

Théorie des Valeurs Extrêmes

ou

Modélisation des Evénements Rares pour la Gestion des Risques

DESS 203 de l'Université Paris IX Dauphine
Marchés Financiers, Marchés des Matières
Premières et Gestion des Risques

Thierry RONCALLI

Groupe de Recherche Opérationnelle
Crédit Lyonnais
Bercy-Expo – Immeuble Bercy Sud – 4ème étage
90, Quai de Bercy — 75613 Paris Cedex 12
France

thierry.roncalli@creditlyonnais.fr

Janvier 2002

Table des matières

1	Introduction	1
1.1	La problématique de la gestion des risques	3
1.1.1	Cadre réglementaire et allocation des fonds propres	3
1.1.2	L'impact des événements rares sur la mesure du risque	5
1.2	L'hypothèse de normalité en finance	8
1.2.1	La distribution gaussienne	8
1.2.2	La dépendance normale	15
1.3	Conclusion	18
2	La théorie des valeurs extrêmes	21
2.1	Les statistiques d'ordre	21
2.2	Le théorème de Fisher-Tippet	31
2.3	Caractérisation des domaines d'attraction	35
2.4	L'approche par processus ponctuels	38
2.5	La distribution GEV	38
2.6	Les modèles à seuil	44
3	Extension au cas multidimensionnel	49
3.1	Une introduction aux copules	49
3.1.1	Représentation canonique d'une distribution bivariée	49
3.1.2	Classe de Fréchet, distributions minimale et maximale et ordre de concordance	52
3.1.3	Copules et variables aléatoires	55
3.1.4	Dépendance et corrélation	57
3.1.5	Extension au cas multidimensionnel	62
3.1.6	Quelques exemples de familles de copules	62
3.2	Les copules de valeurs extrêmes	63
3.3	Caractérisation des domaines d'attraction	69
3.4	L'exemple des copules Archimax	73
3.5	Copules et modélisation financière	75
4	Applications à la gestion des risques	79

4.1	Le risque de marché	79
4.1.1	Construction d'une valeur en risque GEV	79
4.1.2	La mesure de l' <i>Expected Shortfall</i> et GPD	81
4.1.3	Programmes de stress-testing	82
4.1.3.1	Cas unidimensionnel	83
4.1.3.2	Cas multidimensionnel	84
4.1.4	Dépendance de queue et valeur en risque	85
4.2	La prise en compte des événements extrêmes dans le risque opérationnel	86
4.2.1	Modélisation de la distribution de sévérité	87
4.2.2	Aggrégation des risques	89
4.2.3	Méthodes de backtesting	92
4.2.4	LDA et théorie des extrêmes	93
4.3	Le risque de crédit	94
4.3.1	Estimation des probabilités de défaillance	94
4.3.2	Modélisation des défauts joints	94
5	Conclusion	99

1

Introduction

L'intitulé du cours est "Théorie des Valeurs Extrêmes". Le fil conducteur de ce cours est d'utiliser cette théorie pour **comprendre les enjeux majeurs** de la gestion des risques. Ce ne sera donc pas un cours de modélisation de la gestion des risques par la théorie des valeurs extrêmes. La raison principale est que celle-ci n'est sûrement pas le bon cadre d'analyse. C'est pourquoi le sous-titre du cours est "**Modélisation des Événements Rares pour la Gestion des Risques**". La théorie des valeurs extrêmes sera alors un prétexte pour comprendre l'impact des événements rares sur les mesures de risque.

La théorie des valeurs extrêmes est présentée dans le second chapitre. Néanmoins, j'indique ici quelle est la ligne directrice sous-jacente. Considérons n variables aléatoires *iid* (indépendantes et de même loi de distribution \mathbf{F}) que nous notons X_1, \dots, X_n . Nous cherchons alors à caractériser la loi des extrema. Soient χ_n^- et χ_n^+ le minimum et le maximum définis par

$$\chi_n^- := \bigwedge_{i=1}^n X_i = \min(X_1, \dots, X_n) \quad (1.1)$$

et

$$\chi_n^+ := \bigvee_{i=1}^n X_i = \max(X_1, \dots, X_n) \quad (1.2)$$

Nous verrons dans la section 2.1 que nous pouvons parfaitement caractériser la distribution de χ_n^- ou χ_n^+ si nous connaissons la distribution \mathbf{F} . Mais que pouvons-nous dire du comportement de χ_n^- ou χ_n^+ si \mathbf{F} n'est pas connue? La théorie des valeurs extrêmes permet justement de répondre à ce type de question, car elle cherche à caractériser les distributions asymptotiques de χ_n^- et χ_n^+ lorsque n tend vers l'infini. Stuart Coles [3] présente cette méthodologie de la façon suivante :

By definition, extreme values are scarce, meaning that estimates are often required for levels of a process that are much greater than have already been observed. This implies an extrapolation from observed levels to unobserved levels, and extreme value theory provides a class of models to enable such extrapolation. In lieu of an empirical or physical basis, asymptotic argument is used to generate the extreme value models. It is easy to be cynical about this strategy, arguing that extrapolation of models to unseen levels requires a leap of faith, even if the models have an underlying

asymptotic rationale. There is no simple defense against this criticism, except to say that applications demand extrapolation, and that it is better to use techniques that have a rationale of some sort. This argument is well understood and, notwithstanding objections to the general principle of extrapolation, there are no serious competitor models to those provided by extreme value theory (COLES [2001], page vii).

N'en déplaise à ceux qui la supportent, il est malheureusement difficile d'utiliser la théorie des valeurs extrêmes comme une procédure statistique **robuste**. Et nous ne pouvons qu'adhérer aux critiques de Nicolas Bouleau dans son article "Splendeurs et misères des lois de valeurs extrêmes" :

*Ayant observé durant une période d'un siècle dans une certaine région des enregistrements sismiques de magnitude comprise entre 0 et 2, est-il possible d'en déduire avec quelle probabilité se produira dans la même région durant le siècle suivant un séisme de magnitude supérieure à 4 ? A un problème ainsi posé, rares seraient ceux qui répondraient par l'affirmative, néanmoins l'usage de plus en plus répandu dans le milieu des ingénieurs de procédures rapides utilisant les lois des valeurs extrêmes conduit à des affirmations de ce type, dont l'enjeu socio-politique est important notamment par l'habit de scientificité qui leur est donné. Après avoir rappelé les fondements de la théorie des lois des valeurs extrêmes et relevé quelques unes des hypothèses cruciales, difficiles à vérifier en pratique qui la sous-tendent, nous montrons que la méthode qui consiste à caler les paramètres d'une des trois lois de valeurs extrêmes à partir des extrêmes d'un échantillon fini dont la loi est mal connue, est fortement encouragée par la pression sociale de quantifier les risques graves d'autant plus que de tels errements, par la rareté même des événements considérés, sont **peu réfutables**.*

Et la conclusion de Nicolas Bouleau est la suivante :

[...] toute démarche attribuant une valeur numérique précise pour la probabilité d'un phénomène rare est suspecte, sauf si les lois physiques régissant le phénomène sont explicitement et exhaustivement connues.

Dans ce cas, quel est l'intérêt d'utiliser cette théorie pour la gestion des risques ? Nous pouvons avancer deux arguments :

- tout d'abord, la théorie des valeurs extrêmes va nous permettre de comprendre quelles sont les hypothèses sous-jacentes des modèles de risque, et comment les événements rares vont se comporter dans ces types de modèles ;
- ensuite, elle fournit tout un cadre statistique d'étude des événements rares en échantillon fini.

La démarche que j'adopterai est en fait beaucoup plus proche du cours de GUMBEL [1954], que des textes académiques plus récents trop mathématiques.

Avant d'aborder la théorie des valeurs extrêmes proprement dite, je vais rappeler brièvement la problématique générale de la gestion des risques, et illustrer pourquoi l'hypothèse traditionnelle de normalité est inadaptée dans ce contexte-ci.

Remarque 1 *Ce cours peut être complété par celui de l'ENSAI portant sur "La Théorie des Extrêmes et la Gestion des Risques de Marché" qui est disponible sur le site internet du GRO :*

<http://gro.creditlyonnais.fr>

1.1 La problématique de la gestion des risques

L'orientation générale du cours concerne la gestion des risques. Même si des rappels seront rapidement faits, il est indispensable de connaître les grandes lignes directrices de la gestion des risques et les enjeux sous-jacents. Si vous n'êtes pas familiers avec ces thèmes, vous pouvez consulter les notes du cours *Introduction à la Gestion des Risques*, qui est aussi disponible sur le site internet du GRO.

1.1.1 Cadre réglementaire et allocation des fonds propres

Les établissements de crédit sont soumis à une réglementation prudentielle, qui impose une exigence minimale de fonds propres pour toutes les opérations bancaires et financières qui présentent un risque. C'est la fameuse **CAD (Capital Adequacy Directive)** de la Commission Européenne. Celle-ci est une transcription réglementaire de l'Accord de Bâle de 1988 dont le principal intérêt est de fixer un ratio de solvabilité (ratio Cooke). Celui-ci impose que le rapport des fonds propres (tier one + tier two) sur les encours pondérés de crédit soit supérieur à 8%. Les encours pondérés de crédit doivent s'interpréter comme la mesure de risque crédit. Cette directive a été modifiée en 1998 pour prendre en compte le risque de marché. En janvier 2001, le Comité de Bâle propose une refonte de l'Accord de 1988 et un nouveau ratio de solvabilité, le ratio McDonough. Et l'implémentation du Nouvel Accord de Bâle devrait se faire en janvier 2005.

Trois types de risque sont concernés par le Nouvel Accord :

- le risque de crédit ;
- le risque de marché ;
- et le risque opérationnel.

Le risque de crédit correspond aux risques encourus par l'établissement financier qui sont liés à la défaillance d'une contrepartie ou/et à la dégradation de la qualité du crédit (c'est-à-dire à la volatilité du mark-to-market due à une modification du rating). Le risque de marché est le risque de pertes sur les positions du bilan et du hors-bilan à la suite de variations des prix de marché. C'est donc le risque de volatilité des facteurs de marché (taux d'intérêt, prix des actions, cours de change, cours de matières premières, smile de volatilité, etc.). Enfin, le risque opérationnel se définit comme le risque de pertes dues à une inadéquation ou à une défaillance des procédures, personnels, systèmes internes ou à des événements extérieurs. Plus simplement, le risque opérationnel correspond aux pertes, qui ne relèvent pas du risque de crédit ou de marché (mais il ne comprend pas les risques de réputation). La répartition initiale de ces risques pour l'industrie (c'est-à-dire pour l'ensemble des établissements financiers de l'OCDE) que propose le Comité de Bâle pour le Nouvel Accord est la suivante :

Type de risque	Répartition
Crédit	75%
Marché	5%
Opérationnel	20%

Remarque 2 Contrairement au risque de taux du portefeuille de négociation (*trading book*), le risque de taux d'intérêt dans le portefeuille bancaire (*banking book*) ne fait pas partie de l'exigence minimale en fonds propres (premier pilier). Celui-ci est traité dans le second pilier, c'est-à-dire au niveau du processus de surveillance.

Par rapport au ratio Cooke, le ratio McDonough élargit le périmètre des risques retenus sans modifier le niveau d'exigence minimale :

$$\frac{\text{Fonds Propres}}{\text{Risque de Crédit} + \text{Risque de Marché} + \text{Risque Opérationnel}} \geq 8\%$$

La référence utilisée par le Comité de Bâle pour définir la mesure des risques est la Valeur en Risque (**Value-at-Risk** ou **VaR**). Même si cette mesure a été fortement critiquée d'un point de vue académique, elle correspond à une vraie logique économique pour la banque, qui est basée sur le niveau de couverture exigé par les actionnaires. D'un point de vue réglementaire, les niveaux de couverture sont les suivants :

Type de risque	Proposition de janvier 2001	Proposition de septembre 2001
Crédit	99.5%	99.9%
Marché	99%	99%
Opérationnel	99%	99.9%

Les établissements financiers n'ont pas attendu la proposition de janvier 2001 (et n'attendent pas l'implémentation de janvier 2005) pour

1. mesurer le risque opérationnel¹,
2. calculer des fonds propres économiques pour le risque de crédit,
3. augmenter le niveau de couverture de la VaR marché,
4. et mettre en place une démarche intégrée de mesure des risques.

Ces démarches sont notamment nécessaires dans les projets d'allocation de fonds propres. L'allocation de fonds propres consiste à répartir de façon **optimale** le capital entre les différentes activités. Celle-ci peut se faire à un niveau agrégé (retail, middle market, corporate banking, insurance, asset management, etc.) et on parle alors de méthode **top-down**. Elle peut aussi se faire à un niveau beaucoup plus fin (c'est-à-dire au niveau de la ligne métier) et dans ce cas, c'est une approche **bottom-up**. Le critère naturel d'allocation est bien sûr la rentabilité du capital (**Return on Equity** ou **ROE**). Néanmoins, ce critère ignore une composante importante de la gestion d'une banque : le risque. C'est pourquoi on préfère utiliser un critère de rentabilité ajustée du risque (**Risk-Adjusted Performance Measurement** ou **RAPM**). On distingue généralement les fonds propres réglementaires des fonds propres économiques (calculés avec un modèle interne). Selon une agence de conseil américaine, la répartition des fonds propres économiques pour l'industrie pourrait être la suivante :

Type de risque	Répartition
Crédit	65-75%
Marché	10-15%
Opérationnel	15-20%

Contrairement aux fonds propres réglementaires, le capital économique dépend du niveau de couverture que l'établissement financier souhaite mettre en place. Ce niveau dépend du rating actuel de la banque ou/et de l'objectif de rating :

Rating	BBB	A	AA	AAA
Niveau de couverture	99.75%	99.9%	99.95%	99.97%

¹Enfin, pour l'instant, le nombre de banques dans le monde qui le font est (**très ?**) faible.

Quelle que soit la mesure de risque (réglementaire ou économique), on constate que les niveaux de couverture sont particulièrement élevés. Avec ces niveaux de quantile, on manipule des événements qui ont une faible probabilité d'occurrence. De plus, ces mesures de risque sont généralement complétées par des programmes de scénarios de crise (**stress-testing**). Dans ce cas, on fait face à des événements très rares.

1.1.2 L'impact des événements rares sur la mesure du risque

Avant d'aborder l'incidence des événements rares sur la gestion des risques, il est important de pouvoir caractériser ce qu'est un événement rare :

Un événement rare est un événement dont la probabilité d'occurrence est petite.

Présentée de cette façon, il est difficile d'utiliser cette définition pour comparer deux événements. Par exemple, est-ce qu'une probabilité de défaillance de 0,7% pour un horizon d'un an est plus petite qu'une probabilité de défaillance de 1,9% pour un horizon de 3 ans? C'est pourquoi les financiers utilisent une notion plus pertinente pour caractériser la rareté d'un événement : **le temps de retour (ou return period)**. GUMBEL [1958] introduit ce concept de la façon suivante :

Consider first a discontinuous variate as in the case of the classical die, from which so many problems of the calculus of probability arose. For an unbiased die, the probability that a certain point occurs is $\frac{1}{6}$. Therefore, we may expect to get this point, in the long run and on the average, once in $T = 6$ trials. For a continuous variate, there is no probability for a certain value x , but only a density of probability. However, there is a probability $1 - F(x)$ of a value to be equalled or exceeded by x . Its reciprocal

$$\mathcal{T}(x) = \frac{1}{1 - \mathbf{F}(x)} \quad (1.3)$$

*is called the **return period**. It is the number of observations such that, on the average, there is one observation equalling or exceeding x . Obviously, the return period increases with the variate.*

Une autre façon de caractériser le temps de retour est de considérer la loi du premier excès ou dépassement $X \geq x$. Notons $p = 1 - \mathbf{F}(x)$ la probabilité de dépasser le seuil x . Le premier excès

$$\kappa(x) := \min \{i \geq 1 : X_i \geq x\} \quad (1.4)$$

est alors une variable aléatoire de distribution **géométrique** G_p . Le temps de retour des événements $\{X_i \geq x\}$ est la moyenne du temps de premier excès :

$$\begin{aligned} \mathcal{T}(x) &= \mathbb{E}[\kappa(x)] \\ &= \frac{1}{p} \\ &= \frac{1}{1 - \mathbf{F}(x)} \end{aligned} \quad (1.5)$$

$\mathcal{T}(x)$ est donc le temps d'attente moyen entre deux dépassements. Appliquons ce concept à la gestion des risques. Soit \mathbf{F} la distribution des pertes. La valeur en risque pour un niveau de couverture α est le quantile de la distribution :

$$\text{VaR} = \mathbf{F}^{-1}(\alpha) \quad (1.6)$$

L'expression du temps de retour \mathcal{T} (VaR) est donc

$$\begin{aligned}\mathcal{T}(\text{VaR}) &= \frac{1}{1 - \mathbf{F}(\mathbf{F}^{-1}(\alpha))} \\ &= \frac{1}{1 - \alpha}\end{aligned}\tag{1.7}$$

C'est le temps moyen que nous observons entre deux dépassements **successifs** de la valeur en risque. Voici par exemple le temps de retour implicite des différents ratings pour des périodes de détention de 1 jour et 10 jours de trading :

Rating	Période de détention	Risque de marché	BBB	A	AA	AAA
Niveau de couverture		99%	99.75%	99.9%	99.95%	99.97%
Temps de retour	1 jour	100 jours	400 jours	4 ans	8 ans	13,3 ans
	10 jours	4 ans	16 ans	40 ans	80 ans	133 ans

La réglementation impose une procédure de backtesting journalière pour la valeur en risque de période de détention un jour. Cela veut dire qu'une couverture **effective** à 99% implique de dépasser la valeur en risque tous les 100 jours en moyenne. La procédure de contrôle ex-post des autorités de surveillance est basée sur cette propriété.

Reprenons l'exemple des défaillances. Dans le premier cas, le temps de retour est

$$\begin{aligned}\mathcal{T} &= \frac{1}{0.007} \times 1 \text{ an} \\ &= 143 \text{ ans}\end{aligned}$$

Dans le second cas, nous obtenons

$$\begin{aligned}\mathcal{T} &= \frac{1}{0.019} \times 3 \text{ ans} \\ &= 158 \text{ ans}\end{aligned}$$

Le premier événement ayant un temps de retour plus faible que le second événement, il est de ce fait "moins rare". D'où une nouvelle définition d'un événement rare :

Un événement rare est un événement dont le temps de retour est grand.

Nous voyons que la couverture devient aussi plus compréhensible. Couvrir des risques à 1 an, 10 ans ou 50 ans est beaucoup plus parlant que couvrir des risques avec un seuil de confiance de 99%, 99.5% ou 99.9%. Car dans cette seconde expression, il manque une dimension importante : la **période de détention** (ou **holding period**).

Dans la suite du cours, nous utiliserons donc plus volontiers le temps de retour pour caractériser la pertinence d'une mesure de risque. Voyons maintenant comment les événements rares ont une incidence importante sur la mesure de risque. Pour cela, considérons un portefeuille de lignes de crédit négociables (**Mark-to-market portfolio**). Soit $\Pi = (\pi_{i,j})$ la matrice de transition d'un état de la nature à un autre état de la nature. Par exemple, nous reportons dans la table 1.1 la matrice de transition S&P à un an (voir KAVVATHAS [1999]). Nous notons \mathcal{E} l'ensemble des huit états de la nature :

$$\mathcal{E} = \{\text{AAA, AA, A, BBB, BB, B, CCC, D}\}\tag{1.8}$$

(Π, \mathcal{E}) définit une chaîne de Markov à temps discret homogène. La probabilité de passer de l'état de nature i à l'état de nature j entre les dates t_1 et t_2 est donc donnée par la formule suivante :

$$\pi(t_1, i; t_2, j) = \mathbf{e}_i^\top \Pi^{(t_2 - t_1)} \mathbf{e}_j\tag{1.9}$$

Notation initiale	Notation finale							
	AAA	AA	A	BBB	BB	B	CCC	D
AAA	92.54	6.48	0.86	0.06	0.06	0.00	0.00	0.00
AA	0.63	91.87	6.64	0.65	0.06	0.11	0.04	0.00
A	0.08	2.26	91.66	5.11	0.61	0.23	0.01	0.04
BBB	0.05	0.27	5.84	87.74	4.74	0.98	0.16	0.22
BB	0.04	0.11	0.64	7.85	81.14	8.27	0.89	1.06
B	0.00	0.11	0.30	0.42	6.75	83.07	3.86	5.49
CCC	0.19	0.00	0.38	0.75	2.44	12.03	60.71	23.50
D	0	0	0	0	0	0	0	100

TABLE 1.1. Matrice de transition S&P d'horizon temporel 1 an (en %)

Nous pouvons alors construire les fonctions de survie correspondantes. Notons T_i la variable de survie sachant que la notation initiale à la date t_0 est l'état de la nature i . Nous avons

$$\begin{aligned} \mathbf{S}_i(t) &= \mathbb{E}[\mathbf{1}\{T_i \geq t\} \mid \mathcal{F}_{t_0}] \\ &= 1 - \pi(t_0, i; t, D) \end{aligned} \quad (1.10)$$

Dans le cadre d'une allocation de fonds propres dont l'horizon de couverture est l'année (et si les fonds propres ne servent à couvrir que le risque de défaillance et non le risque de dégradation de la créance), nous obtenons des temps de retour² très élevés :

Rating	CCC	B	BB	BBB	A	AA	AAA
Temps de retour	4 ans	18 ans	94 ans	455 ans	2500 ans		

Dans ce cas, devons-nous couvrir seulement les défaillances des lignes de créances notées CCC ou B? Si nous prenons les lignes de créances de façon individuelle, cela peut se comprendre. Néanmoins, un portefeuille de crédit d'une banque comprend des milliers (ou des millions) de lignes de créances. L'événement isolé qui nous semble rare peut présenter un risque important lorsque nous raisonnons de façon globale. C'est l'une des grandes problématiques de la gestion des risques financiers :



Nous rencontrons alors deux problèmes majeurs de modélisation :

1. la distribution des facteurs de risques individuels;
2. le choix de la dépendance entre les facteurs de risques individuels.

La notion de dépendance permet de passer du cadre unidimensionnel (gestion des risques individuels) au cadre multidimensionnel (gestion des risques multiples). Le choix de la dépendance est alors crucial et conditionne fortement la mesure de risque. Considérons par exemple une ligne de crédit notée CCC qui est garantie par un collatéral noté lui aussi CCC. Le temps de retour du défaut de la créance pour une couverture à un an est alors compris entre 4 ans et ∞ ! La rareté de cet événement dépend donc de la dépendance jointe.

²Nous avons

$$\mathcal{T}_i = \frac{1}{1 - \mathbf{S}_i(1)} \quad (1.11)$$

t (en années)	AAA	AA	A	BBB	BB	B	CCC
1	100.00	100.00	99.96	99.78	98.94	94.51	76.50
2	100.00	99.98	99.89	99.44	97.40	88.97	61.55
3	99.99	99.94	99.79	98.99	95.53	83.69	51.76
4	99.99	99.89	99.64	98.43	93.45	78.79	45.13
5	99.97	99.81	99.45	97.78	91.26	74.32	40.47
6	99.95	99.72	99.22	97.04	89.02	70.28	37.04
7	99.92	99.61	98.95	96.22	86.77	66.64	34.41
8	99.88	99.47	98.63	95.34	84.56	63.35	32.31
9	99.83	99.32	98.27	94.41	82.40	60.39	30.58
10	99.77	99.14	97.87	93.44	80.32	57.71	29.11
15	99.27	97.88	95.31	88.25	71.14	47.52	23.98
20	98.36	96.02	92.08	82.99	63.92	40.83	20.74
25	97.00	93.61	88.43	78.00	58.22	36.13	18.46
50	84.44	77.31	69.59	58.35	41.03	24.18	12.53
100	53.45	47.24	41.68	34.46	23.91	13.94	7.26

TABLE 1.2. Fonction de survie (en %)

1.2 L'hypothèse de normalité en finance

Dans cette section, nous poursuivons l'analyse des événements rares au regard de l'hypothèse de normalité qui est largement exploitée en finance. L'idée sous-jacente n'est pas de montrer que cette hypothèse n'est pas appropriée pour la gestion des risques, mais de comprendre les implications de celle-ci sur une mesure de risque.

1.2.1 La distribution gaussienne

Une distribution est gaussienne si l'expression de sa densité est

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x - \mu)^2}{2\sigma^2} \quad (1.12)$$

Nous notons alors $X \sim \mathcal{N}(\mu, \sigma)$ et nous avons les propriétés

$$\mathbb{E}[X] = \mu \quad (1.13)$$

et

$$\text{var}[X] = \sigma^2 \quad (1.14)$$

Lorsque μ et σ valent respectivement 0 et 1, nous parlons de distribution normale centrée et réduite $\mathcal{N}(0, 1)$, et nous notons Φ la fonction de répartition. Le lien entre les lois de probabilité $\mathcal{N}(\mu, \sigma)$ et $\mathcal{N}(0, 1)$ est donné par la relation suivante :

$$\mathcal{N}(\mu, \sigma) = \mu + \sigma\mathcal{N}(0, 1) \quad (1.15)$$

L'importance de la distribution gaussienne s'explique en partie par le fait que c'est la distribution naturelle des erreurs de mesure. En particulier, elle intervient dans le théorème de limite centrale. Soit X_1, \dots, X_n une séquence de variables aléatoires *iid* de moyenne μ et de variance $\sigma^2 < +\infty$. Nous avons

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad (1.16)$$

En finance, la distribution gaussienne est la distribution “classique” pour modéliser les rendements des actifs. Dans le modèle de Black et Scholes, le prix de l’actif $S(t)$ est un mouvement brownien géométrique dont la représentation SDE est

$$dS(t) = \mu S(t) dt + \sigma S(t) dW(t) \quad (1.17)$$

avec $W(t)$ un processus de Wiener (ou un mouvement brownien). Soit $r(t)$ le rendement logarithmique de l’actif. Nous avons

$$r(t) = \ln S(t) \quad (1.18)$$

et

$$dr(t) = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW(t) \quad (1.19)$$

La solution explicite de cette équation différentielle stochastique est un processus de diffusion

$$r(t) | \mathcal{F}_{t_0} = r(t_0) + \left(\mu - \frac{1}{2}\sigma^2 \right) (t - t_0) + \sigma (W(t) - W(t_0)) \quad (1.20)$$

dont le distribution est gaussienne :

$$\Pr \{r(t) | \mathcal{F}_{t_0} \leq r\} = \Phi \left(\frac{r - r(t_0) - (\mu - \frac{1}{2}\sigma^2) (t - t_0)}{\sigma \sqrt{t - t_0}} \right) \quad (1.21)$$

Ce modèle est à la base de toute la théorie financière moderne. C’est un modèle qui fonctionne pas trop mal et il a le mérite d’être une référence ou un point de comparaison.

Nous considérons les indices CAC 40 et Dow Jones. Nous avons représenté leur évolution sur les graphiques 1.1 et 1.2. Soit $\tilde{r}(t)$ le rendement journalier “standardisé”, c’est-à-dire

$$\tilde{r}(t) = \frac{r(t) - \hat{\mu}}{\hat{\sigma}} \quad (1.22)$$

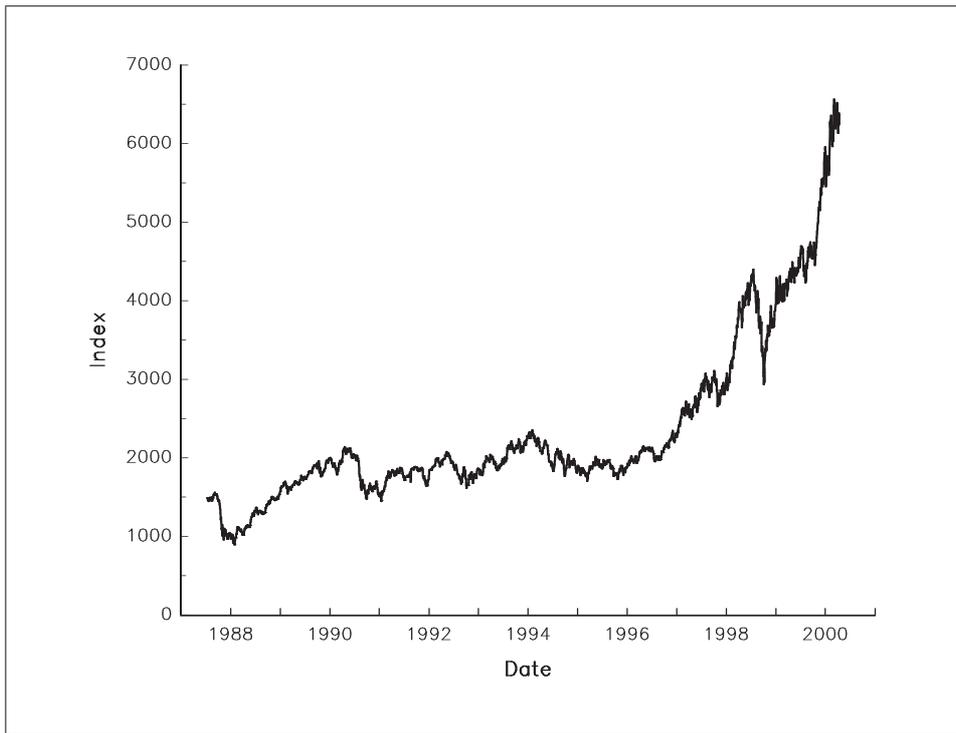
avec

$$r(t) = \ln \frac{S(t)}{S(t-1)}$$

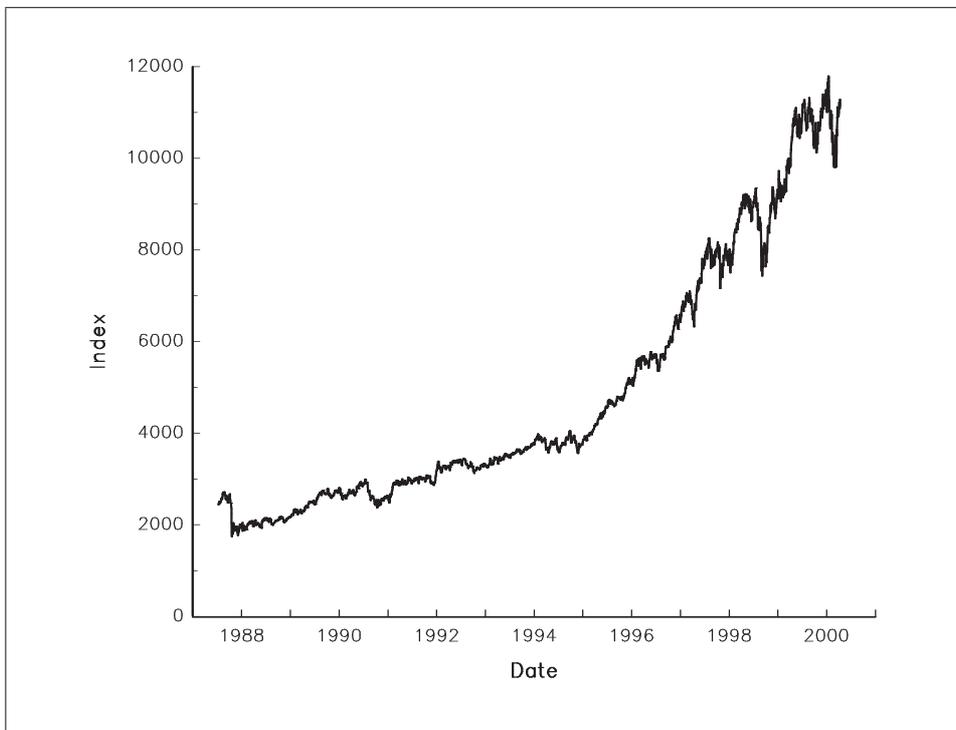
$\hat{\mu}$ et $\hat{\sigma}$ sont les estimateurs ML de la moyenne et de l’écart-type de $r(t)$. Il est naturel de postuler que

$$\tilde{r}(t) \sim \mathcal{N}(0, 1) \quad (1.23)$$

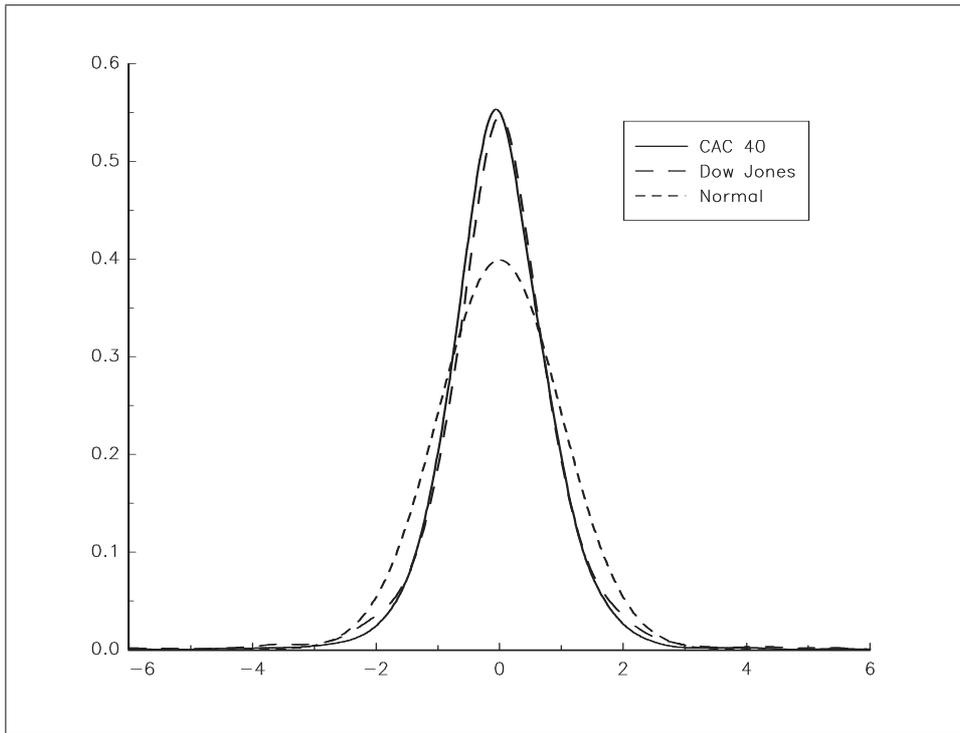
Si nous comparons l’estimation non paramétrique de la densité de $\tilde{r}(t)$ avec la densité gaussienne, nous remarquons que celles-ci sont relativement éloignées (voir le graphique 1.3). En particulier, les rendements journaliers standardisés présentent des queues beaucoup plus épaisses (voir le graphique 1.4). D’ailleurs, si nous calculons les temps de retour des événements les plus extrêmes, nous obtenons des valeurs peu plausibles. **Des temps de retour de plusieurs millions d’année, voire de plusieurs milliards de milliards d’années, sont difficilement acceptables.** Ils correspondent à des événements trop rares (au regard de l’hypothèse gaussienne), alors que ceux-ci se sont bien produits. Bien sûr, un événement extrêmement rare



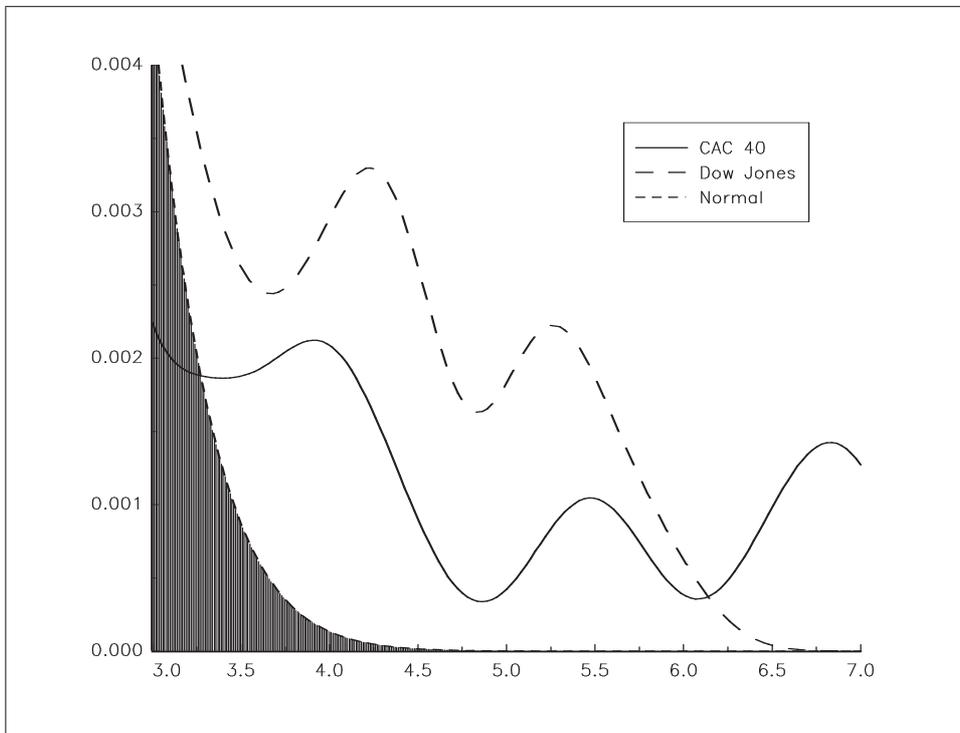
Graphique 1.1. Evolution de l'indice CAC 40



Graphique 1.2. Evolution de l'indice Dow Jones



Graphique 1.3. Comparaison des distributions non paramétriques des rendements avec la distribution gaussienne



Graphique 1.4. Comparaison des queues de distribution

	Date	$r(t)$ (en %)	Temps de retour	Date	$r(t)$ (en %)	Temps de retour
#1	1994/05/12	-16.72	6.4×10^{24}	1994/05/10	16.15	3.8×10^{22}
#2	1994/04/14	-12.33	7.8×10^{12}	1994/05/18	13.63	4.5×10^{15}
#3	1994/05/05	-11.63	2.1×10^{11}	1994/04/19	12.81	4.3×10^{13}
#4	1992/01/02	-9.18	3.2×10^6	1992/03/03	12.34	3.6×10^{12}
#5	1987/10/28	-9.11	2.4×10^6	1994/04/12	10.86	2.4×10^9
#6	1992/01/08	-6.94	1094	1992/02/20	10.62	7.6×10^8
#7	1987/10/13	-6.80	719	1994/04/06	10.15	9.3×10^7
#8	1994/06/09	-5.99	69	1994/07/19	8.71	2.7×10^5
#9	1998/09/17	-5.63	27	1992/01/01	8.28	57294
#10	1994/04/01	-5.56	22	1991/01/17	6.81	452

TABLE 1.3. Temps de retour (en années) des rendements les plus extrêmes du CAC 40

peut toujours se réaliser, mais nous avons du mal à concevoir qu'un nombre important de tels événements se soit réalisés pendant ces dix dernières années.

La notion de queue épaisse est souvent reliée en finance au caractère leptokurtique de la distribution de probabilité (d'ailleurs de façon impropre, car les deux notions sont différentes). La kurtosis d'une variable aléatoire X se définit comme le rapport du moment centré d'ordre 4 sur le carré du moment centré d'ordre 2 :

$$\kappa = \frac{\mathbb{E} \left[(X - \mu)^4 \right]}{\mathbb{E}^2 \left[(X - \mu)^2 \right]} \quad (1.24)$$

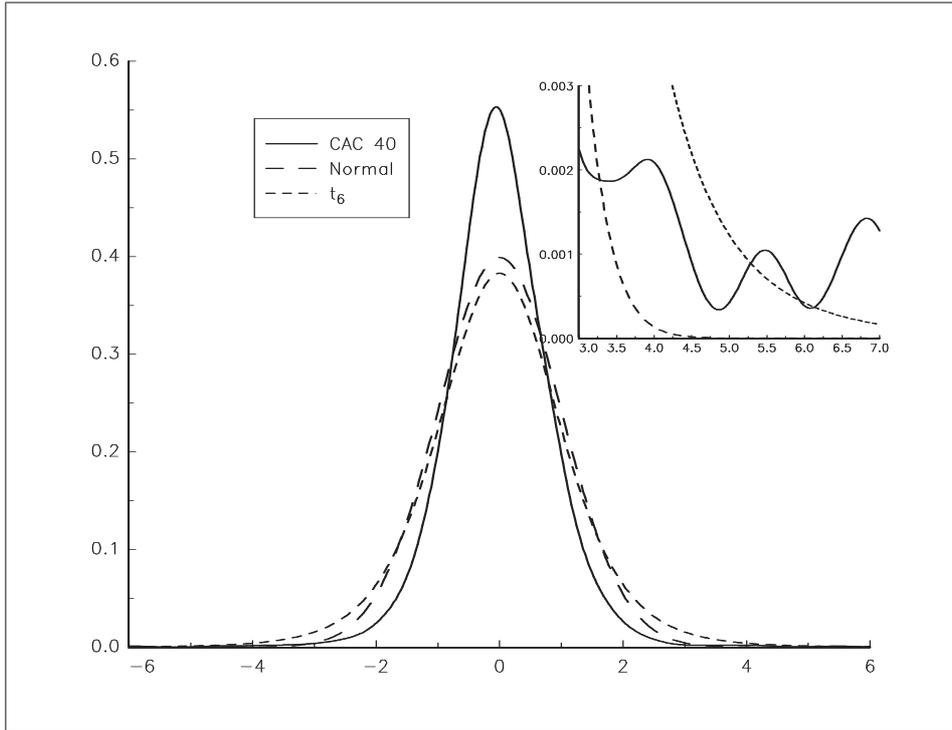
Pour la distribution gaussienne, κ est égal à 3. Lorsque $\kappa \geq 3$, la distribution est dite leptokurtique et présente (souvent) des queues épaisses. C'est par exemple le cas de la distribution de student t_ν . Considérons maintenant l'hypothèse

$$\tilde{r}(t) \sim t_\nu \quad (1.25)$$

Sur le graphique 1.5, nous représentons la densité de t_6 . Nous remarquons qu'elle est assez éloignée de la distribution non paramétrique des rendements standardisés du CAC 40. Néanmoins, par rapport à la distribution gaussienne, elle présente des queues plus proches de celles observées. Aussi les temps de retour des rendements extrêmes sont plus acceptables (il suffit de comparer les résultats de la table 1.4 avec ceux de la table 1.3).

	Date	$r(t)$ (en %)	Temps de retour	Date	$r(t)$ (en %)	Temps de retour
#1	1994/05/12	-16.72	157.13	1994/05/10	16.15	121.79
#2	1994/04/14	-12.33	28.27	1994/05/18	13.63	46.24
#3	1994/05/05	-11.63	20.53	1994/04/19	12.81	32.52
#4	1992/01/02	-9.18	5.79	1992/03/03	12.34	26.45
#5	1987/10/28	-9.11	5.56	1994/04/12	10.86	13.08
#6	1992/01/08	-6.94	1.43	1992/02/20	10.62	11.53
#7	1987/10/13	-6.80	1.30	1994/04/06	10.15	9.02
#8	1994/06/09	-5.99	0.72	1994/07/19	8.71	4.01
#9	1998/09/17	-5.63	0.55	1992/01/01	8.28	3.10
#10	1994/04/01	-5.56	0.52	1991/01/17	6.81	1.17

TABLE 1.4. Temps de retour (en années) des rendements les plus extrêmes du CAC 40 lorsque la distribution paramétrique est t_6



Graphique 1.5. Modélisation du rendement du CAC 40 avec une distribution t_6

Le fait que la distribution gaussienne modélise assez mal les temps de retour des rendements extrêmes se voit assez facilement si nous considérons les rendements implicites à un temps de retour donné. Nous avons

$$\mathcal{T} = \frac{1}{1 - \mathbf{F}(r)} \quad (1.26)$$

ou encore

$$r = \mathbf{F}^{-1} \left(\frac{\mathcal{T} - 1}{\mathcal{T}} \right) \quad (1.27)$$

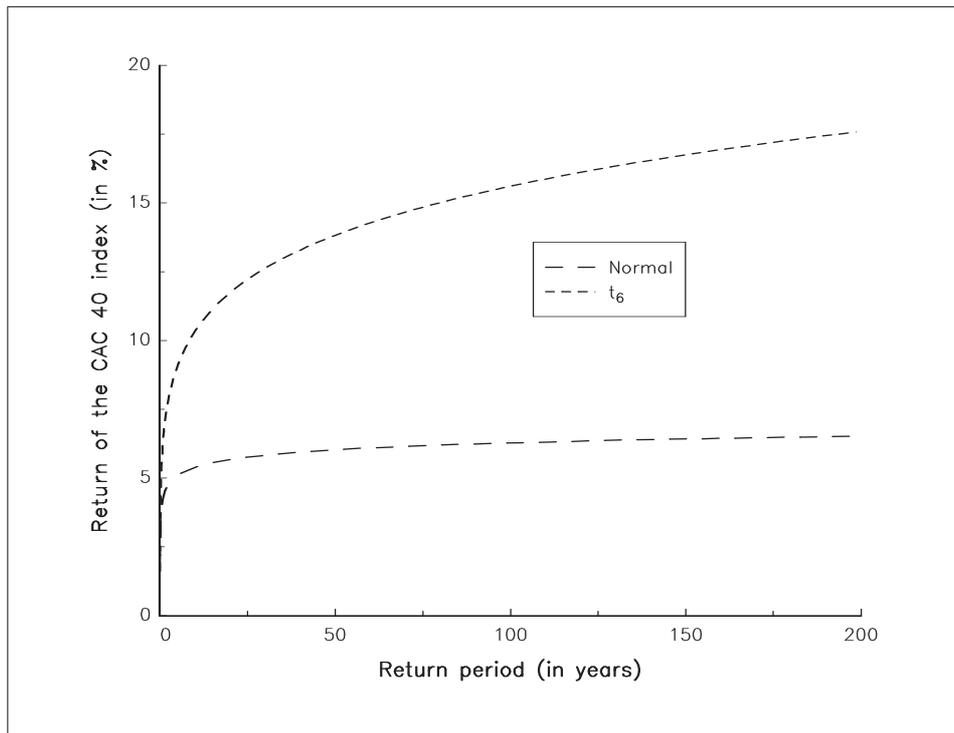
Sur le graphique 1.6, nous remarquons que la distribution gaussienne présente des rendements implicites trop faibles par rapport à ceux observés dans la réalité, ce qui n'est pas le cas de la distribution t_6 . Notons aussi que nous pouvons construire des rendements implicites non paramétriques avec la formule

$$r = \hat{\mathbf{F}}^{-1} \left(\frac{\mathcal{T} - 1}{\mathcal{T}} \right) \quad (1.28)$$

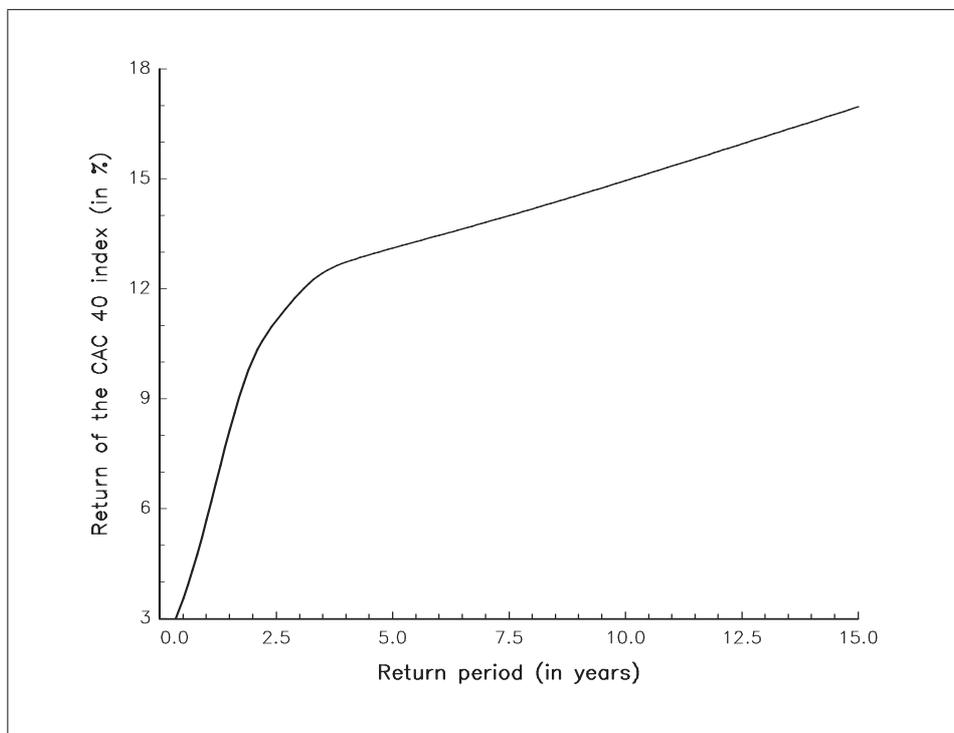
Le graphique 1.7 présente l'estimation non paramétrique des rendements implicites du CAC 40. Nous constatons qu'effectivement les temps de retour sont beaucoup plus faibles.

Remarque 3 *L'exemple de ce paragraphe ne montre pas que la distribution gaussienne n'est pas appropriée pour modéliser les prix des actifs. Il indique seulement que dans un contexte de gestion des risques, elle peut conduire à une sous-évaluation importante de la mesure des risques.*

Remarque 4 *Le problème de la modélisation ne peut se poser sans faire référence au problème donné. Dans la cadre de la gestion des risques, celui-ci concerne l'estimation d'un quantile relativement élevé. Nous pouvons très bien modéliser le centre de la distribution, mais cela n'a aucun intérêt pour définir la mesure de risque. Les procédures statistiques classiques (par exemple, les tests d'adéquation de type Kolmogorov-Smirnov) ont alors très peu de pertinence pour le choix du bon modèle.*



Graphique 1.6. Rendement implicite du CAC 40 pour un temps de retour donné



Graphique 1.7. Rendement implicite du CAC 40 pour un temps de retour donné (estimation non paramétrique)

1.2.2 La dépendance normale

Dans le paragraphe précédent, nous n'avons considéré qu'une seule direction de risque. Dans le cas de plusieurs directions de risque, il est important de prendre en compte la dépendance entre celles-ci. Une première solution est de les considérer indépendantes :

$$\mathbf{F}(x_1, \dots, x_n) = \prod_{i=1}^n \mathbf{F}_i(x_i) \quad (1.29)$$

Si les facteurs de risque ne sont pas indépendants, on peut toujours les rendre indépendants grâce à une analyse factorielle (sous une certaine condition, que nous allons expliciter). La solution la plus utilisée en finance consiste à postuler que la distribution des facteurs de risque X_1, \dots, X_n est une distribution normale multidimensionnelle. Soit le vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$. Nous avons

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (1.30)$$

On dit alors que \mathbf{X} est un vecteur gaussien de moyenne $\boldsymbol{\mu}$ et de matrice de covariance Σ .

Remarque 5 *Le fait que chaque composante du vecteur aléatoire \mathbf{X} est gaussienne n'est pas une condition suffisante pour que \mathbf{X} soit un vecteur aléatoire gaussien. Il faut **aussi** que toute combinaison **linéaire** de X_1, \dots, X_n soit gaussienne.*

L'hypothèse de **normalité jointe** est donc une hypothèse particulièrement forte. Par exemple, lorsque nous faisons la régression **linéaire** $Y = X\beta + u$, nous faisons l'hypothèse que la distribution de u est gaussienne pour mener à bien l'inférence statistique. Derrière cette hypothèse se cache donc l'hypothèse que le vecteur aléatoire (X, Y) est gaussien (hypothèse difficile à vérifier dans de nombreux travaux économétriques, ce qui explique que les économètres préfèrent tester la normalité des résidus...). Considérons maintenant la régression sur les variables centrées et réduites. Nous avons

$$\tilde{Y} = \tilde{X}\beta + u \quad (1.31)$$

La solution de ce problème est donnée par la corrélation entre X et Y :

$$\beta = \rho(X, Y) \quad (1.32)$$

Cette corrélation est une mesure de la dépendance entre X et Y puisqu'elle indique comment décomposer la variable aléatoire standard \tilde{Y} en deux variables aléatoires indépendantes \tilde{X} et u :

$$\tilde{Y} = \tilde{X}\rho(X, Y) + u \quad \text{avec} \quad \tilde{X} \perp u \quad (1.33)$$

Il existe d'autres mesures de corrélation, mais celle-ci est la plus fréquemment utilisée. Le nom exact de $\rho(X, Y)$ est coefficient de corrélation de Pearson ou coefficient de corrélation **linéaire**.

Nous remarquons que le caractère **linéaire** est systématiquement associé au caractère gaussien d'un vecteur aléatoire. C'est pourquoi on parle d'un monde linéaire lorsque nous modélisons des variables financières en utilisant la distribution normale. Cela simplifie énormément les calculs puisqu'on se ramène à des calculs sur la distribution gaussienne unidimensionnelle. Par exemple, nous avons

$$\mathbf{a}^\top \mathbf{X} \sim \mathcal{N}\left(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \Sigma \mathbf{a}\right) \quad (1.34)$$

et

$$\text{VaR}(\mathbf{a}^\top \mathbf{X}; \alpha) = \mathbf{a}^\top \boldsymbol{\mu} + \Phi^{-1}(\alpha) \sqrt{\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}} \quad (1.35)$$

De même, la corrélation est associée au caractère linéaire de la régression, et non au caractère gaussien du vecteur aléatoire. Elle n'a donc de sens que dans un monde linéaire ou gaussien.

Nous avons vu dans le paragraphe précédent qu'une modélisation gaussienne n'était pas la mieux adaptée pour calculer une mesure de risque dans un cadre unidimensionnel. Nous allons donc retrouver les mêmes problèmes dans le cas multivarié, puisque les marginales de la distribution multidimensionnelle sont gaussiennes. Au delà de ces limites, quelles sont les autres implications de l'hypothèse de normalité jointe sur la mesure de risque? Considérons la probabilité conditionnelle que nous dépassons la valeur en risque dans une direction sachant que nous l'avons déjà dépassée dans l'autre direction :

$$\begin{aligned} \lambda(\alpha) &= \Pr\{X_2 > \text{VaR}_2(\alpha) \mid X_1 > \text{VaR}_1(\alpha)\} \\ &= \Pr\{X_2 > \mathbf{F}_2^{-1}(\alpha) \mid X_1 > \mathbf{F}_1^{-1}(\alpha)\} \\ &= \frac{\Pr\{\mathbf{F}_2(X_2) > \alpha, \mathbf{F}_1(X_1) > \alpha\}}{\Pr\{\mathbf{F}_1(X_1) > \alpha\}} \\ &= \frac{1 - 2\alpha + \mathbf{F}(\mathbf{F}_1^{-1}(\alpha), \mathbf{F}_2^{-1}(\alpha))}{1 - \alpha} \end{aligned} \quad (1.36)$$

Puisque $\lambda(\alpha)$ est une probabilité, nous avons bien sûr

$$0 \leq \lambda(\alpha) \leq 1 \quad (1.37)$$

La limite, si elle existe, est notée λ :

$$\lambda = \lim_{\alpha \rightarrow 1^+} \lambda(\alpha) \quad (1.38)$$

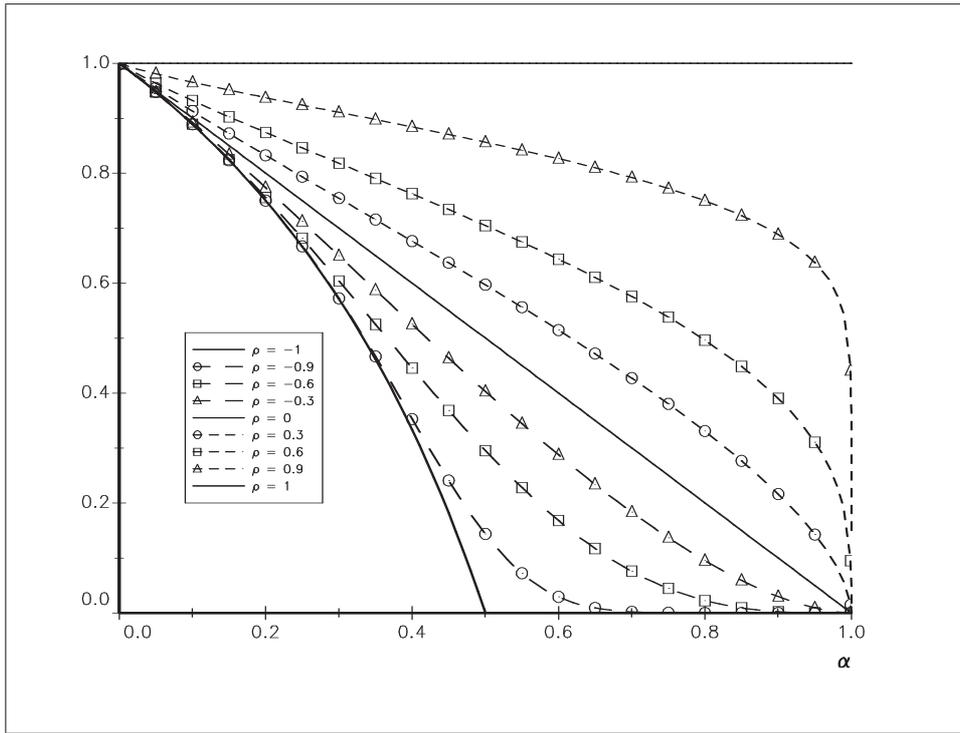
C'est la probabilité d'avoir un extrême dans une direction sachant que nous avons déjà un extrême dans l'autre direction. Si λ est égal à 0, alors les extrêmes sont indépendants, et si λ est égal à 1, les extrêmes sont parfaitement "corrélés". λ est donc une "mesure de corrélation" des extrêmes. C'est le **coefficient de dépendance de queue**. Pour la distribution normale multidimensionnelle, λ vaut systématiquement 0 sauf lorsque la corrélation est égale à 1 (voir le graphique 1.8). La distribution normale ne présente pas de dépendance de queue, ce qui n'est pas le cas de la distribution de Student multidimensionnelle (voir le graphique 1.9).

Les extrêmes d'une distribution normale sont donc indépendants. Cette propriété pose problème dans un cadre de gestion des risques. En effet, il est reconnu que les situations de *bear markets* ou de *bull markets* présentent justement des dépendances très fortes. Si le Dow Jones chute de 5%, une baisse du CAC 40 est fortement probable. Nous pouvons étudier les extrêmes d'une distribution normale à partir de l'analyse des régions de confiance — la terminologie anglaise est *confidence regions analysis* (DAVIDSON and MACKINNON [1993]). La région de confiance est l'ensemble des points \mathbf{x} tels que

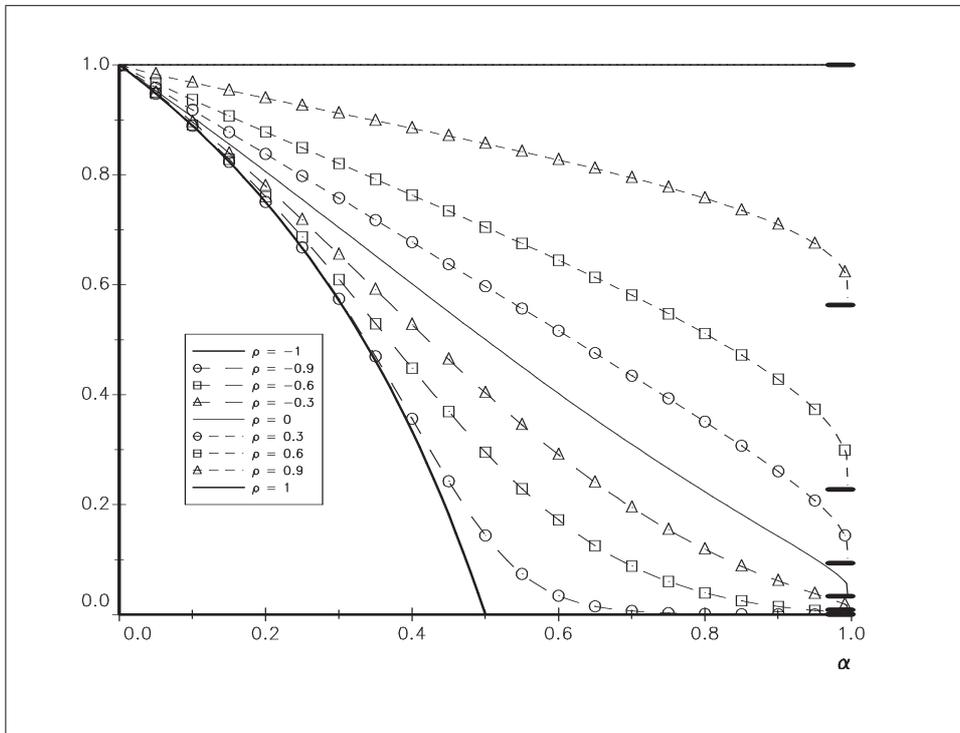
$$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) < c_\alpha \quad (1.39)$$

avec $c_\alpha = \chi_n^{-1}(\alpha)$ — pour être juste, la distribution de la forme quadratique est une combinaison linéaire de chi-deux non centrés. Dans le cas bivarié, les lignes de niveau de probabilité c_α dans le plan (x_1, x_2) sont des ellipses d'équation :

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = c_\alpha (1 - \rho^2) \quad (1.40)$$

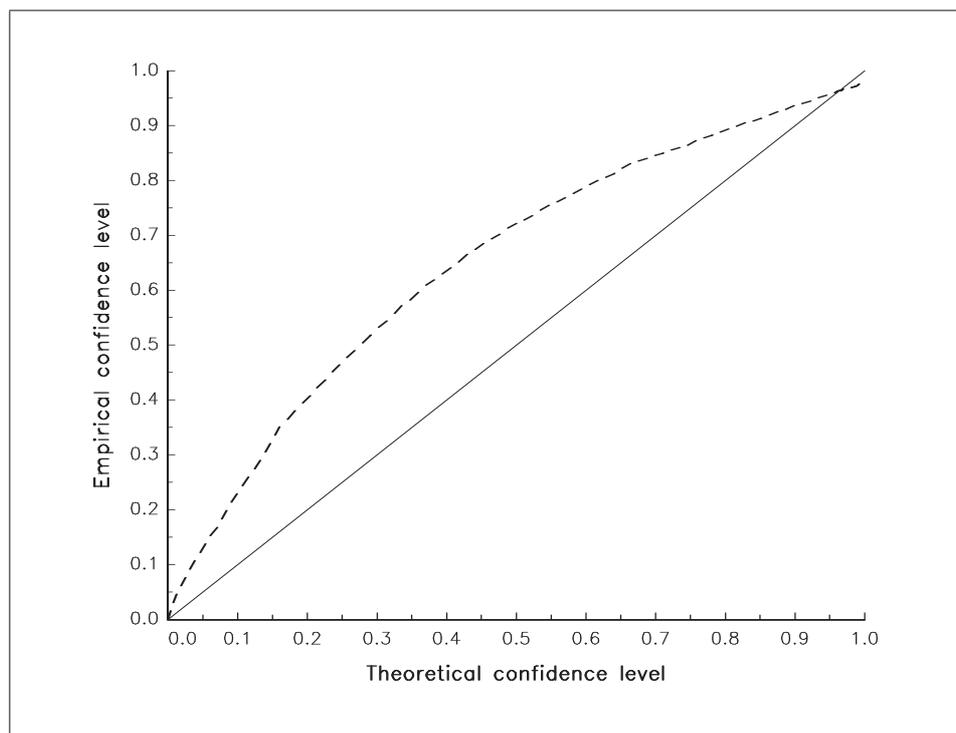


Graphique 1.8. Dépendance de queue de la distribution normale



Graphique 1.9. Dépendance de queue de la distribution de Student à 6 degrés de liberté

Le graphique 1.10 présente une analyse quantile-quantile de l'ellipse de covariance. Nous remarquons que pour $\alpha < 95\%$, la distribution normale sous-estime le niveau de confiance. Pour des quantiles élevés, ce niveau de confiance est sur-estimé. Par exemple, les observations hors de l'ellipse de covariance 99% représentent 2.5% du nombre total d'observations (voir le graphique 1.11). Cette sous-estimation des extrêmes est d'autant plus évidente que la dimension de la distribution est grande. Par exemple, nous reportons le graphique 1.12 de [2] qui compare les QQ-plot de l'ellipse de covariance pour deux actifs du LME (**L**ondon **M**etal **E**xchange) et pour cinq actifs du LME.



Graphique 1.10. QQ-plot de l'ellipse de covariance

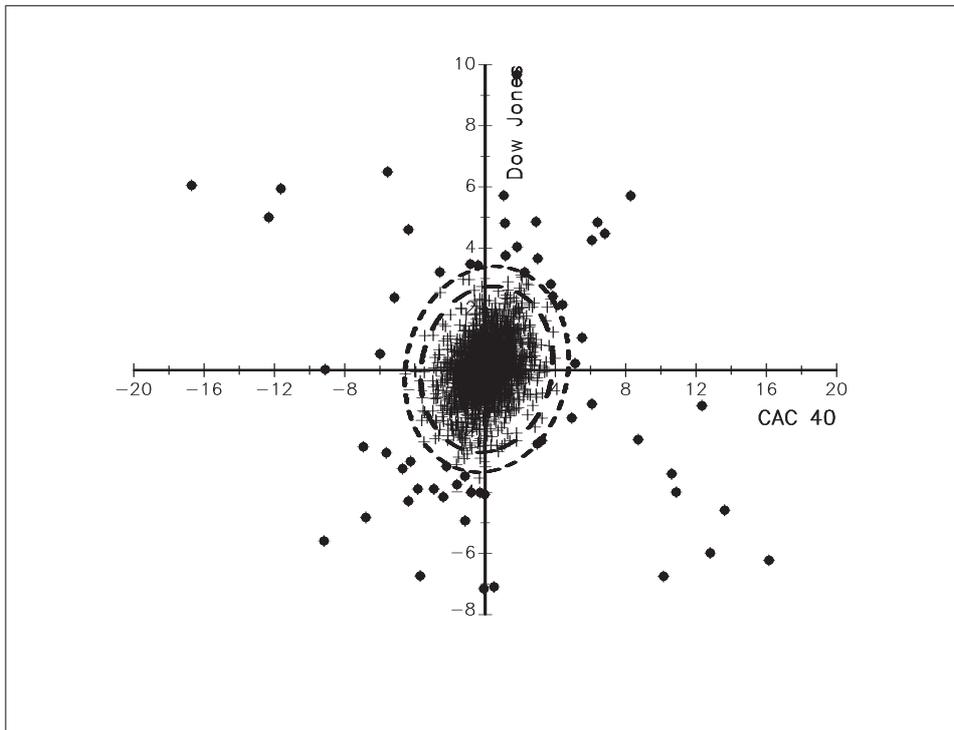
Remarque 6 *Sur le graphique 1.11, nous avons l'impression que les extrêmes ne sont pas corrélés. Ceci s'explique par le fonctionnement psychologique des marchés. Une forte baisse à New York entraîne une forte baisse à Paris. Le lendemain, New York connaît souvent un phénomène correcteur tout comme Paris. Le décalage des ouvertures des bourses explique que les baisses à New York et à Paris ne sont pas forcément enregistrées le même jour calendaire.*

1.3 Conclusion

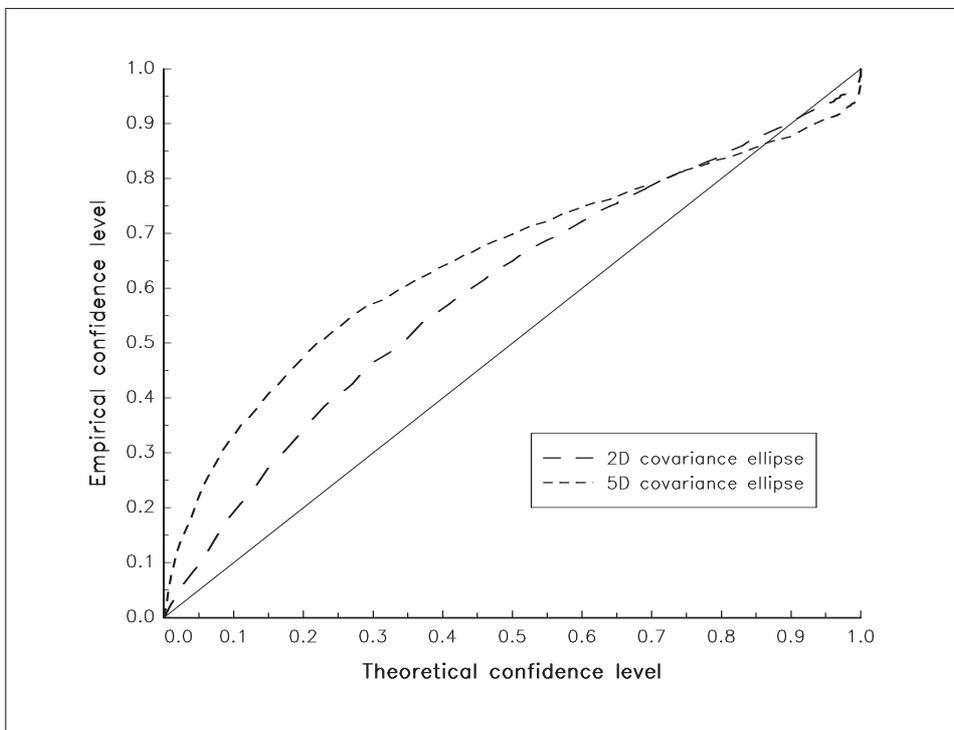
Pour définir une mesure de risque (extrême), il est important de bien modéliser les extrêmes (queue de distribution) et la dépendance entre ces extrêmes (dépendance de queue).

Références

- [1] BOULEAU, N [1991], Splendeurs et misères des lois de valeurs extrêmes, *Risques*, **3**, 85-92
- [2] BOUYÉ, E., V. DURRLEMAN, A. NIKEGHBALI, G. RIBOULET et T. RONCALLI [2000], Copulas for finance : a reading guide and some applications, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*



Graphique 1.11. Ellipses de covariance 95% et 99%



Graphique 1.12. QQ-plot de l'ellipse de covariance (2 et 5 actifs du LME)

- [3] COLES, S. [2001], An Introduction to Statistical Modeling of Extremes Values, *Springer Series in Statistics*, Springer-Verlag, Londres
- [4] GUMBEL, E.J. [1954], Statistical Theory of Extreme Values and Some Pratical Applications, *Applied Mathematics Series*, **33**, National Bureau of Standards, Washington
- [5] GUMBEL, E.J. [1958], Statistics of Extremes, Columbia University Press, Columbia
- [6] KAVVATHAS, D. [1999], Estimating credit transition probabilities for corporate bonds, University of Chicago, *Working Paper*
- [7] RONCALLI, T. [2001], Introduction à la gestion des risques, Cours ENSAI de 3^{ème} année, [http ://gro.creditlyonnais.fr/content/wp/gdr.zip](http://gro.creditlyonnais.fr/content/wp/gdr.zip)

2

La théorie des valeurs extrêmes

Dans ce chapitre, je vais présenter de façon très classique la théorie des extrêmes. Cette présentation s’inspire très largement des textes de COLES [2001], EMBRECHTS, KLÜPPELBERG et MIKOSCH [1997] et KOTZ et NADARAJAH [2000]. Néanmoins, je n’aborderai pas immédiatement la théorie telle qu’elle est présentée dans ces livres. En effet, mon point de départ sera les statistiques d’ordre. Celles-ci sont très peu enseignées dans les cours de gestion des risques, alors qu’elles sont pour moi un outil essentiel de modélisation des risques. En particulier, elles sont indispensables pour construire une valeur en risque pertinente pour le risque opérationnel (FRACHOT, GEORGES et RONCALLI [2001]).

2.1 Les statistiques d’ordre

Soient n variables aléatoires *iid* X_1, \dots, X_n . Rangeons ces variables aléatoires par “ordre croissant de grandeur”. Pour cela, nous introduisons la notation $X_{i:n}$ avec

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n-1:n} \leq X_{n:n} \quad (2.1)$$

$X_{i:n}$ est donc la i -ième statistique d’ordre (ou statistique d’ordre i) dans un échantillon de taille n . Soient n observations x_1, \dots, x_n . La valeur observée de $X_{i:n}$ est notée $x_{i:n}$ et nous avons

$$x_{i:n} = \mathbf{e}_i^\top \text{sort}(\mathbf{x}) \quad (2.2)$$

où sort est la fonction de tri ascendant et \mathbf{x} le vecteur des observations x_1, \dots, x_n . Deux statistiques d’ordre sont particulièrement intéressantes pour l’étude des événements extrêmes. Ce sont les statistiques d’ordre extrême qui correspondent à la plus petite statistique d’ordre $X_{1:n}$ (ou statistique du minimum)

$$X_{1:n} = \min(X_1, \dots, X_n) \quad (2.3)$$

et à la plus grande statistique d’ordre $X_{n:n}$ (ou statistique du maximum)

$$X_{n:n} = \max(X_1, \dots, X_n) \quad (2.4)$$

Dans le chapitre précédent, nous les avons notées χ_n^- et χ_n^+ . Enfin, nous introduisons l’écart des statistiques d’ordre $W_{i,j:n} = X_{j:n} - X_{i:n}$ ($i < j$). L’écart $W = W_{1,n:n} = X_{n:n} - X_{1:n}$ est appelée la déviation extrême.

Remarque 7 Même si les variables aléatoires X_1, \dots, X_n sont indépendantes, les statistiques d'ordre ne sont pas indépendantes (par définition).

DAVID [1970] et BALAKRISHNAN et CLIFFORD COHEN [1991] montrent que l'expression de la distribution de $X_{i:n}$ est

$$\begin{aligned} \mathbf{F}_{i:n}(x) &= \Pr\{X_{i:n} \leq x\} \\ &= \sum_{r=i}^n \binom{n}{r} [\mathbf{F}(x)]^r [1 - \mathbf{F}(x)]^{n-r} \end{aligned} \quad (2.5)$$

Nous en déduisons que la fonction de densité est

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} [\mathbf{F}(x)]^{i-1} [1 - \mathbf{F}(x)]^{n-i} f(x) \quad (2.6)$$

Pour les statistiques d'ordre extrême, nous obtenons des expressions très simples. En effet, nous avons

$$\mathbf{F}_{1:n}(x) = 1 - [1 - \mathbf{F}(x)]^n \quad (2.7)$$

et

$$f_{1:n}(x) = n [1 - \mathbf{F}(x)]^{n-1} f(x) \quad (2.8)$$

pour la statistique du minimum et

$$\mathbf{F}_{n:n}(x) = [\mathbf{F}(x)]^n \quad (2.9)$$

et

$$f_{n:n}(x) = n [\mathbf{F}(x)]^{n-1} f(x) \quad (2.10)$$

pour la statistique du maximum.

Remarque 8 Les expressions de $\mathbf{F}_{1:n}$ et $\mathbf{F}_{n:n}$ peuvent s'obtenir très facilement en considérant les relations

$$\begin{aligned} \{X_{1:n} \geq x\} &\Leftrightarrow \{\min(X_1, \dots, X_n) \geq x\} \\ &\Leftrightarrow \bigcap_{i=1}^n \{X_i \geq x\} \end{aligned} \quad (2.11)$$

et

$$\begin{aligned} \{X_{n:n} \leq x\} &\Leftrightarrow \{\max(X_1, \dots, X_n) \leq x\} \\ &\Leftrightarrow \bigcap_{i=1}^n \{X_i \leq x\} \end{aligned} \quad (2.12)$$

En utilisant la propriété d'indépendance des variables aléatoires X_1, \dots, X_n , nous en déduisons que

$$\begin{aligned}
\mathbf{F}_{1:n}(x) &= \Pr\{X_{1:n} \leq x\} \\
&= 1 - \Pr\{X_{1:n} > x\} \\
&= 1 - \Pr\left\{\bigcap_{i=1}^n \{X_i > x\}\right\} \\
&= 1 - \prod_{i=1}^n \Pr\{X_i > x\} \\
&= 1 - \prod_{i=1}^n [1 - \Pr\{X_i \leq x\}] \\
&= 1 - [1 - \mathbf{F}(x)]^n
\end{aligned} \tag{2.13}$$

et

$$\begin{aligned}
\mathbf{F}_{n:n}(x) &= \Pr\{X_{n:n} \leq x\} \\
&= \Pr\left\{\bigcap_{i=1}^n \{X_i \leq x\}\right\} \\
&= \prod_{i=1}^n \Pr\{X_i \leq x\} \\
&= [\mathbf{F}(x)]^n
\end{aligned} \tag{2.14}$$

Considérons quelques exemples lorsque \mathbf{F} est la distribution gaussienne standard. Le graphique 2.1 montre comment évolue la densité de la statistique du maximum $X_{n:n}$ lorsque n varie. La distribution correspondante $\mathbf{F}_{n:n}$ est représentée sur le graphique 2.2. Nous vérifions l'ordre stochastique

$$n > m \Rightarrow X_{n:n} \geq X_{m:m} \tag{2.15}$$

qui est équivalent à $\mathbf{F}_{n:n} \succ \mathbf{F}_{m:m}$ (c'est-à-dire que $\mathbf{F}_{n:n}(x) \geq \mathbf{F}_{m:m}(x)$ pour tout $x \in \mathbb{R}$). Les graphiques 2.3 et 2.4 concernent la statistique d'ordre $X_{i:5}$ pour $i = 1, \dots, 5$.

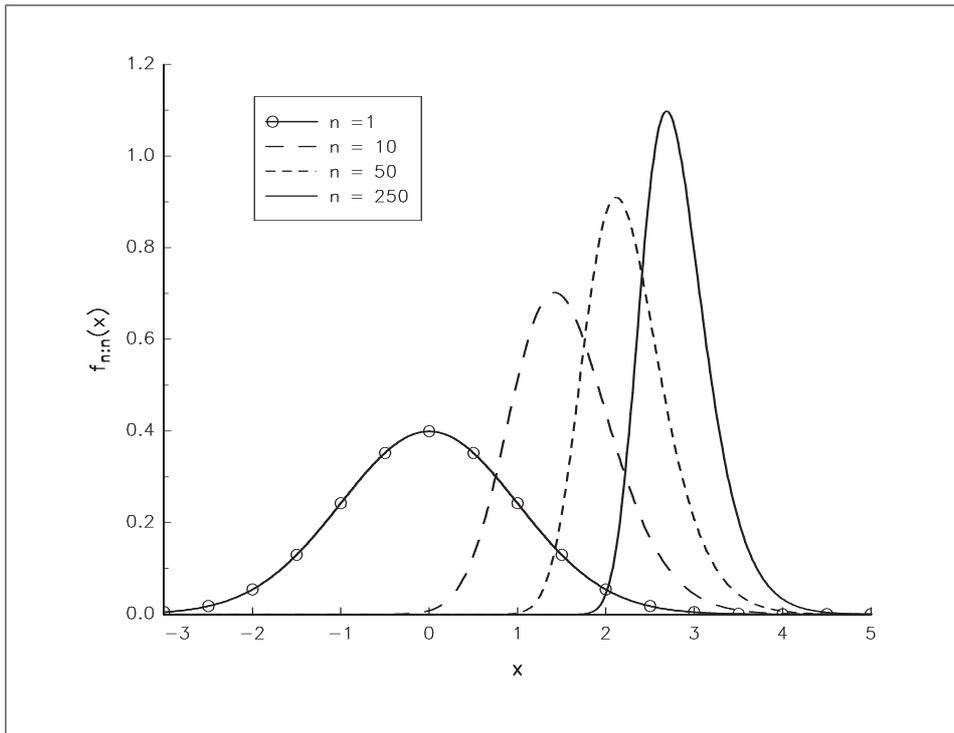
Voyons maintenant l'influence des queues de distribution sur les statistiques d'ordre. Reprenons l'exemple du CAC 40. Nous supposons que les rendements standardisés sont gaussiens

$$\mathbf{F}(r(t)) = \Phi\left(\frac{r(t) - \hat{\mu}}{\hat{\sigma}}\right) \tag{2.16}$$

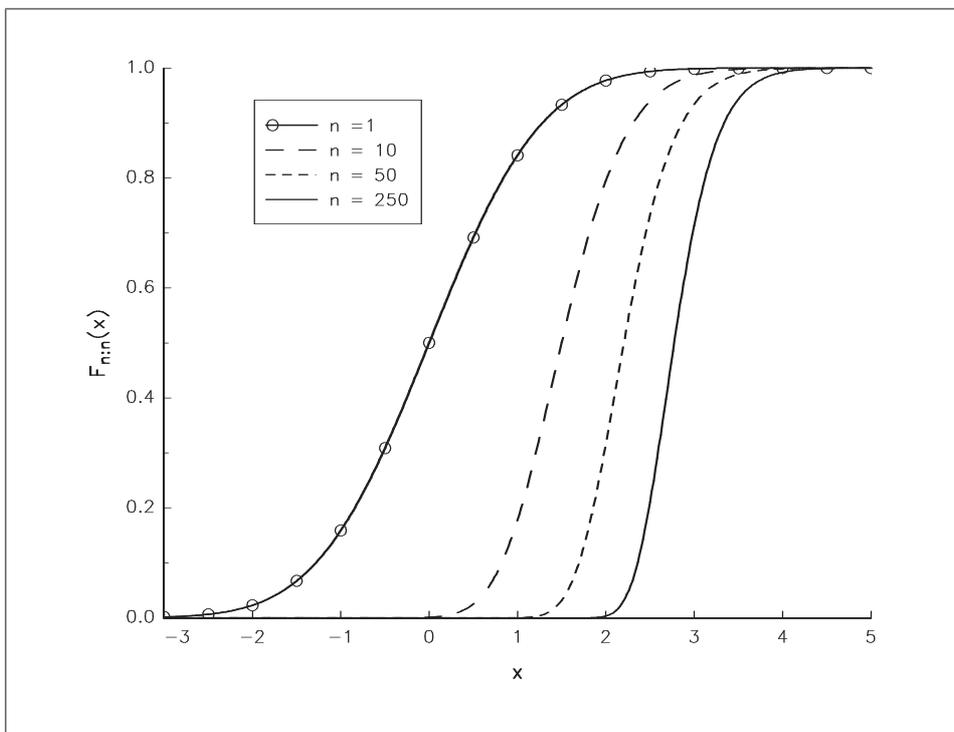
ou suivent une distribution de Student

$$\mathbf{F}(r(t)) = \mathbf{t}_\nu\left(\frac{r(t) - \hat{\mu}}{\hat{\sigma}\sqrt{\frac{\nu-2}{\nu}}}\right) \tag{2.17}$$

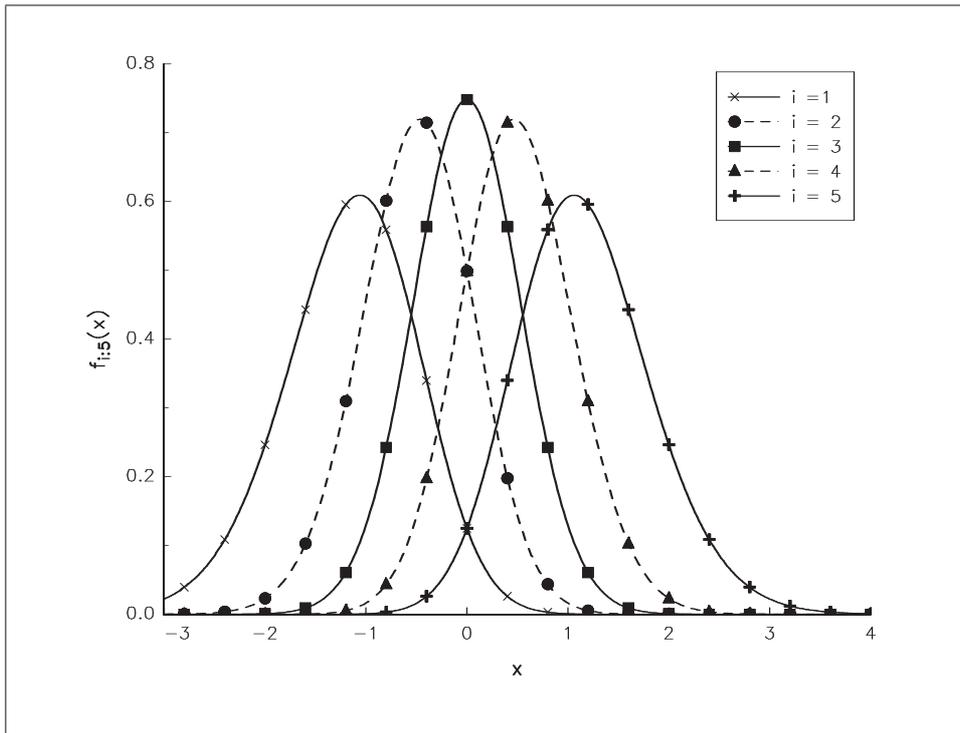
Dans cette dernière expression, nous corrigeons l'écart-type estimé $\hat{\sigma}$ par le facteur $\sqrt{\frac{\nu-2}{\nu}}$ afin que la variance de $r(t)$ soit bien égale à $\hat{\sigma}^2$ — la variance de la distribution \mathbf{t}_ν est $\nu/(\nu-2)$.



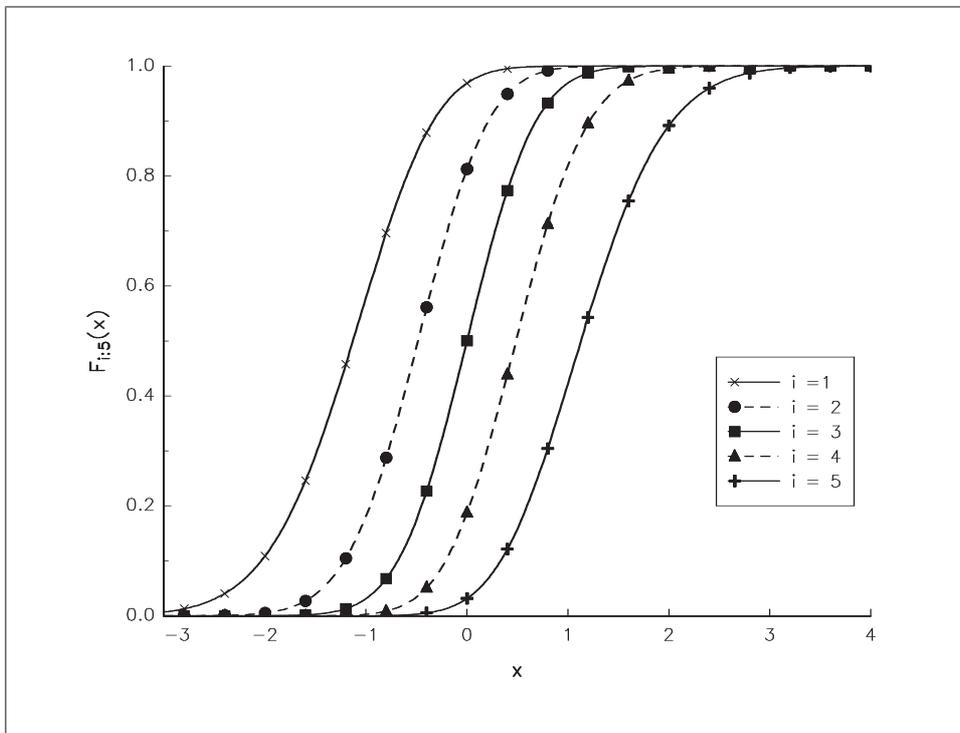
Graphique 2.1. Fonction de densité $f_{n:n}(x)$ lorsque X est une variable aléatoire gaussienne



Graphique 2.2. Fonction de probabilité $F_{n:n}(x)$ lorsque X est une variable aléatoire gaussienne



Graphique 2.3. Fonction de densité $f_{i:5}(x)$ lorsque X est une variable aléatoire gaussienne



Graphique 2.4. Fonction de probabilité $F_{i:5}(x)$ lorsque X est une variable aléatoire gaussienne

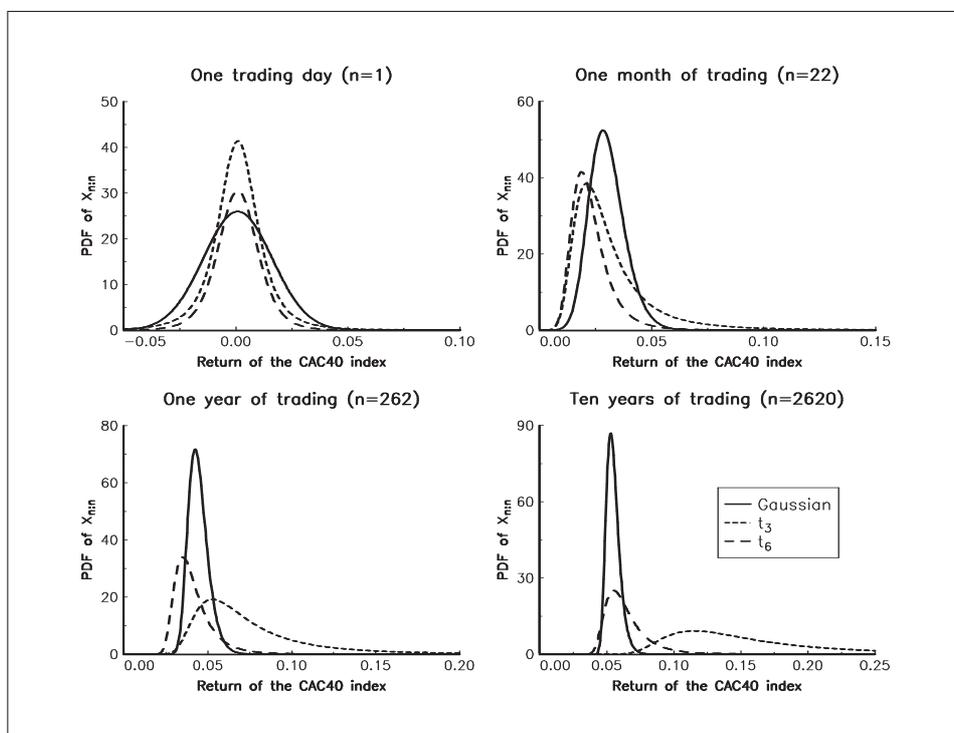
Nous représentons les densités du maximum pour différents horizons (c'est-à-dire pour différentes tailles d'échantillon) sur le graphique 2.5. Nous voyons clairement l'influence de la queue de distribution (en particulier avec la distribution \mathbf{t}_3). Pour obtenir des résultats encore plus marquants, nous considérons la fonction quantile $\mathbf{F}_{n:n}^{-1}$ définie par

$$\mathbf{F}_{n:n}^{-1}(\alpha) = \mathbf{F}^{-1}\left(\sqrt[n]{\alpha}\right) \quad (2.18)$$

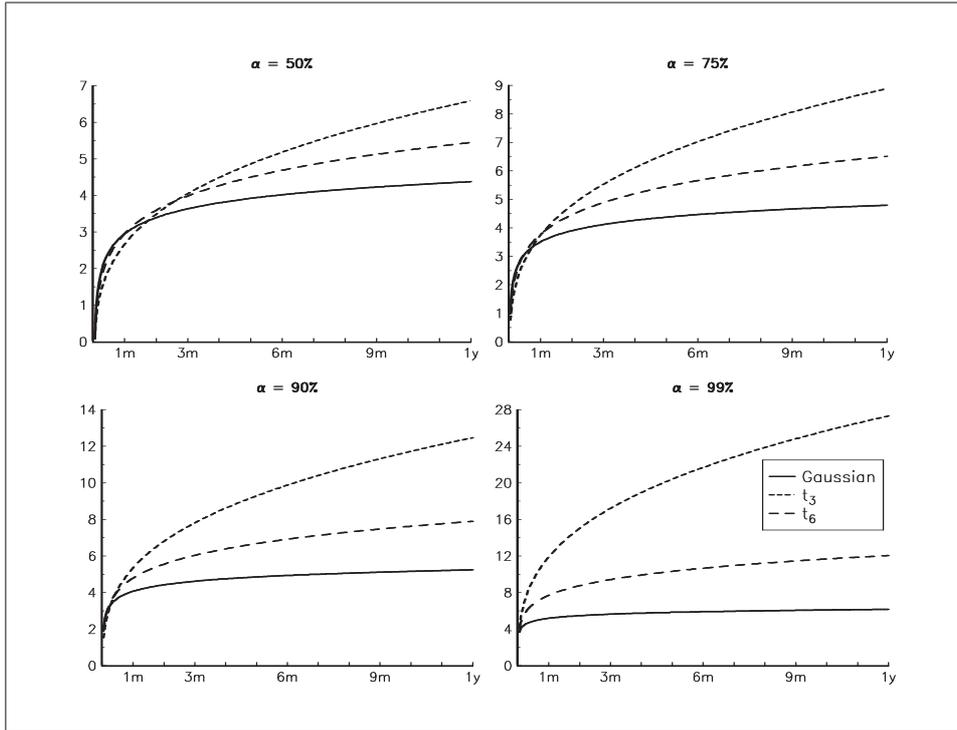
Nous représentons cette fonction sur le graphique 2.6. Maintenant, nous fixons n égal à la taille de la base de données (ce qui correspond à un horizon de 13 années). Nous calculons les rendements les plus extrêmes $r_{(m)}$ que nous observons pendant ces 13 années. Les probabilités $\Pr\{\chi_n^+ > r_{(m)}\}$ sont les suivantes

m	$r_{(m)}$	$\mathcal{N}(0, 1)$	\mathbf{t}_3	\mathbf{t}_6
1	16.15%	0%	35.2%	1.6%
2	13.63%	0%	51.3%	4.3%
3	12.81%	0%	58.0%	6.1%
4	12.34%	0%	62.1%	7.5%
5	10.87%	0%	75.7%	14.9%
10	6.81%	1.4%	99.6%	87.0%
15	5.12%	70.8%	100%	100%
20	3.86%	100%	100%	100%

La lecture de ce tableau est le suivant. Par exemple, le 10^{ème} rendement le plus extrême que nous avons observé ces 13 dernières années est égal à 6.81%. La probabilité que nous observons pour les 13 prochaines années un rendement supérieur à 6.81% est respectivement 1.4% pour la modélisation gaussienne, 99.6% pour la modélisation \mathbf{t}_3 et 87% pour la modélisation \mathbf{t}_6 .



Graphique 2.5. Densité du maximum du rendement du CAC 40 pour différents horizons



Graphique 2.6. Graphe de la fonction quantile $\mathbf{F}_{n:n}^{-1}$

Nous finissons cette section par quelques remarques sur l'utilisation des statistiques d'ordre dans un cadre d'inférence statistique. Supposons que nous cherchons à estimer les paramètres μ et σ du modèle gaussien suivant :

$$\mathbf{F}(r(t)) = \Phi\left(\frac{r(t) - \mu}{\sigma}\right) \quad (2.19)$$

La log-vraisemblance individuelle est égale dans ce cas-là à l'expression suivante :

$$\ell_t(r(t); \mu, \sigma) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \frac{1}{2} \left(\frac{r(t) - \mu}{\sigma}\right)^2 \quad (2.20)$$

Les estimateurs du maximum de vraisemblance sont

$$\hat{\mu}_{\text{ML}} = \frac{1}{T} \sum_{t=1}^T r(t) \quad (2.21)$$

et

$$\hat{\sigma}_{\text{ML}} = \sqrt{\frac{1}{T} \sum_{t=1}^T r^2(t) - \hat{\mu}_{\text{ML}}^2} \quad (2.22)$$

Nous pourrions aussi vouloir estimer ces paramètres à partir de la loi du maximum. Dans ce cas, nous construisons à partir de l'échantillon $\{r(t), t = 1, \dots, T\}$ l'échantillon des maxima

$$r^+(k) = \bigvee_{t=1+(k-1)n}^{k \cdot n} r(t) \quad (2.23)$$

n/i	1	2	3	4	5	6	7	8	9	10
1	0.0008									
2	0.0021	-0.0011								
3	0.0037	0.0005	-0.0033							
4	0.0060	0.0006	0.0004	-0.0058						
5	0.0078	0.0010	0.0006	-0.0000	-0.0080					
6	0.0097	0.0017	0.0005	0.0004	-0.0004	-0.0104				
7	0.0121	0.0026	0.0003	0.0005	0.0003	-0.0005	-0.0132			
8	0.0116	0.0060	0.0011	0.0004	0.0003	0.0002	-0.0018	-0.0145		
9	0.0154	0.0035	0.0013	0.0008	0.0004	0.0002	-0.0004	-0.0025	-0.0166	
10	0.0153	0.0082	0.0019	0.0006	0.0005	0.0004	0.0003	-0.0006	-0.0047	-0.0185

TABLE 2.1. Estimation ML de μ partir des statistiques d'ordre (distribution $\mathcal{N}(0, 1)$)

La log-vraisemblance individuelle du k -ième maximum est égale à

$$\begin{aligned}
\ell_k(r^+(k); \mu, \sigma) &= \ln f_{n:n}(r^+(k)) \\
&= \ln n + (n-1) \ln \mathbf{F}(r^+(k)) + \ln f(r^+(k)) \\
&= \ln n + (n-1) \ln \Phi\left(\frac{r^+(k) - \mu}{\sigma}\right) - \\
&\quad \left[\frac{1}{2} \ln(2\pi) + \frac{1}{2} \ln(\sigma^2) + \frac{1}{2} \left(\frac{r^+(k) - \mu}{\sigma}\right)^2 \right]
\end{aligned} \tag{2.24}$$

Si la distribution est vraiment gaussienne, alors les deux estimations donnent sensiblement les mêmes résultats. De même, nous pourrions considérer toute autre statistique d'ordre. Soit $\{r_{i:n}(k)\}$ l'échantillon de la i -ième statistique d'ordre. Nous avons

$$\begin{aligned}
\ell_k(r_{i:n}(k); \mu, \sigma) &= \ln f_{i:n}(r_{i:n}(k)) \\
&= \ln n! - \ln(i-1)! - \ln(n-i)! + \\
&\quad (i-1) \ln \Phi\left(\frac{r_{i:n}(k) - \mu}{\sigma}\right) + (n-i) \ln \Phi\left(\frac{\mu - r_{i:n}(k)}{\sigma}\right) - \\
&\quad \left[\frac{1}{2} \ln(2\pi) + \frac{1}{2} \ln(\sigma^2) + \frac{1}{2} \left(\frac{r_{i:n}(k) - \mu}{\sigma}\right)^2 \right]
\end{aligned} \tag{2.25}$$

A titre d'exemple, nous considérons l'indice CAC 40. Nous estimons les paramètres μ et σ pour différentes valeurs de i (ordre de la statistique) et de n (taille des blocs). Les valeurs sont données dans les tables 2.1 et 2.2. Pour i et n égal à 1, nous obtenons les estimations classiques du maximum de vraisemblance. Nous remarquons que les valeurs prises par σ sont généralement plus élevées pour les statistiques d'ordre extrême que pour les statistiques d'ordre médian. Cela s'explique bien sûr par la sous-estimation des extrêmes par la distribution gaussienne. Si nous utilisons maintenant la distribution \mathbf{t}_6 :

$$\mathbf{F}(r(t)) = \mathbf{t}_6\left(\frac{r(t) - \mu}{\sigma}\right) \tag{2.26}$$

nous remarquons que les paramètres sont beaucoup plus stables (voir les tables 2.3 et 2.4). Cela est particulièrement vrai pour le paramètre σ . Dans le cas où la distribution est \mathbf{t}_1 , nous observons le phénomène opposé : les valeurs prises par σ sont **moins** élevées pour les statistiques d'ordre extrême que pour les statistiques d'ordre médian (voir les tables 2.5 et 2.6). La distribution \mathbf{t}_1 a donc tendance à sur-estimer les extrêmes.

n/i	1	2	3	4	5	6	7	8	9	10
1	0.0154									
2	0.0149	0.0161								
3	0.0162	0.0116	0.0180							
4	0.0178	0.0112	0.0115	0.0198						
5	0.0188	0.0125	0.0112	0.0122	0.0212					
6	0.0198	0.0132	0.0110	0.0111	0.0128	0.0228				
7	0.0212	0.0141	0.0112	0.0108	0.0113	0.0131	0.0243			
8	0.0204	0.0182	0.0122	0.0104	0.0108	0.0117	0.0146	0.0248		
9	0.0231	0.0145	0.0121	0.0114	0.0106	0.0111	0.0124	0.0151	0.0260	
10	0.0224	0.0196	0.0131	0.0115	0.0110	0.0108	0.0109	0.0130	0.0175	0.0269

TABLE 2.2. Estimation ML de σ partir des statistiques d'ordre (distribution $\mathcal{N}(0,1)$)

n/i	1	2	3	4	5	6	7	8	9	10
1	0.0006									
2	0.0010	0.0001								
3	0.0010	0.0005	0.0001							
4	0.0015	0.0006	0.0003	-0.0002						
5	0.0014	0.0006	0.0006	0.0002	-0.0000					
6	0.0016	0.0008	0.0005	0.0004	0.0001	-0.0004				
7	0.0025	0.0013	0.0002	0.0004	0.0004	0.0002	-0.0008			
8	0.0016	0.0018	0.0009	0.0004	0.0003	0.0004	-0.0000	-0.0007		
9	0.0019	0.0017	0.0010	0.0008	0.0004	0.0002	-0.0001	-0.0005	-0.0004	
10	0.0017	0.0026	0.0013	0.0005	0.0005	0.0004	0.0004	-0.0001	-0.0002	-0.0002

TABLE 2.3. Estimation ML de μ partir des statistiques d'ordre (distribution t_6)

n/i	1	2	3	4	5	6	7	8	9	10
1	0.0102									
2	0.0102	0.0105								
3	0.0101	0.0103	0.0106							
4	0.0105	0.0100	0.0103	0.0109						
5	0.0102	0.0102	0.0102	0.0105	0.0108					
6	0.0103	0.0103	0.0100	0.0102	0.0107	0.0111				
7	0.0107	0.0108	0.0100	0.0099	0.0103	0.0109	0.0111			
8	0.0102	0.0111	0.0107	0.0095	0.0099	0.0104	0.0111	0.0111		
9	0.0104	0.0110	0.0106	0.0104	0.0099	0.0102	0.0110	0.0115	0.0109	
10	0.0100	0.0117	0.0112	0.0104	0.0102	0.0100	0.0099	0.0111	0.0111	0.0107

TABLE 2.4. Estimation ML de σ partir des statistiques d'ordre (distribution t_6)

n/i	1	2	3	4	5	6	7	8	9	10
1	0.0003									
2	0.0000	0.0008								
3	-0.0002	0.0003	0.0012							
4	-0.0003	0.0003	0.0003	0.0016						
5	-0.0010	0.0001	0.0005	0.0006	0.0021					
6	-0.0012	0.0000	0.0004	0.0005	0.0008	0.0024				
7	-0.0009	-0.0001	-0.0000	0.0005	0.0006	0.0011	0.0027			
8	-0.0022	-0.0002	0.0003	0.0003	0.0004	0.0007	0.0014	0.0033		
9	-0.0026	-0.0005	0.0003	0.0005	0.0004	0.0003	0.0005	0.0014	0.0045	
10	-0.0032	-0.0005	0.0001	0.0002	0.0004	0.0003	0.0006	0.0010	0.0025	0.0052

TABLE 2.5. Estimation ML de μ partir des statistiques d'ordre (distribution t_1)

n/i	1	2	3	4	5	6	7	8	9	10
1	0.0064									
2	0.0063	0.0065								
3	0.0056	0.0070	0.0060							
4	0.0053	0.0068	0.0070	0.0056						
5	0.0046	0.0063	0.0072	0.0066	0.0052					
6	0.0043	0.0059	0.0070	0.0071	0.0066	0.0048				
7	0.0042	0.0059	0.0066	0.0070	0.0070	0.0062	0.0044			
8	0.0036	0.0054	0.0069	0.0067	0.0071	0.0068	0.0058	0.0040		
9	0.0032	0.0052	0.0065	0.0072	0.0072	0.0070	0.0068	0.0058	0.0036	
10	0.0028	0.0050	0.0063	0.0069	0.0074	0.0071	0.0065	0.0063	0.0049	0.0032

TABLE 2.6. Estimation ML de σ partir des statistiques d'ordre (distribution t_1)

Remarque 9 *L'application précédente est un bon exemple de travail de statistique en gestion des risques. Cela n'a aucun intérêt d'utiliser des modélisations complexes (et qui sont non implémentables). Vous pouvez donc oublier les outils de type processus ARCH, processus fractionnaires, distributions stables, etc. L'important est de construire une modélisation cohérente qui est implémentable et dont on comprend les implications sous-jacentes.*

2.2 Le théorème de Fisher-Tippet

Ce théorème permet de caractériser la loi de distribution des extrêmes :

Théorème 1 (Th. 3.2.3 de Embrechts, Klüppelberg et Mikosch (1997)) *Supposons n variables aléatoires X_1, \dots, X_n indépendantes et de même loi de distribution \mathbf{F} . S'il existe des constantes a_n et b_n et une distribution limite non dégénérée \mathbf{G} telles que*

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{\chi_n^+ - b_n}{a_n} \leq x \right) = \mathbf{G}(x) \quad \forall x \in \mathbb{R} \quad (2.27)$$

alors \mathbf{G} appartient à l'un des trois types suivants de distribution :

Type I	(Gumbel)	$\mathbf{G}(x) = \exp(-e^{-x})$	$x \in \mathbb{R}$	$g(x) = \exp(-x - e^{-x})$
Type II	(Frechet)	$\mathbf{G}(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases}$	$x \leq 0$ $x > 0$	$g(x) = \begin{cases} 0 & x \leq 0 \\ \alpha x^{-(1+\alpha)} \exp(-x^{-\alpha}) & x > 0 \end{cases}$
Type III	(Weibull)	$\mathbf{G}(x) = \begin{cases} \exp(-(-x)^\alpha) & x \leq 0 \\ 1 & x > 0 \end{cases}$	$x \leq 0$ $x > 0$	$g(x) = \begin{cases} \alpha (-x)^{\alpha-1} \exp(-(-x)^\alpha) & x \leq 0 \\ 0 & x > 0 \end{cases}$

Ce théorème présente un intérêt important, car si l'ensemble des distributions est 'grand', l'ensemble des distributions de valeurs extrêmes est lui très petit. Stuart Coles fait le parallèle suivant entre le théorème de limite centrale et celui de Fisher-Tippet :

The remarkable feature of this result is that the three types of extreme value distributions are the only possible limits for the distribution of the $\frac{\chi_n^+ - b_n}{a_n}$, regardless of the distribution \mathbf{F} for the population. It is in this sense that the theorem provides an extreme value analog of the central limit theorem (COLES [2001], page 47).

Bien sûr, ce théorème n'est valable que si les séquences $\{a_n\}$ et $\{b_n\}$ existent et admettent une limite. C'est par exemple le cas de la distribution exponentielle (COLES [2001], section 3.1.5). Nous avons

$$\mathbf{F}(x) = 1 - \exp(-x) \quad (2.28)$$

Nous remarquons que

$$\begin{aligned} \Pr \left(\frac{\chi_n^+ - b_n}{a_n} \leq x \right) &= \Pr(\chi_n^+ \leq a_n x + b_n) \\ &= \mathbf{F}^n(a_n x + b_n) \\ &= [1 - \exp(-a_n x - b_n)]^n \end{aligned} \quad (2.29)$$

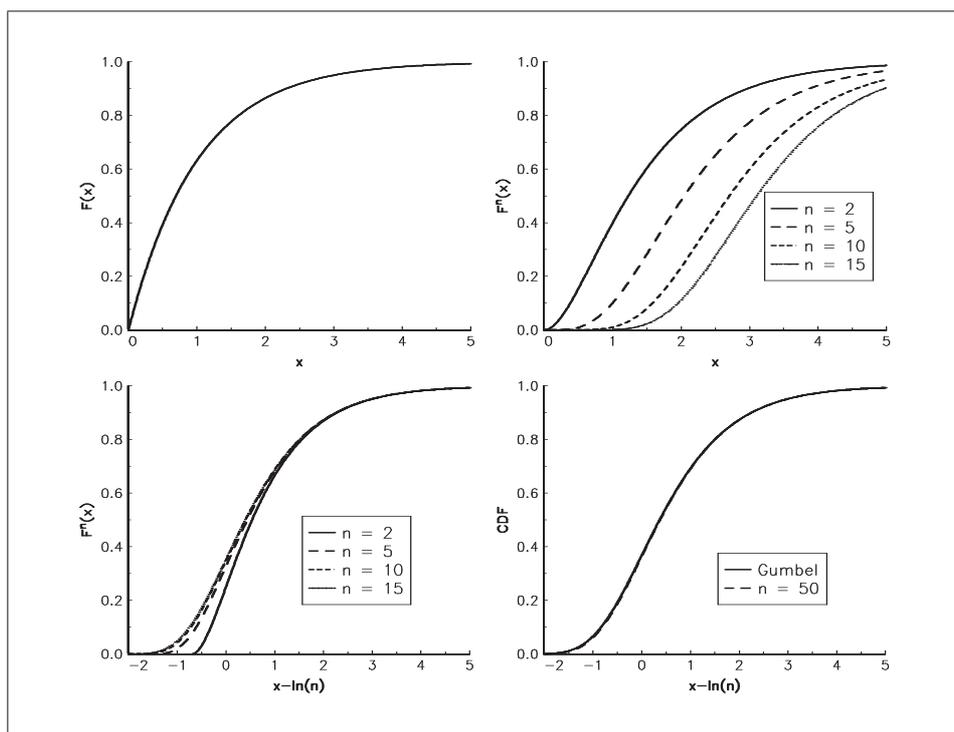
Si nous posons $a_n = 1$ et $b_n = \ln n$, nous avons

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr \left(\frac{\chi_n^+ - b_n}{a_n} \leq x \right) &= \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n} \exp(-x) \right)^n \\ &= \exp(-\exp(-x)) \end{aligned} \quad (2.30)$$

car

$$\begin{aligned} \exp(x) &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n \end{aligned} \quad (2.31)$$

A titre d'illustration, nous montrons sur le graphique 2.7 le convergence de $\mathbf{F}^n(x - \ln n)$ vers la distribution Gumbel.



Graphique 2.7. Max-convergence de la distribution exponentielle vers la distribution Gumbel

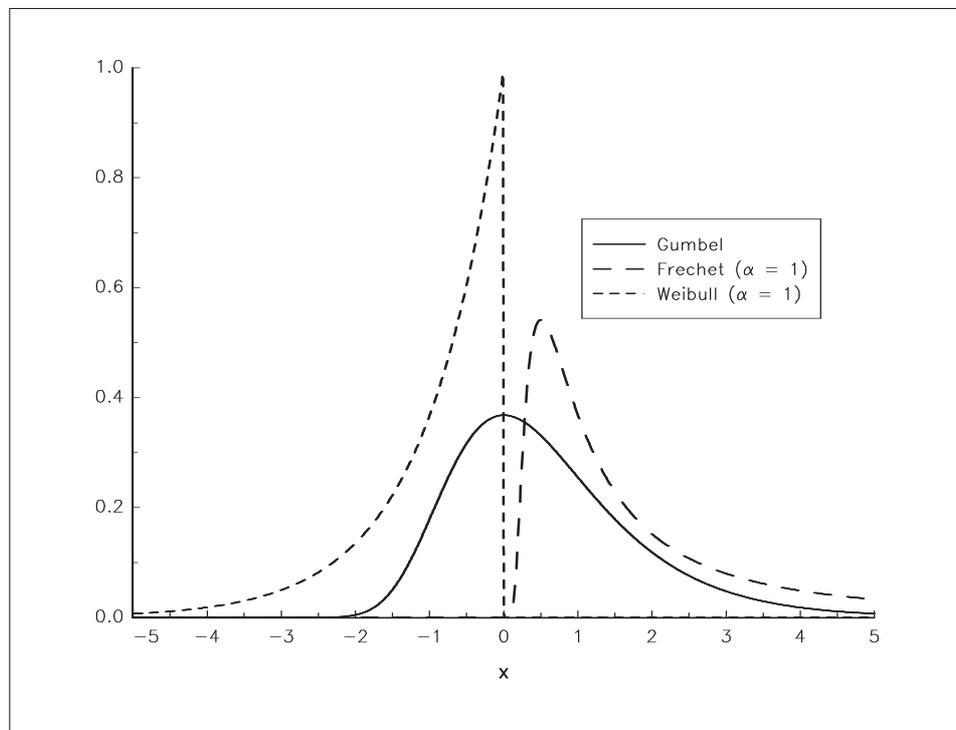
Remarque 10 Pour distinguer les trois distributions, on utilise généralement les notations suivantes : Λ pour la distribution Gumbel, Φ_α pour la distribution Frechet et Ψ_α pour la distribution Weibull (RESNICK [1987]).

Remarque 11 Les distributions Λ , Φ_α et Ψ_α sont appelées les distributions de valeurs extrêmes et les variables aléatoires correspondantes sont les variables aléatoires **extrémales**.

Nous considérons maintenant les différences entre les trois distributions Λ , Φ_α et Ψ_α . Sur le graphique 2.8, nous représentons les fonctions de densité de Λ , Φ_1 et Ψ_1 . Nous remarquons que ces distributions sont très différentes. Néanmoins, la distribution Weibull est liée à la distribution Frechet par la relation suivante :

$$\Psi_\alpha(x) = \Phi_\alpha \left(-\frac{1}{x} \right) \quad (2.32)$$

Notons que par construction la distribution Weibull concerne des variables aléatoires bornées à droite. Elle a donc peu d'intérêt en finance. Concernant la distribution Fréchet, l'allure de la densité dépend du paramètre α (voir le graphique 2.9). Comme nous le verrons plus loin, ce paramètre α influence la queue de distribution.



Graphique 2.8. Densité des distributions Λ , Φ_1 et Ψ_1

On ne pouvait pas finir cette section sans évoquer les fondements du théorème de Fisher-Tippett. La preuve de celui-ci a été formulée plusieurs fois. Le premier à donner une version complète est GNEDENKO [1943]. Vous trouverez une démonstration moderne dans la section 0.3 de RESNICK [1987]. Nous indiquons ici les lignes directrices. Nous avons

$$\lim_{n \rightarrow \infty} \mathbf{F}^n(a_n x + b_n) = \mathbf{G}(x) \quad (2.33)$$

Soit $[t]$ le plus grand entier inférieur ou égal à t . Pour tout $t > 0$, nous en déduisons que

$$\lim_{n \rightarrow \infty} \mathbf{F}^{[nt]}(a_{[nt]}x + b_{[nt]}) = \mathbf{G}(x) \quad (2.34)$$

et aussi

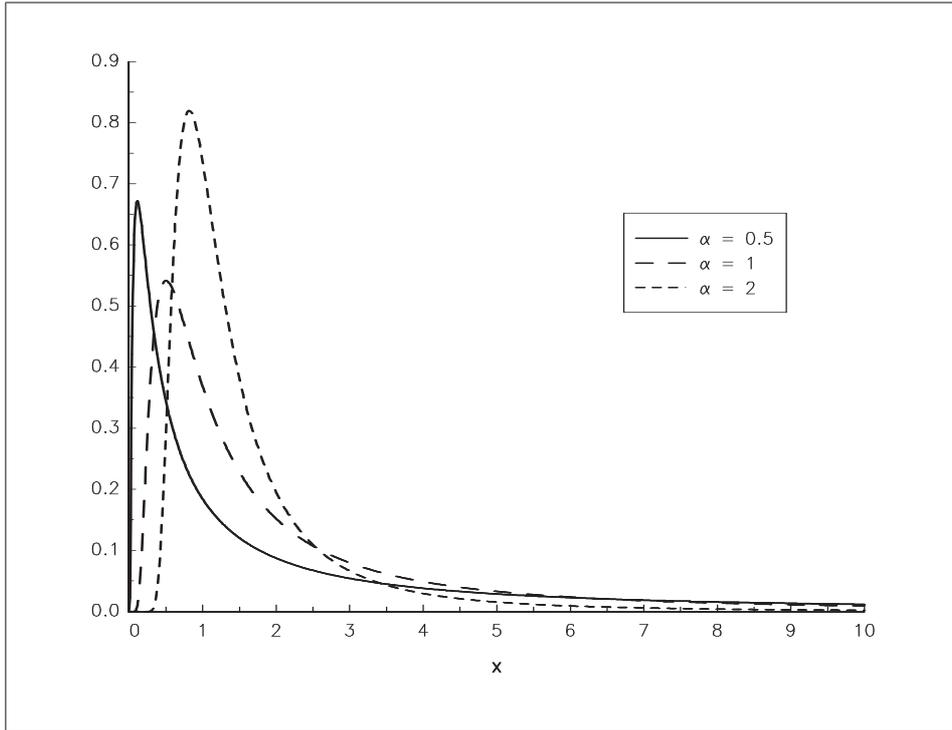
$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{F}^{[nt]}(a_n x + b_n) &= \lim_{n \rightarrow \infty} [\mathbf{F}^n(a_n x + b_n)]^{[nt]/n} \\ &= \mathbf{G}^t(x) \end{aligned} \quad (2.35)$$

Notons

$$\alpha(t) = \lim_{n \rightarrow \infty} a_n / a_{[nt]} \quad (2.36)$$

et

$$\beta(t) = \lim_{n \rightarrow \infty} (b_n - b_{[nt]}) / a_{[nt]} \quad (2.37)$$



Graphique 2.9. Densité de la distribution Fréchet

En utilisant le théorème de convergence des types¹ (RESNICK [1987], Proposition 0.2), nous savons que ces limites existent et nous en déduisons que

$$\mathbf{G}^t(x) = \mathbf{G}(\alpha(t)x + \beta(t)) \quad (2.39)$$

Le théorème de Fisher-Tippett dérive alors de la résolution de cette équation fonctionnelle. Nous avons

$$\mathbf{G}^{ts}(x) = \mathbf{G}(\alpha(ts)x + \beta(ts)) \quad (2.40)$$

et aussi

$$\begin{aligned} \mathbf{G}^{ts}(x) &= [\mathbf{G}^s(x)]^t \\ &= [\mathbf{G}(\alpha(s)x + \beta(s))]^t \\ &= \mathbf{G}(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)) \end{aligned} \quad (2.41)$$

Nous obtenons un système d'équations fonctionnelles :

$$\begin{cases} \alpha(ts) = \alpha(t)\alpha(s) \\ \beta(ts) = \alpha(t)\beta(s) + \beta(t) \end{cases} \quad (2.42)$$

La solution de la première équation fonctionnelle est $\alpha(t) = t^{-\theta}$ avec $\theta \in \mathbb{R}$. Selon les valeurs de θ , nous obtenons les trois types de distribution de valeurs extrêmes :

$$\begin{aligned} \text{Gumbel} & \theta = 0 \\ \text{Fréchet} & \theta > 0 \\ \text{Weibull} & \theta < 0 \end{aligned}$$

¹Les distributions \mathbf{F} et \mathbf{G} sont dites de même type s'il existe des constantes $a > 0$ et $b \in \mathbb{R}$ telles que

$$\mathbf{F}(ax + b) = \mathbf{G}(x) \quad (2.38)$$

2.3 Caractérisation des domaines d'attraction

On dit qu'une distribution \mathbf{F} appartient au max-domaine d'attraction de \mathbf{G} , et on note $\mathbf{F} \in \text{MDA}(\mathbf{G})$ si la distribution du maximum normalisée converge vers \mathbf{G} . Dans la section précédente, nous avons vu que l'exponentielle appartient au max-domaine d'attraction de Λ . Voici d'autres exemples :

- MDA (Λ)
Exponentielle, Gaussienne, Gamma, Lognormale, etc.
- MDA (Φ_α)
Cauchy, Pareto, α -stable ($\alpha < 2$), etc.
- MDA (Ψ_α)
Uniforme, Beta, etc.

Nous voyons que les trois distributions de valeurs extrêmes sont très différentes en terme de max-domaine d'attraction :

1. Dans le max-domaine d'attraction de la distribution **Gumbel**, nous trouvons des distributions qui n'ont pas de queues épaisses (mais qui peuvent être leptokurtiques).
2. Dans le max-domaine d'attraction de la distribution **Frechet**, nous trouvons des distributions qui ont des queues épaisses.
3. Dans le max-domaine d'attraction de la distribution **Weibull**, nous trouvons des distributions à support fini, ce qui implique que le support du maximum soit borné à droite.

Cela a des implications d'un point de vue financier, puisque la problématique va être de choisir entre les distributions Λ et Φ_α (et dans ce cas, quelle valeur pour α ?) pour modéliser le maximum.

Sachant la distribution \mathbf{F} , nous voudrions connaître à quel max-domaine d'attraction elle appartient et quelles sont les constantes de normalisation. La réponse est relativement complexe et n'est pas unique. Nous indiquons ici les critères les plus utilisés.

Définition 1 Une fonction f est dite à variation régulière d'indice α et on note $f \in \text{RV}_\alpha$ si pour tout $x > 0$, on a

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha \quad (2.43)$$

Théorème 2 $\mathbf{F} \in \text{MDA}(\Phi_\alpha)$ si et seulement si $1 - \mathbf{F} \in \text{RV}_{-\alpha}$. Dans ce cas, $a_n = \mathbf{F}^{-1}(1 - n^{-1})$ et $b_n = 0$.

Ce théorème formulé par GNEDENKO [1943] permet de caractériser très simplement les distributions $\mathbf{F} \in \text{MDA}(\Phi_\alpha)$. En effet, elles doivent vérifier

$$\lim_{t \rightarrow \infty} \frac{1 - \mathbf{F}(tx)}{1 - \mathbf{F}(t)} = x^{-\alpha} \quad (2.44)$$

Prenons par exemple le cas de la distribution **Pareto**. Nous avons

$$\mathbf{F}(x) = 1 - x^{-1/\gamma} \quad (2.45)$$

Nous en déduisons que

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1 - \mathbf{F}(tx)}{1 - \mathbf{F}(t)} &= \lim_{t \rightarrow \infty} \frac{(tx)^{-1/\gamma}}{t^{-1/\gamma}} \\ &= x^{-1/\gamma} \end{aligned}$$

Donc $1 - \mathbf{F} \in \text{RV}_{-1/\gamma}$ et $\mathbf{F} \in \text{MDA}(\Phi_{1/\gamma})$. Remarquons que $a_n = \mathbf{F}^{-1}(1 - n^{-1}) = n^\gamma$. Bien sûr, nous vérifions

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{F}^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} \mathbf{F}^n(n^\gamma x) \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{x^{-1/\gamma}}{n}\right) \\ &= \exp(-x^{-1/\gamma}) \end{aligned} \quad (2.46)$$

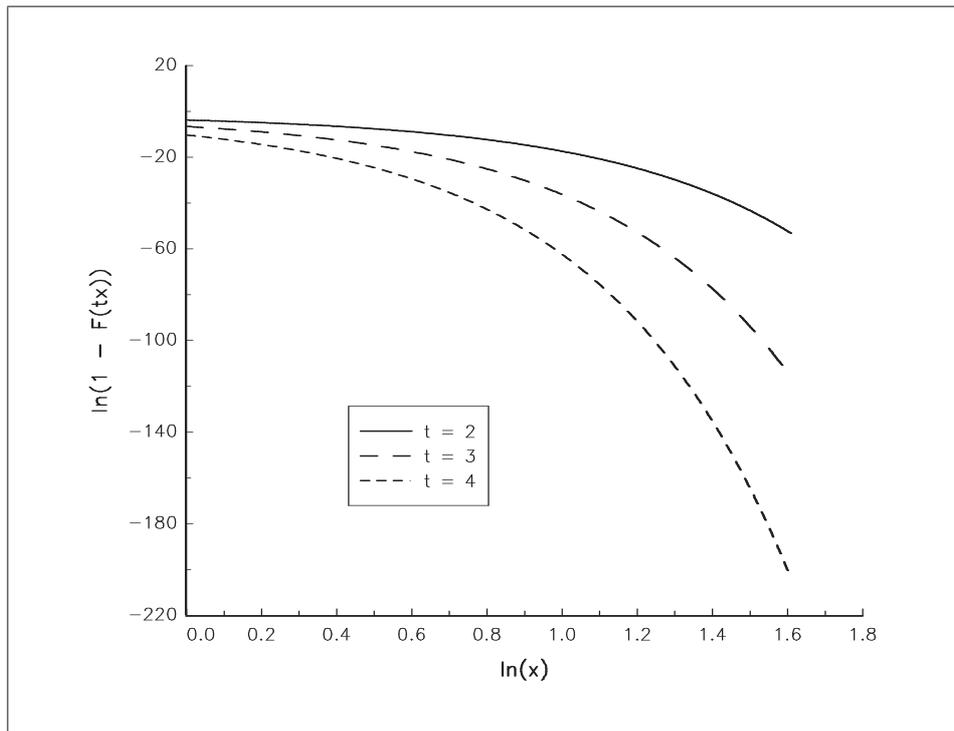
Le théorème précédent suggère une vérification graphique de la propriété $\mathbf{F} \in \text{MDA}(\Phi_\alpha)$. Nous avons

$$\frac{1 - \mathbf{F}(tx)}{1 - \mathbf{F}(t)} \simeq x^{-\alpha} \quad (2.47)$$

pour t suffisamment grand. Dans ce cas, nous vérifions que

$$\ln(1 - \mathbf{F}(tx)) \simeq \ln(1 - \mathbf{F}(t)) - \alpha \ln x \quad (2.48)$$

Par exemple, nous vérifions graphiquement que la distribution gaussienne n'appartient pas au max-domaine d'attraction de la distribution Fréchet (voir le graphique 2.10).



Graphique 2.10. Vérification de la propriété de variation régulière de la distribution gaussienne

Voyons maintenant comment nous pouvons caractériser $\text{MDA}(\Psi_\alpha)$ et $\text{MDA}(\Lambda)$.

Théorème 3 $\mathbf{F} \in \text{MDA}(\Psi_\alpha)$ si et seulement si $1 - \mathbf{F}(x_0 - x^{-1}) \in \text{RV}_{-\alpha}$ et $x_0 < \infty$. Dans ce cas, $a_n = x_0 - \mathbf{F}^{-1}(1 - n^{-1})$ et $b_n = x_0$.

Considérons la distribution uniforme. Nous avons $x_0 = 1$.

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1 - \mathbf{F}(1 - t^{-1}x^{-1})}{1 - \mathbf{F}(t^{-1})} &= \lim_{t \rightarrow \infty} \frac{t^{-1}x^{-1}}{t^{-1}} \\ &= x^{-1} \end{aligned} \quad (2.49)$$

Donc $\mathbf{F} \in \text{MDA}(\Psi_1)$. Remarquons que $a_n = 1 - \mathbf{F}^{-1}(1 - n^{-1}) = n^{-1}$. Bien sûr, nous vérifions

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{F}^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} \mathbf{F}^n(n^{-1}x + 1) \\ &= \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \\ &= \exp(x) \end{aligned} \quad (2.50)$$

La caractérisation de $\text{MDA}(\Lambda)$ est la plus difficile à énoncer. Tout d'abord, il n'existe pas une condition **nécessaire** et suffisante relativement simple. Généralement, les auteurs présentent une condition suffisante basée sur la représentation d'une fonction *Von Mises*. Cependant, cette représentation est difficile à manipuler, tout comme la CNS formulée par GNEDENKO [1943]. Si la distribution \mathbf{F} est de classe C^2 , une condition suffisante relativement simple à vérifier est la suivante

$$\lim_{x \rightarrow \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial^2 \mathbf{F}(x)}{(\partial \mathbf{F}(x))^2} = -1 \quad (2.51)$$

Prenons l'exemple de la distribution exponentielle. Nous avons $\mathbf{F}(x) = 1 - \exp(-x)$, $\partial \mathbf{F}(x) = \exp(-x)$ et $\partial^2 \mathbf{F}(x) = -\exp(-x)$. Nous vérifions que

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial^2 \mathbf{F}(x)}{(\partial \mathbf{F}(x))^2} &= \lim_{x \rightarrow \infty} -\frac{\exp(-x) \cdot \exp(-x)}{(\exp(-x))^2} \\ &= -1 \end{aligned} \quad (2.52)$$

Si nous considérons la distribution gaussienne, nous avons $\mathbf{F}(x) = \Phi(x)$, $\partial \mathbf{F}(x) = \phi(x)$ et $\partial^2 \mathbf{F}(x) = -x\phi(x)$. En utilisant la règle de L'Hospital, nous en déduisons que

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{(1 - \mathbf{F}(x)) \cdot \partial^2 \mathbf{F}(x)}{(\partial \mathbf{F}(x))^2} &= \lim_{x \rightarrow \infty} -\frac{x\phi(-x)}{\phi(x)} \\ &= -1 \end{aligned} \quad (2.53)$$

Les précédents éléments suggèrent l'importance du comportement de la distribution au point droit du support pour caractériser la distribution des extrêmes. Il peut alors être intéressant de définir des classes d'équivalence. Deux distributions \mathbf{F}_1 et \mathbf{F}_2 sont équivalentes en queue (*tail equivalent*) si elles ont le même point final x_0 et si

$$\lim_{x \rightarrow x_0} \frac{1 - \mathbf{F}_1(x)}{1 - \mathbf{F}_2(x)} = c < \infty$$

Cette classe d'équivalence est utilisée dans le théorème suivant (RESNICK [1987], page 67) :

Théorème 4 Soient \mathbf{F}_1 et \mathbf{F}_2 deux distributions de probabilité. Supposons que $\mathbf{F}_1 \in \text{MDA}(\mathbf{G}_1)$ et que

$$\lim_{n \rightarrow \infty} \mathbf{F}_1^n(a_n x + b_n) = \mathbf{G}_1(x) \quad (2.54)$$

Alors

$$\lim_{n \rightarrow \infty} \mathbf{F}_2^n(a_n x + b_n) = \mathbf{G}_2(x) \quad (2.55)$$

si et seulement si pour $a > 0$ et $b \in \mathbb{R}$, nous avons

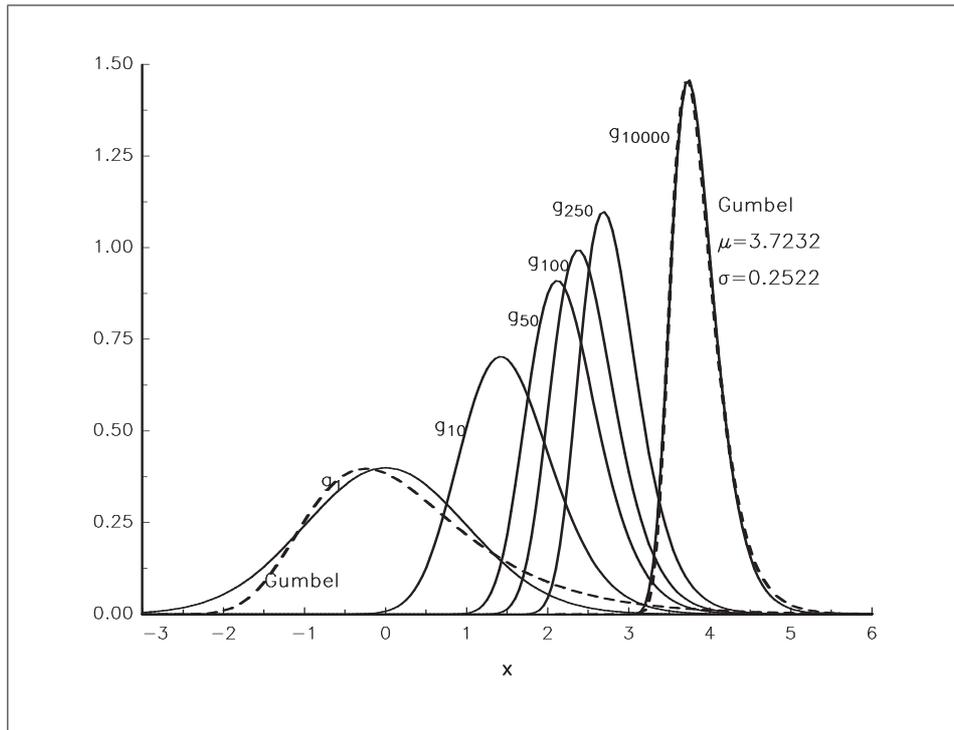
$$\mathbf{G}_2(x) = \mathbf{G}_1(ax + b) \quad (2.56)$$

\mathbf{F}_1 et \mathbf{F}_2 sont équivalentes en queue et si

- (i) $\mathbf{G}_1 = \Lambda$, alors $a = 1$ et $c = \exp(b)$;
1. $\mathbf{G}_1 = \Phi_\alpha$, alors $b = 0$ et $c = a^\alpha$;
 2. $\mathbf{G}_1 = \Psi_\alpha$, alors $b = 0$ et $c = a^{-\alpha}$.

RESNICK [1987] utilise ce théorème pour trouver les constantes de normalisation des distributions gaussienne et Gamma.

Nous finissons cette section par le problème de la vitesse de convergence, c'est-à-dire à partir de quelle valeur de n pouvons nous "approcher" $\mathbf{F}^n(a_n x + b_n)$ par sa limite $\mathbf{G}(x)$. Il n'y a pas de réponse générale. Tout va dépendre de la loi \mathbf{F} . Par exemple, les distributions exponentielle ou Pareto convergent assez rapidement, ce qui n'est pas le cas des Student et α -stable.

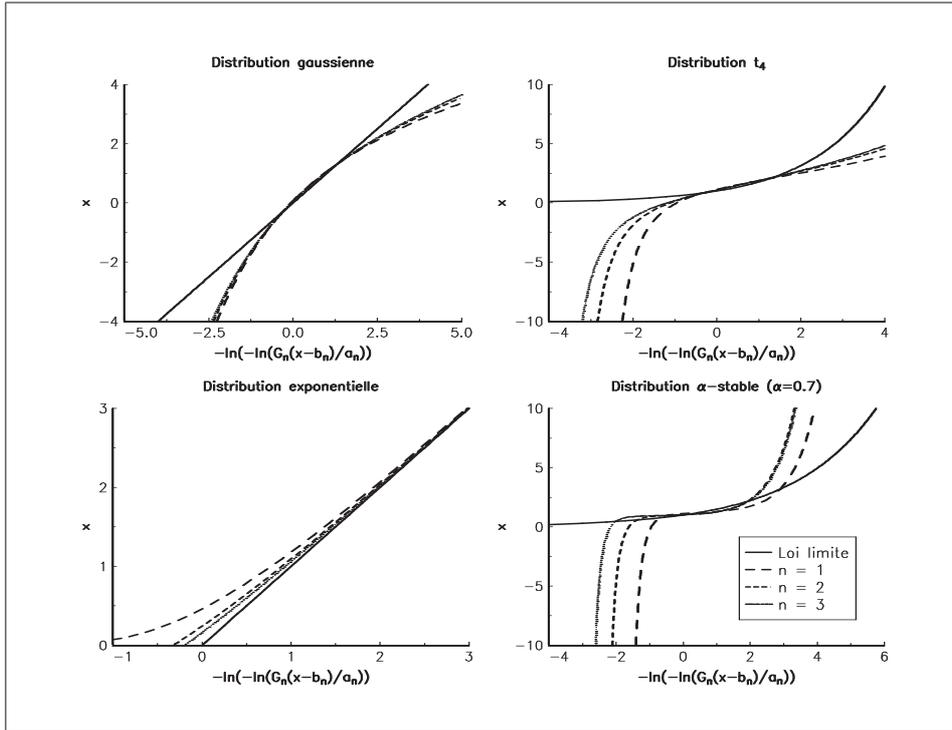


Graphique 2.11. Max-convergence de la distribution gaussienne vers la distribution Gumbel

2.4 L'approche par processus ponctuels

2.5 La distribution GEV

The three types of limits that arise in Theorem 1 have distinct forms of behavior, corresponding to the different forms of tail behaviour for the distribution function F of the X_i . This can be made precise by considering the behavior of the limit distribution G at x_+ , its upper end-point. For the Weibull distribution x_+ is finite, while for both the Frechet and Gumbel distributions $x_+ = \infty$. However, the density of G decays exponentially for the Gumbel distribution and polynomially for the Frechet distribution, corresponding to relatively different rates of decay in the tail of F . It follows that in applications the three different families give quite different representations of extreme value behaviour. In early applications of extreme value



Graphique 2.12. Comparaison de Max-convergence vers la distribution Fréchet

theory, it was usual to adopt one of the three families, and then to estimate the relevant parameters of that distribution. But there are two weakness : first, a technique is required to choose which of the three families is most appropriate for the data at hand ; second, once such a decision is made, subsequent inferences presume this choice to be correct, and do not allow for the uncertainty such a selection involves, even though this uncertainty may be substantial. (COLES [2001], page 47).

En fait, nous pouvons caractériser ces trois types de distribution par une distribution unique :

$$\mathbf{G}(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (2.57)$$

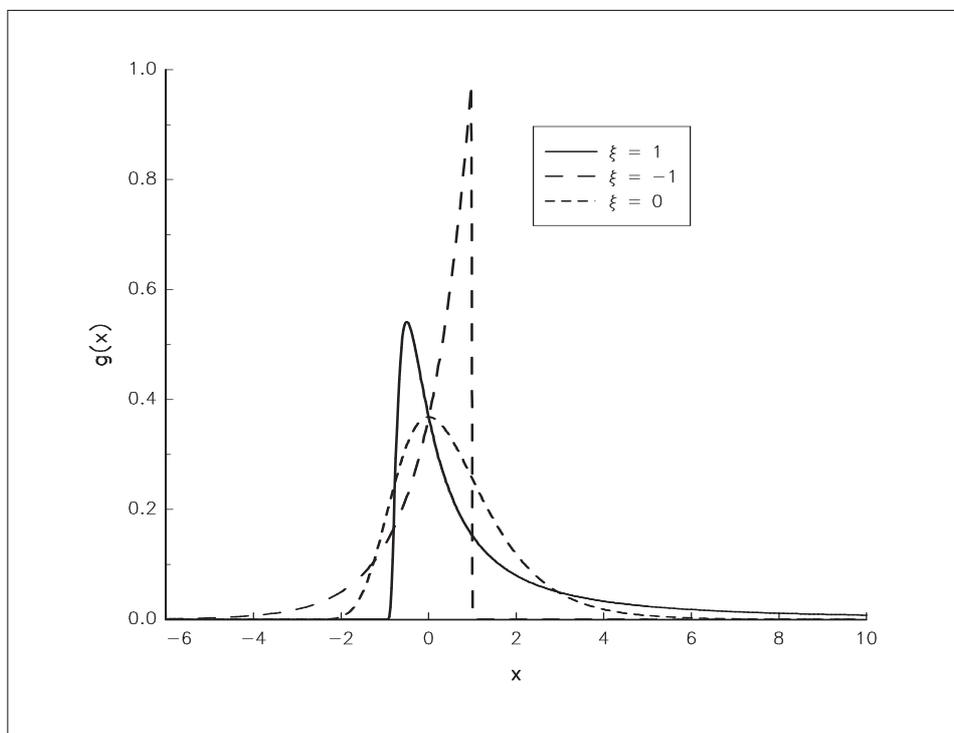
défini sur le support $\Delta = \{x : 1 + \xi \sigma^{-1} (x - \mu) > 0\}$ et que nous notons $\mathcal{GEV}(\mu, \sigma, \xi)$. Cette fonction de distribution correspond à la loi de probabilité de von Mises-Jenkinson, mais elle est plus connue sous le nom de “Generalized Extreme Value distribution” (**GEV**). Nous avons alors les correspondances suivantes :

Fréchet	$\xi = \alpha^{-1} > 0$
Weibul	$\xi = -\alpha^{-1} < 0$
Gumbel	$\xi \rightarrow 0$

Remarquons que les paramètres μ et σ sont en fait les limites de b_n et a_n . A titre d’illustration, voyons quelques exemples de distribution **GEV**. Pour cela, nous pouvons facilement montrer que la fonction de densité correspondante est

$$g(x) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (2.58)$$

Les graphiques 2.13, 2.14 et 2.15 présentent des fonctions de densité pour différentes valeurs de paramètre. Lorsque μ varie, nous remarquons une translation des fonctions. μ est donc un paramètre de **localisation**. σ joue le rôle d'une variance, c'est pourquoi nous le considérons comme un paramètre de **dispersion**. Enfin, le paramètre ξ est lié à la queue de distribution. C'est pourquoi on lui donne généralement le nom d'indice de queue lorsque $\xi > 0$ ou d'indice de valeur extrême.



Graphique 2.13. Densité de la distribution $\mathcal{GEV}(0, 1, \xi)$ (I)

Remarque 12 *Le quantile α de la distribution $\mathcal{GEV}(\mu, \sigma, \xi)$ est donné par la formule suivante :*

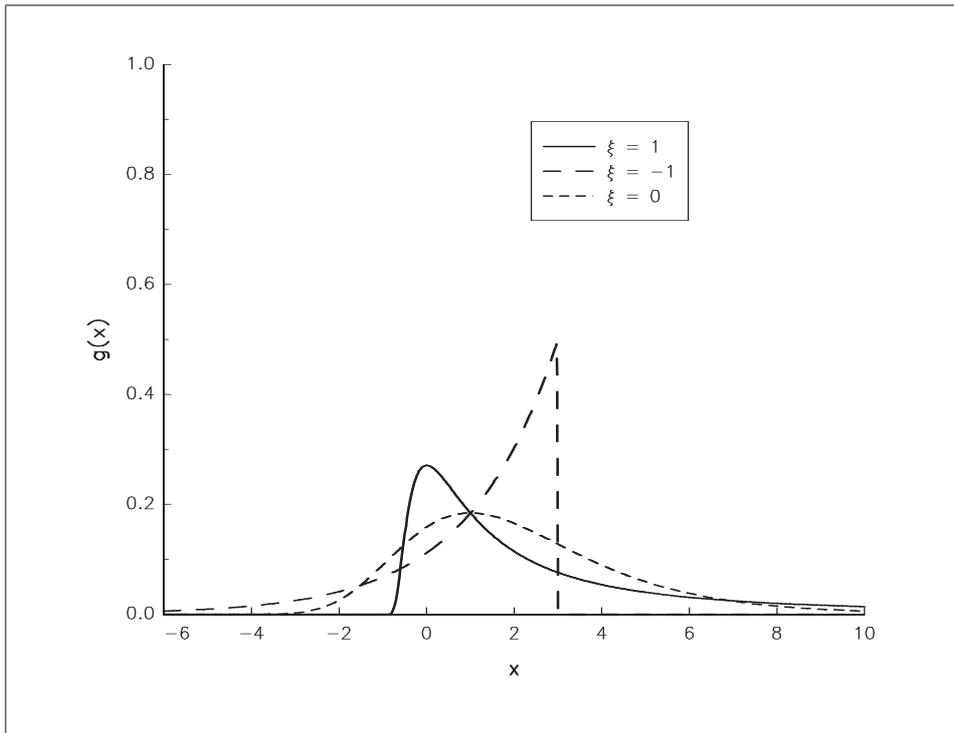
$$\mathbf{G}^{-1}(\alpha) = \mu - \sigma \xi^{-1} \left[1 - (-\ln \alpha)^{-\xi} \right] \quad (2.59)$$

Ce quantile est donc fortement influencé² par les deux paramètres σ et ξ . A titre d'illustration, le graphique 2.16 présente différentes courbes de quantiles en fonction de ξ . Intuitivement, nous comprenons que plus ξ est grand, plus le quantile est élevé.

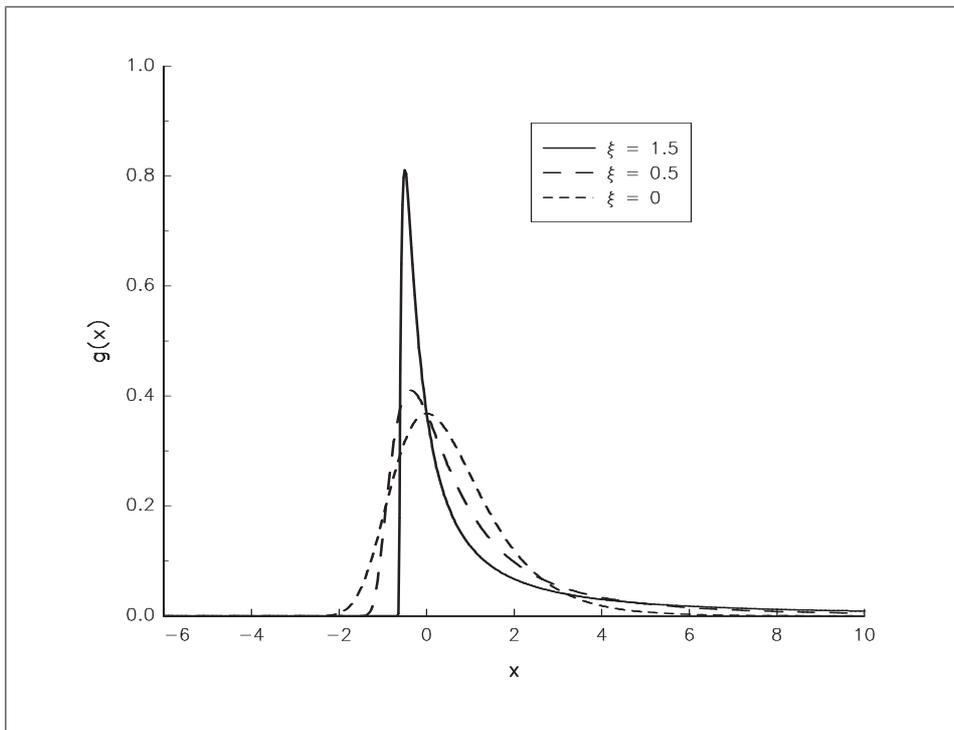
Remarque 13 *Nous pouvons montrer que le mode de la distribution **GEV** est atteint pour la valeur $\mu + \sigma(1 - \xi)^\xi$. Quant à la moyenne, nous pouvons la calculer facilement en utilisant une procédure d'intégration numérique de type Gauss-Legendre ou Gauss-Hermite.*

²En effet, nous avons

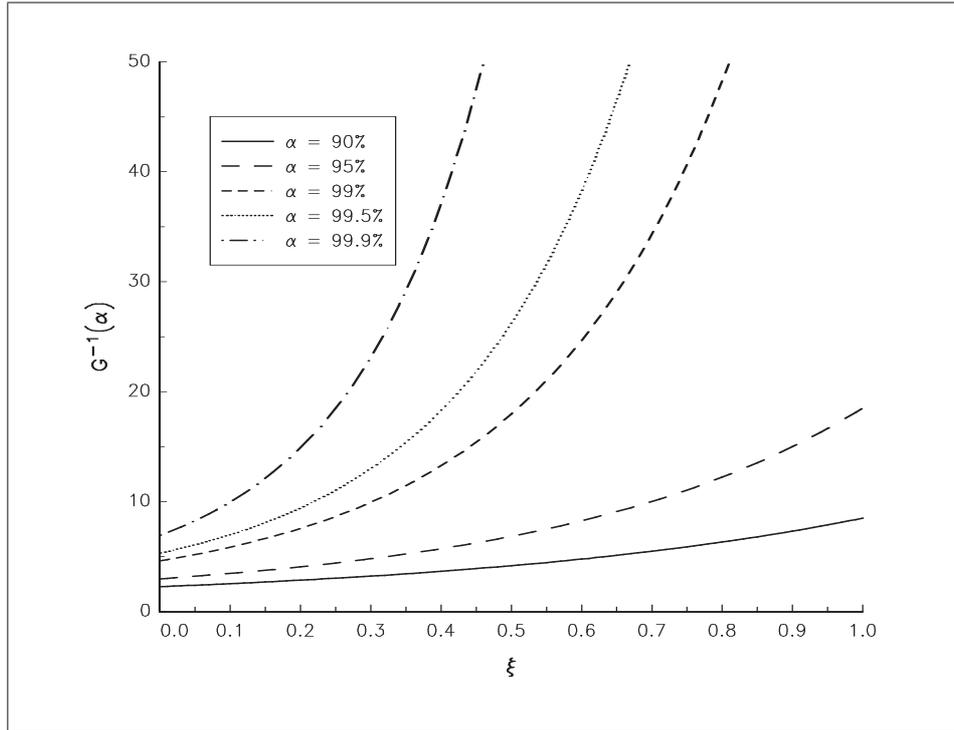
$$\partial_\alpha \mathbf{G}^{-1}(\alpha) = \frac{\sigma}{\alpha (-\ln \alpha)^{1+\xi}} \quad (2.60)$$



Graphique 2.14. Densité de la distribution $\mathcal{GEV}(0, 2, \xi)$ (II)



Graphique 2.15. Densité de la distribution $\mathcal{GEV}(0, 1, \xi)$ (III)



Graphique 2.16. Quantile de la distribution $\mathcal{GEV}(0, 1, \xi)$

Il existe plusieurs méthodes pour estimer les paramètres de la distribution $\mathcal{GEV}(\mu, \sigma, \xi)$. Nous pouvons par exemple citer les méthodes d'estimation de l'indice de queue (de type Hill ou Pickands — voir DREES, DE HAAN et RESNICK [1998]) ou la méthode des moments. Mais celle qui reste la plus populaire et qui sous certaines conditions est la plus efficace est la méthode du maximum de vraisemblance. Soit θ le vecteur des paramètres. Nous avons

$$\theta = \begin{bmatrix} \mu \\ \sigma \\ \xi \end{bmatrix}$$

Nous considérons un échantillon de données $\{X_t\}$ de dimension $T = Kn$ avec $K \in \mathbb{N}$. Nous divisons cet échantillon en K blocs et nous définissons X_k^+ de la façon suivante

$$X_k^+ = \bigvee_{t=1+(k-1)n}^{k \cdot n} X_t \quad (2.61)$$

L'expression de la vraisemblance de l'observation k est donc

$$L(X_k^+; \theta) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{X_k^+ - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[1 + \xi \left(\frac{X_k^+ - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (2.62)$$

Nous en déduisons l'expression suivante pour la log-vraisemblance :

$$\ell(X_k^+; \theta) = -\ln \sigma - \left(\frac{1+\xi}{\xi} \right) \ln \left(1 + \xi \left(\frac{X_k^+ - \mu}{\sigma} \right) \right) - \left[1 + \xi \left(\frac{X_k^+ - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \quad (2.63)$$

L'estimateur du maximum de vraisemblance correspond alors à

$$\hat{\theta}_{\text{ML}} = \arg \max_{\theta \in \Theta} \sum_{k=1}^K \ell(X_k^+; \theta) \quad (2.64)$$

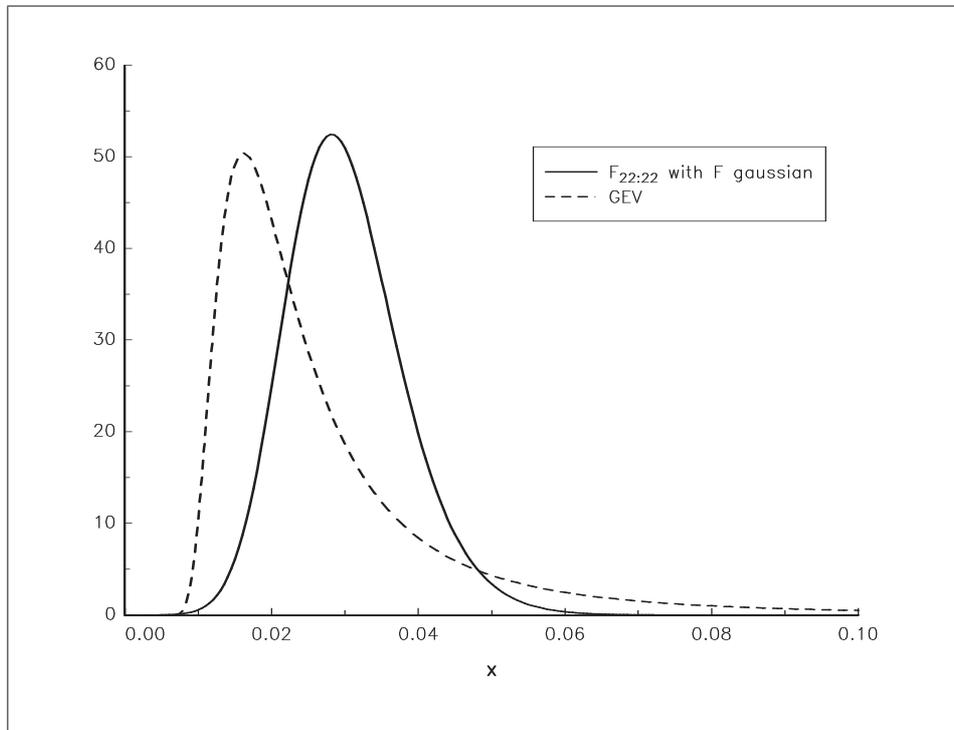
avec Θ l'espace des paramètres. L'estimation de l'expression (2.64) est relativement aisée à condition de prendre quelques précautions, notamment pour la spécification de Θ (à cause de la singularité au point $\xi = 0$). Généralement, l'optimisation s'effectue avec un algorithme de type quasi-Newton (par exemple, l'algorithme Bryoden-Fletcher-Goldfarb-Shanno) en utilisant un jacobien analytique :

$$\begin{aligned}\partial_{\mu} \ell(X_k^+; \theta) &= \frac{1 + \xi - \omega^{-\frac{1}{\xi}}}{\sigma \omega} \\ \partial_{\sigma} \ell(X_k^+; \theta) &= \frac{\left((1 + \xi) - \omega^{-\frac{1}{\xi}} \right) (X_k^+ - \mu) - \sigma \omega}{\sigma^2 \omega} \\ \partial_{\xi} \ell(X_k^+; \theta) &= \left(1 - \omega^{-\frac{1}{\xi}} \right) \left(\frac{1}{\xi^2} \ln \omega - \frac{(X_k^+ - \mu)}{\xi \sigma \omega} \right) - \frac{(X_k^+ - \mu)}{\sigma \omega}\end{aligned}\quad (2.65)$$

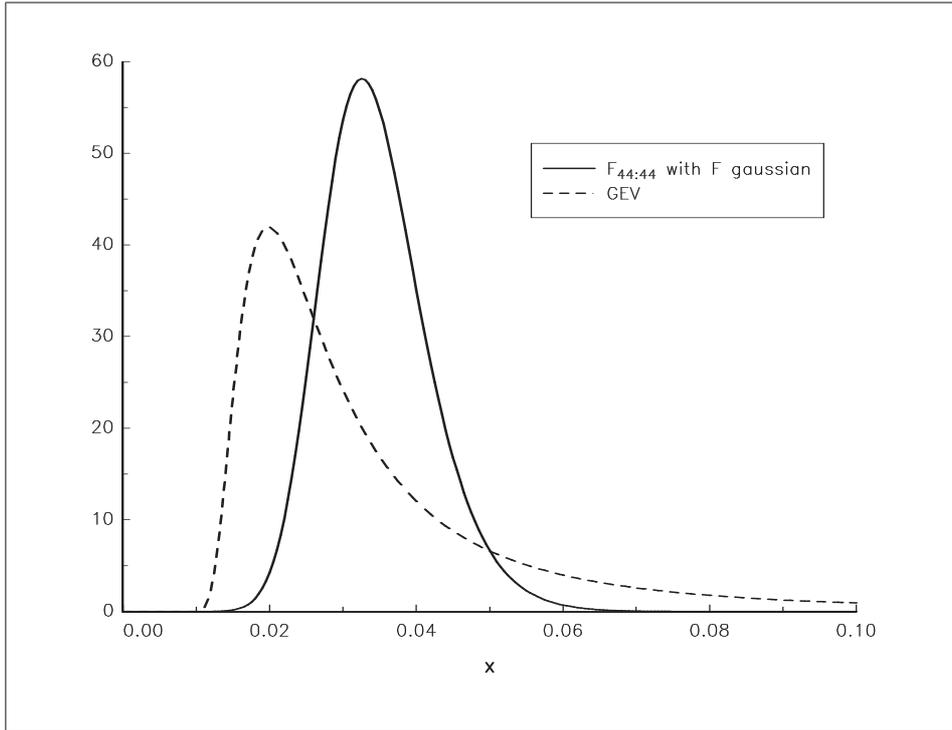
avec

$$\omega = 1 + \xi \left(\frac{X_k^+ - \mu}{\sigma} \right) \quad (2.66)$$

A titre d'exemple, nous estimons les paramètres de la distribution $\mathcal{GEV}(\mu, \sigma, \xi)$ à partir des rendements logarithmiques du CAC 40 en considérant respectivement des blocs de 22 et 44 jours de trading. Nous comparons les distributions obtenues avec celles issues de l'hypothèse gaussienne (voir les graphiques 2.17 et 2.18).



Graphique 2.17. Densité de la distribution du maximum du rendement logarithmique du CAC 40 (22 jours de trading)



Graphique 2.18. Densité de la distribution du maximum du rendement logarithmique du CAC 40 (44 jours de trading)

2.6 Les modèles à seuil

L'estimation des paramètres de la distribution **GEV** est une estimation dite “block component-wise”. Cela veut dire qu'à partir d'un échantillon, on construit un échantillon de maxima en formant des blocs de même dimension. Cela implique donc une perte de certaines informations. En particulier, certains blocs peuvent contenir plusieurs extrêmes, alors que d'autres blocs peuvent ne pas en contenir.

Dans les modèles à seuil (threshold models), nous nous intéressons à la probabilité conditionnelle suivante :

$$\Pr \{X > u + y \mid X > u\} = \frac{1 - \mathbf{F}(u + y)}{1 - \mathbf{F}(u)} \quad (2.67)$$

Pour cela, nous introduisons la fonction de distribution des excès par rapport à un seuil u :

$$\mathbf{F}_u(y) = \Pr \{X - u \leq y \mid X > u\} \quad (2.68)$$

pour $0 \leq y < x_0 - u$ avec $x_0 = \sup \{x \in \mathbb{R} : \mathbf{F}(x) < 1\}$. La fonction de distribution des excès représente la probabilité que la variable aléatoire X dépasse le seuil u d'au plus une quantité y , sachant qu'elle dépasse u . Remarquons que l'on peut écrire cette fonction de distribution sous la forme

$$\begin{aligned} \mathbf{F}_u(y) &= 1 - \left(\frac{1 - \mathbf{F}(u + y)}{1 - \mathbf{F}(u)} \right) \\ &= \frac{\mathbf{F}(u + y) - \mathbf{F}(u)}{1 - \mathbf{F}(u)} \end{aligned} \quad (2.69)$$

Supposons que l'approximation de la distribution du maximum par une distribution **GEV** soit satisfaisante :

$$\Pr(\chi_n^+ \leq x) \approx \mathbf{G}(x) \quad (2.70)$$

Alors, pour u assez grand, $\mathbf{F}_u(y)$ est approximativement

$$\mathbf{H}(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \quad (2.71)$$

Cette distribution est définie sur $\Delta = \{y : 1 + \xi \tilde{\sigma}^{-1} y > 0\}$ avec

$$\tilde{\sigma} = \sigma + \xi(u - \mu) \quad (2.72)$$

C'est en fait une distribution Pareto généralisée (*Generalized Pareto Distribution* ou **GPD**) :

$$\mathbf{H}(x) = 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} \quad (2.73)$$

que nous notons $\mathcal{GPD}(\beta, \xi)$. Nous avons donc une certaine dualité entre les distributions **GPD** et **GEV** :

[...] if block maxima have approximating distribution G , then threshold excesses have a corresponding approximate distribution within the generalized Pareto family. Moreover, the parameters of the generalized Pareto distribution of threshold excesses are uniquely determined by those of the associated GEV distribution of block maxima. In particular, the parameter ξ is equal to that of the corresponding GEV distribution. Choosing a different, but still large, block size size n would affect the values of the GEV parameters, but not those of the corresponding generalized Pareto distribution of threshold excesses : ξ is invariant to block size, while the calculation of $\tilde{\sigma}$ is unperturbed by the changes in μ and σ which are self-compensating. (COLES [2001], page 75).

Pour établir le résultat précédent, nous remarquons que

$$\mathbf{F}^n(x) \approx \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (2.74)$$

Nous en déduisons que

$$n \ln \mathbf{F}(x) \approx - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \quad (2.75)$$

Or $\ln \mathbf{F}(x) \approx -(1 - \mathbf{F}(x))$ pour x grand. Nous avons donc

$$1 - \mathbf{F}(x) \approx \frac{1}{n} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \quad (2.76)$$

Nous obtenons le résultat final

$$\begin{aligned} \Pr\{X > u + y \mid X > u\} &= \frac{1 - \mathbf{F}(u + y)}{1 - \mathbf{F}(u)} \\ &= \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \end{aligned} \quad (2.77)$$

Nous pouvons vérifier le résultat sur différents exemples. Si \mathbf{F} est une distribution exponentielle, nous avons

$$\frac{1 - \mathbf{F}(u + y)}{1 - \mathbf{F}(u)} = \exp(-y) \quad (2.78)$$

Dans ce cas, $\xi = 0$ et $\tilde{\sigma} = 1$. Si \mathbf{F} est maintenant une distribution uniforme, nous avons

$$\frac{1 - \mathbf{F}(u + y)}{1 - \mathbf{F}(u)} = 1 - \frac{y}{1 - u} \quad (2.79)$$

Dans ce cas, $\xi = -1$ et $\tilde{\sigma} = 1 - u$.

L'estimation des paramètres de la distribution **GPD** pose le problème de la détermination du seuil u . u doit être suffisamment grand pour que l'on puisse appliquer le résultat précédent, mais u ne doit pas être trop grand afin d'avoir suffisamment de données pour obtenir des estimateurs de bonne qualité. Généralement, u est déterminé graphiquement en exploitant le résultat suivant :

$$e(u) = \mathbb{E}[X - u \mid X > u] = \frac{\beta + \xi u}{1 - \xi} \quad (2.80)$$

si $\xi < 1$. Si l'approximation **GPD** est valide pour un seuil u_0 , alors elle est valide pour $u > u_0$. Donc, pour $u > u_0$, la fonction moyenne d'excès $e(u)$ est linéaire en u . Pour déterminer u_0 , on utilise donc un graphique appelé **mean residual life plot** qui consiste à représenter $\hat{e}(u)$ en fonction de u avec

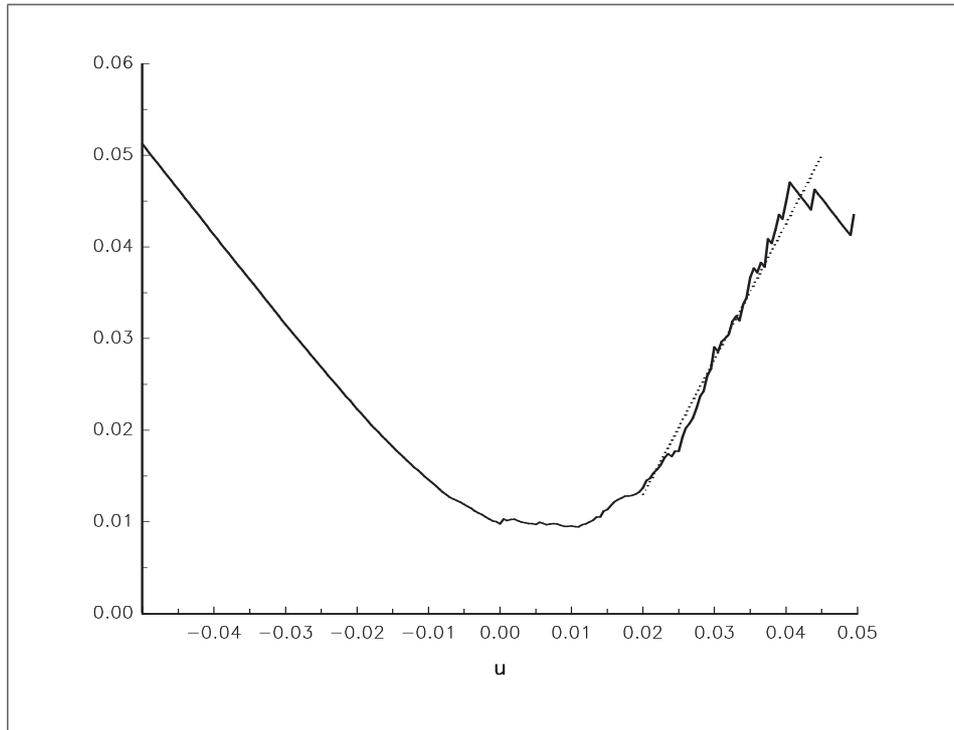
$$\hat{e}(u) = \frac{\sum (X_i - u)^+}{\sum \mathbf{1}\{X_i > u\}} \quad (2.81)$$

Lorsque u_0 est déterminé, on peut estimer les paramètres β et ξ par maximum de vraisemblance.

L'estimation graphique de u_0 n'est pas aisée dans la pratique, même lorsque nous travaillons avec des données simulées. Considérons par exemple le rendement logarithmique du CAC 40. Nous représentons sur le graphique 2.19 la fonction $\hat{e}(u)$. La détermination de u_0 consiste à déterminer les plages de linéarité. Nous avons une première plage de $u = -0.05$ à $u = -0.01$. Cette plage n'est pas acceptable car il y a ensuite un changement de pente alors que le nombre d'observations reste important. La seconde plage de linéarité va de $u = 0$ à $u = 1.5\%$. Pour la même raison, cette plage n'est pas retenue (il y a 241 observations qui présentent un rendement supérieur à 1.5%). Enfin, une dernière plage de linéarité concerne le segment $[0.02, 0.045]$. La pente de la droite est approximativement égale à 1.5, ce qui implique une valeur de ξ égale à 0.6. A titre de comparaison, les estimations ML de ξ étaient 0.45 et 0.56 pour 22 et 44 jours de trading.

Références

- [1] BALAKRISHNAN, N. and A. CLIFFORD COHEN [1991], Order Statistics and Inference, Academic Press, San Diego
- [2] COLES, S. [2001], An Introduction to Statistical Modeling of Extremes Values, *Springer Series in Statistics*, Springer-Verlag, Londres
- [3] DAVID, H. A. [1981], Order Statistics, *Wiley Series in Probability and Mathematical Statistics*, seconde édition, John Wiley & Sons, New York (première édition : 1970)



Graphique 2.19. **Mean residual life plot** du rendement logarithmique du CAC 40

- [4] DREES, H., L. DE HAAN et S. RESNICK [1998], How to make a Hill plot, *University of Cologne*, Technical report
- [5] FRACHOT, A., P. GEORGES et T. RONCALLI [2001], Loss Distribution Approach for operational risk, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [6] GNEDENKO, B. [1943], Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**, 423-453
- [7] EMBRECHTS, P., C. KLÜPPELBERG et T. MIKOSCH [1997], Modelling Extremal Events for Insurance and Finance, Springer-Verlag, Berlin
- [8] GALAMBOS, J. [1987], The Asymptotic Theory of Extreme Order Statistics, seconde édition, Krieger (première édition : John Wiley & Sons, 1978)
- [9] GUMBEL, E.J. [1954], Statistical Theory of Extreme Values and Some Practical Applications, *Applied Mathematics Series*, **33**, National Bureau of Standards, Washington
- [10] GUMBEL, E.J. [1958], Statistics of Extremes, Columbia University Press, Columbia
- [11] KOTZ, S. et S. NADARAJAH [2000], Extreme Value Distributions, Imperial College Press, London
- [12] RESNICK, S.I. [1987], Extreme Values, Point Processes and Regular Variation, Springer-Verlag, New York

3

Extension au cas multidimensionnel

Les textes consacrés à la théorie des extrêmes dans le cas multidimensionnel sont généralement difficiles d'accès. Dans ce chapitre, j'aborderai cette théorie en utilisant un outil statistique, la copule, qui permet de réduire considérablement les difficultés mathématiques.

3.1 Une introduction aux copules

Le texte de base sur le sujet est l'ouvrage de NELSEN [1999], qui peut être complété par le livre de JOE [1997]. Concernant les applications financières, vous pouvez vous référer aux différents travaux du GRO sur ce domaine :

http://gro.creditlyonnais.fr/content/rd/home_copulas.htm

3.1.1 Représentation canonique d'une distribution bivariée

Une copule est un outil statistique relativement ancien introduit par SKLAR [1959].

Définition 2 Une copule bidimensionnelle (ou 2-copula) est une fonction \mathbf{C} qui possède les propriétés suivantes :

1. $\text{Dom } \mathbf{C} = [0, 1] \times [0, 1]$;
2. $\mathbf{C}(0, u) = \mathbf{C}(u, 0) = 0$ et $\mathbf{C}(u, 1) = \mathbf{C}(1, u) = u$ pour tout u dans $[0, 1]$;
3. \mathbf{C} est 2-increasing :

$$\mathbf{C}(v_1, v_2) - \mathbf{C}(v_1, u_2) - \mathbf{C}(u_1, v_2) + \mathbf{C}(u_1, u_2) \geq 0 \quad (3.1)$$

pour tout $(u_1, u_2) \in [0, 1]^2$, $(v_1, v_2) \in [0, 1]^2$ tel que $0 \leq u_1 \leq v_1 \leq 1$ et $0 \leq u_2 \leq v_2 \leq 1$.

Cette définition veut dire que \mathbf{C} est une distribution avec des marges uniformes. Nous dirons que \mathbf{C} est une distribution uniforme multidimensionnelle.

Soient \mathbf{F}_1 et \mathbf{F}_2 deux distributions de probabilité. Il est évident que $\mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2))$ définit une distribution de probabilité bidimensionnelle $\mathbf{F}(x_1, x_2)$ dont les marges sont données.

En effet, nous savons que $U_i = \mathbf{F}_i(X_i)$ définit une transformation uniforme¹ (*probability integral transformation* ou **PIT**). De plus, nous vérifions que

$$\begin{aligned} \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(\infty)) &= \mathbf{C}(\mathbf{F}_1(x_1), 1) \\ &= \mathbf{F}_1(x_1) \end{aligned} \tag{3.3}$$

Les copules sont donc un outil très puissant pour construire des distributions multidimensionnelles dont les marges sont données (“**Distributions with Given Marginals**”).

Remarque 14 *Les copules permettent donc de construire des modèles multidimensionnels **compatibles** avec les modèles unidimensionnels. Dans l’industrie financière, il n’est pas rare qu’une banque utilise un modèle sophistiqué dans le cas à une dimension (par exemple, elle peut utiliser un modèle à volatilité stochastique pour pricer les options equities), et un modèle plus simple et dégradé dans le cas à plusieurs dimensions (par exemple, elle utilisera le modèle Black et Scholes pour pricer des produits structurés). Le fait que le modèle multidimensionnel ne soit pas compatible avec le modèle unidimensionnel peut conduire à des risques très importants (par exemple, un trader peut arbitrer les deux modèles).*

Remarque 15 *La construction de modèles multidimensionnels non gaussiens est très difficile. Par exemple, si nous prenons l’ouvrage classique de KOTZ, BALAKRISHNAN et JOHNSON [2000] sur les distributions multivariées continues, nous avons très peu de familles de probabilité (exponentielle, Gamma, Dirichlet, Liouville, logistique, Pareto, exponentielle naturelle). De plus, ces différentes distributions de probabilité présentent l’inconvénient que les marges sont de même type. Il n’est donc pas possible d’avoir une distribution avec une marge gaussienne, une marge uniforme, une marge Pareto, etc. Les copules permettent de résoudre ce problème. Par exemple, nous représentons sur le graphique 3.1 une distribution bivariée dont les marges sont une Inverse Gaussienne et une Beta.*

SKLAR [1959] a prouvé un résultat encore plus intéressant :

Théorème 5 *Soit \mathbf{F} une fonction de distribution bidimensionnelle dont les marges sont \mathbf{F}_1 et \mathbf{F}_2 . Alors \mathbf{F} admet une représentation copule :*

$$\mathbf{F}(x_1, x_2) = \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \tag{3.4}$$

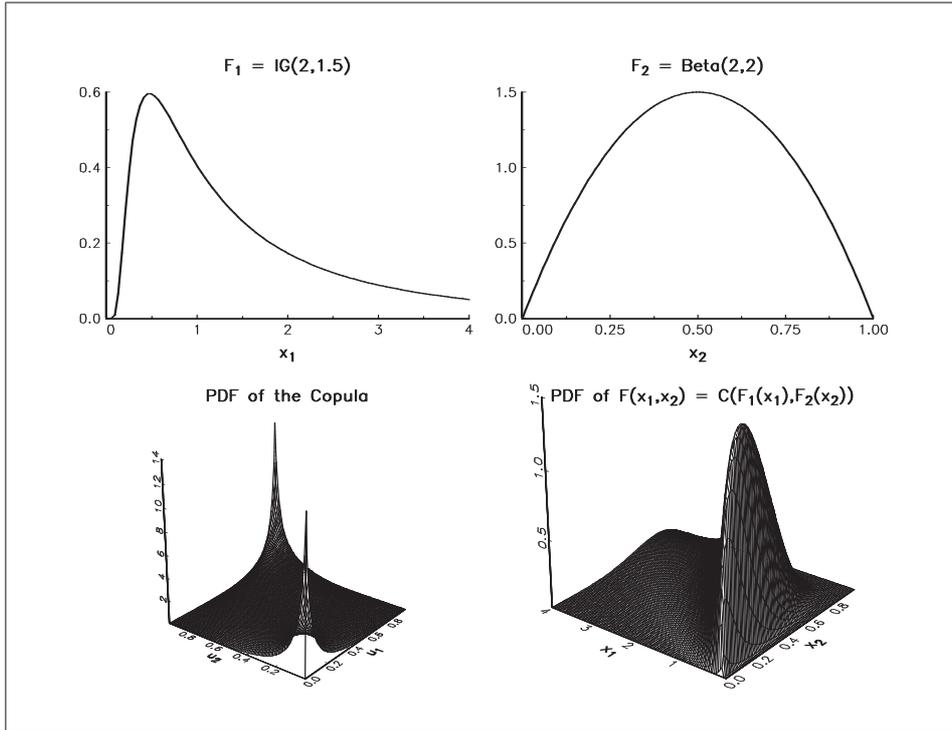
La copule \mathbf{C} est unique si les marges sont continues. Sinon, seule la sous-copule est déterminée de façon unique sur $\text{Ran } \mathbf{F}_1 \times \text{Ran } \mathbf{F}_2$.

Ce théorème est très important, puisque nous pouvons associer à chaque distribution bidimensionnelle une copule (qui peut être unique). Nous avons donc une représentation canonique de la distribution : d’un côté, nous avons les marges \mathbf{F}_1 et \mathbf{F}_2 , c’est-à-dire les directions unidimensionnelles ; d’un autre côté, nous avons la copule qui permet de relier ces marges, elle exprime la dépendance entre les directions unidimensionnelles.

Le théorème de Sklar peut être utilisée pour construire des fonctions copules à partir des distributions bidimensionnelles. Considérons par exemple le distribution logistique bivariée de

¹Soit $U = \mathbf{F}(X)$. Le support de U est évidemment $[0, 1]$. Et nous vérifions que la distribution de U est uniforme :

$$\begin{aligned} \Pr\{U \leq u\} &= \Pr\{\mathbf{F}(X) \leq u\} \\ &= \Pr\{X \leq \mathbf{F}^{-1}(u)\} \quad (\text{car } \mathbf{F} \text{ est une fonction croissante}) \\ &= \mathbf{F}(\mathbf{F}^{-1}(u)) \\ &= u \end{aligned} \tag{3.2}$$



Graphique 3.1. Un exemple de distribution bivariable avec des marges données

Gumbel

$$\mathbf{F}(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}$$

définie sur \mathbb{R}^2 . Nous pouvons montrer que les marges sont $\mathbf{F}_1(x_1) \equiv \mathbf{F}(x_1, \infty) = (1 + e^{-x_1})^{-1}$ et $\mathbf{F}_2(x_2) = (1 + e^{-x_2})^{-1}$. Les fonctions inverses sont donc $\mathbf{F}_1^{-1}(u_1) = \ln u_1 - \ln(1 - u_1)$ et $\mathbf{F}_2^{-1}(u_2) = \ln u_2 - \ln(1 - u_2)$. Nous en déduisons que la fonction copule est

$$\begin{aligned} \mathbf{C}(u_1, u_2) &= \mathbf{F}(\mathbf{F}_1^{-1}(u_1), \mathbf{F}_2^{-1}(u_2)) \\ &= \left(1 + \frac{1 - u_1}{u_1} + \frac{1 - u_2}{u_2}\right)^{-1} \\ &= \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2} \end{aligned} \quad (3.5)$$

Cette copule est appelée la copule logistique de Gumbel (Gumbel Logistic copula). Dans certains cas, nous ne pouvons pas obtenir une expression analytique de la copule. C'est par exemple le cas de la copule Normale :

$$\mathbf{C}(u_1, u_2) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)$$

Si la distribution bivariable est absolument continue, alors elle admet une densité et nous avons

$$f(x_1, x_2) = c(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \times f_1(x_1) \times f_2(x_2) \quad (3.6)$$

avec $c(u_1, u_2)$ la densité de la copule \mathbf{C} . Notons que la condition

$$\mathbf{C}(v_1, v_2) - \mathbf{C}(v_1, u_2) - \mathbf{C}(u_1, v_2) + \mathbf{C}(u_1, u_2) \geq 0 \quad (3.7)$$

est équivalente à la positivité de la densité

$$c(u_1, u_2) = \partial_{1,2}^2 \mathbf{C}(u_1, u_2) \geq 0 \quad (3.8)$$

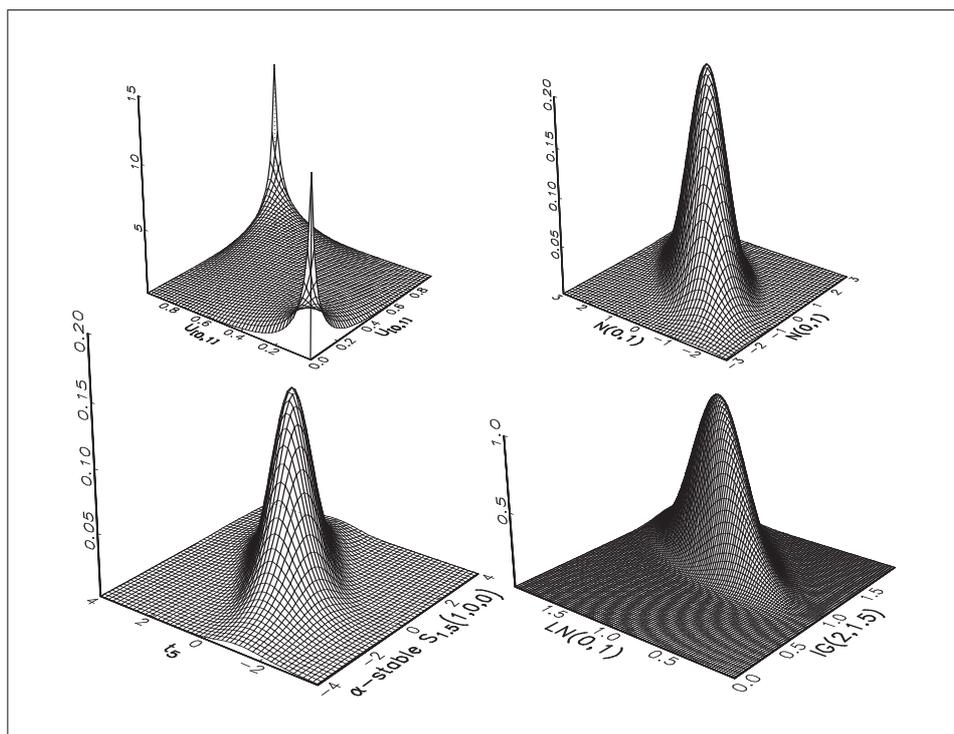
lorsque celle-ci existe. Nous pouvons utiliser la relation (3.6) pour caractériser la densité de la copule Normale. Nous avons

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} [x_1^2 + x_2^2 - 2\rho x_1 x_2]\right) = c(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x_1^2\right) \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x_2^2\right) \quad (3.9)$$

d'où

$$c(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} [x_1^2 + x_2^2 - 2\rho x_1 x_2] + \frac{1}{2} [x_1^2 + x_2^2]\right) \quad (3.10)$$

A titre d'illustration, le graphique 3.2 présente quatre distributions bivariées avec une copule Normale de paramètre $\rho = 0,7071$. **Ces quatre distributions ont donc la même structure de dépendance!** Notons aussi qu'il est très facile de construire des distributions bivariées dont les marges sont gaussiennes mais dont la distribution jointe n'est pas gaussienne (voir le graphique 3.3). **Un vecteur aléatoire dont les composantes sont gaussiennes n'est donc pas nécessairement un vecteur aléatoire gaussien!**

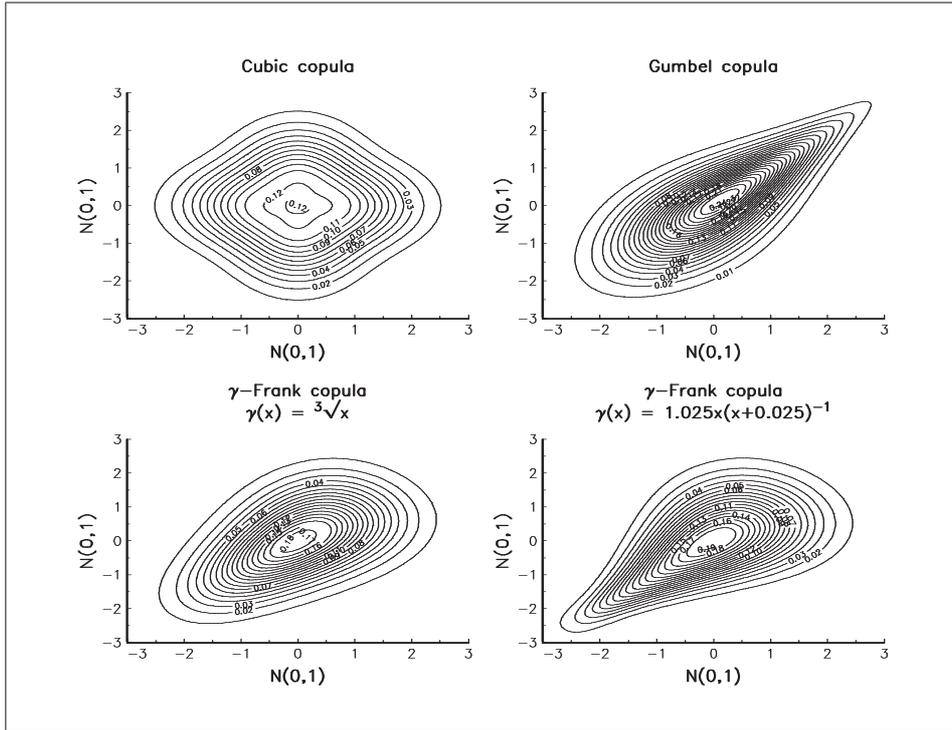


Graphique 3.2. Distributions bivariées avec une copule Normale

3.1.2 Classe de Fréchet, distributions minimale et maximale et ordre de concordance

Définition 3 \mathbf{F} appartient à la classe de Fréchet $(\mathbf{F}_1, \mathbf{F}_2)$ et on note $\mathbf{F} \in \mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ si et seulement si les marges de \mathbf{F} sont \mathbf{F}_1 et \mathbf{F}_2 :

$$\mathbf{F}(x_1, \infty) = \mathbf{F}_1(x_1) \quad (3.11)$$



Graphique 3.3. Distributions bivariées avec des marges gaussiennes

et

$$\mathbf{F}(\infty, x_2) = \mathbf{F}_2(x_2) \quad (3.12)$$

Caractériser la classe de Fréchet $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ revient à trouver l'ensemble \mathcal{C} des fonctions copules puisque nous avons

$$\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2) = \{\mathbf{F} : \mathbf{F}(x_1, x_2) = \mathbf{C}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)), \mathbf{C} \in \mathcal{C}\} \quad (3.13)$$

Le problème de la caractérisation de la classe de Fréchet $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ est donc indépendante des marges \mathbf{F}_1 et \mathbf{F}_2 .

Les distributions extrémales \mathbf{F}^- et \mathbf{F}^+ de $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ sont

$$\mathbf{F}^-(x_1, x_2) = \max(\mathbf{F}_1(x_1) + \mathbf{F}_2(x_2) - 1, 0) \quad (3.14)$$

et

$$\mathbf{F}^+(x_1, x_2) = \min(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \quad (3.15)$$

\mathbf{F}^- et \mathbf{F}^+ sont appelées respectivement les bornes basse et haute de Fréchet. Les copules associées sont

$$\mathbf{C}^-(u_1, u_2) = \max(u_1 + u_2 - 1, 0) \quad (3.16)$$

et

$$\mathbf{C}^+(u_1, u_2) = \min(u_1, u_2) \quad (3.17)$$

Ce sont les copules extrémales de \mathcal{C} . Nous avons

$$\mathbf{C}^-(u_1, u_2) \leq \mathbf{C}(u_1, u_2) \leq \mathbf{C}^+(u_1, u_2) \quad (3.18)$$

pour tout $(u_1, u_2) \in \mathbb{R}^2$ et $\mathbf{C} \in \mathcal{C}$.

Remarque 16 *Considérons la classe de Fréchet $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$ avec $\mathbf{F}_1 \stackrel{d}{=} \mathbf{F}_2 \stackrel{d}{=} \mathcal{N}(0, 1)$. Ainsi, la distribution normale à deux dimensions de corrélation ρ appartient à $\mathcal{F}(\mathbf{F}_1, \mathbf{F}_2)$. Néanmoins, beaucoup d'autres distributions non normales appartiennent à cette classe de Fréchet. C'est par exemple le cas de la distribution suivante*

$$\mathbf{F}(x_1, x_2) = \frac{\Phi(x_1)\Phi(x_2)}{\Phi(x_1) + \Phi(x_2) - \Phi(x_1)\Phi(x_2)} \quad (3.19)$$

Considérons la distribution normale

$$\mathbf{F}(x_1, x_2; \rho) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}[x_1^2 + x_2^2 - 2\rho x_1 x_2]\right) dx_1 dx_2 \quad (3.20)$$

Nous avons

$$\mathbf{F}^-(x_1, x_2) := \mathbf{F}(x_1, x_2; -1) = \max(\Phi(x_1) + \Phi(x_2) - 1, 0) \quad (3.21)$$

et

$$\mathbf{F}^+(x_1, x_2) := \mathbf{F}(x_1, x_2; +1) = \min(\Phi(x_1), \Phi(x_2)) \quad (3.22)$$

Le cas $\rho = 0$ correspond à l'indépendance des variables aléatoires et nous avons

$$\mathbf{F}^\perp(x_1, x_2) := \mathbf{F}(x_1, x_2; 0) = \Phi(x_1)\Phi(x_2) \quad (3.23)$$

Nous en déduisons l'expression suivante de la copule produit :

$$\mathbf{C}^\perp(u_1, u_2) = u_1 u_2 \quad (3.24)$$

Nous allons maintenant introduire un ordre partiel sur les copules.

Définition 4 *Soient \mathbf{C}_1 et \mathbf{C}_2 deux copules. On dit que \mathbf{C}_1 est plus petite que \mathbf{C}_2 et on note $\mathbf{C}_1 \prec \mathbf{C}_2$ si et seulement si $\mathbf{C}_1(u_1, u_2) \leq \mathbf{C}_2(u_1, u_2)$ pour tout $(u_1, u_2) \in [0, 1]^2$.*

Cette relation d'ordre est **partielle** car on ne peut pas comparer toutes les copules entre elles. Néanmoins, nous avons toujours

$$\mathbf{C}^- \prec \mathbf{C} \prec \mathbf{C}^+ \quad (3.25)$$

Nous en déduisons qu'une structure de dépendance **positive** est une fonction copule \mathbf{C} qui vérifie l'inégalité suivante :

$$\mathbf{C}^\perp \prec \mathbf{C} \prec \mathbf{C}^+ \quad (3.26)$$

De même, une structure de dépendance **négative** est une fonction copule \mathbf{C} qui vérifie l'inégalité suivante :

$$\mathbf{C}^- \prec \mathbf{C} \prec \mathbf{C}^\perp \quad (3.27)$$

Cependant, comme cette relation d'ordre est partielle, il existe des fonctions copules \mathbf{C} telles que $\mathbf{C} \not\prec \mathbf{C}^\perp$ et $\mathbf{C} \not\prec \mathbf{C}^+$. Il existe donc des fonctions copules qui ne sont ni des structures de dépendance **positive**, ni des structures de dépendance **négative**.

Remarque 17 *L'ordre \prec est appelé l'ordre de concordance. Il correspond à la dominance stochastique du premier ordre sur les fonctions de distribution.*

Remarque 18 Considérons la copule $\mathbf{C}(u_1, u_2; \theta)$ avec θ un paramètre. Nous notons $\mathbf{C}_\theta(u_1, u_2) = \mathbf{C}(u_1, u_2; \theta)$. La famille de copules $\{\mathbf{C}_\theta\}$ est dite *totallement ordonnée* si pour tout $\theta_2 \geq \theta_1$ nous vérifions

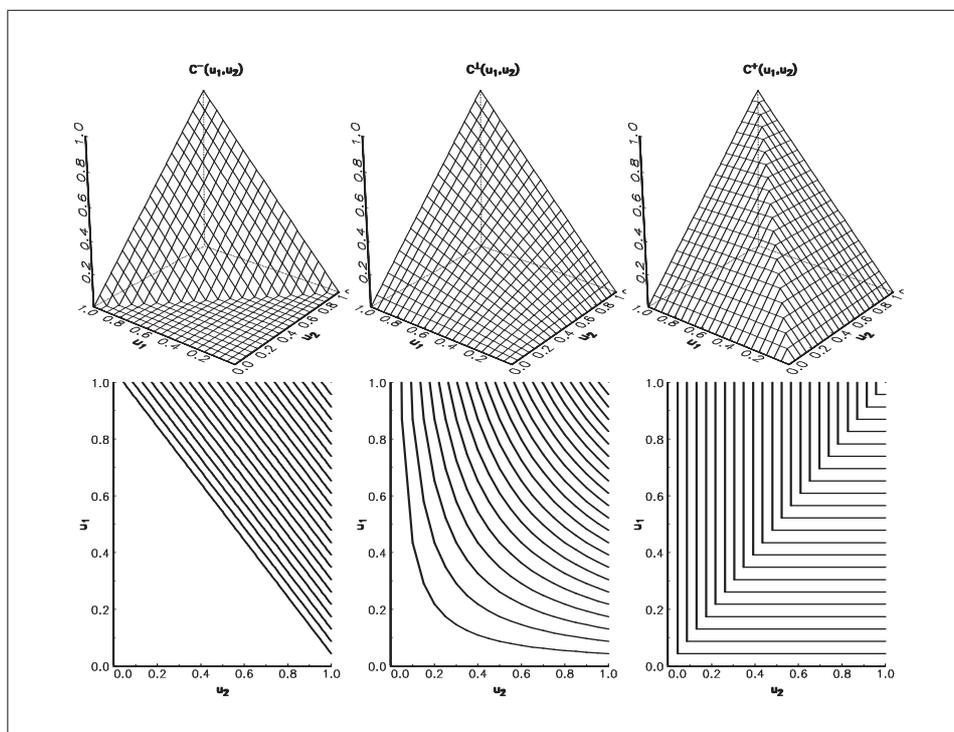
$$\mathbf{C}_{\theta_2} \succ \mathbf{C}_{\theta_1} \quad (3.28)$$

(famille positivement ordonnée) ou

$$\mathbf{C}_{\theta_2} \prec \mathbf{C}_{\theta_1} \quad (3.29)$$

(famille négativement ordonnée). Par exemple, la famille des copules Normales $\{\mathbf{C}_\rho\}$ est positivement ordonnée.

Les courbes de niveau des copules \mathbf{C}^- , \mathbf{C}^\perp et \mathbf{C}^+ sont représentées sur le graphique 3.4. Nous vérifions graphiquement que la copule Frank (définie plus loin) est une famille positivement ordonnée.



Graphique 3.4. Les trois copules \mathbf{C}^- , \mathbf{C}^\perp et \mathbf{C}^+

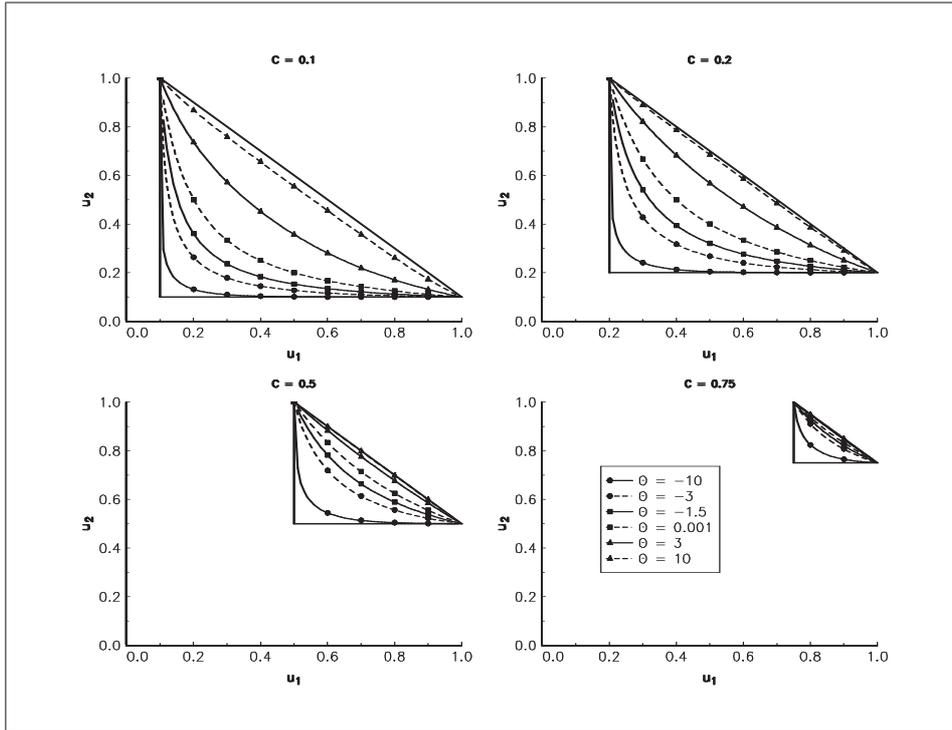
3.1.3 Copules et variables aléatoires

Considérons deux variables aléatoires X_1 et X_2 de distributions \mathbf{F}_1 et \mathbf{F}_2 . Nous notons \mathbf{F} la distribution du vecteur aléatoire $\mathbf{X} = (X_1, X_2)$. On définit alors la copule du vecteur aléatoire (X_1, X_2) par la copule de \mathbf{F} :

$$\mathbf{F}(x_1, x_2) = \mathbf{C}_{\langle X_1, X_2 \rangle}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \quad (3.30)$$

Nous pouvons donner une nouvelle interprétation des copules \mathbf{C}^- , \mathbf{C}^\perp et \mathbf{C}^+ :

- deux variables aléatoires X_1 et X_2 sont contre-monotones (**countermonotonic**) — ou $\mathbf{C}_{\langle X_1, X_2 \rangle} = \mathbf{C}^-$ — s'il existe une variable aléatoire X telle que $X_1 = f_1(X)$ et $X_2 = f_2(X)$ avec f_1 une fonction non-croissante et f_2 une fonction non-décroissante ;



Graphique 3.5. Courbes de niveau de la copule Frank

- deux variables aléatoires X_1 et X_2 sont **indépendantes** si la structure de dépendance est la copule produit \mathbf{C}^\perp ;
- deux variables aléatoires X_1 et X_2 sont co-monotones (**comonotonic**) — ou $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^+$ — s'il existe une variable aléatoire X telle que $X_1 = f_1(X)$ et $X_2 = f_2(X)$ avec f_1 et f_2 deux fonctions non-décroissantes.

Considérons un vecteur aléatoire gaussien $\mathbf{X} = (X_1, X_2)$ de corrélation ρ . $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^-$ correspond au cas

$$X_2 = -X_1 \quad (3.31)$$

alors que $\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^+$ correspond au cas

$$X_2 = X_1 \quad (3.32)$$

Si $\mathbf{X} = (X_1, X_2)$ est un vecteur aléatoire dont les marges sont lognormales. Alors nous avons

$$\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^- \Leftrightarrow X_2 = \exp(-\ln X_1) = \frac{1}{X_1} \quad (3.33)$$

et

$$\mathbf{C}\langle X_1, X_2 \rangle = \mathbf{C}^- \Leftrightarrow X_2 = X_1 \quad (3.34)$$

Nous donnons maintenant un des théorèmes les plus importants des copules :

Théorème 6 Soient deux variables aléatoires continues X_1 et X_2 de marges \mathbf{F}_1 et \mathbf{F}_2 et de copule $\mathbf{C}\langle X_1, X_2 \rangle$. Si h_1 et h_2 sont deux fonctions strictement croissantes sur $\text{Im}(X_1)$ et $\text{Im}(X_2)$ respectivement, alors

$$\mathbf{C}\langle h_1(X_1), h_2(X_2) \rangle = \mathbf{C}\langle X_1, X_2 \rangle \quad (3.35)$$

Ainsi la copule est invariante par transformations strictement croissantes des variables aléatoires.

Nous pouvons montrer ce théorème en utilisant les propriétés des distributions. Notons respectivement \mathbf{F} et \mathbf{G} les distributions des vecteurs aléatoires (X_1, X_2) et $(h_1(X_1), h_2(X_2))$. Nous remarquons que les marges de \mathbf{G} sont

$$\begin{aligned}\mathbf{G}_1(x_1) &= \Pr\{h_1(X_1) \leq x_1\} \\ &= \Pr\{X_1 \leq h_1^{-1}(x_1)\} \quad (\text{car } h_1 \text{ est une fonction croissante}) \\ &= \mathbf{F}_1(h_1^{-1}(x_1))\end{aligned}\tag{3.36}$$

et $\mathbf{G}_2(x_2) = \mathbf{F}_2(h_2^{-1}(x_2))$. Nous avons donc $\mathbf{G}_1^{-1}(u_1) = h_1(\mathbf{F}_1^{-1}(u_1))$ et $\mathbf{G}_2^{-1}(u_2) = h_2(\mathbf{F}_2^{-1}(u_2))$. Nous en déduisons le résultat :

$$\begin{aligned}\mathbf{C}\langle h_1(X_1), h_2(X_2) \rangle(u_1, u_2) &= \mathbf{G}(\mathbf{G}_1^{-1}(u_1), \mathbf{G}_2^{-1}(u_2)) \\ &= \Pr\{h_1(X_1) \leq \mathbf{G}_1^{-1}(u_1), h_2(X_2) \leq \mathbf{G}_2^{-1}(u_2)\} \\ &= \Pr\{X_1 \leq h_1^{-1}(\mathbf{G}_1^{-1}(u_1)), X_2 \leq h_2^{-1}(\mathbf{G}_2^{-1}(u_2))\} \\ &= \Pr\{X_1 \leq \mathbf{F}_1^{-1}(u_1), X_2 \leq \mathbf{F}_2^{-1}(u_2)\} \\ &= \mathbf{C}\langle X_1, X_2 \rangle(u_1, u_2)\end{aligned}\tag{3.37}$$

Par exemple, nous avons

$$\begin{aligned}\mathbf{C}\langle X_1, X_2 \rangle &= \mathbf{C}\langle \ln X_1, X_2 \rangle \\ &= \mathbf{C}\langle \ln X_1, \ln X_2 \rangle \\ &= \mathbf{C}\langle X_1, \exp X_2 \rangle \\ &= \mathbf{C}\langle \sqrt{X_1}, \exp X_2 \rangle\end{aligned}\tag{3.38}$$

L'application de transformations croissantes ne modifie donc pas la copule, seulement les marges. Ainsi, la copule de la distribution lognormale bivariée est la même que celle de la distribution normale bivariée.

3.1.4 Dépendance et corrélation

Nous pouvons interpréter la copule du vecteur aléatoire $\mathbf{X} = (X_1, X_2)$ comme une reparamétrisation ou une normalisation de la distribution jointe \mathbf{F} après avoir éliminé les effets des marges \mathbf{F}_1 et \mathbf{F}_2 . De façon tout à fait indépendante, Paul Deheuvels a redécouvert l'outil copule à la fin des années 70 et lui a donné le nom de **fonction de dépendance**. La copule est en fait la structure de dépendance de \mathbf{X} . De plus, c'est une statistique **exhaustive** de cette dépendance. Considérons une statistique (**non exhaustive**) de la dépendance, c'est-à-dire une mesure de dépendance. Celle-ci sera effectivement une mesure de dépendance si elle s'exprime exclusivement à partir de la copule.

Parmi ces statistiques, nous pouvons considérer de façon générale les mesures de concordance.

Définition 5 (Nelsen (1999), page 136) Une mesure numérique κ d'association entre deux variables aléatoires continues X_1 et X_2 dont la copule est \mathbf{C} est une mesure de concordance si elle satisfait les propriétés suivantes :

1. κ est définie pour toute paire (X_1, X_2) de variables aléatoires continues ;
2. $-1 = \kappa_{X, -X} \leq \kappa_{\mathbf{C}} \leq \kappa_{X, X} = 1$;
3. $\kappa_{X_1, X_2} = \kappa_{X_2, X_1}$;
4. si X_1 et X_2 sont indépendantes, alors $\kappa_{X_1, X_2} = \kappa_{\mathbf{C}^\perp} = 0$;
5. $\kappa_{-X_1, X_2} = \kappa_{X_1, -X_2} = -\kappa_{X_1, X_2}$;

6. si $\mathbf{C}_1 \prec \mathbf{C}_2$, alors $\kappa_{\mathbf{C}_1} \leq \kappa_{\mathbf{C}_2}$;

7. si $\{(X_{1,n}, X_{2,n})\}$ est une séquence de variables aléatoires continues dont la copule est \mathbf{C}_n , et si $\{\mathbf{C}_n\}$ converge vers \mathbf{C} (pointwise convergence), alors $\lim_{n \rightarrow \infty} \kappa_{\mathbf{C}_n} = \kappa_{\mathbf{C}}$.

Parmi toutes les mesures de concordance, trois mesures très célèbres jouent un rôle important en statistiques non paramétriques : le tau de Kendall, le rho de Spearman et l'indice de Gini. Elles peuvent toutes les trois s'exprimer à l'aide de la copule, et nous avons (SCHWEITZER et WOLFF [1981])

$$\begin{aligned}\tau &= 4 \iint_{\mathbf{I}^2} \mathbf{C}(u_1, u_2) d\mathbf{C}(u_1, u_2) - 1 \\ \varrho &= 12 \iint_{\mathbf{I}^2} u_1 u_2 d\mathbf{C}(u_1, u_2) - 3 \\ \gamma &= 2 \iint_{\mathbf{I}^2} (|u_1 + u_2 - 1| - |u_1 - u_2|) d\mathbf{C}(u_1, u_2)\end{aligned}\quad (3.39)$$

Remarque 19 Considérons un échantillon de n observations $\{(x_1, y_1), \dots, (x_n, y_n)\}$ du vecteur aléatoire (X, Y) . Le tau de Kendall est la probabilité de concordance — $(X_i - X_j)(Y_i - Y_j) > 0$ — moins la probabilité de discordance des paires — $(X_i - X_j)(Y_i - Y_j) < 0$. Le rho de Spearman est la corrélation des statistiques de rang R_X et R_Y :

$$\varrho = \text{cor}(\mathbf{F}_X(X), \mathbf{F}_Y(Y)) \quad (3.40)$$

Considérons le coefficient de corrélation linéaire de Pearson. Est-ce une mesure de concordance ? Nous rappelons que l'expression de ce coefficient est

$$\rho \langle X_1, X_2 \rangle = \frac{\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1] \mathbb{E}[X_2]}{\sigma[X_1] \sigma[X_2]} \quad (3.41)$$

En utilisant les travaux de TCHEN [1980] sur les fonctions superadditives, nous pouvons montrer les résultats suivants :

- (4) Si la copule de (X_1, X_2) est \mathbf{C}^\perp , alors $\rho \langle X_1, X_2 \rangle = 0$;
(6) ρ est une mesure croissante par rapport à l'ordre de concordance :

$$\mathbf{C}_1 \succ \mathbf{C}_2 \Rightarrow \rho_1 \langle X_1, X_2 \rangle \geq \rho_2 \langle X_1, X_2 \rangle \quad (3.42)$$

(2') $\rho \langle X_1, X_2 \rangle$ est bornée

$$\rho^- \langle X_1, X_2 \rangle \leq \rho \langle X_1, X_2 \rangle \leq \rho^+ \langle X_1, X_2 \rangle \quad (3.43)$$

et les bornes sont atteintes pour les copules de Fréchet \mathbf{C}^- et \mathbf{C}^+ .

Il est évident que les propriétés (1), (3), (5) et (7) sont aussi vérifiées. Pour la propriété (5), nous avons en effet

$$\begin{aligned}\rho \langle -X_1, X_2 \rangle &= \frac{\mathbb{E}[-X_1 X_2] - \mathbb{E}[-X_1] \mathbb{E}[X_2]}{\sigma[-X_1] \sigma[X_2]} \\ &= \frac{\mathbb{E}[-X_1 X_2] - \mathbb{E}[X_1] \mathbb{E}[-X_2]}{\sigma[X_1] \sigma[-X_2]} \\ &= \rho \langle X_1, -X_2 \rangle \\ &= -\frac{\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1] \mathbb{E}[X_2]}{\sigma[X_1] \sigma[X_2]} \\ &= -\rho \langle X_1, X_2 \rangle\end{aligned}\quad (3.44)$$

Néanmoins, si les propriétés (1)-(7) sont vérifiées, alors nous avons

$$\kappa_{\mathbf{C}^-} = -1 \quad (3.45)$$

et

$$\kappa_{\mathbf{C}^+} = -1 \quad (3.46)$$

Si nous utilisons l'interprétation des bornes de Fréchet en terme de variables aléatoires, nous avons

$$\rho^-(X_1, X_2) = \rho^+(X_1, X_2) = \frac{\mathbb{E}[f_1(X) f_2(X)] - \mathbb{E}[f_1(X)] \mathbb{E}[f_2(X)]}{\sigma[f_1(X)] \sigma[f_2(X)]} \quad (3.47)$$

Janos Aczel a montré que la solution de l'équation $\rho^-(X_1, X_2) = -1$ est

$$f_1(x) = a_1x + b \quad (3.48)$$

et

$$f_2(x) = a_2x + b \quad (3.49)$$

avec $a_1a_2 < 0$. Pour l'équation $\rho^+(X_1, X_2) = 1$, la condition est $a_1a_2 > 0$. En dehors des variables aléatoires gaussiennes, il existe très peu de variables aléatoires capables de satisfaire ces conditions. Notons aussi que si le coefficient de corrélation de Pearson est une mesure de concordance, alors **il doit être invariant par transformations croissantes** :

$$\rho\langle X_1, X_2 \rangle = \rho\langle f_1(X_1), f_2(X_2) \rangle \quad (3.50)$$

Une nouvelle fois, la solution de cette équation est $f_1(x) = a_1x + b$ et $f_2(x) = a_2x + b$ avec $a_1 > 0$ et $a_2 > 0$. Nous comprenons beaucoup mieux que cette mesure est dite **linéaire**.

A titre d'illustration, nous considérons le modèle de Black et Scholes avec deux actifs. Les prix des actifs $S_1(t)$ and $S_2(t)$ sont des processus de diffusion dont la représentation SDE est

$$\begin{cases} dS_1(t) = \mu_1 S_1(t) dt + \sigma_1 S_1(t) dW_1(t) \\ dS_2(t) = \mu_2 S_2(t) dt + \sigma_2 S_2(t) dW_2(t) \end{cases} \quad (3.51)$$

avec $W_1(t)$ et $W_2(t)$ deux \mathcal{F}_t^- mouvements browniens avec

$$\mathbb{E}[W_1(t) W_2(t) | \mathcal{F}_{t_0}] = \rho(t - t_0) \quad (3.52)$$

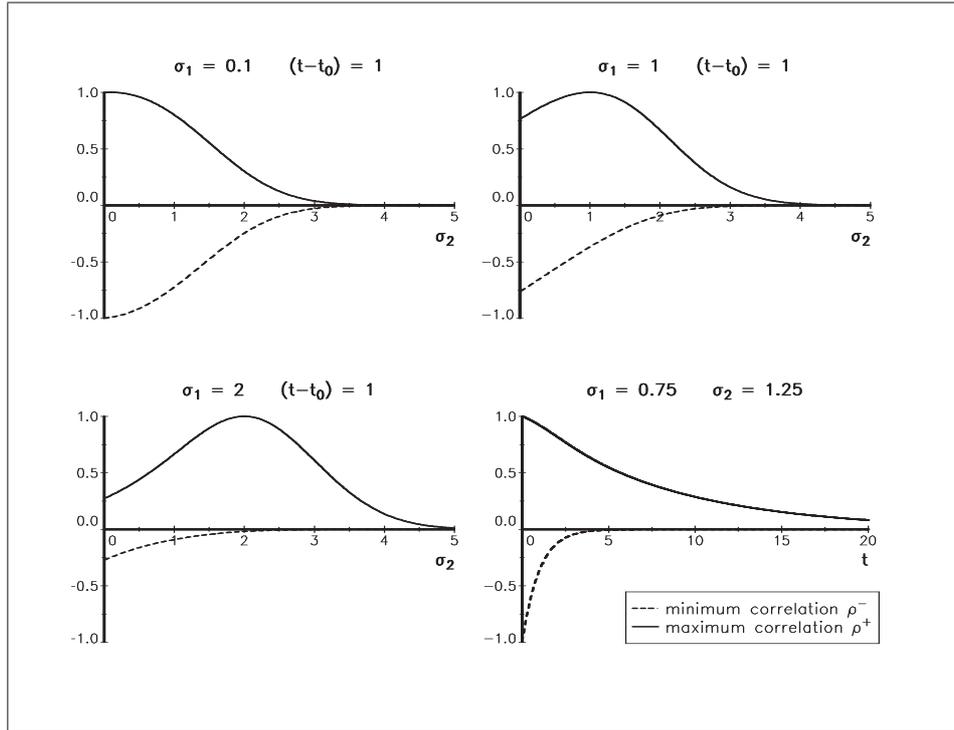
Nous pouvons montrer que

$$\rho^- \leq \rho\langle S_1(t), S_2(t) \rangle \leq \rho^+ \quad (3.53)$$

avec

$$\rho^\pm = \frac{\exp(\pm\sigma_1\sigma_2(t-t_0)) - 1}{\sqrt{\exp(\sigma_1^2(t-t_0)) - 1} \cdot \sqrt{\exp(\sigma_2^2(t-t_0)) - 1}} \quad (3.54)$$

$\rho(S_1(t), S_2(t)) = \rho^-$ (resp. ρ^+) si et seulement $\mathbf{C}\langle S_1(t), S_2(t) \rangle = \mathbf{C}^-$ ou encore $\rho\langle W_1(t), W_2(t) \rangle =$



Graphique 3.6. Intervalle de corrélation $[\rho^-, \rho^+]$ lorsque $(S_1(t), S_2(t))$ est un processus GBM bidimensionnel

-1 (resp. $\mathbf{C} \langle S_1(t), S_2(t) \rangle = \mathbf{C}^+$ ou encore $\rho \langle W_1(t), W_2(t) \rangle = +1$). Par exemple, si nous fixons $\sigma_1 = 1$, $\sigma_2 = 3$ et $t - t_0 = 1$, nous obtenons les résultats suivants

Copule	$\rho \langle S_1(t), S_2(t) \rangle$	$\tau \langle S_1(t), S_2(t) \rangle$	$\varrho \langle S_1(t), S_2(t) \rangle$
\mathbf{C}^-	-0.008	-1	-1
$\rho = -0.7$	$\simeq 0$	-0.49	-0.68
\mathbf{C}^\perp	0	0	0
$\rho = 0.7$	$\simeq 0.10$	0.49	0.68
\mathbf{C}^+	0.16	1	1

Définition 6 (Nelsen (1999), page 170) Une mesure numérique δ d'association entre deux variables aléatoires continues X_1 et X_2 dont la copule est \mathbf{C} est une mesure de dépendance si elle satisfait les propriétés suivantes :

1. δ est définie pour toute paire (X_1, X_2) de variables aléatoires continues ;
2. $0 = \delta_{\mathbf{C}^\perp} \leq \delta_{\mathbf{C}} \leq \delta_{\mathbf{C}^+} = 1$;
3. $\delta_{X_1, X_2} = \delta_{X_2, X_1}$;
4. $\delta_{X_1, X_2} = \delta_{\mathbf{C}^\perp} = 0$ si et seulement si X_1 et X_2 sont indépendantes ;
5. $\delta_{X_1, X_2} = \delta_{\mathbf{C}^+} = \delta_{\mathbf{C}^-} = 1$ si et seulement si X_1 est une fonction strictement monotone de X_2 ;
6. si h_1 et h_2 sont des fonctions strictement monotones respectivement sur $\text{Im } X_1$ et $\text{Im } X_2$, alors

$$\delta_{h_1(X_1), h_2(X_2)} = \delta_{X_1, X_2}$$

7. si $\{(X_{1,n}, X_{2,n})\}$ est une séquence de variables aléatoires continues dont la copule est \mathbf{C}_n , et si $\{\mathbf{C}_n\}$ converge vers \mathbf{C} (pointwise convergence), alors $\lim_{n \rightarrow \infty} \delta_{\mathbf{C}_n} = \delta_{\mathbf{C}}$.

SCHWEITZER et WOLFF [1981] donnent différentes mesures qui vérifient cette définition :

$$\begin{aligned}\sigma &= 12 \iint_{\mathbf{I}^2} |\mathbf{C}(u_1, u_2) - \mathbf{C}^\perp(u_1, u_2)| \, du_1 \, du_2 \\ \Phi^2 &= 90 \iint_{\mathbf{I}^2} |\mathbf{C}(u_1, u_2) - \mathbf{C}^\perp(u_1, u_2)|^2 \, du_1 \, du_2\end{aligned}\quad (3.55)$$

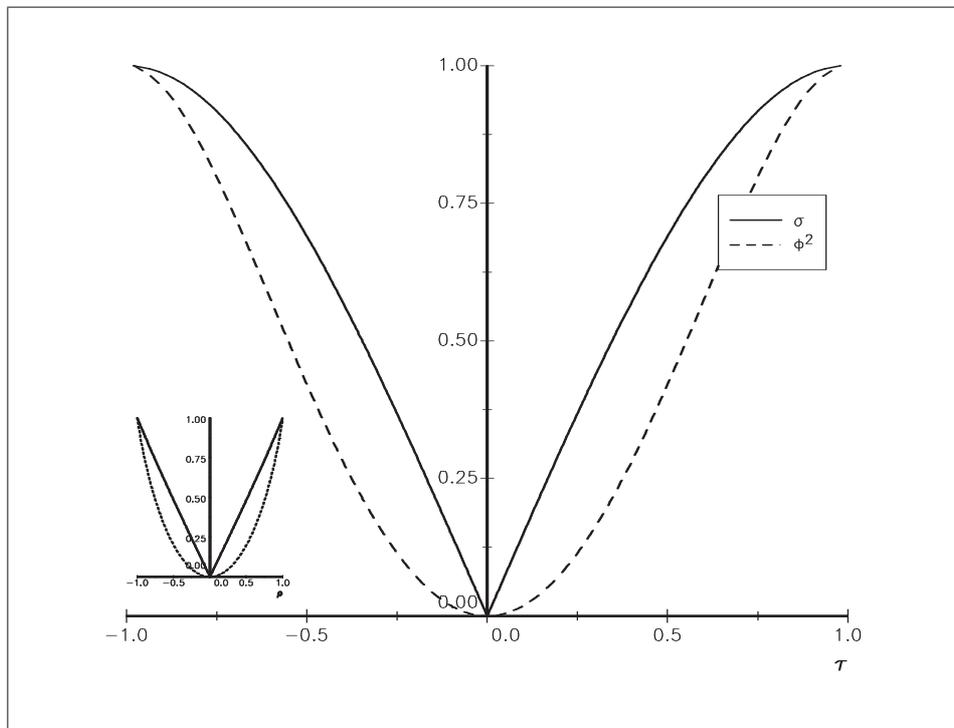
Remarque 20 Nous pouvons interpréter les mesures de concordance comme une généralisation de la corrélation linéaire dans le cas où la distribution bivariable n'est pas gaussienne. Elles prennent ainsi la valeur -1 pour la structure de dépendance la plus négative \mathbf{C}^- (qui correspond à une corrélation linéaire égale à -1 dans le cas gaussien) et +1 pour la structure de dépendance la plus positive \mathbf{C}^+ (qui correspond à une corrélation linéaire égale à +1 dans le cas gaussien). Ces deux copules représentent la forme de dépendance la plus extrême, c'est-à-dire que les variables aléatoires X_1 et X_2 sont parfaitement dépendantes puisque

$$X_2 = f(X_1) \quad (3.56)$$

avec f une fonction déterministe. Néanmoins, ces mesures de dépendance ne prennent la valeur 1 que si cette dépendance est **monotone**. Donc, il existe de nombreuses copules qui correspondent à des dépendances parfaites et dont δ est strictement inférieur à 1 (les copules les plus célèbres sont les "shuffles of min").

Remarque 21 Pour la famille des copules Normales, nous vérifions que

$$\delta \langle \mathbf{C}_\rho \rangle = \delta \langle \mathbf{C}_{-\rho} \rangle \quad (3.57)$$



Graphique 3.7. Mesures de dépendance σ et Φ^2 pour la copule Normale

Remarque 22 $\delta \langle \mathbf{C} \rangle = 0$ si et seulement si $\mathbf{C} = \mathbf{C}^\perp$. Cette propriété n'est pas vérifiée par les mesures de concordance et la corrélation linéaire.

3.1.5 Extension au cas multidimensionnel

Les résultats précédents se généralisent au cas multidimensionnel de façon assez directe. Néanmoins, le cas multidimensionnel présente quelques difficultés par rapport au cas bidimensionnel. La principale est la notion de dépendance négative. Par exemple, comment pouvons-nous caractériser la dépendance négative extrême ? Prenons par exemple le cas d'un vecteur gaussien (X_1, X_2, X_3) . Supposons que X_1 soit parfaitement anti-corrélé avec X_2 et que X_2 soit parfaitement anti-corrélé avec X_3 . Alors X_1 est parfaitement corrélé avec X_3 . L'expression de la matrice de corrélation est donc

$$\rho = \begin{bmatrix} 1 & -1 & 1 \\ & 1 & -1 \\ & & 1 \end{bmatrix} \quad (3.58)$$

Cela veut dire que

$$\begin{bmatrix} 1 & -1 & -1 \\ & 1 & -1 \\ & & 1 \end{bmatrix} \notin \rho \langle X_1, X_2, X_3 \rangle \quad (3.59)$$

Ceci explique que la borne de Fréchet minimale n'existe pas (dans le cas général). Même si nous avons

$$\mathbf{C}^- \prec \mathbf{C} \prec \mathbf{C}^+ \quad (3.60)$$

avec

$$\mathbf{C}^-(u_1, \dots, u_n) = \max \left(\sum_{i=1}^n u_i - n + 1, 0 \right) \quad (3.61)$$

et

$$\mathbf{C}^+(u_1, \dots, u_n) = \min(u_1, \dots, u_n) \quad (3.62)$$

\mathbf{C}^- n'est pas une copule.

3.1.6 Quelques exemples de familles de copules

Nous avons déjà vu que la copule Normale est définie de la façon suivante :

$$\mathbf{C}(u_1, u_2; \rho) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho) \quad (3.63)$$

La copule associée à la distribution de Student bivariée est

$$\mathbf{C}_\nu(u_1, u_2; \rho) = \mathbf{t}_\nu(\mathbf{t}_\nu^{-1}(u_1), \mathbf{t}_\nu^{-1}(u_2); \rho) \quad (3.64)$$

La copule Frank correspond à

$$\mathbf{C}(u_1, u_2; \theta) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right) \quad (3.65)$$

avec $\theta \in \mathbb{R}$.

Remarque 23 Ces trois familles de copules sont dites des familles compréhensives (comprehensive family) car elles contiennent les trois copules \mathbf{C}^- , \mathbf{C}^\perp et \mathbf{C}^+ .

Certaines familles de copules présentent des structures de dépendance plus restreintes. C'est par exemple le cas de la copule Gumbel définie par²

$$\mathbf{C}(u_1, u_2; \theta) = \exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right) \quad (3.66)$$

qui vérifie l'inégalité suivante

$$\mathbf{C}^\perp \prec \mathbf{C} \prec \mathbf{C}^+ \quad (3.67)$$

La copule FGM (Farlie-Gumbel-Morgenstern) a pour expression

$$\mathbf{C}(u_1, u_2; \theta) = u_1 u_2 + \theta u_1 u_2 (1 - u_1)(1 - u_2) \quad (3.68)$$

avec $\theta \in [-1, 1]$. Cette copule est une perturbation de la copule produit et nous vérifions que

$$-\frac{1}{3} \leq \rho\langle \mathbf{C} \rangle \leq \frac{1}{3} \quad (3.69)$$

Le rho de Spearman prend des valeurs dans un intervalle très restreint. Nous ne pouvons donc pas utiliser cette copule pour modéliser des variables aléatoires qui présentent une forte dépendance.

A titre d'illustration, nous construisons différentes distributions avec les copules Normale, Frank et Gumbel. Afin de pouvoir les comparer, les paramètres des copules ont été calibrés de telle façon que le tau de Kendall soit égal à 0.5. Cela veut dire que toutes ces distributions présentent la même dépendance au sens du tau de Kendall. Pourtant, nous voyons qu'elles sont difficilement comparables. Cependant, les distributions d'un même graphe ont la même dépendance puisqu'elles ont la même copule.

Nous pouvons construire des distributions bivariées qui sont encore moins comparables. Même si les copules Normale, Frank et Gumbel sont différentes (en particulier concernant le comportement des extrêmes), elles sont d'un certain point de vue assez "similaires". En fait, l'ensemble de ces trois copules est **très petit** par rapport à l'ensemble des copules \mathcal{C} . Il existe donc des structures de dépendance qui ne ressemblent pas du tout aux trois copules précédentes. Pour illustrer ce problème, nous considérons la région atteignable $\mathcal{B}(\tau, \rho)$ par les statistiques (τ, ρ) qui est définie par

$$(\tau, \rho) \in \mathcal{B}(\tau, \rho) \Leftrightarrow \begin{cases} \frac{3\tau-1}{2} \leq \rho \leq \frac{1+2\tau-\tau^2}{2} & \tau \geq 0 \\ \frac{\tau^2+2\tau-1}{2} \leq \rho \leq \frac{1+3\tau}{2} & \tau \leq 0 \end{cases} \quad (3.70)$$

On peut montrer que ces bornes ne peuvent pas être améliorées et qu'il existe toujours une copule qui correspond à un point de la frontière $\mathcal{B}(\tau, \rho)$. Sur le graphique 3.11, nous représentons la frontière de $\mathcal{B}(\tau, \rho)$, ainsi que la zone atteinte par les dix copules les plus communes B1–B10 (JOE [1997]). Nous voyons que ces copules occupent une toute petite surface de la région $\mathcal{B}(\tau, \rho)$. Du point de vue des statistiques (τ, ρ) , ces dix familles sont relativement similaires. Cela veut donc dire qu'il existe des copules qui ont des comportements très différents de ces dix copules.

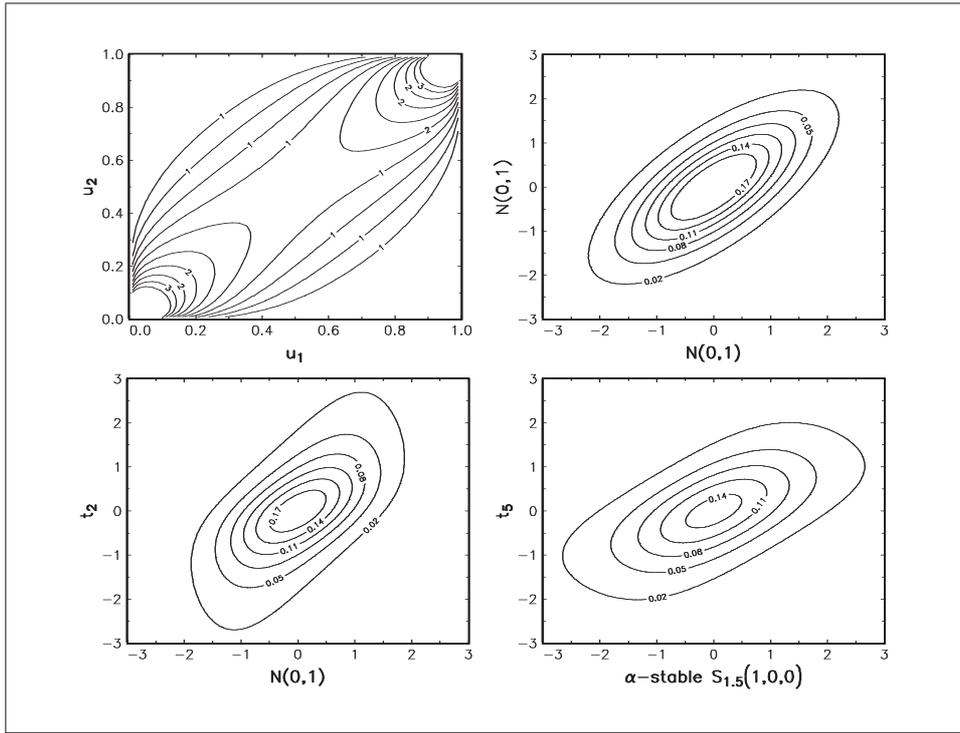
3.2 Les copules de valeurs extrêmes

Définition 7 Une copule de valeurs extrêmes (extreme value copula) vérifie la relation suivante :

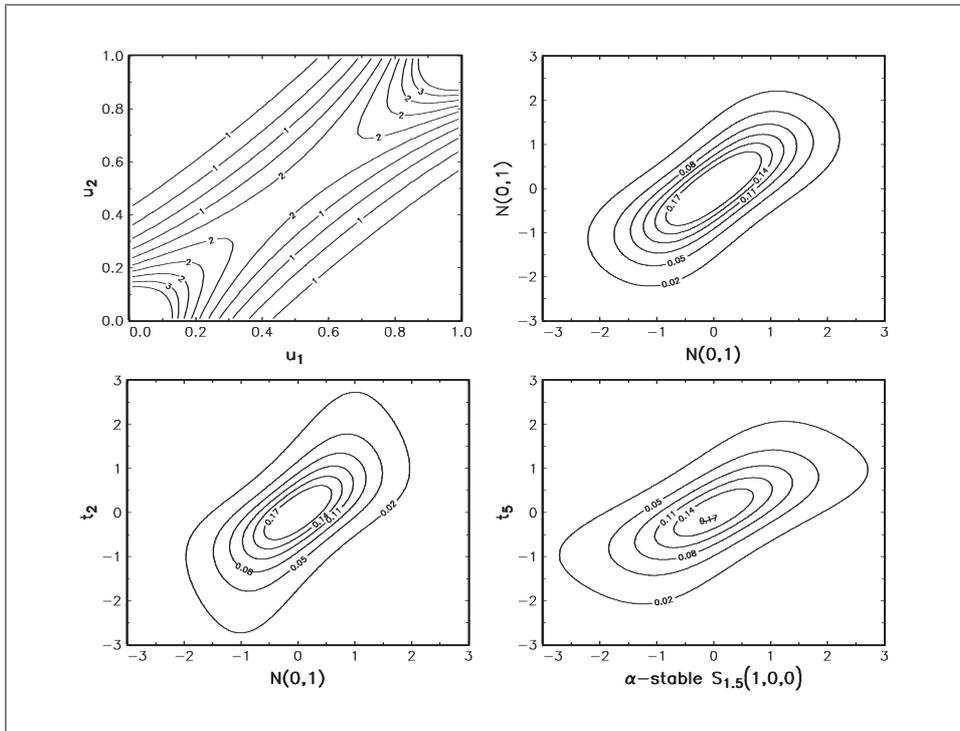
$$\mathbf{C}(u_1^t, \dots, u_n^t) = \mathbf{C}^t(u_1, \dots, u_n) \quad (3.71)$$

pour tout réel t positif.

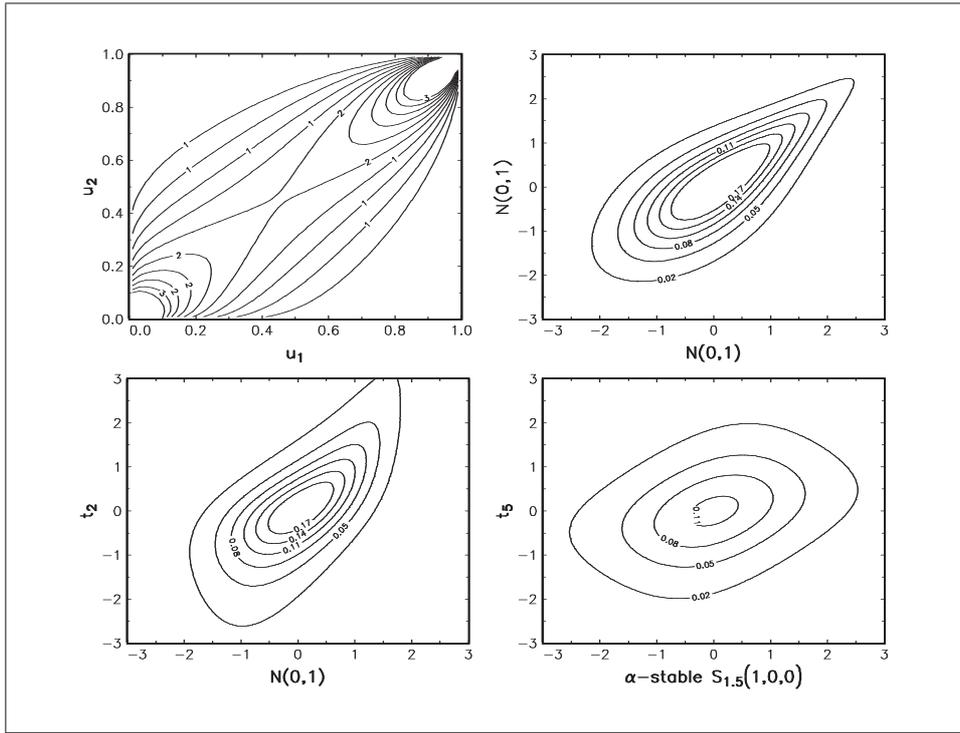
² $\theta \geq 0$.



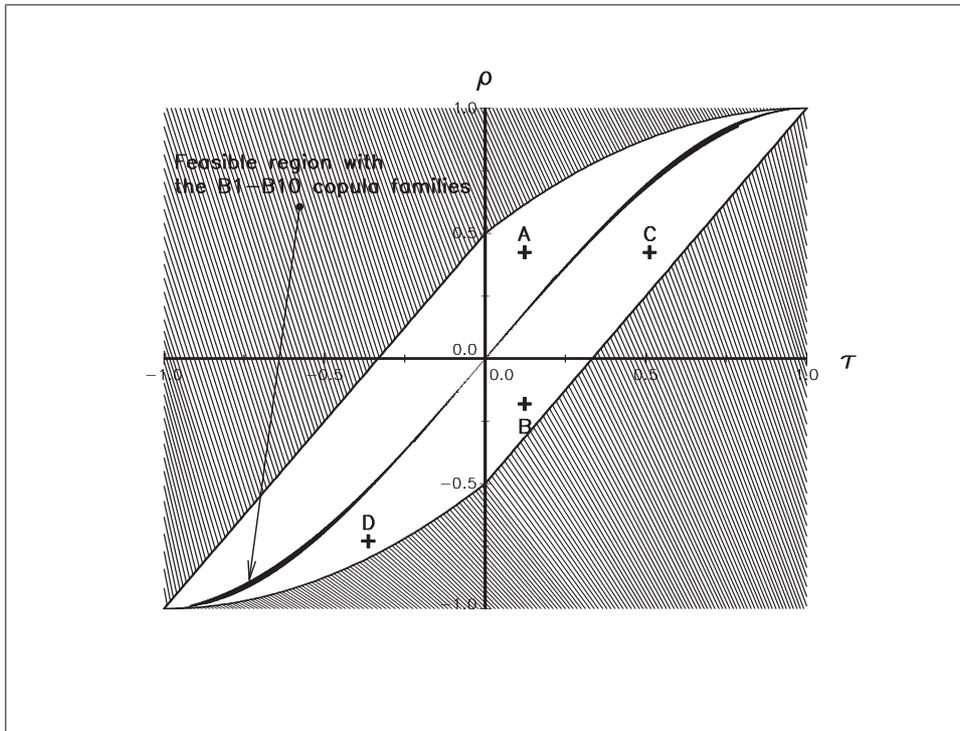
Graphique 3.8. Courbes de niveau de densités bivariées (copule Normale)



Graphique 3.9. Courbes de niveau de densités bivariées (copule Frank)



Graphique 3.10. Courbes de niveau de densités bivariées (copule Gumbel)



Graphique 3.11. Région atteignable des statistiques (τ, ρ)

Par exemple, la copule Gumbel est une copule de valeurs extrêmes :

$$\begin{aligned}
\mathbf{C}(u_1^t, u_2^t; \theta) &= \exp\left(-\left[(-\ln u_1^t)^\theta + (-\ln u_2^t)^\theta\right]^{1/\theta}\right) \\
&= \exp\left(-\left(t^\theta \left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]\right)^{1/\theta}\right) \\
&= \left[\exp\left(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right)\right]^t \\
&= \mathbf{C}^t(u_1, u_2; \theta)
\end{aligned} \tag{3.72}$$

ce qui n'est pas le cas de la copule FGM :

$$\begin{aligned}
\mathbf{C}(u_1^t, u_2^t; \theta) &= u_1^t u_2^t + \theta u_1^t u_2^t (1 - u_1^t) (1 - u_2^t) \\
&= u_1^t u_2^t (1 + \theta - \theta u_1^t - \theta u_2^t + \theta u_1^t u_2^t) \\
&\neq u_1^t u_2^t (1 + \theta - \theta u_1 - \theta u_2 + \theta u_1 u_2)^t \\
&\neq \mathbf{C}^t(u_1, u_2; \theta)
\end{aligned} \tag{3.73}$$

Le nom “extreme value copula” suggère un lien entre la théorie des extrêmes et ces copules. Considérons le vecteur $\boldsymbol{\chi}_n^+$ des statistiques d'ordre maximum. Nous avons

$$\boldsymbol{\chi}_n^+ = \begin{bmatrix} X_{n:n}^{(1)} \\ \vdots \\ X_{n:n}^{(m)} \end{bmatrix} \tag{3.74}$$

Notons $\mathbf{F}_{n:n}$ la distribution de $\boldsymbol{\chi}_n^+$ définie par

$$\begin{aligned}
\mathbf{F}_{n:n}(\mathbf{x}) &= \Pr\{\boldsymbol{\chi}_n^+ \leq \mathbf{x}\} \\
&= \Pr\{X_{n:n}^{(1)} \leq x^{(1)}, \dots, X_{n:n}^{(m)} \leq x^{(m)}\}
\end{aligned} \tag{3.75}$$

La théorie multidimensionnelle des valeurs extrêmes s'intéresse à la loi limite

$$\begin{aligned}
\lim_{n \rightarrow \infty} \Pr\left(\frac{X_{n:n}^{(1)} - b_n^{(1)}}{a_n^{(1)}} \leq x^{(1)}, \dots, \frac{X_{n:n}^{(m)} - b_n^{(m)}}{a_n^{(m)}} \leq x^{(m)}\right) &= \mathbf{G}(x^{(1)}, \dots, x^{(m)}) \\
&= \mathbf{G}(\mathbf{x})
\end{aligned} \tag{3.76}$$

D'après la représentation canonique de la distribution de \mathbf{G} :

$$\mathbf{G}(x^{(1)}, \dots, x^{(m)}) = \mathbf{C}\langle \mathbf{G} \rangle(\mathbf{G}_1(x^{(1)}), \dots, \mathbf{G}_m(x^{(m)})) \tag{3.77}$$

il est évident que les marges de $\mathbf{G}(\mathbf{x})$ vérifient le théorème de Fisher-Tippet. Donc les marges d'une distribution multidimensionnelle de valeurs extrêmes sont des Gumbel, ou/et Frechet ou/et Weibull. Concernant la copule $\mathbf{C}\langle \mathbf{G} \rangle$, nous avons le résultat suivant :

Théorème 7 $\mathbf{C}\langle \mathbf{G} \rangle$ est une copule de valeurs extrêmes.

La représentation copule permet donc de caractériser assez facilement les distributions multidimensionnelles de valeurs extrêmes. Prenons l'exemple du vecteur aléatoire $(X^{(1)}, X^{(2)})$ dont la distribution est

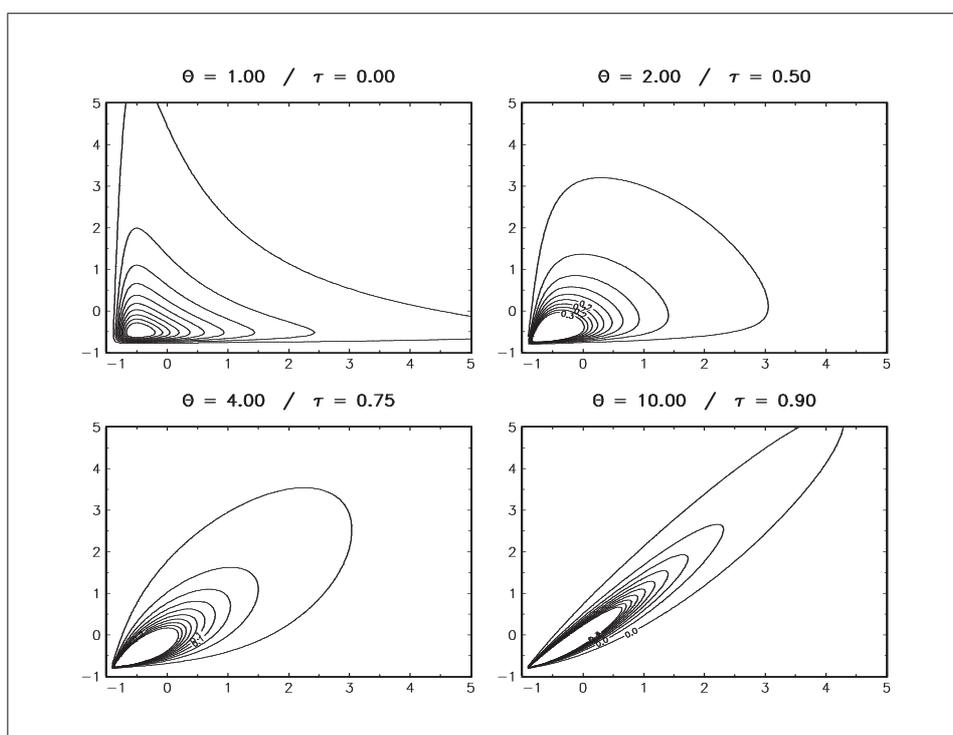
$$\mathbf{F}(x_1, x_2) = \exp\left(-\left[(-\ln \Phi(x_1))^\theta + (-\ln x_2)^\theta\right]^{1/\theta}\right) \tag{3.78}$$

$X^{(1)}$ est donc une variable aléatoire gaussienne et $X^{(2)}$ est une variable aléatoire uniforme. La distribution \mathbf{G} est donc

$$\mathbf{G}(x^{(1)}, x^{(2)}) = \exp\left(-\left[\left(-\ln \Lambda(x^{(1)})\right)^\theta + \left(-\ln \Psi_1(x^{(2)})\right)^\theta\right]^{1/\theta}\right) \quad (3.79)$$

Pour construire une distribution multidimensionnelle de valeurs extrêmes, il suffit donc de coupler des marges GEV avec une copule de valeurs extrêmes. Par exemple, le graphique 3.12 correspond à la distribution suivante :

$$\mathbf{G}(x^{(1)}, x^{(2)}) = \exp\left(-\left[\left(-\ln \mathcal{G}\mathcal{E}\mathcal{V}(0, 1, 1)\right)^\theta + \left(-\ln \mathcal{G}\mathcal{E}\mathcal{V}(0, 1, 1.2)\right)^\theta\right]^{1/\theta}\right) \quad (3.80)$$



Graphique 3.12. Exemples de distribution bivariable de valeurs extrêmes

Dans le cas bivariable, nous pouvons montrer que l'expression de la copule $\mathbf{C}(\mathbf{G})$ est

$$\mathbf{C}(\mathbf{G})(u_1, u_2) = \exp\left[\ln(u_1 u_2) A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)\right] \quad (3.81)$$

avec A une fonction convexe qui vérifie $A(0) = A(1) = 1$ et $\max(w, 1-w) \leq A(w) \leq 1$. Il est alors facile de montrer que³

$$\mathbf{C}^\perp \prec \mathbf{C} \langle \mathbf{G} \rangle \prec \mathbf{C}^+ \quad (3.83)$$

Les extrêmes sont donc **toujours** positivement dépendants. Et cette propriété se généralise facilement au cas multidimensionnel.

Nous pouvons distinguer trois situations :

1. $\mathbf{C} \langle \mathbf{G} \rangle = \mathbf{C}^\perp$;
2. $\mathbf{C} \langle \mathbf{G} \rangle = \mathbf{C}^+$;
3. et $\mathbf{C} \langle \mathbf{G} \rangle$ est une copule comprise entre la copule produit et la copule de la borne haute de Fréchet.

Comment pouvons-nous déterminer a priori quel cas vérifie la copule $\mathbf{C} \langle \mathbf{G} \rangle$? La réponse dépend de la dépendance de queue.

Définition 8 Si une copule bivariée \mathbf{C} est telle que la limite

$$\lim_{u \rightarrow 1} \frac{\bar{\mathbf{C}}(u, u)}{1-u} = \lim_{u \rightarrow 1} \frac{1-2u + \mathbf{C}(u, u)}{1-u} = \lambda \quad (3.84)$$

existe, alors \mathbf{C} a une dépendance de queue si $\lambda \in (0, 1]$ et n'a pas de dépendance de queue si $\lambda = 0$.

Pour la copule produit \mathbf{C}^\perp , nous avons

$$\begin{aligned} \lim_{u \rightarrow 1} \frac{\bar{\mathbf{C}}(u, u)}{1-u} &= \lim_{u \rightarrow 1} \frac{1-2u + u^2}{1-u} \\ &= \lim_{u \rightarrow 1} (1-u) \\ &= 0 \end{aligned} \quad (3.85)$$

Pour la copule \mathbf{C}^+ , nous obtenons

$$\begin{aligned} \lim_{u \rightarrow 1} \frac{\bar{\mathbf{C}}(u, u)}{1-u} &= \lim_{u \rightarrow 1} \frac{1-2u + \min(u, u)}{1-u} \\ &= \lim_{u \rightarrow 1} \frac{1-2u + u}{1-u} \\ &= 1 \end{aligned} \quad (3.86)$$

Notons que pour la copule \mathbf{C}^- , la dépendance de queue est nulle

$$\begin{aligned} \lim_{u \rightarrow 1} \frac{\bar{\mathbf{C}}(u, u)}{1-u} &= \lim_{u \rightarrow 1} \frac{1-2u + \max(2u-1, 0)}{1-u} \\ &= 0 \end{aligned} \quad (3.87)$$

³Nous avons

$$\begin{aligned} &\max(w, 1-w) \leq A(w) \leq 1 \\ \Leftrightarrow &\max\left(\frac{\ln u_1}{\ln(u_1 u_2)}, \frac{\ln u_2}{\ln(u_1 u_2)}\right) \leq A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right) \leq 1 \\ \Leftrightarrow &\min(\ln u_1, \ln u_2) \geq \ln(u_1 u_2) A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right) \geq \ln(u_1 u_2) \\ \Leftrightarrow &\min(u_1, u_2) \geq \exp\left[\ln(u_1 u_2) A\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)\right] \geq u_1 u_2 \\ \Leftrightarrow &\mathbf{C}^+ \succ \mathbf{C} \langle \mathbf{G} \rangle \succ \mathbf{C}^\perp \end{aligned} \quad (3.82)$$

car $\ln(u_1 u_2) \leq 0$.

Nous avons une relation très intéressante entre la dépendance de queue de \mathbf{C} et celle de la copule des valeurs extrêmes $\mathbf{C}\langle\mathbf{G}\rangle$:

Théorème 8 *La dépendance de queue de $\mathbf{C}\langle\mathbf{G}\rangle$ est égale à la dépendance de queue de \mathbf{C} .*

Remarque 24 *Les extrêmes sont donc indépendants si la copule n'a pas de dépendance de queue.*

Remarque 25 *Dans le cas unidimensionnel, la caractérisation des extrêmes dépend de la queue de distribution. Dans le cas multidimensionnel, elle dépend de la dépendance des queues de distribution.*

La mesure λ peut s'interpréter dans le cadre de la valeur en risque. Considérons la probabilité conditionnelle que nous dépassons la valeur en risque dans une direction sachant que nous l'avons déjà dépassée dans l'autre direction :

$$\begin{aligned}
\lambda(\alpha) &= \Pr\{X_2 > \text{VaR}_2(\alpha) \mid X_1 > \text{VaR}_1(\alpha)\} \\
&= \Pr\{X_2 > \mathbf{F}_2^{-1}(\alpha) \mid X_1 > \mathbf{F}_1^{-1}(\alpha)\} \\
&= \frac{\Pr\{\mathbf{F}_2(X_2) > \alpha, \mathbf{F}_1(X_1) > \alpha\}}{\Pr\{\mathbf{F}_1(X_1) > \alpha\}} \\
&= \frac{1 - 2\alpha + \mathbf{F}(\mathbf{F}_1^{-1}(\alpha), \mathbf{F}_2^{-1}(\alpha))}{1 - \alpha} \\
&= \frac{1 - 2\alpha + \mathbf{C}(\alpha, \alpha)}{1 - \alpha}
\end{aligned} \tag{3.88}$$

Nous vérifions que

$$\lambda = \lim_{\alpha \rightarrow 1^+} \lambda(\alpha) \tag{3.89}$$

Autrement dit, λ est la probabilité d'observer un extrême dans une direction sachant que nous avons déjà un extrême dans l'autre direction.

Pour illustrer l'importance de la dépendance de queue, nous considérons les valeurs prises par $\lambda(\alpha)$ pour les copules Normale et Student. Pour la copule Normale, nous avons $\lambda = 0$ si $\rho < 1$. Dans le cas de la copule Student, $\lambda \neq 0$ si $\rho > -1$.

3.3 Caractérisation des domaines d'attraction

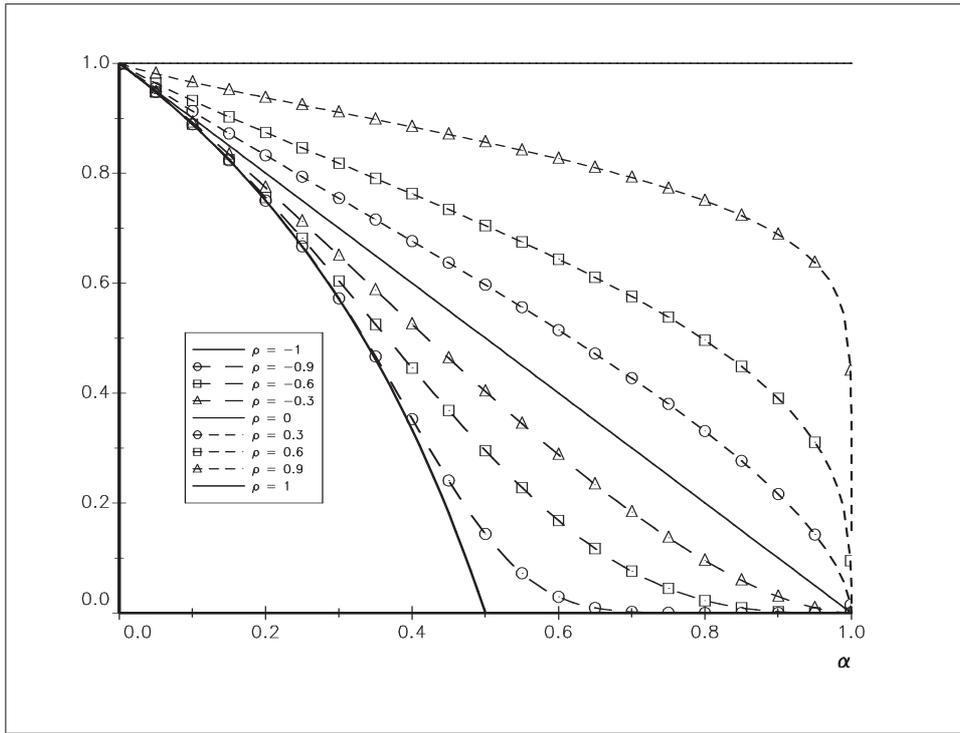
Soit \mathbf{F} la distribution multidimensionnelle dont les marges sont $\mathbf{F}_1, \dots, \mathbf{F}_m$ et dont la copule est \mathbf{C} . Soit \mathbf{G} la distribution multidimensionnelle des valeurs extrêmes. Notons $\mathbf{G}_1, \dots, \mathbf{G}_m$ les marges de \mathbf{G} et $\mathbf{C}\langle\mathbf{G}\rangle$ la copule correspondante. Nous avons alors le théorème suivant :

Théorème 9 $\mathbf{F} \in \text{MDA}(\mathbf{G})$ si et seulement si

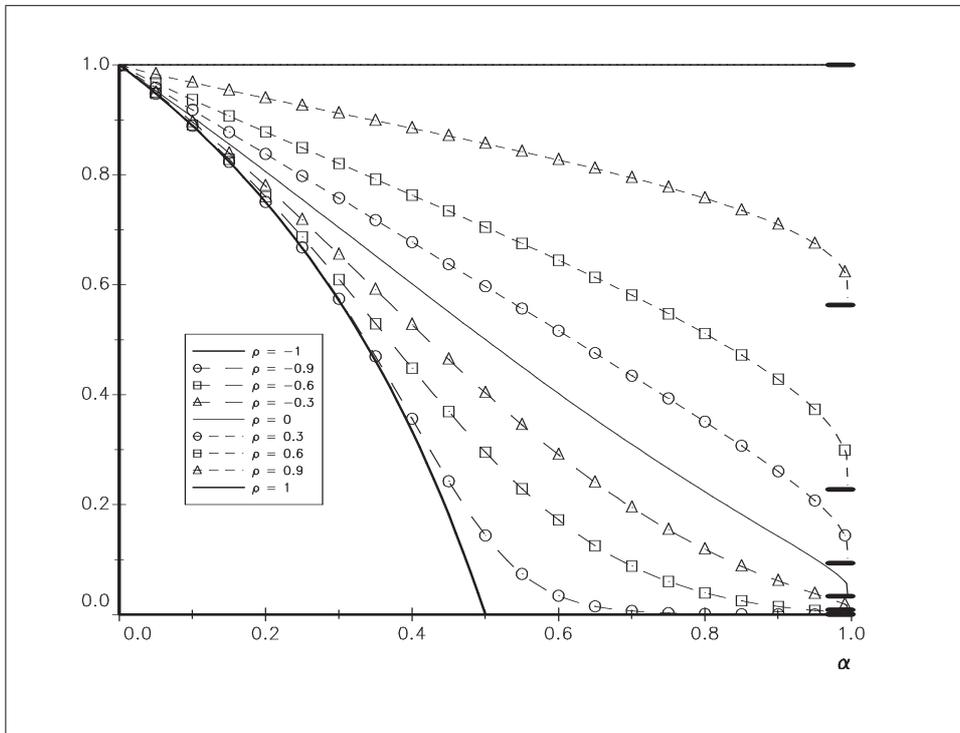
1. $\mathbf{F}_i \in \text{MDA}(\mathbf{G}_i)$ pour tout $i = 1, \dots, m$;
2. $\mathbf{C} \in \text{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$.

Dans le chapitre précédent, nous avons vu comment caractériser le max-domaine d'attraction dans le cas unidimensionnel et aussi comment calculer les coefficients de normalisation. Ces coefficients de normalisation sont valides dans le cas multidimensionnel. Donc, la seule difficulté est la détermination de $\mathbf{C}\langle\mathbf{G}\rangle$:

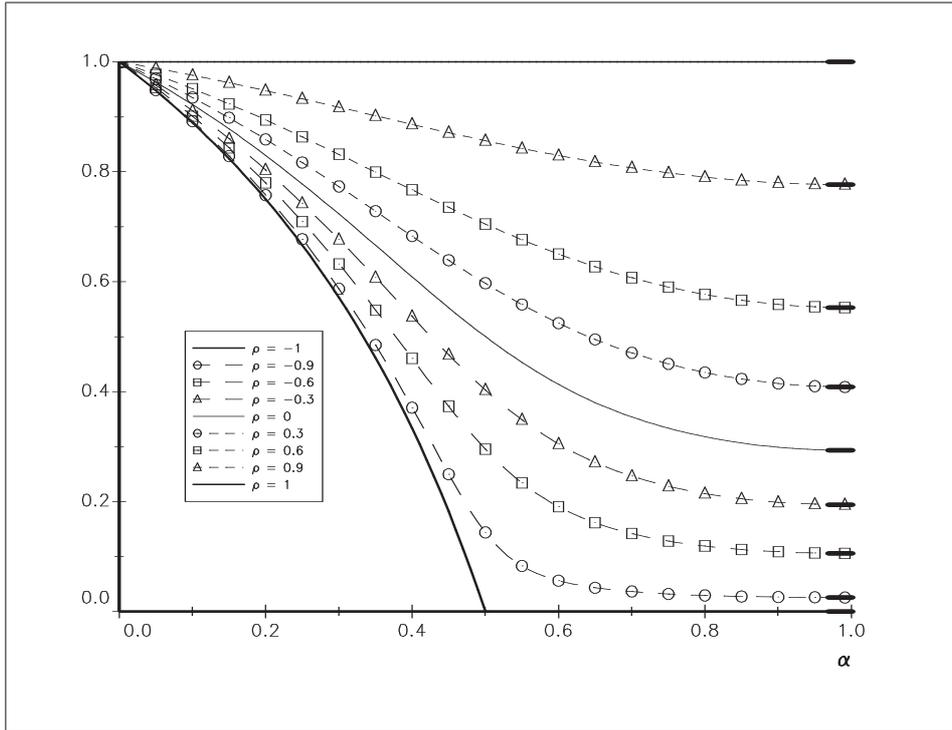
$$\lim_{n \rightarrow \infty} \Pr\left(\frac{X_{n:n}^{(1)} - b_n^{(1)}}{a_n^{(1)}} \leq x^{(1)}, \dots, \frac{X_{n:n}^{(m)} - b_n^{(m)}}{a_n^{(m)}} \leq x^{(m)}\right) = \mathbf{C}\langle\mathbf{G}\rangle\left(\mathbf{G}_1(x^{(1)}), \dots, \mathbf{G}_m(x^{(m)})\right) \tag{3.90}$$



Graphique 3.13. Dépendance de queue de la copule Normale



Graphique 3.14. Dépendance de queue de la copule Student t_6



Graphique 3.15. Dépendance de queue de la copule Student t_1

Théorème 10 $\mathbf{C} \in \text{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$ si \mathbf{C} satisfait la relation suivante :

$$\lim_{t \rightarrow \infty} \mathbf{C}^t \left(u_1^{1/t}, \dots, u_m^{1/t} \right) = \mathbf{C}\langle\mathbf{G}\rangle(u_1, \dots, u_m) \quad (3.91)$$

Remarque 26 Si \mathbf{C} est une copule de valeurs extrêmes, $\mathbf{C} \in \text{MDA}(\mathbf{C})$. En effet, nous avons

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{C}^t \left(u_1^{1/t}, \dots, u_m^{1/t} \right) &= \lim_{t \rightarrow \infty} \mathbf{C} \left(\left(u_1^{1/t} \right)^t, \dots, \left(u_m^{1/t} \right)^t \right) \\ &= \lim_{t \rightarrow \infty} \mathbf{C}(u_1, \dots, u_m) \\ &= \mathbf{C}(u_1, \dots, u_m) \end{aligned} \quad (3.92)$$

Dans le cas bidimensionnel, ce théorème peut s'écrire avec la fonction convexe A :

Théorème 11 $\mathbf{C} \in \text{MDA}(\mathbf{C}\langle\mathbf{G}\rangle)$ avec

$$\mathbf{C}\langle\mathbf{G}\rangle(u_1, u_2) = \exp \left[\ln(u_1 u_2) A \left(\frac{\ln u_1}{\ln(u_1 u_2)} \right) \right] \quad (3.93)$$

si \mathbf{C} satisfait la relation suivante :

$$\lim_{u \rightarrow 0} \frac{1 - \mathbf{C} \left((1-u)^{1-t}, (1-u)^t \right)}{u} = A(t) \quad (3.94)$$

pour tout $t \in [0, 1]$.

Considérons un exemple avec la distribution de $(X^{(1)}, X^{(2)})$ définie par

$$\mathbf{F}(x_1, x_2) = \left((1 - e^{-x_1})^{-\theta} + x_2^{-\theta} - 1 \right)^{-1/\theta} \quad (3.95)$$

sur $[0, \infty] \times [0, 1]$. Les marges de $\mathbf{F}(x_1, x_2)$ sont donc

$$\begin{aligned} \mathbf{F}_1(x_1) &= \mathbf{F}(x_1, 1) \\ &= 1 - e^{-x_1} \end{aligned} \quad (3.96)$$

et

$$\begin{aligned} \mathbf{F}_2(x_2) &= \mathbf{F}(\infty, x_2) \\ &= x_2 \end{aligned} \quad (3.97)$$

$X^{(1)}$ est donc une variable aléatoire exponentielle et $X^{(2)}$ est une variable aléatoire uniforme. Nous cherchons à caractériser la distribution bidimensionnelle des extrêmes :

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{X_{n:n}^{(1)} - b_n^{(1)}}{a_n^{(1)}} \leq x^{(1)}, \frac{X_{n:n}^{(2)} - b_n^{(2)}}{a_n^{(2)}} \leq x^{(2)} \right) = \mathbf{C} \langle \mathbf{G} \rangle \left(\mathbf{G}_1(x^{(1)}), \mathbf{G}_2(x^{(2)}) \right) \quad (3.98)$$

Nous savons que

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{X_{n:n}^{(1)} - \ln n}{1} \leq x^{(1)} \right) = \Lambda(x^{(1)}) \quad (3.99)$$

et

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{X_{n:n}^{(2)} - 1}{n^{-1}} \leq x^{(2)} \right) = \Psi_1(x^{(2)}) \quad (3.100)$$

La copule de \mathbf{F} est la copule Clayton :

$$\mathbf{C}(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1 \right)^{-1/\theta} \quad (3.101)$$

Nous avons

$$\begin{aligned} \lim_{u \rightarrow 0} \frac{1 - \mathbf{C} \left((1-u)^t, (1-u)^{1-t} \right)}{u} &= \lim_{u \rightarrow 0} \frac{1 - \left((1-u)^{-\theta t} + (1-u)^{-\theta(1-t)} - 1 \right)^{-1/\theta}}{u} \\ &= \lim_{u \rightarrow 0} \frac{1 - (1 + \theta u + o(u))^{-1/\theta}}{u} \\ &= \lim_{u \rightarrow 0} \frac{u + o(u)}{u} \\ &= 1 \end{aligned} \quad (3.102)$$

$A(t)$ est donc égal à 1, nous en déduisons que

$$\mathbf{C} \langle \mathbf{G} \rangle = \mathbf{C}^\perp \quad (3.103)$$

Finalement, nous obtenons

$$\begin{aligned} \mathbf{G}(x^{(1)}, x^{(2)}) &= \lim_{n \rightarrow \infty} \Pr \left(X_{n:n}^{(1)} - \ln n \leq x^{(1)}, n \left(X_{n:n}^{(2)} - 1 \right) \leq x^{(2)} \right) \\ &= \Lambda(x^{(1)}) \times \Psi_1(x^{(2)}) \\ &= \exp(-e^{-x^{(1)}}) \exp(x^{(2)}) \end{aligned} \quad (3.104)$$

Si maintenant nous changeons la copule de \mathbf{F} , seule la copule $\mathbf{C}\langle\mathbf{G}\rangle$ est modifiée. Par exemple, si \mathbf{C} est la copule Normale de paramètre $\rho < 1$, alors

$$\mathbf{G}\left(x^{(1)}, x^{(2)}\right) = \exp\left(-e^{-x^{(1)}}\right) \exp\left(x^{(2)}\right) \quad (3.105)$$

Si \mathbf{C} est la copule Normale de paramètre $\rho = 1$, nous avons

$$\mathbf{G}\left(x^{(1)}, x^{(2)}\right) = \min\left(\exp\left(-e^{-x^{(1)}}\right), \exp\left(x^{(2)}\right)\right) \quad (3.106)$$

Si \mathbf{C} est la copule Gumbel, nous obtenons

$$\mathbf{G}\left(x^{(1)}, x^{(2)}\right) = \exp\left(-\left[\exp\left(-\theta x^{(1)}\right) + \left(-x^{(2)}\right)^\theta\right]^{1/\theta}\right) \quad (3.107)$$

3.4 L'exemple des copules Archimax

Remarque 27 *Cette section est extraite du GT ENSAE d'Antoine Bezat et Ashkan Nikeghbali (Rapport disponible sur le site web du GRO).*

Nous considérons une nouvelle famille de copules introduite par CAPÉRAÀ, FOUGÈRES et GENEST [2000] qui englobe la plupart des familles connues des copules, notamment les copules Archimédiennes et **toutes** les copules de valeurs extrêmes. Cette nouvelle famille offre plus de flexibilité pour la modélisation. En effet, nous pouvons connaître a priori les différents max-domaines d'attraction.

Définition 9 *Une fonction bivariée est une copule Archimax si et seulement si elle est de la forme*

$$\mathbf{C}_{\phi, A}(x, y) = \phi^{-1}\left[(\phi(x) + \phi(y)) A\left(\frac{\phi(x)}{\phi(x) + \phi(y)}\right)\right] \quad (3.108)$$

pour tout $0 \leq x, y \leq 1$, avec

1. $A : [0, 1] \rightarrow [1/2, 1]$ tel que $\max(t, 1-t) \leq A(t) \leq 1$ pour tout $0 \leq t \leq 1$;
2. $\phi : (0, 1] \rightarrow [0, \infty)$ est une fonction convexe, décroissante qui vérifie $\phi(1) = 0$. Nous adopterons la convention $\phi(0) = \lim_{t \rightarrow 0^+} \phi(t)$ et $\phi^{-1}(s) = 0$ pour $s \geq \phi(0)$.

Nous voyons aisément que cette nouvelle famille de copules contient toutes les copules Archimédiennes et toutes les copules de valeurs extrêmes. En effet, si $\phi(t) = \ln(1/t)$, nous obtenons

$$\mathbf{C}_{\phi, A}(x, y) = \mathbf{C}_A(x, y) = \exp\left[\ln(xy) A\left(\frac{\ln(y)}{\ln(xy)}\right)\right] \quad (3.109)$$

pour tout $0 \leq x, y \leq 1$. On reconnaît la forme générale des copules bivariées de valeurs extrêmes. Si on choisit maintenant de prendre $A(t) \equiv 1$, on retrouve la forme générale des copules Archimédiennes :

$$\mathbf{C}_{\phi, A}(x, y) = \mathbf{C}_\phi(x, y) = \phi^{-1}(\phi(x) + \phi(y)) \quad (3.110)$$

L'intérêt de la construction (3.108) réside dans le fait qu'avec un choix judicieux de ϕ et A , on obtient une copule qui sera dans le max-domaine d'attraction de n'importe quel "attracteur" \mathbf{C}_{A^*} donné.

Proposition 1 Soit $\mathbf{C}_{\phi,A}$ une copule Archimax telle que $\phi(1-1/t) \in \text{RV}_{-m}$ pour un certain $m \geq 1$. Alors $\mathbf{C}_{\phi,A}$ appartient au max-domaine d'attraction de la copule \mathbf{C}_{A^*} où la fonction A^* est donnée par

$$A^*(t) = [t^m + (1-t)^m]^{1/m} A^{1/m} \left(\frac{t^m}{t^m + (1-t)^m} \right) \quad (3.111)$$

pour $t \in [0, 1]$. Cet attracteur peut être considéré de deux points de vue : \mathbf{C}_{A^*} est une copule Archimax de générateur A^* et $\phi(t) = \ln(1/t)$ ou \mathbf{C}_{A^*} est une copule Archimax de générateurs A et $\phi^*(t) = \ln^m(1/t)$. De plus, A^* et A coïncident si et seulement si $m = 1$.

Remarque 28 Il est alors très facile de montrer que la seule copule Archimédienne et de valeurs extrêmes est la copule Gumbel.

En faisant un raisonnement en sens inverse, on remarque que $\mathbf{C}_{\phi,A}$ appartient au max-domaine d'attraction de \mathbf{C}_{A^*} si $\phi(1-1/t) \in \text{RV}_{-m}$ et

$$A(t) = [t^{1/m} + (1-t)^{1/m}]^m \left\{ A^* \left(\frac{t^{1/m}}{t^{1/m} + (1-t)^{1/m}} \right) \right\}^m \quad (3.112)$$

Maintenant il faut savoir si la relation (3.112) définit bien une fonction de dépendance. En posant $D(t) = \partial_t \ln(A^*(t))$, $\mathbf{H}^*(t) = t + (1-t)D(t)$ et $h^*(t) = \partial_t \mathbf{H}^*(t)$, CAPÉRAÀ, FOUGÈRES et GENEST [2000] montrent que la relation (3.112) est vraie si et seulement si

$$\frac{h^*(t)}{\mathbf{H}^*(t)(1-\mathbf{H}^*(t))} \geq \frac{m}{t(1-t)} \quad (3.113)$$

pour $t \in [0, 1]$. On peut au passage remarquer que cette relation reste valable pour tout $1 \leq m' \leq m$. Ainsi, si $A(t)$ est définie par (3.112), $\mathbf{C}_{\phi,A}$ appartient au max-domaine d'attraction de \mathbf{C}_{A^*} pourvu que $\phi(1-1/t) \in \text{RV}_{-m'}$ pour $1 \leq m' \leq m$. Enfin, on peut aussi remarquer que pour des générateurs A et ϕ quelconques tels que $\phi(1-1/t) \in \text{RV}_{-m}$, on a

$$\lambda = \lim_{u \rightarrow 1^-} \frac{1 - 2u + \mathbf{C}_{\phi,A}(u, u)}{1 - u} = 2 - \{2A(1/2)\}^{1/m} \leq 1 \quad (3.114)$$

Ainsi, les copules Archimax présentent une dépendance au niveau supérieur de la queue de distribution sauf si $m = 1$ et $A \equiv 1$.

On peut donner des exemples de copules Archimax. On peut par exemple considérer le générateur de TAWN [1988] qui est défini pour $0 \leq \theta \leq 1$ par

$$A_\theta(t) = \theta t^2 - \theta t + 1 \quad (3.115)$$

pour $t \in [0, 1]$ et le combiner au générateur de la famille de Clayton⁴

$$\phi_{1,\eta}(t) = \frac{t^{-\eta} - 1}{\eta} \quad (3.116)$$

ou avec le générateur des copules⁵ de GENEST et GHOUDI [1994]

$$\phi_{2,\eta}(t) = (1 - t^\eta)^{1/\eta} \quad (3.117)$$

⁴ $\eta > 0$.

⁵ $0 < \eta \leq 1$.

On peut remarquer que $\phi_{2,\eta}(t) \in \text{RV}_{-1/\eta}$. Ainsi, pour $\eta \neq 1$, $\mathbf{C}_{\phi_{2,\eta},A}(x,y)$ n'est pas dans le domaine d'attraction de \mathbf{C}_A . Enfin supposons que l'on veuille construire une copule dans le max-domaine d'attraction de

$$A_{\alpha,\beta}^*(t) = \left[\alpha^k t^k + \beta^k (1-t)^k \right]^{1/k} + \left[(1-\alpha)^k t^k + (1-\beta)^k (1-t)^k \right]^{1/k} \quad (3.118)$$

pour $t \in [0,1]$ avec $0 \leq \alpha, \beta \leq 1$ et $k \geq 1$. Pour les valeurs $\alpha = 1/10$, $\beta = 3/10$, $m = 3/2$ et $k = 1.6$, la condition (3.113) est vérifiée. La copule originale A est donc celle définie par l'expression (3.112).

3.5 Copules et modélisation financière

Nous finissons ce chapitre en considérant le problème de la modélisation financière. Celle-ci consiste

1. à spécifier un modèle pour chaque facteur,
2. et à corrélérer ces différents modèles pour obtenir une modélisation jointe des différents facteurs.

Traditionnellement, nous faisons ces deux étapes simultanément en considérant un modèle multidimensionnel, qui est généralement gaussien.

L'intérêt des copules est justement de pouvoir spécifier le modèle multidimensionnel en respectant les deux étapes. Il y a plusieurs avantages à procéder de cette façon :

- l'extension de modèles unidimensionnels existants est beaucoup plus facile ;
- le choix de la structure de dépendance (la copule) peut se faire indépendamment du choix des modèles unidimensionnels (les marges) ;
- la représentation copule permet de réduire la complexité de calibration du modèle multidimensionnel, puisque les paramètres de la copule et les paramètres des marges peuvent être estimés séparément ;
- dans une approche directe de modélisation du modèle multidimensionnel, une mauvaise spécification des marges introduit un biais dans l'estimation des paramètres de la copule et inversement ;
- enfin, il est clair qu'une représentation copule permet de mieux comprendre quelles sont les hypothèses sous-jacentes du modèle, surtout dans une optique de gestion des risques.

Nous donnons ici différentes illustrations qui montrent les difficultés ou les erreurs que nous pouvons rencontrer lorsque nous spécifions directement un modèle multidimensionnel.

1. Dans COSTINOT, RONCALLI et TEILETCHE [2000], nous cherchons à modéliser la dépendance jointe des trois indices CAC40, DowJones et NIKKEI. Pour la période allant du 1/4/1988 au 10/9/2000, la matrice de corrélation des rendements est

$$\hat{\rho} = \begin{bmatrix} 1 & 0.158 & 0.175 \\ & 1 & 0.0589 \\ & & 1 \end{bmatrix} \quad (3.119)$$

Si nous supposons que le vecteur aléatoire des rendements est gaussien, nous avons donc

$$\mathbf{C} \langle \text{CAC40}, \text{NIKKEI} \rangle \succ \mathbf{C} \langle \text{CAC40}, \text{DowJones} \rangle \succ \mathbf{C} \langle \text{NIKKEI}, \text{DowJones} \rangle \quad (3.120)$$

Maintenant, si nous supposons que seule la dépendance est Normale (**sans faire aucune hypothèse sur les marges**), nous pouvons estimer la matrice des paramètres ρ de la copule Normale tridimensionnelle et nous obtenons

$$\rho = \begin{bmatrix} 1 & 0.207 & 0.157 \\ & 1 & 0.0962 \\ & & 1 \end{bmatrix} \quad (3.121)$$

Dans ce cas, nous vérifions que

$$\mathbf{C} \langle \text{CAC40}, \text{DowJones} \rangle \succ \mathbf{C} \langle \text{CAC40}, \text{NIKKEI} \rangle \succ \mathbf{C} \langle \text{NIKKEI}, \text{DowJones} \rangle \quad (3.122)$$

L'ordre de dépendance a donc changé.

2. Dans BOUYÉ, DURRLEMAN, NIKEGHBALI, RIBOULET et RONCALLI [2000], nous considérons la distribution hyperbolique généralisée GH (EBERLEIN and KELLER [1995], EBERLEIN [1999], PRAUSE [1999]). La fonction de densité correspondante est

$$\begin{aligned} f(x) &= a(\lambda, \alpha, \beta, \delta) \left(\delta^2 + (x - \mu)^2 \right)^{\frac{1}{4}(2\lambda-1)} \\ &\quad \times \mathbf{K}_{\lambda-\frac{1}{2}} \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right) \exp(\beta(x - \mu)) \end{aligned} \quad (3.123)$$

où \mathbf{K} est la fonction de Bessel modifiée du troisième type et

$$a(\lambda, \alpha, \beta, \delta) = \frac{(\alpha^2 - \beta^2)^{\frac{1}{2}\lambda}}{\sqrt{2\pi} \alpha^{\lambda-\frac{1}{2}} \delta^\lambda \mathbf{K}_\lambda \left(\delta \sqrt{\alpha^2 - \beta^2} \right)} \quad (3.124)$$

Dans un récent article, EBERLEIN [2000] applique le modèle hyperbolique à la gestion du risque. Néanmoins, il ne considère que le cas univarié. Le cas multivarié est traité par PRAUSE [1999]. La densité de la distribution hyperbolique généralisée **multidimensionnelle** est

$$\begin{aligned} f(\mathbf{x}) &= A(\lambda, \alpha, \beta, \delta, \boldsymbol{\rho}) \left(\delta^2 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\rho}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^{\frac{1}{4}(2\lambda-N)} \\ &\quad \times \mathbf{K}_{\lambda-\frac{N}{2}} \left(\alpha \sqrt{\delta^2 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\rho}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right) \exp(\boldsymbol{\beta}^\top (\mathbf{x} - \boldsymbol{\mu})) \end{aligned} \quad (3.125)$$

avec

$$A(\lambda, \alpha, \beta, \delta, \boldsymbol{\rho}) = \frac{(\alpha^2 - \boldsymbol{\beta}^\top \boldsymbol{\rho} \boldsymbol{\beta})^{\frac{1}{2}\lambda}}{(2\pi)^{\frac{N}{2}} \alpha^{\lambda-\frac{N}{2}} \delta^\lambda \mathbf{K}_\lambda \left(\delta \sqrt{\alpha^2 - \boldsymbol{\beta}^\top \boldsymbol{\rho} \boldsymbol{\beta}} \right)} \quad (3.126)$$

Cependant, cette distribution est peu maniable du point de vue de l'implémentation, car elle est difficile à simuler et l'estimation des paramètres est relativement compliquée. Ces problèmes peuvent être évités si nous construisons une distribution "hyperbolique généralisée" multidimensionnelle à partir de marges hyperboliques généralisées et d'une copule. Par exemple, avec une copule Normale, la densité de la distribution est

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{|\boldsymbol{\rho}|^{\frac{1}{2}}} A(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\rho}) \left[\prod_{n=1}^N \left(\delta_n^2 + (x_n - \mu_n)^2 \right)^{\frac{1}{4}(2\lambda_n-1)} \right] \\ &\quad \times \left[\prod_{n=1}^N \mathbf{K}_{\lambda_n-\frac{1}{2}} \left(\alpha_n \sqrt{\delta_n^2 + (x_n - \mu_n)^2} \right) \right] \exp \left(\boldsymbol{\beta}^\top (\mathbf{x} - \boldsymbol{\mu}) - \frac{1}{2} \boldsymbol{\varsigma}^\top (\boldsymbol{\rho}^{-1} - (\mathbb{I}) \otimes \boldsymbol{\mathcal{Q}}) \right) \end{aligned}$$

avec

$$A(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\rho}) = \frac{1}{(2\pi)^{\frac{N}{2}}} \prod_{n=1}^N \frac{(\alpha_n^2 - \beta_n^2)^{\frac{1}{2}\lambda_n}}{\alpha_n^{\lambda_n - \frac{1}{2}} \delta_n^{\lambda_n} \mathbf{K}_{\lambda_n} \left(\delta_n \sqrt{\alpha_n^2 - \beta_n^2} \right)} \quad (3.128)$$

et $\boldsymbol{\varsigma} = (\varsigma_1, \dots, \varsigma_N)^\top$, $\varsigma_n = \Phi^{-1}(\mathbf{F}_n(x_n))$ (\mathbf{F}_n est la distribution GH univariée). Si nous comparons ces deux distributions multidimensionnelles, nous remarquons que la première distribution a $\frac{1}{2}N[(N+3)]+3$ paramètres alors que la seconde distribution a $\frac{1}{2}N[(N+9)]$ paramètres. Par exemple, si $N = 25$, nous avons respectivement 353 et 425 paramètres. Cette dernière distribution possédant plus de paramètres, nous pouvons penser qu'elle est moins facile à estimer. En fait, c'est tout le contraire. Pour la première distribution, les 353 paramètres doivent être estimés simultanément par maximum de vraisemblance. Pour la seconde distribution, les paramètres des marges sont estimés dans une première étape (25 estimations de 5 paramètres par maximum de vraisemblance) alors que les paramètres de la copule est estimée dans une seconde étape (l'estimation des 300 paramètres est directe car il existe un estimateur **omnibus analytique** de la matrice des paramètres d'une copule Normale).

3. Dans COUTANT, DURRLEMAN, RAPUCH et RONCALLI [2001], nous cherchons à estimer la structure de dépendance implicite sur le marché des options sur spread. La calibration est faite en deux étapes :
 - (a) les distributions risque neutre unidimensionnelles sont calibrées à partir des options vanilles ;
 - (b) la copule risque neutre est calibrée à partir des options sur spread en imposant que les marges de la distribution risque neutre multidimensionnelle soient les distributions risque neutre unidimensionnelles calibrées précédemment.
4. Nous considérons deux modèles de pricing. Le premier est le modèle multidimensionnel de Black et Scholes (marges lognormales et copule Normale), alors que le second est une extension bivariée du modèle de Bahra (les marges sont des mélanges de deux distributions lognormales et la copule est Normale). Nous donnons un exemple où le premier modèle conduit à une corrélation implicite égale -0,341 alors que le paramètre implicite de la copule Normale pour le second modèle est égal à 0,767. **En utilisant le modèle de Black et Scholes, nous concluons à une dépendance négative, alors qu'avec le modèle de Bahra, nous obtenons une dépendance positive !**

Références

- [1] BEZAT, A. et A. NIKEGHBALI [2000], La théorie des extrêmes et la gestion des risques de marché, GT ENSAE (document disponible sur le site web du GRO)
- [2] BOUYÉ, E., V. DURRLEMAN, A. NIKEGHBALI, G. RIBOULET et T. RONCALLI [2000], Copulas for finance — a reading guide and some applications, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [3] CAPÉRAÀ, P., A-L. FOUGÈRES et C. GENEST [2000], Bivariate distributions with given extreme value attractor, *Journal of Multivariate Analysis*, **72**, 30-49
- [4] COLES, S., J. CURRIE et J. TAWN [1999], Dependence measures for extreme value analyses, Department of Mathematics and Statistics, Lancaster University, *Working Paper*

- [5] COSTINOT, A., T. RONCALLI et J. TEILETCHE [2000], Revisiting the dependence between financial markets with copulas, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [6] COUTANT, S., V. DURRLEMAN, G. RAPUCH et T. RONCALLI [2001], Copulas, multivariate risk-neutral distributions and implied dependence functions, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [7] EBERLEIN, E. [1999], Application of generalized hyperbolic Lévy motions to finance, Universität Freiburg, *Working Paper*
- [8] EBERLEIN, E. [2000], Recent advances in more realistic market and credit risk management : The hyperbolic model, to appear in *ICBI Report*
- [9] EBERLEIN, E. and U. KELLER [1995], Hyperbolic distributions in finance, *Bernoulli*, **1**, 281-299
- [10] EBERLEIN, E., U. KELLER and K. PRAUSE [1998], New insights into smile, mispricing and value at risk : the hyperbolic model, *Journal of Business*, **71**, 371-405
- [11] GENEST, C. et K. GHOUDI [1994], Une famille de lois bidimensionnelles insolite, *Comptes Rendus de l'Académie des Sciences de Paris*, **318**, série I, 351-354
- [12] JOE, H. [1997], Multivariate Models and Dependence Concepts, Chapman & Hall, London
- [13] KOTZ, S., N. BALAKRISHNAN et N.L. JOHNSON [2000], Continuous Multivariate Distributions — Volume 1 : Models and Applications, John Wiley & Sons, New York
- [14] NELSEN, R.B. [1999], An Introduction to Copulas, *Lectures Notes in Statistics*, **139**, Springer-Verlag, New York
- [15] PRAUSE, K. [1999], The Generalized Hyperbolic Model : Estimation, Financial Derivatives, and Risk Measures, University of Freiburg, *Ph.D. Dissertation*
- [16] SCHWEIZER, B. et E. WOLFF [1981], On nonparametric measures of dependence for random variables, *Annals of Statistics*, **9**, 879-885
- [17] TAWN, J.A. [1988], Bivariate extreme value theory : models and estimation, *Biometrika*, **75**, 397-415
- [18] TCHEN, A.H. [1980], Inequalities for distributions with given marginals, *Annals of Statistics*, **8(4)**, 814-827

4

Applications à la gestion des risques

Dans ce chapitre, nous adoptons le principe de modélisation à deux étapes qui dérive de la décomposition canonique d'une distribution multidimensionnelle. En effet, nous voulons contrôler les sources de risque qui peuvent provenir des directions unidimensionnelles des facteurs, mais aussi de la dépendance de ces directions unidimensionnelles. Prenons un exemple. Sur le graphique 4.1, nous avons construit quatre vecteurs aléatoires qui ont les deux premiers moments identiques. Cela veut dire qu'ils présentent la même valeur en risque gaussienne linéaire. Or, il est évident qu'ils présentent des risques très différents. D'un point de vue unidimensionnel, les extrêmes prennent des valeurs qui ne sont pas comparables. D'un point de vue de la structure de dépendance, ces extrêmes sont "corrélés" très différemment. Par exemple, le risque d'un portefeuille long en X_1 et court en X_2 peut être plus ou moins grand selon la position de ces extrêmes.

4.1 Le risque de marché

Cette section utilise les travaux de BEZAT et NIKEGHBALI [2000] et BOUYÉ, DURRLEMAN, NIKEGHBALI, RIBOULET et RONCALLI [2000,2001].

4.1.1 Construction d'une valeur en risque GEV

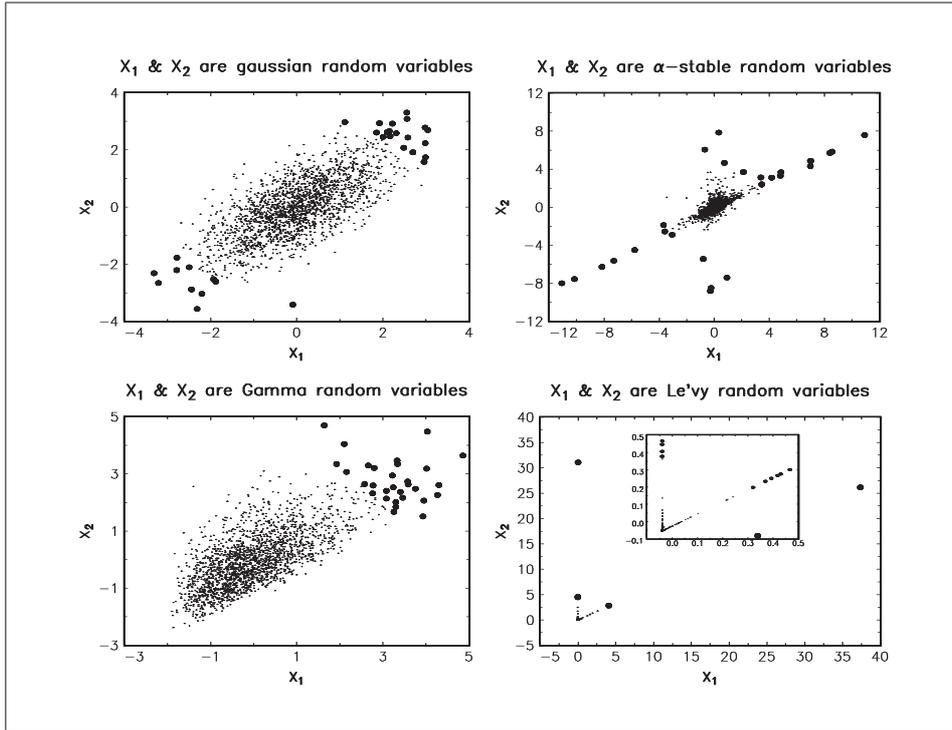
Considérons N actifs de prix $P_n(t)$. Supposons un portefeuille linéaire $P(t; \theta)$ de stratégie $\theta = (\theta_1, \dots, \theta_N)$. Nous avons

$$P(t; \theta) = \sum_{n=1}^N \theta_n P_n(t) \quad (4.1)$$

Notons r le rendement logarithmique du portefeuille défini par

$$r(t; \theta) = \ln P(t; \theta) - \ln P(t-1; \theta) \quad (4.2)$$

Nous pouvons construire une valeur en risque GEV pour le portefeuille linéaire en procédant de la façon suivante :



Graphique 4.1. 4 vecteurs aléatoires présentant la même valeur en risque gaussienne linéaire

1. Soit φ la taille des blocs. Nous définissons le seuil de couverture α_{GEV} équivalent en égalisant les temps de retour :

$$\frac{1}{1 - \alpha_{\text{VaR}}} = \varphi \frac{1}{1 - \alpha_{\text{GEV}}} \quad (4.3)$$

Nous obtenons donc

$$\alpha_{\text{GEV}} = 1 - \varphi(1 - \alpha_{\text{VaR}}) \quad (4.4)$$

2. Nous modélisons alors la loi du maximum de $-r(t; \theta)$ par la distribution GEV. Notons $\hat{\mathbf{G}}$ la distribution estimée.
3. Rappelons que la valeur en risque (journalière) est définie par

$$\text{VaR}(\alpha_{\text{VaR}}) = \sup_x \{ \Pr \{ P(t+1; \theta) - P(t; \theta) \leq x \mid \mathcal{F}_t \} \leq 1 - \alpha_{\text{VaR}} \} \quad (4.5)$$

Nous en déduisons que

$$\text{VaR}(\alpha_{\text{VaR}}) = P(t; \theta) \times \hat{\mathbf{G}}^{-1}(\alpha_{\text{GEV}}) \quad (4.6)$$

La valeur en risque est donc égale à la valeur actuelle du portefeuille multipliée par le quantile correspondant du rendement logarithmique. Pour des raisons de robustesse, nous préférons modéliser $-r(t; \theta)$ et non $r(t; \theta)$, car nous cherchons à calculer directement le quantile α de la perte, et non le quantile $1 - \alpha$ du gain.

Nous reportons ici l'exemple donné par RONCALLI [2000]. Nous considérons quatre portefeuilles dont les compositions sont les suivantes :

Portefeuille	CAC40	DowJones
P1	1	0
P2	0	1
P3	1	1
P4	1	-1

Nous obtenons alors ces résultats :

VaR Gaussienne			
Portefeuille	VaR 99%	VaR 99.6%	VaR 99.9%
P1	3.780	4.310	5.022
P2	2.896	3.302	3.847
P3	2.473	2.819	3.285
P4	7.604	8.669	10.101

VaR Historique			
Portefeuille	VaR 99%	VaR 99.6%	VaR 99.9%
P1	4.168	7.174	12.221
P2	3.296	5.740	7.408
P3	2.862	4.392	6.773
P4	8.118	14.393	27.726

VaR GEV			
Portefeuille	VaR 99%	VaR 99.6%	VaR 99.9%
P1	3.421	5.153	8.929
P2	2.778	4.482	8.784
P3	2.375	3.691	6.883
P4	6.578	10.501	19.774

Vous trouverez d'autres exemples dans BEZAT et NIKEGHBALI [2000] (tables 2.1 et 2.2 pages 21–22). Ces auteurs discutent de la pertinence de la valeur en risque GEV. **Le principal intérêt est de calculer des mesures de risque pour des quantiles extrêmement élevés. Néanmoins, la robustesse des résultats dépend de la longueur de l'historique.**

4.1.2 La mesure de l'Expected Shortfall et GPD

Remarque 29 *A titre de curiosité, je mentionne des résultats académiques concernant le lien entre la théorie des valeurs extrêmes et la mesure de l'expected shortfall. Cependant, cette mesure a peu d'intérêt pour les professionnels.*

Nous utilisons les présentations de BEZAT et NIKEGHBALI [2000] et KËLLEZI et GILLI [2000]. Nous rappelons que

$$\mathbf{F}_u(y) = \frac{\mathbf{F}(u+y) - \mathbf{F}(u)}{1 - \mathbf{F}(u)} \approx \mathbf{G}(y) \quad (4.7)$$

avec

$$\mathbf{G} \sim \mathcal{GPD}(\beta, \xi) \quad (4.8)$$

Soit $x = u + y$. Nous avons donc

$$\mathbf{F}(x) = \mathbf{F}(u) + (1 - \mathbf{F}(u)) \mathbf{F}_u(y) \quad (4.9)$$

Considérons un échantillon de taille n . Notons N_u le nombre d'observations au-dessus du seuil u . Un estimateur non-paramétrique de $\mathbf{F}(u)$ est alors

$$\hat{\mathbf{F}}(u) = 1 - \frac{N_u}{n} \quad (4.10)$$

Nous en déduisons un estimateur semi-paramétrique de $\mathbf{F}(x)$:

$$\begin{aligned}\hat{\mathbf{F}}(x) &= \left(1 - \frac{N_u}{n}\right) + \frac{N_u}{n} \left(1 - \left(1 + \frac{\hat{\xi}(x-u)}{\hat{\beta}}\right)^{-1/\hat{\xi}}\right) \\ &= 1 - \frac{N_u}{n} \left(1 + \frac{\hat{\xi}(x-u)}{\hat{\beta}}\right)^{-1/\hat{\xi}}\end{aligned}\quad (4.11)$$

$\hat{\xi}$ et $\hat{\beta}$ sont les estimateurs des paramètres de la distribution $\mathcal{GPD}(\beta, \xi)$. Nous interprétons $\hat{\mathbf{F}}(x) = \hat{\mathbf{F}}(u+y)$ comme un estimateur de la queue de distribution. **Il peut être vu comme un estimateur historique amélioré par la théorie des extrêmes.** Un estimateur de la VaR est obtenu en inversant $\hat{\mathbf{F}}(x)$:

$$\text{VaR}(\alpha) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1-\alpha) \right)^{-\hat{\xi}} - 1 \right) \quad (4.12)$$

L'expected shortfall (ou la valeur en risque conditionnelle) est définie par

$$\begin{aligned}\text{ES}(\alpha) &= \mathbb{E}[X \mid X > \text{VaR}(\alpha)] \\ &> \text{VaR}(\alpha)\end{aligned}\quad (4.13)$$

C'est donc l'espérance de la perte au delà de la valeur en risque. Une autre écriture de l'expected shortfall est

$$\text{ES}(\alpha) = \text{VaR}(\alpha) + \mathbb{E}[X - \text{VaR}(\alpha) \mid X > \text{VaR}(\alpha)] \quad (4.14)$$

Nous avons

$$\begin{aligned}\text{ES}(\alpha) &= \text{VaR}(\alpha) + \frac{\beta + \xi(\text{VaR}(\alpha) - u)}{1 - \xi} \\ &= \frac{\text{VaR}(\alpha)}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}\end{aligned}\quad (4.15)$$

Un estimateur semi-paramétrique de $\text{ES}(\alpha)$ est donc

$$\begin{aligned}\text{ES}(\alpha) &= \frac{u}{1 - \hat{\xi}} + \frac{\hat{\beta}}{\hat{\xi}(1 - \hat{\xi})} \left(\left(\frac{n}{N_u} (1-\alpha) \right)^{-\hat{\xi}} - 1 \right) + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}} \\ &= u - \frac{\hat{\beta}}{\hat{\xi}} + \frac{\hat{\beta}}{\hat{\xi}(1 - \hat{\xi})} \left(\frac{n}{N_u} (1-\alpha) \right)^{-\hat{\xi}}\end{aligned}\quad (4.16)$$

4.1.3 Programmes de stress-testing

Pour une vue réglementaire des méthodes de stress-testing, vous pouvez vous référer au chapitre 7 de RONCALLI [2001]. Les programmes de stress-testing correspondent à des simulations de crise. Le but est de calculer la perte **maximale** auquel la banque doit faire face si ces scénarios de crise se réalisaient. Ces programmes de stress-testing sont demandés par la Commission Bancaire et font partie du processus de validation de la VaR.

A titre d'information, nous fournissons les résultats d'une enquête menée par Deloitte & Touche en 1999 auprès de 200 banques dans le monde sur le Risk Management et publiée en mai 2000 dans la revue *Banque Magazine* :

L'utilisation du stress testing (en % des établissements interrogés)			
		Actuels	Envisagés
1	Scénarios reposant sur la politique et l'économie	40%	45%
2	Scénarios chocs marchés	70%	55%
3	Chocs historiques	65%	58%
4	Stress testing global	30%	50%
5	Micro-stress testing	44%	29%
6	Théorie des valeurs extrêmes	5%	24%

Seulement 5% des banques en 1999 utilisaient la théorie des valeurs extrêmes pour construire ses scénarios de crises. Néanmoins, 24% des banques interrogées envisageaient de l'utiliser dans le futur. En fait, la théorie des valeurs extrêmes n'est peut-être pas la méthode la plus robuste pour créer des simulations de crise. Cependant, elle permet de quantifier ces scénarios, c'est-à-dire de les probabiliser et de leur donner un temps de retour implicite. Elle permet donc dans une certaine mesure d'associer un degré de sévérité/gravité à chaque scénario et donc de pouvoir les comparer d'un point de vue probabiliste.

4.1.3.1 Cas unidimensionnel

Nous nous donnons un temps de retour \mathcal{T} et nous cherchons l'événement implicite à ce temps de retour. Soit \mathbf{F} la distribution du rendement journalier d'un actif. Nous avons

$$r = \mathbf{F}^{-1} \left(1 - \frac{1}{\mathcal{T}} \right) \quad (4.17)$$

Généralement, \mathcal{T} prend des valeurs très fortes (5 ans, 10 ans, etc.), ce qui implique des seuils de confiance pour les quantiles extrêmement faibles. Par exemple, si \mathcal{T} est égal à 5 ans, et si nous supposons 250 jours de trading par année en moyenne, le seuil de confiance α est égal à

$$\alpha = 1 - \frac{1}{5 \times 250} = 99.92\% \quad (4.18)$$

Pour \mathcal{T} égal à 50 ans, nous avons $\alpha = 99.992\%$. Ces seuils de confiance étant très élevés, l'utilisation de la théorie des valeurs extrêmes peut apparaître plus pertinente. Soit φ la taille des blocs. Nous avons

$$\alpha_{\text{GEV}} = 1 - \frac{\varphi}{\mathcal{T}} \quad (4.19)$$

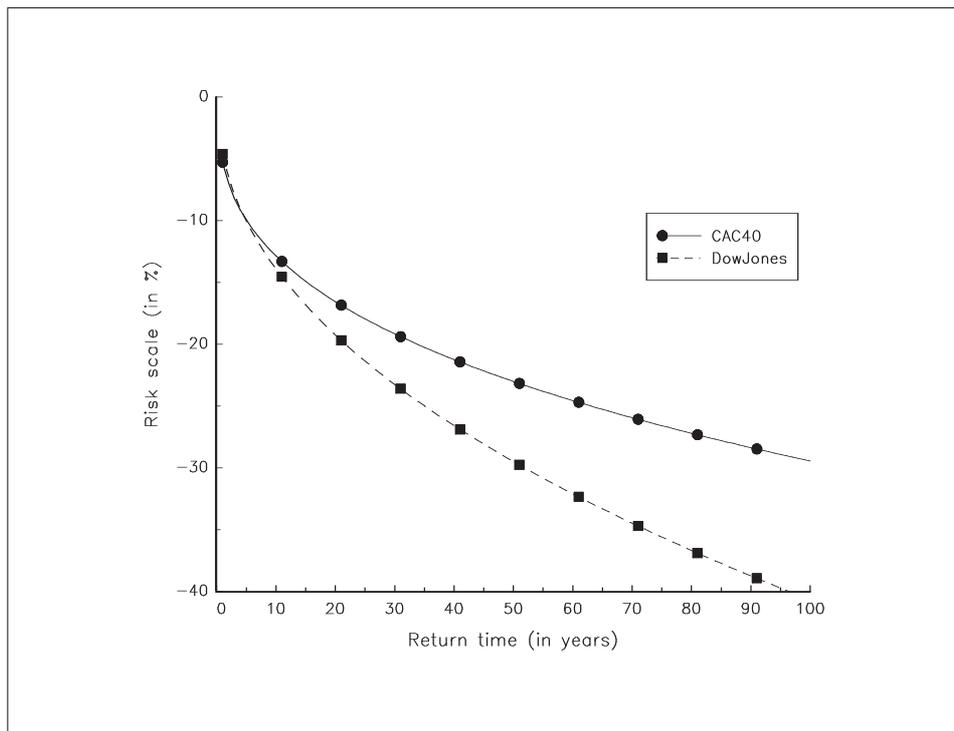
Par exemple, pour \mathcal{T} égal à 5 ans, α_{GEV} prend la valeur de 98%.

Nous considérons l'exemple donné dans RONCALLI [2001]. Les échelles de risque des indices CAC40 et DowJones sont les suivantes :

Echelles de Risque pour des positions longues (en %)		
Temps de Retour	CAC40	DowJones
5	-9.93	-10.04
10	-12.85	-13.91
25	-17.95	-21.35
50	-23.01	-29.48
75	-26.59	-35.59
100	-29.44	-40.67

Echelles de Risque pour des positions courtes (en %)		
Temps de Retour	CAC40	DowJones
5	9.95	6.13
10	13.06	7.38
25	18.66	9.36
50	24.41	11.13
75	28.55	12.30
100	31.91	13.19

Grâce à ces tableaux, le risk manager peut évaluer les temps de retour implicites d'un scénario de crise unidimensionnel (voir le graphique 4.2).



Graphique 4.2. Temps de retour pour les positions longues

4.1.3.2 Cas multidimensionnel

Dans le cas unidimensionnel, nous avons utilisé la théorie des valeurs extrêmes pour construire des scénarios de crise et/ou pour calculer des temps de retour implicites pour des scénarios de crise donnés (c'est-à-dire pour probabiliser ces scénarios de crise). Dans le cas multidimensionnel, la construction de scénarios de crise à partir de la théorie multivariée des valeurs extrêmes est difficile. C'est pourquoi nous utiliserons la théorie des extrêmes uniquement pour calculer des temps de retour implicites.

Les scénarios bivariés de crise peuvent être vus comme la représentation d'une zone de défaillance. Soient χ_1^+ et χ_2^+ les statistiques du maximum des variables aléatoires X_1 et X_2 . Nous avons

$$\begin{aligned}
\Pr(\chi_1^+ > \chi_1, \chi_2^+ > \chi_2) &= 1 - \Pr(\chi_1^+ \leq \chi_1) - \Pr(\chi_2^+ \leq \chi_2) + \Pr(\chi_1^+ \leq \chi_1, \chi_2^+ \leq \chi_2) \\
&= 1 - \mathbf{F}_1(\chi_1) - \mathbf{F}_2(\chi_2) + \mathbf{C}(\mathbf{F}_1(\chi_1), \mathbf{F}_2(\chi_2)) \\
&= \overline{\mathbf{C}}(\mathbf{F}_1(\chi_1), \mathbf{F}_2(\chi_2))
\end{aligned} \tag{4.20}$$

Soit \mathcal{T} le temps de retour. La zone de défaillance est représentée par

$$\left\{ (\chi_1, \chi_2) \in \mathbb{R}^2 \mid \overline{\mathbf{C}}(\mathbf{F}_1(\chi_1), \mathbf{F}_2(\chi_2)) < \frac{1}{\mathcal{T}} \right\} \quad (4.21)$$

Reprenons l'exemple du CAC40 et du DowJones. En utilisant une copule Gumbel et des marges GEV, nous pouvons définir la zone de défaillance à 5 ans (voir le graphique 4.3). Notons que pour un scénario de crise donné, nous pouvons facilement calculer le temps de retour implicite en utilisant la relation suivante :

$$\mathcal{T} = \frac{1}{\Pr(\chi_1^+ > \chi_1, \chi_2^+ > \chi_2)} = \frac{1}{\overline{\mathbf{C}}(\mathbf{F}_1(\chi_1), \mathbf{F}_2(\chi_2))} \quad (4.22)$$

La structure de dépendance joue alors un rôle crucial pour le calcul de \mathcal{T} . Supposons une simulation de crise correspondant à une baisse simultanée de 10% du CAC40 et 10% du DowJones. Voici les temps de retour implicites unidimensionnels :

	CAC40	DowJones
\mathcal{T}	5.090	4.959

Si les deux actifs sont parfaitement dépendants, alors le temps de retour du couple (CAC40, DowJones) est le maximum des temps de retour unidimensionnels. Mais si la dépendance n'est pas parfaite, alors nous obtenons des temps de retour beaucoup plus élevés. Par exemple, nous obtenons un temps de retour égal à 304 années dans le cas de l'indépendance. En l'absence d'information sur la dépendance, les bornes minimales et maximales du temps de retour sont donc

$$5.090 \leq \mathcal{T} \leq 304.12$$

Un programme de stress-testing doit être un scénario extrême susceptible de se réaliser. Il est totalement absurde de faire des scénarios de crise présentant des temps de retour de plusieurs centaines d'années. En général, nous privilégions des temps de retour de 5 ans, 10 ans voire 25 ans. C'est pourquoi il est important d'avoir une estimation de la fonction de dépendance. Dans notre exemple, le temps de retour \mathcal{T} est égal à 10.779 années. Il est deux fois plus grand que la borne inférieure, mais beaucoup plus faible que les 304 années !

Remarque 30 *Vous trouverez d'autres exemples dans BOUYÉ, DURRLEMAN, NIKEGHBALI, RIBOULET et RONCALLI [2000].*

4.1.4 Dépendance de queue et valeur en risque

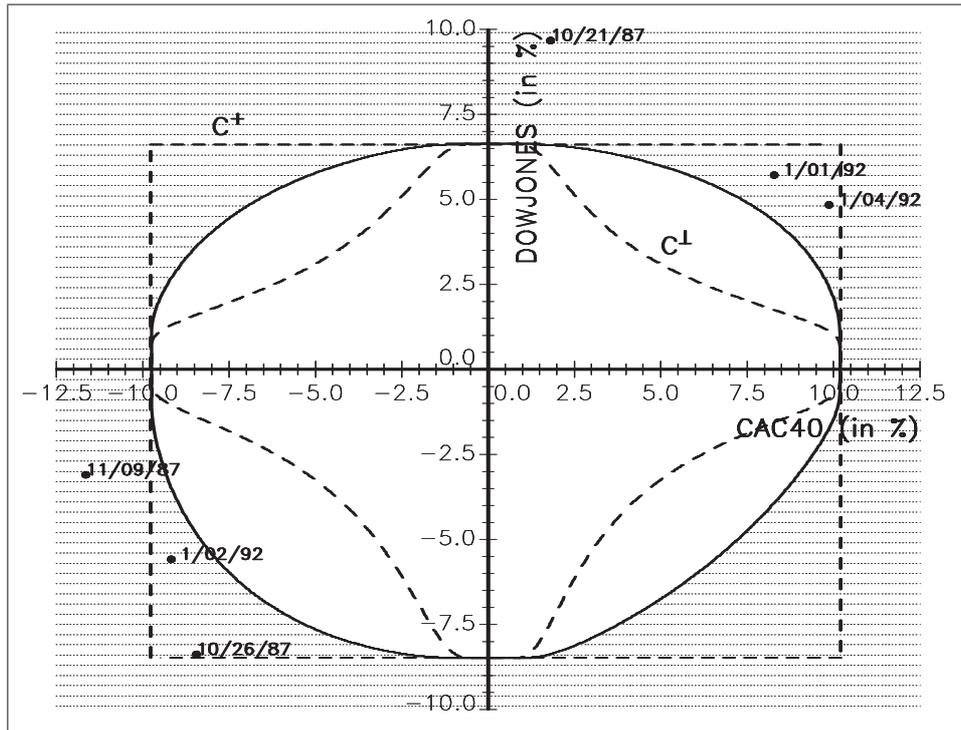
Nous finissons cette section avec un exemple donné par BOUYÉ, DURRLEMAN, NIKEGHBALI, RIBOULET et RONCALLI [2001]. Nous considérons trois portefeuilles composés d'actifs du London Metal Exchange :

	AL	AL-15	CU	NI	PB
P_1	1	1	1	1	1
P_2	-1	-1	-1	1	1
P_3	2	1	-3	4	5

Pour illustrer l'impact des queues de distribution et de la dépendance de queue sur le calcul de la valeur en risque, nous considérons différents cas :

- marges gaussiennes et copule Normale

	90%	95%	99%	99.5%	99.9%
P_1	7.26	9.33	13.14	14.55	17.45
P_2	4.04	5.17	7.32	8.09	9.81
P_3	13.90	17.82	25.14	27.83	33.43



Graphique 4.3. Zone de défaillance pour le couple (CAC40,DowJones) et un temps de retour de 5 ans

– marges Student ($\nu = 4$) et copule Normale

	90%	95%	99%	99.5%	99.9%
P_1	6.51	8.82	14.26	16.94	24.09
P_2	3.77	5.00	7.90	9.31	13.56
P_3	12.76	17.05	27.51	32.84	49.15

– marges gaussiennes et copule Student ($\nu = 1$)

	90%	95%	99%	99.5%	99.9%
P_1	5.69	7.95	13.19	15.38	20.06
P_2	3.82	5.55	9.75	11.65	16.41
P_3	13.41	19.36	34.16	40.55	54.48

Pour des seuils de confiance élevés, la valeur en risque dépend fortement des queues de distribution et de la dépendance de queue. Le choix d'une modélisation n'est donc pas neutre sur la mesure de risque.

4.2 La prise en compte des événements extrêmes dans le risque opérationnel

Le Nouvel Accord de Bâle prévoit la prise en compte des risques opérationnels dans le calcul des fonds propres réglementaires. Les risques opérationnels correspondent aux pertes qui ne relèvent pas du risque de marché ou du risque de crédit. Une définition plus précise est la suivante :

«Les risques opérationnels se définissent comme le risque de pertes dues à une inadéquation ou à une défaillance des procédures, personnels, systèmes internes ou à des événements extérieurs» (THORAVAL [2001]).

Dans son dernier document consultatif [2], le Comité de Bâle propose un éventail de méthodes pour mesurer le capital réglementaire au titre du risque opérationnel : BIA, SA et AMA. Parmi les méthodes avancées, il est possible d'utiliser un modèle interne pour calculer la charge en capital. Ce modèle interne est normalisé puisque les choix méthodologiques ne relèvent pas de la banque. La méthodologie retenue provient de la théorie du risque en Assurances (Collective Risk Theory) et correspond à l'approche des distributions de perte (**Loss Distribution Approach** ou LDA). Cette approche est intensivement exposée dans FRACHOT, GEORGES et RONCALLI [2001].

Dans la méthode LDA, nous cherchons à modéliser la perte totale des risques opérationnels (pour un type de risque donné et une ligne de métier donnée) qui est perçue comme une perte agrégée (**aggregate loss distribution**) de différents événements. Celle-ci se définit donc par

1. le nombre de pertes individuelles,
2. et le montant de chaque pertes individuelles.

A priori, nous ne connaissons pas le nombre de pertes ainsi que le montant des pertes pour la future année. C'est pourquoi nous allons les considérer aléatoires. Nous allons donc modéliser ce nombre de pertes N par un processus de comptage, et la distribution \mathbf{P} de ce nombre de pertes est appelée la distribution de la fréquence des pertes (**loss frequency distribution**). Nous supposons aussi que les pertes individuelles sont indépendantes et identiquement distribuées. La distribution du montant d'une perte individuelle ζ est appelée la distribution de la sévérité des pertes \mathbf{F} (**loss severity distribution**). Dans ce cas là, la perte agrégée ϑ est la somme **aléatoire** (car N est aléatoire) des pertes individuelles :

$$\vartheta = \sum_{n=1}^N \zeta_n \quad (4.23)$$

La distribution \mathbf{G} de la perte agrégée est donc une distribution composée (**compound distribution**) :

$$\begin{aligned} \mathbf{G}(x) = & \Pr\{N = 1\} \times \mathbf{F}(x) + \Pr\{N = 2\} \times \mathbf{F}^{2*}(x) + \\ & \Pr\{N = 3\} \times \mathbf{F}^{3*}(x) + \dots \end{aligned} \quad (4.24)$$

\mathbf{F}^{n*} est la distribution de la somme de n pertes ζ_1, \dots, ζ_n . La charge en capital réglementaire (Capital-at-Risk) correspond à la valeur en risque au seuil de confiance α :

$$\text{CaR}(\alpha) = \mathbf{G}^{-1}(\alpha) \quad (4.25)$$

Pour information, le Comité de Bâle a fixé α égal à 99.9% dans son dernier document consultatif.

4.2.1 Modélisation de la distribution de sévérité

La grande difficulté de modélisation de la méthode LDA est la spécification de la distribution \mathbf{F} des distributions des pertes, ainsi que son estimation. En pratique, nous disposons d'une base de données historiques de pertes individuelles. Pour certains types de risque et/ou certaines lignes de métier, nous observons des pertes allant de quelques euro à plusieurs centaines de millions d'euro. **Le support de la distribution \mathbf{F} doit donc être extrêmement grand.** Une autre caractéristique du risque opérationnel est le caractère **extrême** de certains événements. Pour de nombreux types de risque, quelques pertes suffisent à représenter plus de 90% des pertes totales. Donc le capital économique sert à couvrir ces quelques pertes extrêmes. Pour ces types de risque, la fréquence a une faible influence. Ce qui est important est la queue de distribution

de la sévérité. A noter que c'est le contraire pour certains types de risque. La fréquence peut être parfois un élément clé de la mesure en risque. C'est par exemple le cas des fraudes sur cartes bleues ou sur chèques. Dans ce cas, la distribution de la sévérité est "bornée". C'est pourquoi nous devons distinguer les risques opérationnels selon la classification **risque de fréquence/risque de sévérité**.

Précisons ce que peut être un risque de sévérité. Pour cela, nous utilisons les données réelles du Crédit Lyonnais. Pour un type de risque, nous calculons le rapport entre la perte maximale observée et le quantile α empirique. Nous obtenons les résultats suivants :

α	Maximum/Quantile
50%	$\simeq 160000$
75%	$\simeq 27000$
90%	$\simeq 13825$

Bien sûr, ce type de risque présente des rapports Maximum/Quantile peu communs. Néanmoins, cela illustre parfaitement la nature de nombreux types de risque opérationnel. A titre de comparaison, le rapport du quantile 99.9999% de la distribution gaussienne sur le quantile 90% de la distribution gaussienne est beaucoup plus faible :

$$\frac{\Phi^{-1}(99.9999\%)}{\Phi^{-1}(90\%)} = 3.7 \quad (4.26)$$

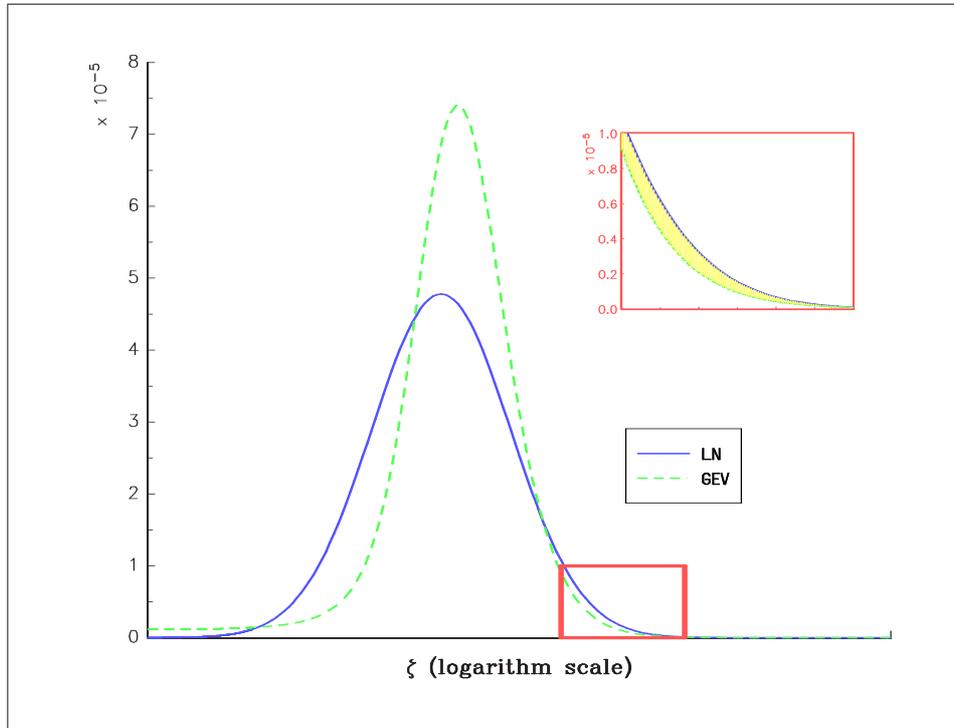
En fait, le risque opérationnel est sûrement le risque qui présente le plus d'événements rares. Le choix de la distribution de sévérité est donc crucial. Généralement, les distributions les plus utilisées sont les lois lognormale, Pareto, Weibull, GEV, etc.

Considérons un type de risque opérationnel qui correspond à **un risque de sévérité**. La grande difficulté de modélisation est de fournir une estimation de la distribution \mathbf{F}

- qui ne sous-estime pas les risques extrêmes afin de ne pas sous-estimer la charge en capital, et
- qui ne sur-estime pas les risques extrêmes afin de ne pas sur-estimer la charge en capital.

Nous cherchons à obtenir la mesure de risque la plus juste possible. Dans FRACHOT, GEORGES et RONCALLI [2001], nous présentons des méthodes qui permettent de vérifier que la distribution \mathbf{F} ne sous-estime pas et ne sur-estime pas les risques extrêmes. Ces méthodes sont basées sur les statistiques d'ordre. L'idée sous-jacente est de comparer la fonction de survie $\mathbf{S}_{i:n}$ des statistiques d'ordre les plus grandes avec les pertes observées les plus importantes. Afin que la comparaison ait un sens, n est fixé au nombre d'observations de la base de données. D'une certaine manière, ces probabilités de survie $\mathbf{S}_{i:n}(x) = 1 - \mathbf{F}_{i:n}(x) = \Pr\{\zeta_{i:n} \geq x\}$ indiquent les probabilités d'avoir des pertes plus grandes que celles qui ont été observées pour les n prochaines pertes. A titre d'exemple, le graphique 4.4 présente l'estimation de la distribution \mathbf{F} pour le type de risque "vols/fraudes" avec les données internes Crédit Lyonnais. Pour des raisons de confidentialité, nous avons supprimé l'échelle de l'axe des pertes. Nous avons considéré deux distributions, la distribution lognormale et la distribution généralisée des valeurs extrêmes. Sur le graphique 4.5, nous reportons $\mathbf{S}_{n:n}$ ainsi que les pertes observées les plus grandes. Dans le cas de la distribution \mathcal{GEV} , la probabilité d'avoir une perte supérieure à la première perte est égale à 1. Donc, nous avons une probabilité certaine d'avoir un maximum plus grand si n nouvelles pertes se réalisent. Il est clair que ceci est inacceptable et que la distribution \mathcal{GEV} estimée sur-estime (le mot est faible) les événements extrêmes. Concernant la distribution \mathcal{LN} , les résultats sont "acceptables". Supposons maintenant que nous utilisons une distribution gaussienne pour modéliser la distribution des pertes pour le type de risque "vols/fraudes". Dans ce cas, la probabilité d'avoir

un maximum supérieur aux cinq pertes observées les plus importantes est inférieure à 10^{-5} . L'intérêt des statistiques d'ordre est donc de fournir une puissante "loupe" pour analyser la queue de distribution afin de repérer assez facilement une sur-estimation ou une sous-estimation des événements extrêmes.



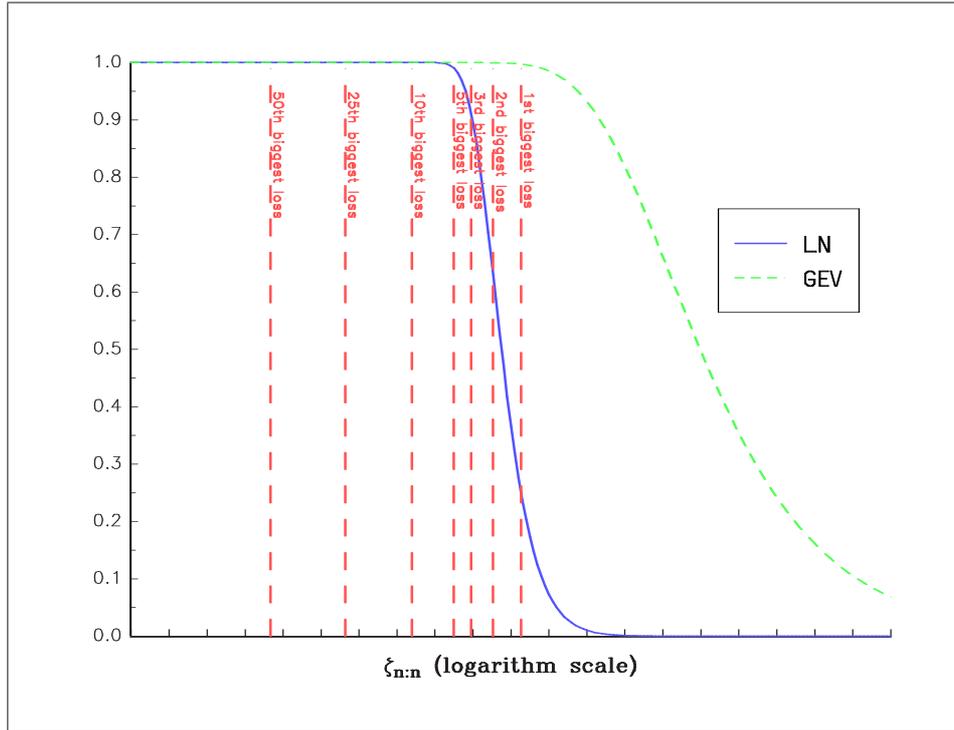
Graphique 4.4. Estimation des distributions \mathcal{LN} and \mathcal{GEV} pour le risque "vols/fraudes"

4.2.2 Aggrégation des risques

L'aggrégation est un problème relativement difficile dans le cas du risque opérationnel. Bien sûr, nous pouvons faire l'hypothèse que les différents types de risque sont indépendants. Dans ce cas, le problème est résolu. Si nous voulons prendre en compte la dépendance entre les risques, nous pouvons

- soit corrélérer les fréquences ;
- soit corrélérer les pertes agrégées.

La première méthode a été proposée par BOUYÉ, DURRLEMAN, NIKEGHBALI, RIBOULET et RONCALLI [2000,2001]. L'idée est de construire des distributions multidimensionnelles de variables



Graphique 4.5. Fonction de survie de la statistique du maximum et pertes observées les plus importantes aléatoires de Poisson à partir des copules :

$$\begin{aligned}
 \Pr \{N_1 = n_1, N_2 = n_2\} &= \mathbf{C} \left(\sum_{n=0}^{n_1} \frac{\lambda_1^n e^{-\lambda_1}}{n!}, \sum_{n=0}^{n_2} \frac{\lambda_2^n e^{-\lambda_2}}{n!} \right) - \\
 &\quad \mathbf{C} \left(\sum_{n=0}^{n_1-1} \frac{\lambda_1^n e^{-\lambda_1}}{n!}, \sum_{n=0}^{n_2} \frac{\lambda_2^n e^{-\lambda_2}}{n!} \right) - \\
 &\quad \mathbf{C} \left(\sum_{n=0}^{n_1} \frac{\lambda_1^n e^{-\lambda_1}}{n!}, \sum_{n=0}^{n_2-1} \frac{\lambda_2^n e^{-\lambda_2}}{n!} \right) + \\
 &\quad \mathbf{C} \left(\sum_{n=0}^{n_1-1} \frac{\lambda_1^n e^{-\lambda_1}}{n!}, \sum_{n=0}^{n_2-1} \frac{\lambda_2^n e^{-\lambda_2}}{n!} \right) \quad (4.27)
 \end{aligned}$$

A titre d'illustration, les tables 4.1 et 4.2 correspondent aux probabilités $p_{i,j} = \Pr \{N_1 = i, N_2 = j\}$ de la distribution de Poisson bivariable avec $\lambda_1 = 1$, $\lambda_2 = 1$ et une copule Normale de paramètre ρ .

Néanmoins, les simulations que nous avons faites montrent que le paramètre ρ de la copule Normale a très peu d'influence sur la mesure de risque. Les explications sont les suivantes :

- Le choix de la copule Normale ne permet pas de mettre de la dépendance sur les extrêmes. Une copule de valeurs extrêmes a une influence plus importante sur la mesure de risque.
- L'absence de dépendance sur les extrêmes est accentuée par deux phénomènes : les fréquences sont discrètes (ce qui veut dire par exemple que la mesure de risque sera plus sensible à la copule si les fréquences de Poisson sont importantes, car le caractère discret de la distribution est moins accentué) et la mesure de risque porte sur la distribution agrégée. Nous pouvons très bien avoir deux extrêmes N_1 et N_2 , cela ne veut pas dire que nous obtenons nécessairement un extrême pour la somme des deux distributions composées, car le risque

$p_{i,j}$	0	1	2	3	4	5	...	$p_{i\cdot}$
0	0.0945	0.133	0.0885	0.0376	0.0114	0.00268		0.368
1	0.0336	0.1	0.113	0.0739	0.0326	0.0107		0.368
2	0.00637	0.0312	0.0523	0.0478	0.0286	0.0123		0.184
3	0.000795	0.00585	0.0137	0.0167	0.013	0.0071		0.0613
4	7.28E-005	0.000767	0.00241	0.00381	0.00373	0.00254		0.0153
5	5.21E-006	7.6E-005	0.000312	0.000625	0.000759	0.000629		0.00307
⋮								
$p_{\cdot,j}$	0.135	0.271	0.271	0.18	0.0902	0.0361		1

TABLE 4.1. Loi de Poisson bidimensionnelle avec $\rho = 0.5$

$p_{i,j}$	0	1	2	3	4	5	...	$p_{i\cdot}$
0	0.0136	0.0617	0.101	0.0929	0.058	0.027		0.368
1	0.0439	0.112	0.111	0.0649	0.026	0.00775		0.368
2	0.0441	0.0683	0.0458	0.0188	0.00548	0.00121		0.184
3	0.0234	0.0229	0.0109	0.00331	0.000733	0.000126		0.0613
4	0.00804	0.00505	0.00175	0.000407	7.06E-005	9.71E-006		0.0153
5	0.002	0.00081	0.000209	3.79E-005	5.26E-006	5.89E-007		0.00307
⋮								
$p_{\cdot,j}$	0.135	0.271	0.271	0.18	0.0902	0.0361		1

TABLE 4.2. Loi de Poisson bidimensionnelle avec $\rho = -0.5$

opérationnel est avant tout **un risque de sévérité** (quelques pertes extrêmes suffisent à représenter 95% du risque opérationnel).

Dans FRACHOT, GEORGES et RONCALLI [2001], nous présentons une seconde méthode pour agréger les risques qui est beaucoup plus acceptable. La dépendance est directement mise sur les pertes. Soient ϑ_1 et ϑ_2 deux pertes correspondant à deux types différents de risque opérationnel. Soit \mathbf{G} la distribution jointe. Nous avons

$$\mathbf{G}(x_1, x_2) = \mathbf{C}(\mathbf{G}_1(x_1), \mathbf{G}_2(x_2)) \quad (4.28)$$

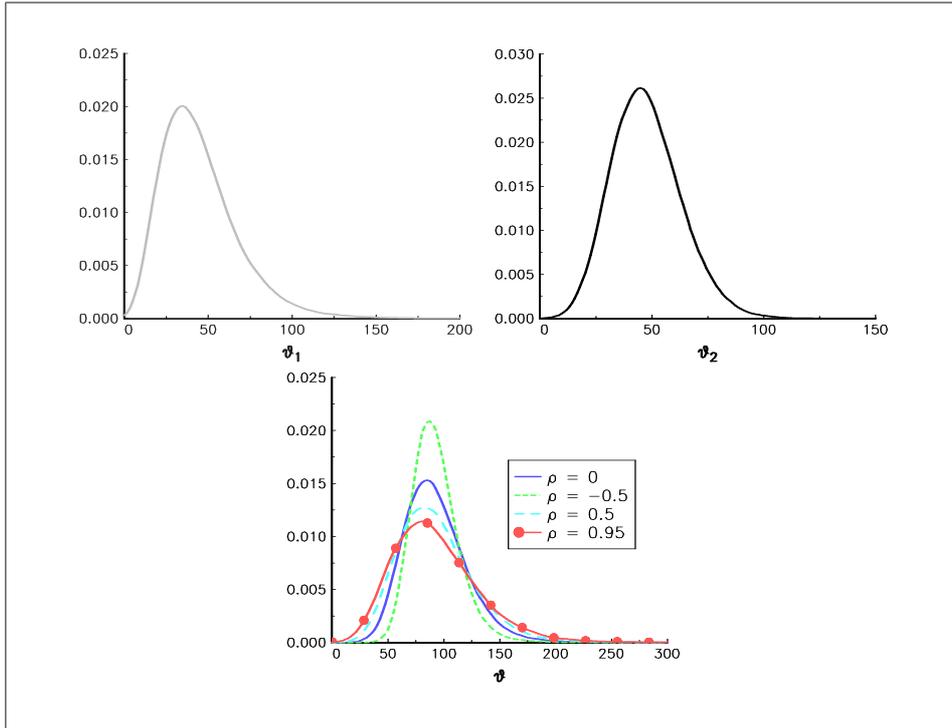
Dans ce cas, \mathbf{G}_1 et \mathbf{G}_2 sont les deux distributions composées qui correspondent aux variables aléatoires ϑ_1 et ϑ_2 . Soit \mathbf{G}_{1+2} la distribution de $\vartheta_1 + \vartheta_2$. Nous avons

$$\begin{aligned} \mathbf{G}_{1+2}(x) &= \Pr\{\vartheta_1 + \vartheta_2 \leq x\} \\ &= \iint_{x_1+x_2 \leq x} d\mathbf{C}(\mathbf{G}_1(x_1), \mathbf{G}_2(x_2)) \end{aligned} \quad (4.29)$$

et

$$\text{CaR}(\alpha) = \mathbf{G}_{1+2}^{-1}(\alpha) \quad (4.30)$$

La grande difficulté de cette méthode est l'aspect numérique pour obtenir $\text{CaR}(\alpha)$. En général, nous n'avons pas d'expression analytique des marges \mathbf{G}_1 et \mathbf{G}_2 . Il est donc difficile de construire la distribution \mathbf{G} et encore plus difficile d'obtenir \mathbf{G}_{1+2} et $\text{CaR}(\alpha)$. Néanmoins, nous pouvons contourner ces difficultés en utilisant une méthode de Monte Carlo. Dans ce cas, la simulation de \mathbf{G} se fait par la méthode des quantiles empiriques. L'exemple suivant est celui donné par



Graphique 4.6. Impact de la fonction de dépendance sur la distribution de la perte totale

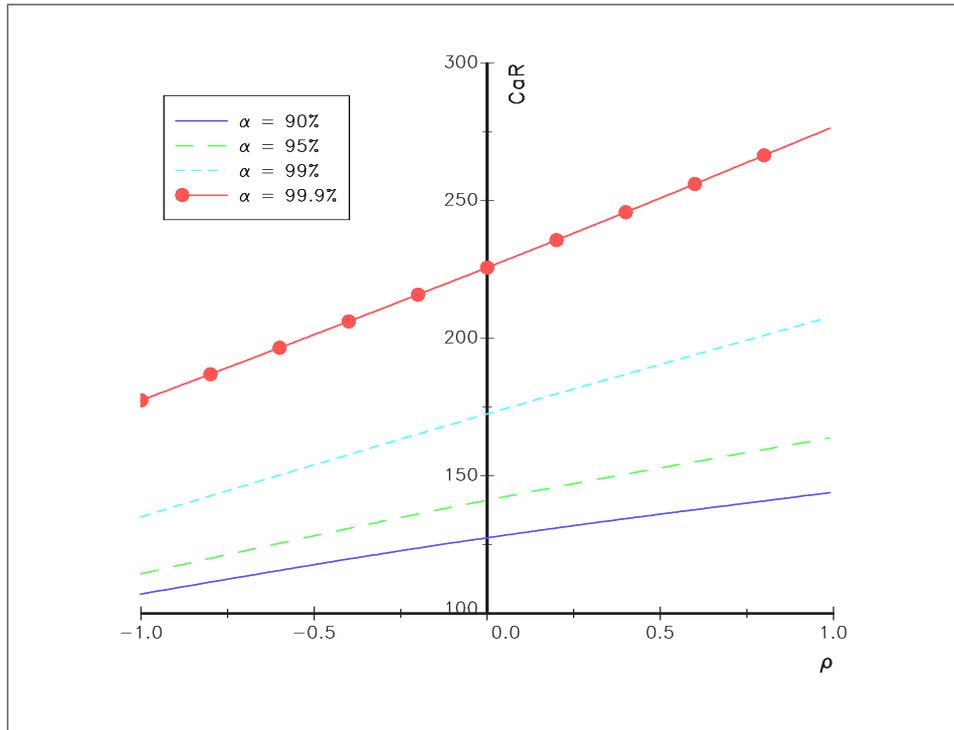
FRACHOT, GEORGES et RONCALLI [2001]. Nous avons $\zeta_1 \sim \mathcal{LN}(1, 1)$ et $\zeta_2 \sim \mathcal{LN}(1.25, 0.5)$. Les distributions de fréquence sont respectivement $\mathcal{P}(10)$ et $\mathcal{P}(12)$. Sur le graphique 4.6, nous représentons la distribution de la perte totale $\vartheta = \vartheta_1 + \vartheta_2$ pour différentes copules Normales. La charge en capital correspondante est reportée sur le graphique 4.7. Il est intéressant de noter que le *Capital-at-Risk* semble être une fonction linéaire du paramètre de ρ de la copule. **Encore une fois, cette propriété provient du caractère gaussien de la copule.** De plus, ceci n'est plus forcément vrai si les distributions lognormales ont un paramètre σ plus élevé.

4.2.3 Méthodes de backtesting

Il est important de disposer d'outils pour vérifier la robustesse des estimations. Une nouvelle fois, nous pouvons utiliser la théorie des extrêmes en échantillon fini (c'est-à-dire les statistiques d'ordre extrémales). Prenons un exemple. Nous pouvons calculer la fonction de survie du maximum de N pertes de loi de probabilité \mathbf{F} . Nous avons

$$p_{\max} := 1 - [\mathbf{F}(x)]^N \quad (4.31)$$

Cette probabilité a une interprétation très intuitive. Supposons que N pertes surviennent. p_{\max} est la probabilité que la perte maximale soit supérieure au montant x . Le Risk Management a des idées (ou des benchmarks internes et externes) sur la perte maximale à un an, deux ans, etc. Il peut donc confronter ces propres croyances avec les résultats donnés par le modèle. Supposons que 100 pertes surviennent en moyenne par an. Supposons aussi que la distribution de la sévérité est une distribution log-normale $\mathcal{LN}(\mu, \sigma)$. Une première estimation à partir des données historiques donne $\mu = 10$ et $\sigma = 2, 50$. Sur la table 4.3, nous avons reporté les valeurs correspondantes de p_{\max} pour différentes valeurs de N et x . Par exemple, la probabilité d'observer une perte maximale sur 5 ans supérieure à 100 millions est de 17,2%. Si l'estimation a été faite avec un historique de 5 ans, et que nous observons dans la base de données plusieurs pertes supérieures à 100 millions, nous pouvons penser que cette distribution sous-estime largement



Graphique 4.7. Impact of the parameter ρ on the *Capital-at-Risk*

N	τ	1 million	10 millions	100 millions	1 milliard	5 milliards
100	1 an	99.8	51.4	3.7	0.08	0.00
200	2 ans	99.9	76.4	7.2	0.17	0.00
300	3 ans	100.0	88.5	10.7	0.26	0.01
400	4 ans	100.0	94.4	14.0	0.35	0.01
500	5 ans	100.0	97.3	17.2	0.44	0.02
1000	10 ans	100.0	99.9	31.4	0.89	0.04
5000	50 ans	100.0	100.0	84.9	4.38	0.20

TABLE 4.3. $\mu = 10$ and $\sigma = 2.50$

les pertes exceptionnelles. Si $\mu = 10$ et $\sigma = 2,75$, nous obtenons la table 4.4. Dans ce cas, la probabilité est égale à 42,2%. La statistique du maximum permet donc de construire des tableaux de probabilités de perte maximale qui sont très utiles pour le Risk Management.

4.2.4 LDA et théorie des extrêmes

Il existe une littérature qui veut absolument construire une mesure du risque opérationnel à partir de la théorie des extrêmes (CESKE, et HERNÁNDEZ [1999], CESKE, HERNÁNDEZ et SÁNCHEZ [2000], CRUZ [2001], EBNÖTHER, VANINI, MCNEIL et ANTOLINEZ-FEHR [2001], MEDOVA et KYRIACOU [2001]). Bien sûr, nous pouvons toujours utiliser une distribution GEV pour modéliser la distribution de sévérité. Mais cela ne relève pas de la théorie des extrêmes. L'utilisation de la théorie des extrêmes implique de modéliser le distribution du maximum des pertes. La difficulté avec le risque opérationnel est que la perte totale provient de plusieurs pertes individuelles. Les auteurs sont cependant très "flous" sur l'incorporation de la fréquence sur le risque opérationnel. Bien sûr, une solution consiste à ne considérer qu'une perte annuelle. Ce qui n'est pas faux pour certains types de risque pour lesquels la perte totale est principalement déterminée par une ou deux pertes extrêmes. **Néanmoins, à l'heure actuelle, c'est**

N	τ	1 million	10 millions	100 millions	1 milliard	5 milliards
100	1 an	99.9	73.1	10.4	0.48	0.03
200	2 ans	100.0	92.7	19.7	0.95	0.07
300	3 ans	100.0	98.0	28.1	1.43	0.10
400	4 ans	100.0	99.4	35.5	1.91	0.14
500	5 ans	100.0	99.8	42.2	2.38	0.18
1000	10 ans	100.0	99.9	66.7	4.70	0.36
5000	50 ans	100.0	100.0	99.5	21.42	1.80

TABLE 4.4. $\mu = 10$ and $\sigma = 2.75$

totalemment irréaliste de penser que l'on puisse construire une mesure du risque opérationnel sur la théorie des extrêmes car

- les bases de données ne sont pas suffisamment importantes (quelques années à peine de collecte exhaustive) ;
- et les estimations sont vraiment trop instables (ce qui est paradoxal, puisque le risque opérationnel est par nature un risque de valeurs extrêmes, mais un risque de valeurs extrêmes fortement hétérogènes).

4.3 Le risque de crédit

Faute de temps, je ne pourrais pas développer cette section de façon intensive. Je donne donc ici quelques pistes. Vous trouverez des approfondissements de cette section dans les notes de cours de l'ENSAI "Gestion des Risques Multiples (Application des Copules à la Gestion des Risques)".

4.3.1 Estimation des probabilités de défaillance

L'un des inputs les plus importants dans un système de gestion/tarifcation du risque de crédit est la fonction de survie d'un crédit (ou les probabilités de défaillance pour différents horizons). Ces probabilités de défaillance dépendent généralement d'une note issue d'un système de notation interne de contreparties. Selon la grille de notation adoptée, les probabilités de défaillances pour un horizon d'un an peuvent varier de 0.01% à 25%! Pour les contreparties les mieux notées, la probabilité de défaillance est extrêmement faible, même pour des horizons de 25 ans (il n'est pas rare que cette probabilité soit inférieure à 2%). Ces défaillances sont donc des événements très rares. Pour une application de la théorie des valeurs extrêmes à l'estimation des fonctions de survie, vous pouvez consulter HAMILTON, JAMES et WEBBER [2001].

4.3.2 Modélisation des défauts joints

Dans ce paragraphe, nous cherchons à 'corrélér' les défaillances dans une optique de mesure de risque crédit, ou de pricing de basket credit derivatives. Nous considérons I contreparties dont les probabilités de survie sont données par les fonctions de survie $\mathbf{S}_1, \dots, \mathbf{S}_I$. Pour corrélér les temps de défaillance, nous pouvons utiliser la représentation canonique d'une fonction de survie multidimensionnelle :

$$\mathbf{S}(t_1, \dots, t_I) = \check{\mathbf{C}}(\mathbf{S}_1(t_1), \dots, \mathbf{S}_I(t_I)) \quad (4.32)$$

avec $\check{\mathbf{C}}$ une fonction copule de survie. Cette méthode a été suggérée par LI [2000a,2000b]. Du fait que les probabilités de défaillance sont très petites, il convient de choisir la copule appropriée

Č afin d'avoir une dépendance de queue (lower tail dependence) — voir GEORGES, LAMY, NICOLAS, QUIBEL et RONCALLI [2001]. La caractéristique de la copule peut expliquer certaines différences de mesure de risque crédit lorsque le portefeuille contient des contreparties très bien notées. Par exemple, la copule implicite de CreditMetrics est la copule Normale alors que la copule implicite de CreditRisk+ est le copule Cook-Johnson (COUTANT, MARTINEU, MESSINES, RIBOULET et RONCALLI [2001]).

Une seconde approche a été proposée récemment (GIESECKE [2001a,2001b], SCHÖNBUCHER et SCHUBERT [2001]). Elle est basée sur les modèles à intensité ou modèles de forme réduite. Dans ce cas, le temps de défaillance est le premier instant de passage du processus d'intensité intégré au dessus d'une barrière stochastique :

$$\tau_i := \inf \left\{ t \geq 0 : \int_0^t \lambda_s^i ds \geq \theta_i \right\} \quad (4.33)$$

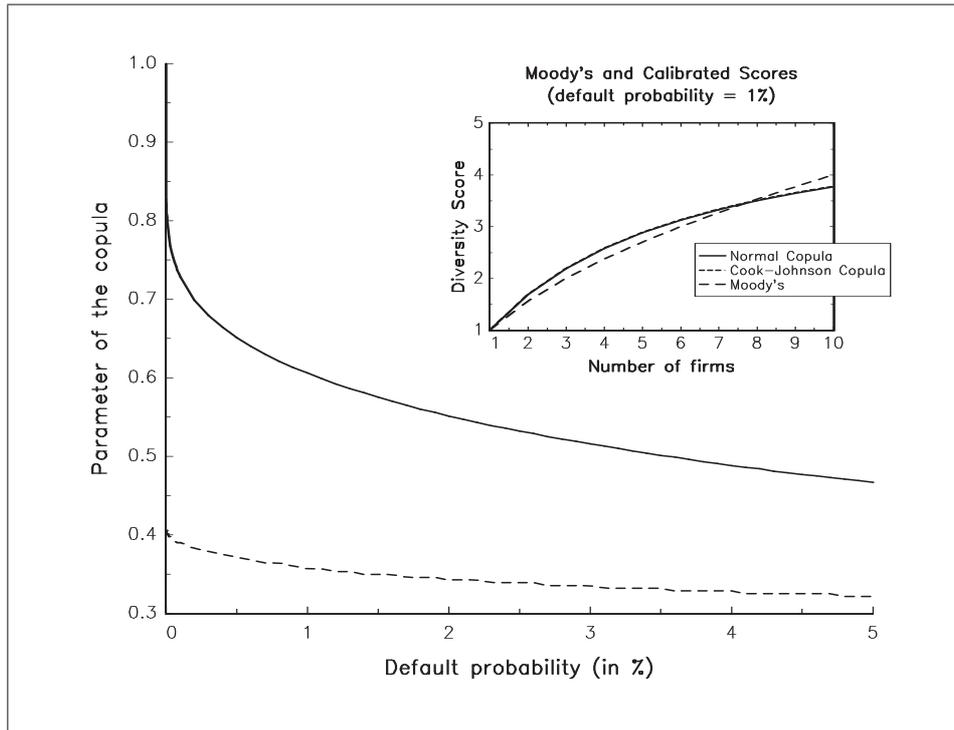
où λ^i est le processus d'intensité du crédit i . Généralement, θ_i est une variable aléatoire exponentielle standard. Pour corrélérer les défaillances, nous pouvons corrélérer les processus d'intensités et les barrières stochastiques. Dans ce cas, nous lions directement les seuils $\boldsymbol{\theta} = (\theta_1, \dots, \theta_I)$ avec une copule \mathbf{C} . Sous certaines conditions (intensités déterminites), les deux approches sont équivalentes.

Encore une fois, le choix de la copule est crucial (MASHAL et NALDI [2002], SCHMIDT et WARD [2002]). Par exemple, nous pouvons montrer que la dépendance de queue a une influence importante sur les changements de spread conditionnels à une défaillance et sur les prix des baskets credit derivatives. Si nous cherchons à calibrer la fonction copule à partir du score de diversification de Moody's, la stabilité de valeur du paramètre de la copule dépend aussi de la dépendance de queue (JOUANIN, RAPUCH, RIBOULET et RONCALLI [2001]). A titre d'illustration, nous reportons le graphique 4.8 de JOUANIN, RIBOULET et RONCALLI [2001] qui concluent :

[...] as far as the choice of the copula is concerned, we first acknowledge a possible non negligible model risk in the choice of the copula family. In KMV or CreditMetrics methodologies, it is usually assumed that latent variables are correlated through a Normal copula. As we mentioned before, the calibration will not succeed for this family, since this copula behaves like the independent copula in the range of very low probabilities. A more realistic family may be the Cook-Johnson family (which is the implicit copula in CreditRisk+). In Figure 4.8, we can notice that the Cook-Johnson copula is more robust with respect to default probability than the Normal copula. Indeed, the calibration procedure leads to a quite steady optimal parameter when p ranges from 1 to 500 bps, whereas the implied correlation of the Normal copula tends to 1 when p tends to 0.

Références

- [1] Basel Committee on Banking Supervision, Operational Risk — Consultative Document, Supporting document to the New Basel Capital Accord, January 2001
- [2] Basel Committee on Banking Supervision, Working Paper on the Regulatory Treatment of Operational Risk, september 2001
- [3] BEZAT, A. et A. NIKEGHBALI [2000], La théorie des extrêmes et la gestion des risques de marché, GT ENSAE (document disponible sur le site web du GRO)



Graphique 4.8. Influence de la probabilité de défaillance et du choix de la copule

- [4] BOUYÉ, E., V. DURRLEMAN, A. NIKEGHBALI, G. RIBOULET et T. RONCALLI [2000], Copulas for finance — a reading guide and some applications, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [5] BOUYÉ, E., V. DURRLEMAN, A. NIKEGHBALI, G. RIBOULET et T. RONCALLI [2001], Copulas : an open field for risk management, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [6] CESKE, R. et J.V. HERNÁNDEZ [1999], Where theory meets practice, *Risk Magazine (Operational Risk Report)*, **12**, November, 17-20
- [7] CESKE, R., J.V. HERNÁNDEZ et L.M. SÁNCHEZ [2000], Quantifying event risk : the next convergence, *The Journal of Risk Finance*, **1(3)**, 9-23
- [8] COUTANT, S., P. MARTINEU, J. MESSINES, G. RIBOULET et T. RONCALLI [2001], Credit risk modelling with copulas, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [9] CRUZ, M. [2001], Extreme value theory : a useful framework for modeling extreme OR events, Risk Methodology Development and Quantitative Analysis, *Slides*
- [10] EBNÖTHER, S., P. VANINI, A. MCNEIL et P. ANTOLINEZ-FEHR [2001], Modelling operational risk, article soumis à *Journal of Risk*
- [11] FRACHOT, A., P. GEORGES et T. RONCALLI [2001], Loss Distribution Approach for operational risk, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [12] GEORGES, P., A-G. LAMY, E. NICOLAS, G. QUIBEL et T. RONCALLI [2001], Multivariate survival modelling : a unified approach with copulas, Groupe de Recherche Opérationnelle, Crédit Lyonnais, *Working Paper*
- [13] GIESECKE, K. [2001a], Structural modeling of correlated defaults with incomplete information, Humboldt-Universität zu Berlin, *Working Paper*
- [14] GIESECKE, K. [2001b], Correlated default with incomplete information, Humboldt-Universität zu Berlin, *Working Paper*

- [15] HAMILTON, D., J. JAMES et N. WEBBER [2001], Copula methods and the analysis of credit risk, University of Warwick, *Working Paper*
- [16] JOUANIN, J-F., G. RAPUCH, G. RIBOULET et T. RONCALLI [2001], Modelling dependence for credit derivatives with copulas, Groupe de Recherche Opérationnelle, Crédit Lyonnais, Working Paper
- [17] JOUANIN, J-F., G. RIBOULET et T. RONCALLI [2001], Beyond conditionally independent defaults, Groupe de Recherche Opérationnelle, Crédit Lyonnais, Working Paper
- [18] KËLLEZI, E. et M. GILLI [2000], Extreme value theory for tail-related risk measures, Working Paper présenté à la conférence Computational Finance 2000, 31 mai – 2 juin 2000, London Business School
- [19] LI, D.X. [2000a], On default correlation : a copula function approach, *Journal of Fixed Income*, **9(4)**, 43-54
- [20] LI, D.X. [2000b], The valuation of basket credit derivatives : a copula function approach, *Workshop on Options in Financial Products : Approaches to Valuation*, 8 December, Toronto
- [21] MASHAL, R. et M. NALDI [2002], Pricing multiname credit derivatives : heavy tailed hybrid approach, Graduate School of Business, Colombia University, *Working Paper*
- [22] MEDOVA, E.A. et M.N. KYRIACOU [2001], Extremes in operational risk management, dans
- [23] RONCALLI, T [2000], La Théorie des Extrêmes et la Gestion des Risques de Marché, *notes de cours ENSAI 3^{ème} année, document disponible sur le site internet du GRO*
- [24] RONCALLI, T [2001], Introduction à la Gestion des Risques, *notes de cours ENSAI 3^{ème} année, document disponible sur le site internet du GRO*
- [25] RONCALLI, T [2002], Gestion des Risques Multiples (Application des Copules à la Gestion des Risques), *notes de cours ENSAI 3^{ème} année, document bientôt disponible sur le site internet du GRO*
- [26] SCHMIDT, W. et I. WARD [2002], Pricing default baskets, *Risk*, 111-114, January
- [27] SCHÖNBUCHER, P.J. et D. SCHUBERT [2001], Copula-dependent default risk in intensity models, Department of Statistics, Bonn University, *Working Paper* (available from <http://www.finasto.uni-bonn.de/~schonbuc/>)

5

Conclusion

La gestion du risque prend une part importante (et de plus en plus prépondérante) dans la gestion d'un établissement financier. Un certain nombre de risques peuvent être qualifiés de "risques extrêmes", car ils présentent une probabilité d'occurrence très faible (ou un temps de retour très élevé). Ils correspondent à des événements rares. A première vue, nous pouvons penser que la théorie des extrêmes va permettre de modéliser ces types d'événements. En fait, il faut être très prudent avec cette théorie. Dans de nombreux cas, l'estimation est fortement instable (même si certains auteurs prétendent le contraire). La théorie des extrêmes n'est donc pas une méthode miracle pour mesurer le risque de ces événements rares.

L'intérêt de la théorie des extrêmes est beaucoup plus en amont. Elle permet de comprendre quelles sont les implications du choix d'une modélisation sur une mesure de risques extrêmes. Elle indique quels sont les facteurs susceptibles de faire varier cette mesure de façon significative. Les deux concepts importants sont alors la notion de queue de distribution et la notion de dépendance de queue. En échantillon fini, la théorie des valeurs extrêmes correspond aux statistiques du maximum et du minimum. Ces statistiques d'ordre peuvent être utilisées dans la gestion des risques pour estimer certains paramètres, construire des backtesting, etc.