lemme 1. D'où le résultat.

4.3.2 Démonstration du théorème 4.2.2

La démonstration du théorème 4.2.2 nécessite le résultat suivant:

Lemme 2

Soient $X_1, ..., X_n$ n variables aléatoires i.i.d. positives de f.r. commune F.

Supposons que $\lim_{n\to\infty} n(1 - F(b_n x^{a_n})) = u(x)$, $\forall x$ avec $a_n > 0$ et $b_n > 0$, alors

$$\lim_{n\to\infty} P(X_{n,n} \leq b_n x^{a_n}) = e^{-u(x)}.$$

Ce résultat a été démontré par Galambos ([1978], Corollaire 1.3.1) dans le cas d'une normalisation linéaire. La démonstration reste valable dans le cas non linéaire.

Montrons le théorème 4.2.2.

Pour **(i),** en effet, si $F \in DA(\Phi_1) \Leftrightarrow F_1 \in DA(\Lambda)$, en utilisant un résultat classique caractérisant le $DA(\Phi_\alpha)$ dans le cas linéaire(voir Embrecht [1997]), nous avons :

$$x_F = \infty, \quad 1 - F(x) = x^{-1} L_1(x),$$

où L_1 est une fonction à variation lente VL_∞.

(ii) Lorsque $F \in DA(H_{1,\alpha}) \Leftrightarrow F_1 \in DA(\Phi_\alpha)$, alors

$1 - F_1(x) = x^{-\alpha} l(x)$ (voir Embrecht [1997], Theorem 3.3.7 p 131) où l est une fonction à variation lente, donc

$$1 - F(x) = 1 - F_1(\log x) = (\log x)^{-\alpha} l(\log x) = (\log x)^{-\alpha} L_2(x),$$

où L_2 est aussi une fonction à variation lente, d'où le résultat.

Nous avons aussi $x_{F_1} = \log x_F$, donc $x_F = +\infty$ puisque $x_{F_1} = +\infty$.

(iii) Lorsque $F \in DA(H_{2,\alpha}) \Leftrightarrow F_1 \in DA(\Psi_\alpha)$, donc

$$1 - F_1(x_{F_1} - x^{-1}) = x^{-\alpha} L_3(x)$$

où L_3 est une fonction à variation lente, donc d'après un résultat de (Embrecht[1997] theorme 3.3.12 p135), nous obtenons

$$1 - F(x_F \exp(-x^{-1})) = x^{-\alpha} L_3(x)$$

et $x_F < \infty$, d'où le résultat.

4.3.3 Démonstration du théorème 4.2.3.

(i) Supposons que$F \in DA(\Phi_1)$.

En utilisant un résultat classique caractérisant le domaine d'attraction $DA(\Phi_\alpha)$ avec $\alpha = 1$, dans le cas linéaire (voir Embrecht [1997], Thm 3.3.7, p 131), nous obtenons : $a_n = F^{\leftarrow}(1 - n^{-1})$ et $b_n = 1$. D'ou le résultat

(ii) Nous avons

$F \in DA(H_{1,\alpha}) \Leftrightarrow F_1 \in DA(\Phi_\alpha)$. Un choix possible des constantes de normalisation dans le cas linéaire (voir Embrecht [1997], Théorème 3.3.7 p 131) est $a_n = F_1^{\leftarrow}(1-n^{-1})$ et $\log b_n = 0$, dans le cas non linéaire nous pouvons choisir : $a_n = \log F^{\leftarrow}(1-n^{-1})$ et $b_n = 1$. D'où le résultat.

(iii) $F \in DA(H_{2,\alpha}) \Leftrightarrow F_1 \in DA(\Psi_\alpha)$. Un choix possible des constantes de normalisation dans le cas linéaire (voir Embrecht [1997], Théorème 3.3.12 p 135) $a_n = x_{F_1} - F_1^{\leftarrow}(1-n^{-1})$ et $\log b_n = x_{F_1}$, dans le cas non linéaire nous pouvons choisir : $a_n = \log x_F - \log F^{\leftarrow}(1-n^{-1})$ et $b_n = x_F$. D'où le résultat.

Exemples

1)

Soit X une variable aléatoire qui suit la loi de Pareto de paramètre 1 telle que

$$1 - F(x) = x^{-1}, x > 1,$$

en prenant $b_n = 1$ et $a_n = F^{\leftarrow}(1-n^{-1}) = n$, nous obtenons

$$F^n(b_n x^{a_n}) = (1-(b_n x^{a_n})^{-1})^n = (1-(nx)^{-1})^n \to \Phi_1(x), \text{ quand } n \to \infty.$$

Ainsi $F \in DA(\Phi_1)$.

2)

Soit F une fonction de répartition définie par $1-F(x) = (\log x)^{-\alpha} 1_{[e,\infty[}(x)$. En prenant $a_n = n^{\frac{1}{\alpha}}$ et $b_n = 1$, nous obtenons pour $x > e$,

$$\begin{aligned}\lim_{n\to\infty} F^n(b_n x^{a_n}) &= \lim_{n\to\infty}(1-(\log b_n + a_n \log x)^{-\alpha}))^n \\ &= \exp(-(\log x)^{-\alpha}) \\ &= H_{1,\alpha}(x),\end{aligned}$$

ainsi $F \in DA(H_{1,\alpha})$.

3)

Soit F une fonction de répartition définie par : $1-F(x) = (-\log x)^{\alpha} 1_{[\frac{1}{e},\infty[}(x)$, en prenant $a_n = n^{\frac{1}{\alpha}}$ et $b_n = 1$, on obtient pour $\frac{1}{e} < x < 1$,

$$\lim_{n\to\infty} F^n(b_n x^{a_n}) = \lim_{n\to\infty}(1-(-\log b_n - a_n \log x)^{\alpha})^n = H_{2,\alpha},$$

ainsi $F \in DA(H_{2,\alpha})$.

4.4 Lois limites d'une statistique d'ordre centrale

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition (f.r.) commune $F(x) = P(X \leq x)$. Soient $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ les statistiques d'ordre associées à $X_1, ..., X_n$ et $F_{k(n)}(x) = P(X_{k,n} \leq x)$ la f.r. de $X_{k,n}$. Il est classique de dire que $X_{k,n}$ est une statistique d'ordre centrale lorsque $k = k(n)$, vérifie $\frac{k(n)}{n} \to \lambda \in \,]0,1[$, pour $n \to \infty$. Notre but dans cette section est la caractérisation des fonctions de répartition non dégénérée $L(x)$ telles que pour des constantes $a_n > 0$ et $b_n > 0$, $P(X_{k,n} \leq b_n x^{a_n}) \to L(x)$, $x > 0$ avec $k = k(n)$ satisfaisant la condition $\sqrt{n}(\frac{k(n)}{n} - \lambda) \to 0$, quand $n \to \infty$. Nous y incluons aussi les conditions nécessaires et suffisantes d'appartenance aux domaines d'attraction des f.r. $L(x)$. Nous commençons d'abord par rappeler un résultat de Smirnov [1962] sur les lois limites d'une statistique d'ordre centrale dans le cas d'une normalisation linéaire.

Théorème A

Si $k(n)$ vérifié $k = k(n)$, $n - k(n) \to \infty$, $\frac{k(n)}{n} \to \lambda \in]0,1[$ avec $\sqrt{n}(\frac{k(n)}{n} - \lambda) \to 0$, et si $a_n > 0$ et $b_n > 0$ sont des constantes réelles telles que $P(X_{k,n} \leq a_n x + b_n) \to M(x)$, lorsque $n \to \infty$, alors M appartient à l'un des quatre types suivants :

$$\begin{aligned}
M_1(x) &= \begin{cases} 0 & \text{si } x < 0 \\ \phi(cx^{\alpha}) & \text{si } x \geq 0 \end{cases} ; \\
M_2(x) &= \begin{cases} \phi(-c|x|^{\alpha}) & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} ; \\
M_3(x) &= \begin{cases} \phi(-c_1|x|^{\alpha}) & \text{si } x < 0 \\ \phi(c_2 x^{\alpha}) & \text{si } x \geq 0 \end{cases} ; \\
M_4(x) &= \begin{cases} 0 & \text{si } x < -1 \\ \frac{1}{2} & \text{si } -1 < x \leq 1 \\ 1 & \text{si } x > 1 \end{cases} .
\end{aligned}$$

où c, c_1, c_2 et α sont des réels strictement positifs et $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$
.

Notre résultat principal pour une statistique d'ordre centrale est le suivant :

Théorème 4.4.1

Supposons que $k = k(n)$ vérifie les conditions $k = k(n)$, $n - k(n) \to \infty$, $\frac{k(n)}{n} \to \lambda \in]0,1[$ avec $\sqrt{n}(\frac{k(n)}{n} - \lambda) \to 0$, et si $a_n > 0$, $b_n > 0$ telles que $P(X_{k,n} \leq b_n(x)^{a_n}) \to L(x)$, $x > 0$ quand $n \to \infty$, où L est une f.r. non dégénérée, alors L appartient à l'un des quatre types suivants :

$$\begin{aligned}
L_1(x) &= \begin{cases} 0 & \text{si } x < 1 \\ \phi(c(\log x)^{\alpha}) & \text{si } x \geq 1 \end{cases} ; \\
L_2(x) &= \begin{cases} \phi(-c|\log x|^{\alpha} & \text{si } x < 1 \\ 1 & \text{si } x \geq 1 \end{cases} ; \\
L_3(x) &= \begin{cases} \phi(-c_1|\log x|^{\alpha} & \text{si } x < 1 \\ \phi(c_2(\log x)^{\alpha}) & \text{si } x \geq 1 \end{cases} ; \\
L_4(x) &= \begin{cases} 0 & \text{si } x < e^{-1} \\ \frac{1}{2} & \text{si } e^{-1} \leq x \leq e \\ 1 & \text{si } x > e \end{cases} .
\end{aligned}$$

où ϕ est la f.r. d'une loi $N(0,1)$ et c, c_1, c_2 et $\blacksquare$ sont des réels strictement positifs.

L'autre but de ce paragraphe est de donner les (C.N.S.) d'appartenance aux $DA(L)$ où $L \in \{L_1, L_2, L_3, L_4\}$. En notant $g \in VR_\alpha$, pour une fonction g à variation régulière d'indice [redacted] au voisinage de ∞ c'est-à-dire $\lim_{x\to\infty} \frac{g(tx)}{g(x)} = t^\alpha$, $t > 0$, nous avons

Théorème 4.4.2

(i)

$F \in DA(L_1)$ si et seulement s'il existe $y_0 \geq 0$ telle que

1^0) $F(y_0 y_1) = \lambda$ et $\forall k > 1$, $F(ky_0) > \lambda$ où $y_1 \to 1^+$.

2^0) $\frac{F(yy_0)+\lambda}{\lambda - F(y_0 y^{-1})} \to 0$, $y \to 1^+$.

3^0) $F(y_0 y) - \lambda \in VR_\alpha(0)$.

(ii)

$F \in DA(L_2)$ si et seulement s'il existe $y_0 \geq 0$ telle que

1^0) $F(y_0 y_2) = \lambda$ et pout tout $0 < k < 1$, $F(ky_0) < \lambda$ où $y_2 \to 1^-$.

2^0) $\frac{\lambda - F(y_0 y^{-1})}{F(y_0 y) - \lambda} \to 0$, $y \to 1^+$.

3^0) $\lambda - F(y_0 e^x) \in VR_\alpha(0)$.

(iii)

$F \in DA(L_3)$ si et seulement s'il existe $y_0 \geq 0$ telle que

1^0) $F(y_0) = \lambda$ et pour tout $k > 1$, $F(y_0/k) < F(y_0) < F(ky_0)$.

2^0) $\frac{F(y_0 y) - F(y_0)}{F(0y) - F(y_0 y^{-1})} \to A$, $y \to 1^+$.

3^0) $F(y_0 e^x) - F(y_0) \in VR_\alpha(0)$.

(vi)

$F \in DA(L_4)$ si et seulement s'il existe

1^0) $\underline{a}_\lambda < \overline{a}_\lambda$.

2^0) $\underline{a}_\lambda = \overline{a}_\lambda = a_\lambda$, $F(y_1 e^{a_\lambda}) = \lambda, F(y_2 e^{a_\lambda}) < \lambda$

et

$\forall \tau > 0$ $\frac{F(e^{a_\lambda} e^{\tau x}) - \lambda}{F(e^{a_\lambda} e^{x}) - \lambda)} \to \infty$, quand $y_1 \to 1^+$ et $y_2 \to 1^-$.

3^0) $\underline{a}_\lambda = \overline{a}_\lambda = a_\lambda$,

$F(y_1 e^{a_\lambda}) = F(y_2 e^{a_\lambda}) = \lambda$

et

$$\sqrt{n}\left\{F(y^{\frac{\alpha_n+\beta_n}{2}} e^{\frac{\alpha_n-\beta_n}{2}} e^{a_\lambda} - \lambda\right\} \to \left\{\begin{array}{lll} \quad -\infty & \text{si} & y < e^{-1} \\ 0 & \text{si} & e^{-1} \leq y \leq e \\ \quad +\infty & \text{si} & y > e \end{array}\right. ,$$

α_n et β_n sont définis par :

α_n=inf $\left\{y > 0 :\ F(y^{(1-0)} e^{a_\lambda}) - \lambda \leq \frac{1}{\sqrt{n}} \leq F(y^{(1+0)} e^{\lambda}) - \lambda\right\}$,

β_n=inf $\left\{y > 0 : \lambda - F(y^{(1-0)} e^{a}{}_\lambda) \leq \frac{1}{\sqrt{n}} \leq \lambda - F(y^{(1+0)} e^{\lambda})\right\}$,

les constantes $\underline{a}_\lambda$ et $\overline{a}_\lambda$ sont définiées ainsi :

$\overline{a}_\lambda = \inf\{y : F(y) > \lambda\}$ et $\underline{a}_\lambda = \sup\{y : F(y) < \lambda\}$.

4.5 Démonstrations

4.5.1 Démonstration du Théorème 4.4.1

Nous notons $Y = \log X$, nous avons

$$\begin{aligned} P(X_{k,n} &\leq b_n x^{a_n}) = P(Y_{k,n} \leq \log(b_n x^{a_n})), x > 0 \\ &= P(Y_{k,n} \leq a_n y + \log b_n),\ y \in IR. \end{aligned}$$

D'après le résultat de Smirnov [1962], nous avons

$$P(X_{k,n} \leq b_n x^{a_n}) = P(Y_{k,n} \leq a_n y + \log b_n) \rightarrow M(y),$$

avec $M \in \{M_1, M_2, M_3, M_4\}$,en prenant $L(x) = M(\log x)$.
D'où le résultat.

4.5.2 Démonstration du Théorème 4.4.2

(i) Si $F \in DA(L_1) \Leftrightarrow F_1 \in DA(M_1)$, en utilisant le théorème 4.4.1. et un résultat classique caractérisant le domaine d'attraction $DA(M_1)$ dans le cas linéaire (voir Smirnov [1962], theoreme 8a p 106) en notant $y_0 = e^{x_0}$, $k = e^{\epsilon}$ et $y = e^x$; nous avons

$F_1 \in DA(M_1)$

$\Longleftrightarrow$

$1^0)$ $\exists x_0$ telle que $F_1(x_0 + 0) = \lambda$ et $\forall \epsilon > 0$, $F_1(x_0 + \epsilon) > \lambda$

$2^0)$ $\exists x_0$ telle que $\frac{F_1(x+x_0)-\lambda}{\lambda - F_1(x_0-x)} \rightarrow 0$, $x \rightarrow 0^+$

$3^0)$ $\forall t > 0$ $\frac{F_1(x_0+tx)-\lambda}{F_1(x_0+x)-\lambda} \rightarrow t^\alpha$, $x \rightarrow 0^+$

$\Longleftrightarrow$

$1^0)$ $\exists x_0$ telle que $F(\exp(x_0 + 0)) = \lambda$ et $\forall \epsilon > 0$, $F(\exp(x_0 + \epsilon)) > \lambda$

$2^0)$ $\exists x_0$ telle que $\frac{F(\exp(x+x_0))-\lambda}{\lambda - F(\exp(x_0-x))} \rightarrow 0$, $x \rightarrow 0^+$

$3^0)$ $\forall t > 0$, $\frac{F(\exp(x_0+tx))-\lambda}{F(\exp(x_0+x))-\lambda} \rightarrow t^\alpha$, $x \rightarrow 0^+$

$\Longleftrightarrow$

$1^0)$ $\exists y_0 \geq 0$ telle que $F(y_0 y_1) = \lambda$ et $\forall k > 1$, $F(ky_0) > \lambda$, $y \rightarrow 1^+$

$2^0)$ $\exists y_0 \geq 0$ telle que $\frac{F(yy_0)-\lambda}{\lambda - F(y_0 y^{-1})} \rightarrow 0$, $y \rightarrow 1^+$

$3^0)$ $\forall t > 0$, $\frac{F(y_0 e^{xt})-\lambda}{F(y_0 e^x)-\lambda} \rightarrow t^\alpha$, $x \rightarrow 0^+$,donc $G(x) = F(y_0 e^x) - \lambda \in VR_\alpha(0)$

(ii) Si $F \in DA(L_2) \Leftrightarrow F_1 \in DA(M_2)$, en utilisant le Théorème 4.4.1 et un résultat classique caractérisant le domaine d'attraction $DA(M_2)$ dans le cas linéaire (voir Smirnov [1962], théorème 8b, p 111), nous avons $F_1 \in DA(M_2)$ si et seulement si

$1^0)$ $\exists x_0$ telle que $F_1(x_0 - 0) = \lambda$ et $\forall \epsilon > 0$, $F_1(x_0 - \epsilon) < \lambda$

$2^0)$ $\exists x_0$ telle que $\frac{F_1(x-x_0)-\lambda}{\lambda - F_1(x_0+x)} \rightarrow 0$, $x \rightarrow 0^+$

$3^0)$ $\forall t > 0$, $\frac{F_1(x_0-tx)-\lambda}{F_1(x_0-x)-\lambda} \rightarrow t^\alpha$, $x \rightarrow 0^+$

$\Leftrightarrow$

$1^0)$ $\exists y_0 \geq 0$ telle que $F(y_0 y_2) = \lambda$ et $\forall k > 1, F(\frac{y_0}{k}) < \lambda$,

$2^0)$ $\exists y_0 \geq 0$ telle que $\frac{F(\frac{y}{y_0})-\lambda}{\lambda - F(y_0 y)} \rightarrow 0$, $y \rightarrow 1^+$

$3^0)$ $\forall t > 0$, $\frac{F(y_0 e^{-tx})-\lambda}{F(y_0 e^{-x})-\lambda} \rightarrow t^\alpha$, $x \rightarrow 0^+$, alors $G(x) = F(y_0 e^{-x}) - \lambda \in VR_\alpha(0)$

iii) Si $F \in DA(L_3) \Leftrightarrow F_1 \in DA(M_3)$, en utilisant le théorème 4.4.1 et un résultat classique caractérisant le domaine d'attraction $DA(M_3)$ dans le cas linéaire (voir Smirnov [1962], théorème 8c, p 112), nous avons

$F_1 \in DA(M_3)$

$\Longleftrightarrow$

$1^0)$ $\exists x_0$ telle que $F_1(x_0) = \lambda$ et $\forall \epsilon > 0, F_1(x_0 - \epsilon) < F_1(x_0) < F_1(x_0 + \epsilon)$

$2^0)$ $\exists x_0$ telle que $\frac{F_1(x+x_0)-F_1(x_0)}{F_1(x_0)-F_1(x_0-x)} \rightarrow A$, $x \rightarrow 0^+$ où A est un réel positif

3^0) $\forall t > 0$, $\frac{F_1(x_0+tx)-F_1(x_0)}{F_1(x_0+x)-F_1(x_0)} \to t^\alpha$, $x \to 0^+$
$\Longleftrightarrow$
1^0) $\exists y_0 \geq 0$ telle que $F(y_0) = \lambda$ et $\forall k > 1, F(\frac{y_0}{k}) < F(y_0) < F(ky_0)$
2^0) $\exists y_0 \geq 0$ telle que $\frac{F(yy_0)-F(y_0)}{F(y_0)-F(y_0y^{-1})} \to A$, $y \to 1^+$
3^0) $\forall t > 0$, $\frac{F(y_0e^{tx})-F(y_0)}{F(y_0e^x)-F(y_0)} \to t^\alpha$, $x \to 0^+$, alors $F(y_0e^x) - F(y_0) \in VR_\alpha(0)$

vi) Si $F \in DA(L_4) \Leftrightarrow F_1 \in DA(M_4)$, en utilisant le théorème 4.4.1 et un résultat classique caractérisant le domaine d'attraction $DA(M_4)$ dans le cas linéaire (voir Smirnov [1962], théorème 9, p 123), nous avons
$F_1 \in DA(M_4)$
$\Longleftrightarrow$
1^0) $\underline{a}_\lambda < \overline{a}_\lambda$
2^0) $\underline{a}_\lambda = \overline{a}_\lambda = a_\lambda$, $F_1(a_\lambda + 0) = \lambda$, $F_1(a_\lambda - 0) < \lambda$
et
$\forall \tau > 0 \frac{F_1(a_\lambda+tx)-\lambda}{F_1(a_\lambda+x)-\lambda} \to \infty$, quand $x \to 0^+$
3^0)$\underline{a}_\lambda = \overline{a}_\lambda = a_\lambda$, $F_1(a_\lambda - 0) = F_1(a_\lambda + 0) = \lambda$
et

$$\sqrt{n}\left\{F_1(x\frac{\alpha_n+\beta_n}{2} + \frac{\alpha_n-\beta_n}{2} + a_\lambda) - \lambda\right\} \to \begin{cases} -\infty & \text{si } x < -1 \\ 0 & \text{si} -1 \leq x \leq 1 \\ +\infty & \text{si } x > 1 \end{cases}$$

$\Longleftrightarrow$
1^0) $\underline{a}_\lambda < \overline{a}_\lambda$
2^0) $\underline{a}_\lambda = \overline{a}_\lambda = a_\lambda$, $F(y_1e^{a_\lambda}) = \lambda, F(y_2e^{a_\lambda}) < \lambda$
et
$\forall \tau > 0$ $\frac{F(e^{a_\lambda}e^{\tau x})-\lambda}{F(e^{a_\lambda}e^x-\lambda}) \to \infty$, quand $y_1 \to 1^+$ et $y_2 \to 1^-$
3^0) $\underline{a}_\lambda = \overline{a}_\lambda = a_\lambda$, $F(y_1e^{a_\lambda}) = F(y_2e^{a_\lambda}) = \lambda$ et

$$\sqrt{n}\left\{F(y^{\frac{\alpha_n+\beta_n}{2}}e^{\frac{\alpha_n-\beta_n}{2}}e^{a_\lambda}) - \lambda\right\} \to \begin{cases} -\infty & \text{si } y < e^{-1} \\ 0 & \text{si } e^{-1} \leq y \leq e \\ +\infty & \text{si } y > e. \end{cases}$$

Conclusion et Perspectives

Soient $X_1, X_2, ..., X_n$ n variables aléatoires indépendantes et identiquement distribuées de fonction de répartition commune $F(x) = P(X_1 \leq x)$.

Dans le chapitre 2, nous avons montré que, lorsque $F \in DA(\Lambda)$,

$$\lim_{n\to\infty} P(\frac{\max_{1\leq i_1<...<i_k\leq n}(X_{i_1} + X_{i_2} + ... + X_{i_k}) - kb_n}{a_n} \leq x) = H_k(x) \qquad (*)$$

avec $a_n > 0$ et b_n des constantes de normalisation.

Y'aurait-il possibilité de trouver des limites convenables pour $\max_{1\leq i_1<...<i_k\leq n}(X_{i_1} + X_{i_2} + ... + X_{i_k})$ moyennant des constants de normalisation lorsque F appartiendrait aux domaines d'attraction $DA(\Phi_\alpha)$ et $DA(\Psi_\alpha)$?

Nous avons, dans notre cas, utilisé la fonction

$h : (x_{i_1}, x_{i_2}, ..., x_{i_k}) \to (x_{i_1} + x_{i_2} + ... + x_{i_k})$ de IR^k vers IR, pour trouver $(*)$.

On pourrait généraliser ce résultat en utilisant une fonction $h : IR^k \to IR$ quelconque et en prenant $F \in \{DA(\Phi_\alpha), DA(\Psi_\alpha), DA(\Lambda)\}$.

Nous avons aussi établi, dans le chapitre 4, des conditions nécessaires et suffisantes d'appartenance aux domaines d'attraction des lois limites $\{\Phi_1, H_{1,\alpha}, H_{2,\alpha}\}$ dans le cas d'une normalisation non linéaire. Il serait intéressant de chercher un estimateur pour l'indice α dans $H_{1,\alpha}$ et $H_{2,\alpha}$ et d'en étudier les propriétés usuelles : convergences forte et faible et convergence en loi.

Ces estimateurs potentiels auront-ils des liens avec les estimateurs de Hill, de Pickands et celui des moments si populaires pour l'évaluation de l'indice des lois extrêmes dans le cas d'une normalisation linéaire?

Abstract : In this thesis, we first study the asymptotic limit laws for maxima of sums of i.i.d random variables in the case of a linear normalization. We are particularly interested in the laws belonging to the dominant attraction of Gumbel. Thereafter we establish limits results of central statistics order in the case of independent but not identically distributed random variables. The case of a non-linear normalization is studied in the last chapter where we characterize the different limit laws for i.i.d. random variables.

Résumé : Dans ce mémoire, nous étudions d'abord des lois limites asymptotiques pour les maxima de sommes de variables aléatoires i.i.d. dans le cas d'une normalisation linéaire adéquate. Nous nous y intéressons particulièrement à des lois appartenant au domaine d'attraction de Gumbel. Par la suite nous établissons des résultats limites pour les statistiques d'ordre centrales dans le cas de variables indépendantes mais non identiquement distribuées. Le cas d'une normalisation non linéaire est étudié dans le dernier chapitre où nous caractérisons les différentes lois limites du maximum pour les variables aléatoires i.i.d.

Références

Arnold, B.C, Balakrishnan, N. and Nagaraga, H.N. (1998), Record. JohnWiley & Sons. Inc.

El arrouchi M.(2007), Inférence statistique des extrêmes en termes des records, thèse, Université Abdelmalek Essaâdi de Maroc.

Balkema, A. A. et de Haan, L., (1978a), Limit distributions for order statistics, I, Theory Probab. Appl. 23, 77-92

Balkema, A. A. et de Haan, L., (1978b), Limit distributions for order statistics, II, Theory Probab. Appl. 23, 341-358.

Balkema, A.A, de HAAN, L., "Residual Life at Great Age", Annals of Probability 2, 1974, p. 792-804.

Barbut M.(2003), "Homme moyen ou homme extrême", Journal de la Société Française de la Statistique 144, p.1-2.

Barbut M.(1998), « Une famille de distributions : des paretiennes aux conta-paretiennes, applications à l'étude de la concentration urbaine et de son évolution", Mathématiques, Informatique et Sciences humaines 141, p. 43-72.

Barry,J., Kang,J., Yongcheng,Q., (2007), Limit distribution of the sum and maximum from multivariate Gaussian sequences. Journal of Multivariate Analysis 98, 517-532.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., and Ferro, C. (2004), Statistics of Extremes : Theory and Applications.Wiley, New York.

Belbachir, M. Hanafi, A. ,(2011), Limit law for maximum of random variables in the domain of attraction of the Gumbel law. International Journal of Tomography & Statistics, Vol. 17, No. 511.

Bingham,N.H, Goldie, C.M. and Teugels, J.L. (1987), Regular Variation. Cambridge University Press.

Castillo, E. (1988), Extreme value theory in engineering. Statistical Modeling and Decision Science. Academic Press, Inc.

COLES S.G., WALSHAW, D.(1994), "Directional modeling of extreme wind speeds", Appl. Statist. 43 , p. 139–157.

Csörg■o, S., Deheuvels, P. and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, Ann. Statist, 13, 1050-1077.

DANIELSSON J., DE VRIES C. G.(1997), "Tail index estimation with very high frequency data", Journal of Empirical Finance 4, p. 241-257.

DAVISON A. C., SMITH R. L.(1990), "Models for exceedances over high thresholds", Journal of Royal Statistic Society Ser. B 52, p. 393-442.

de Haan, L. (1970), On Regular Variation and its Applications to the Weak Convergence of Sample Extremes, Math. Centre Tract 32, Amsterdam.

de Haan, L. (1984), Slow variation and characterization of domains of attraction. In Statistical Extremes and Applications (J. Tiago de Oliveira, ed.), 31-48. Reidel, Dordrecht.

Deheuvels, P. (1988), The strong approximations of extremal processus (II), Z.Wahrsch. verw. Geb, 62, 7-15.

Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989), A moment estimator for the index of an extreme value distribution. Annals of Statistics 17, 1833-1855.

Dekkers, A. L. M. and de Haan, L. (1993), Optimal choice of sample fraction in extreme-value estimation, J. Multivariate Analysis, 47, 173-195.

Dingcheng, W., Qihe, T.(2004)**,** Maxima of sums and random sums for negatively associated random variables with heavy tails, Statistics and Probability Letters, 68, 287-295.

Dwass, M. (1966), Extremal Processes, II, Illinois Journal of Mathematics10, 381-391.

Embrechts, P., Klöppelberg, C. et Mikosch, T. (1997), Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin.

Finkenstädt, B. & Rootzén, H., eds (2004), Extreme Values in Finance, Telecommunications and the Environment, Chapman and Hall, Boca Raton.

Fisher, R. A. & Tippett, L. H. C. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proceedings of the Cambridge Philosophical Society 24, 180-190.

Fréchet, M. (1927), Sur la loi de probabilité de l'écart maximum, Annales de la Société Polonaise de Mathématiques de Cracovie 6, 93-116.

Galambos, J. (1987), The Asymptotic Theory of Extreme Order Statistics 2nd edition, Krieger, Malabar, Florida.

GENÇAY R., SELÇUK F.(2004), "Extreme value theory and value-at-risk: Relative performance in emerging markets", International Journal of Forecasting 20 , p. 287-303.

Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, Annals of Mathematics, 44, 423-453. Translated and reprinted in : Breakthroughs in Statistics, Vol. I, 1992, eds. S. Kotz and N. L. Johnson, Springer-Verlag, pp. 195-225.

Gumbel, E.J. (1954), Statistical Theory of Extreme Values and Some Pratical Applications, Applied Mathematics Series, 33, National Bureau of Standards, Washington.

Gumbel, E.J. (1958), Statistics of extremes. Columbia University Press.

Hall, P. (1982), On some simple estimates of an exponent of regular variation. J. Roy. Statist. Soc. Ser. B 44, 37-42.

Hill, B. M. (1975), A simple general approach to inference about the tail of a distribution. Ann. Statist. 3, 1163-1174.

Hosking, J. R. M. & Wallis, J. R. (1987), Parameter and quantille estimation for the generalized Pareto distribution, Technometrics 29, 339-349.

Hsing, T. (1995), A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables, Ann. Probab, 23, 938-947.

KLAJNMIC H.(2003), « Niveaux de retour pour les vitesses extrêmes des vents », XXXVes Journées de Statistique, Lyon.

Leadbetter, M. R., Lindgren, G. & Rootzén, H. (1983), Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, New York.

Leadbetter, M.R., (1978), Extreme value theory under weak mixing conditions in MAA Studies in Mathematics (M, Rosenblatt, ed.) 18, 46-110.

LONGIN F. M.(1998), « Value at risk: Une nouvelle approche fondée sur les valeurs extrêmes », Annales d'économie et de statistique 52, p. 23-51.

McNEIL A. J., FREY R.(2000), "Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach", Journal of Empirical Finance 7, p. 271-300.

McNEIL A. J., SALADIN T.(1997), "The peaks over thresholds method for estimating high quantiles of loss lois", Department of Mathematic ETH Zurich.

McNEIL A. J.(1998), Calculating quantile risk measures for financial time series using extreme value theory, Department of Mathematics, ETH. Swiss Federal Technical University E-Collection.

McNEIL A. J.(1998), Calculating quantile risk measures for financial time series using extreme value theory, Department of Mathematics, ETH. Swiss Federal Technical University E-Collection.

Mejzler, D., (1984), Asymptotic behaviour of the extreme order statistics in the non identically distributed case in Statistical Extremes and Applications(J. T. de Olivera, ed.), Reided, Dordrecht, 535-547.

Mejzler, D. et Weissman, I. (1969)**.** On some results of N.V. Smirnov concerning limit distributions for variational sries. Ann. Math. Stat. Statist. 40, 480-491.

Nevzorov, V. B. (1987), Records. Theory of Probability and Applications 32(2), 201-228 (English translation).

Pancheva, E.,(1984), limit theorems for extreme order statistics under nonlinear normalization, lecture Notes Math, 1155, 284-309.

Pickands, J . III. (1975), Statistical inference using extreme order statistics. Ann. Statist. 3, 119-131.

Reiss R., Thomas M.(2001), Statistical Analysis of Extreme Value with Applications to Assurance, Finance, Hydrology and Other Fields, Basel, Birkhauser, Verlag.

Raggad B.(2009), Fondements de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier, math. sci. hum / mathematics and social sciences (47e année, n° 186, (2), p. 29-63) .

Resnick, S. I. (1973), Record values and maxima, Annals of Probability 4, 650-662.

Resnick, S. I. (1987), Extreme Values, Regular Variation and Point processes. Springer-Verlag, New York.

Rootzen H., Tajvidi N., "Extreme value statistics and wind storm losses: a case study", Scandinavian Actuarial Journal, 1997, p. 70-94.

Smirnov. N. V., (1962), Limit distributions for the terms of a variational series, Am. Math. Soc. Transl. 11, 82-143.

SMITH R. L. (2001), “Extreme value statistics in meteorology and the environment. In environmental Statistics”,Chapter 8, NSF–CBMS conference notes, , p. 300–357

Ravi,S. Praveena, A.S. (2011) On von Mises type conditions for p-max stable laws,rates of convergence and generalized log Pareto distributions. Journal of Statistical Planning and Inference 141, 3021–3034.

THATCHER R. A.(1999), “The Long-Term Pattern of Adult Mortality and the Highest Attained Age”, [with discussion], Journal of the Royal Statistical Society Series A 162, p. 5–43.

Von Bortkiewicz, L. (1922), Variantionsbreite und mittelerer Fehler, Sitzungsber. Berli. Math. Ges. 21, 3-11.

Von Mises, R. (1936), La distribution de la plus grande de n valeurs, Rev. Math. Union Interbalk. 1, 141-160. Reproduced in Selected Papers of Richard von Mises, II (1954), pp. 271-294, Amer. Math. Soc.

Weissman, I. (1975). Multivariate extremal processes generated by independent, non identically distributed random variables. J. App1. Probab. 12, 477-487.

Wu, C, (1966), The types of limit distributions for some terms of variational series, SCi.Sinica 15, 749-762.