Modélisation de la dépendance et estimation du risque agrégé

Andres Cuberos

UNIVERSITÉ CLAUDE BERNARD LYON 1
Institut de Science Financière et d'Assurances

Ecole Doctorale SEG
Sciences Economiques et de Gestion

T H È S E

pour obtenir le titre de

Docteur de l'Université Claude Bernard Lyon 1

Spécialité : Mathématiques Appliquées

Présentée par

Andrés CUBEROS

Modélisation de la dépendance et estimation du risque agrégé

soutenue le 18 décembre 2015

Jury :

Rapporteurs :	Stéphane Girard (DR INRIA, Université de Grenoble)
	Marie Kratz (Prof. ESSEC)
Examinateur :	Jean-Noël Bacro (Prof. Université de Montpellier)
Directeur :	Véronique Maume-Deschamps (Prof. Université Lyon 1)
Co-Directeur :	Esterina Masiello (MCF. Université Lyon 1)

Remerciements

Je souhaite remercier vivement Marie Kratz et Stéphane Girard d'avoir accepté d'être rapporteurs de cette thèse. Merci également à Jean-Noël Bacro pour avoir accepté de faire partie du jury. Je suis très honoré.

Je remercie chaleureusement mes deux directrices de thèse, Véronique Maume-Deschamps et Esterina Masiello pour tout le temps que vous avez consacré à ce travail. Après chaque réunion, j'ai eu toujours droit, même si parfois je ne le méritais pas, à des paroles aimables et encourageantes. Cette thèse n'aurait pu aboutir sans votre aide, soutien et gentillesse.

J'adresse un grand merci à toute l'équipe de Modélisation Actuarielle de la SCOR, Kati, Mathieu, Mihaela, Théo, Laila, Divyani, Philippe, Julie et Alexandra. J'apprécie beaucoup travailler à vos côtés. Si parfois je semblais absent ou tendu, la cause se trouve peut-être cachée quelque part plus loin dans ces pages.

Je tiens à remercier tous les membres de l'ISFA et du laboratoire SAF, à son directeur Christian Robert, à tous les chercheurs et enseignants et au personnel administratif. Une pensée spéciale aux doctorants, et anciens doctorants que j'ai croisé ces dernières années. C'est toujours bien de savoir qu'on n'est pas seul dans cette quête existentielle.

Por último quiero agradecer a mis padres, por el amor incondicional y apoyo infinito que siempre me dan. A Juanquita y la Gordis por ser las mejores personas que conozco. A Turrita, Yeyin, Mauro, Alethiviris, Yeiros, Mabetín, Mats, JeanPis y Liza porque a pesar de la distancia siempre serán los primores, amigos del alma. A Marthica, Roycito, Mimi, Milote, Mauro, Germancito, Ticor, Blan y Mary, Enrique, Michito, Juanita, Janeth, Ruth, Angélica, Emita, gracias por tantos momentos felices. Y a la Peke, gracias por todo.

Table des matières

Chapitre 1

Introduction

Ce travail à été réalisé dans le cadré d'une thèse CIFRE, issue d'une convention de recherche entre la société de réassurance SCOR et l'Université Claude Bernard Lyon 1. Elle porte sur la modélisation et l'estimation de la dépendance des portefeuilles de risques et du risque agrégé.

Plus généralement, pour un portefeuille de risques $\mathbf{X} = (X_1, \ldots, X_d)$, où $X_1, \ldots, X_d$, sont des variables aléatoires représentant des risques individuels (pertes ou bénéfices dans une période spécifique), on s'intéressera dans ce travail à l'estimation de la dépendance qui peut exister entre les différents risques de $\mathbf{X}$ ainsi qu'à l'estimation de la position globale $\Psi(\mathbf{X})$ associée à celui-ci, où $\Psi : \mathbb{R}^d \to \mathbb{R}$ est une fonction, qu'on appellera fonction d'agrégation.

Dans le cas d'une société d'assurance, chaque composante du vecteur de risques $\mathbf{X}$ peut représenter par exemple les réclamations d'une branche particulière dans l'année et $\Psi(\mathbf{X})$ le montant de réclamation total. La différence principale entre le cas multivarié, lorsque le portefeuille dispose de plusieurs risques, et univarié, lorsque le portefeuille est composé d'un seul risque, est la dépendance qui peut être présente entre les différentes composantes du vecteur aléatoire. Par exemple, pour certaines branches d'assurance des pertes peuvent avoir une haute probabilité de se produire simultanément. Une bonne compréhension de la dépendance dans les événements extrêmes peut donc s'avérer importante, d'autant plus que, dans certains cas, la dépendance dans les événements extrêmes est plus forte que dans les événements modérés.

La modélisation et l'estimation de la dépendance d'un vecteur de risques et son agrégation sont demandées aux assureurs et réassureurs dans le cadre de la règlementation européenne comme expliqué par la suite.

Contexte actuel de Solvabilité 2

Depuis quelques années le marché de l'assurance européen a été modifié à la suite de la crise financière 2007-2009. En raison de l'environnement

d'instabilité financière, le régulateur européen a approuvé un nouveau cadre législatif destiné à assurer la stabilité financière et la solvabilité des compagnies d'assurance à travers le contrôle des risques auxquels ils sont exposés.

La publication au Journal Officiel de L'Union européenne de la directive du Parlement européen et du Conseil du 25 novembre 2009 sur l'accès aux activités d'assurance et de réassurance et leur exercice, également connue sous Solvabilité II, 2009/138 / CE, marque le point de départ officiel dans la mise en œuvre de mesures législatives visant à maîtriser les risques dans des sociétés d'assurance. Solvabilité II entraine un changement législatif sur l'approche que les compagnies d'assurance doivent tenir en fonction des risques auxquels elles sont confrontées en raison de leur activité.

La directive Solvabilité II est structurée sur le principe des trois piliers. Ces piliers définissent les critères et les normes de caractère quantitatifs et qualitatifs que les établissements devraient entreprendre pour assurer leur solvabilité et la stabilité financière.

Le Pilier I contient un ensemble de règles qui déterminent les critères pour l'obtention d'exigences de fonds propres qu'une entité doit maintenir à un horizon de temps annuel, en rapport au risque assumé par l'entité, pour assurer un niveau acceptable de solvabilité par l'évaluation économique du bilan d'une entité. Le Pilier I vise à déterminer les exigences financières minimales pour assurer que les actifs disponibles pour les entités sont en quantité suffisante et de qualité, pour répondre aux engagements en vertu d'un certain horizon de temps.

Le Pilier II décrit les exigences qualitatives que les entités doivent respecter, ce qui se traduit par des procédures de communication avec le régulateur.

Enfin, dans le Pilier III sont présentées les mesures qui visent à assurer la transparence et la discipline du marché de l'assurance à travers un ensemble de règles de communication sur l'information de la situation financière et de solvabilité des entités face aux régulateurs locaux ainsi que des normes de communication et de transparence des régulateurs locaux face au régulateur européen.

Calcul du capital de solvabilité

L'objectif des exigences économiques du Pilier I est de couvrir toute perte inattendue qu'une entité pourrait subir en raison de fluctuations défavorables et inattendues dans la sinistralité. Cette partie des exigences de fonds propres est désignée dans la Directive comme capital de solvabilité requis, ou SCR, de l'anglais *Solvency Capital Requirement*.

La Directive stipule que le SCR doit être obtenu en utilisant un modèle qui reflète le profil de risque de l'entité, et qui est adapté en fonction de la nature, de l'ampleur et de la complexité des risques assumés par elle. Le modèle proposé par le régulateur est appelé Formule Standard.

Selon l'article 110 de la Directive, la Formule Standard peut être utilisée par les entités, en utilisant les paramètres définis dans la Directive, ou peut être adaptée au profil de risque de chaque compagnie en estimant de nouveaux paramètres spécifiques basés sur l'expérience historique de l'entité.

Diverses raisons peuvent exister pour qu'un assureur décide de ré-estimer les paramètres fixés par le régulateur. La principale raison est que les paramètres de la Directive peuvent surestimer le véritable profil de risque de l'entité, ce qui conduirait à un SCR supérieur à celui obtenu avec l'utilisation de ses propres paramètres. Par ailleurs, une autre raison peut être le fait que la structure d'entreprise d'une compagnie d'assurance n'est pas compatible avec celle proposée par le régulateur.

La Directive prévoit également que, sous certaines conditions et sous l'autorisation de l'autorité compétente, le SCR peut être obtenu en utilisant un modèle interne. Ce modèle peut considérer tout ou partie des risques auxquels l'entreprise est confrontée. Dans ce dernier cas, le modèle est appelé modèle interne partiel.

Les conditions à remplir par un modèle interne sont liées à la mise en œuvre et au suivi des normes, de sorte que les entités qui choisissent d'utiliser un modèle interne, total ou partiel, pour le calcul des exigences de fonds propres de solvabilité devraient justifier son utilisation et fonctionnement.

La formule standard est un ensemble de formules et méthodologies proposées par le régulateur qui permettent à l'assureur d'obtenir le montant du SCR. Un modèle interne est une procédure proposée par l'assureur qui sert les même buts et objectifs que la formule standard, mais en utilisant un

modèle qui reflète plus fidèlement le profil de risque de la compagnie. Le SCR obtenu avec l'un des modèles autorisés par le régulateur, doit correspondre à la valeur à risque (VaR) du capital de l'entité à un horizon de temps annuel, calculée avec un niveau de 99,5% de confiance.

La Directive propose, pour la formule standard, le niveau de désagrégation des risques suivant : le risque de souscription, le risque de marché, le risque de crédit et le risque opérationnel. Pour chacun d'entre eux, sont pris en compte divers sous-niveaux en essayant de refléter de manière adéquate et générique le profil de risque d'un assureur représentatif.

Le SCR s'obtient en agrégeant les différentes exigences de fonds propres pour chaque module de risque considéré, en deux étapes. D'une part, les capitaux correspondant aux différents sous-modules examinés doivent être agrégés pour chaque risque du module principal, compte tenu des corrélations linéaires. Cela permet de calculer l'exigence de capital correspondant à chaque risque majeur : de souscription, de marché, de crédit et opérationnel. Puis, l'agrégation des exigences de capital pour chaque capital de risque module principal doit être effectuée en tenant compte des correlations linéaires entre elles, et donc finalement l'exigence de capital de solvabilité total, le SCR, est obtenue. Dans la formule standard, la relation entre les différents modules et sous-modules pour agréger les différentes exigences de capital est représentée par les matrices de corrélation linéaire entre les variables représentant les risques des différents modules et sous-modules.

A la différence de la formule standard, le modèle interne donne la liberté à l'assureur de choisir comment modéliser ses risques individuels et d'étudier la structure de dépendance sous-jacente. Le modèle doit être approuvé par le régulateur, et l'assureur doit être capable de démontrer que la structure de dépendance du modèle et la modélisation de risques individuels reflète bien le profil de risque de la compagnie. Si la compagnie opte pour utiliser un modèle interne, et si celui-ci est approuvé par les régulateurs, le SCR s'obtient en calculant la valeur à risque au niveau de $99,5\%$ des risques agrégés du modèle interne.

Modélisation de la dépendance et calcul du risque agrégé

La modélisation de la dépendance d'un portefeuille de risques est déterminant pour estimer le capital de solvabilité d'une compagnie d'assurance ou

plus généralement pour estimer le risque agrégé. Deux concepts importants pour modéliser la dépendance d'un vecteur aléatoire sont les copules et la mesure spectrale. Dans la suite, on présente brièvement ces concepts, avant de proposer quelques estimateurs dans les Chapitres 3 et 4.

Copules

Pour une introduction à la théorie des copules, les textes de [Joe 1997], [Nelsen 1999], ou encore [Frees & Valdez 1998] sont des références classiques. L'idée d'une fonction qui caractérise la structure de dépendance entre variables aléatoires vient de différents travaux de Hoeffding dans les années 1940, [Hoeffding 1940], mais ce fut Sklar en 1959 qui définit et donna le nom de copule à une fonction qui produit une distribution multivariée à partir des distributions marginales univariées [Sklar 1959]. Cependant, sa mise en œuvre dans le cadre des applications à la finance et l'assurance est récente : un premier article fondateur est [Genest & MacKay 1986] et ensuite des travaux de Embrechets et McNeil qui, par exemple, démontrent les limitations du coefficient de corrélation linéaire comme mesure de la dépendance [Embrechts *et al.* 1999]. Par ailleurs [McNeil *et al.* 2005] est un bon résumé du point de vue théorique des applications des copules dans le domaine financier et actuariel.

Les copules permettent de différencier le comportement des distributions marginales et de la dépendance pour une distribution multivariée. Il résulte que c'est un concept très utile pour la modélisation, mais également dans l'estimation ou la simulation. Il nous permet essentiellement, par un moyen relativement simple, d'étendre les modèles traditionnels, qui étaient fondées sur l'hypothèse de normalité conjointe, à des situations plus générales.

Définition. Une *copule* est une fonction $C : [0,1]^d \to [0,1]$ qui satisfait les trois propriétés suivantes :

(i) C est croissante en chaque composant.

(ii) $C(1,\dots,1,u_i,1,\dots,1) = u_i$ pour tout $u_i \in [0,1]$, $i = 1,\dots,d$.

(iii) C est d-croissante : pour tout $a = (a_1,\dots,a_d) \in [0,1]^d$ et tout $b = (b_1,\dots,b_d) \in [0,1]^d$ tel que $a_i \leq b_i, i = 1,\dots,d$, on a

$$\sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} (-1)^{j_1+\cdots+j_d} C(u_{1j_1},\dots,u_{dj_d}) \geq 0,$$

avec $u_{i1} = a_i, u_{i2} = b_i$ pour tout $i = 1,\dots,d$.

Ces trois propriétés caractérisent les fonctions de répartition d-dimensionnelles de marginales uniformes. Comme les copules sont des

distributions jointes, elles induisent des mesures de probabilités. Par la théorie de la mesure, il existe une mesure μ_C sur la tribu borélienne de $[0,1]^d$ telle que $\mu_C(\prod_{i=1}^d [0,a_i[) = C(a_1,\ldots,a_d)$.

Le point de départ pour les applications en finance et assurance de la copule est son interprétation probabiliste, c'est à dire le lien entre les copules et les distributions de variables aléatoires multivariées. Ce lien est essentiellement établi dans le Théorème de Sklar qui nous dit que non seulement les copules permettent de construire des lois jointes avec des marginales fixées à l'avance mais aussi que pour toute distribution multivariée F on peut toujours trouver une copule C qui permet de lier F à ses marginales.

Théorème (**Théorème de Sklar**). *Soit F une fonction de répartition sur $\mathbb{R}^d$ de marginales $F_1,\ldots,F_d$. Alors, il existe une copule C tel que*

$$C(F_1(x_1),\ldots,F_d(x_d)) = F(x_1,\ldots,x_d), \tag{1.1}$$

pour tout $\mathbf{x} \in \mathbb{R}^d$. Si de plus les marginales sont continues alors C est unique. Réciproquement, étant donnés une copule C et un ensemble de fonctions de répartition univariés $F_1,\ldots,F_d$ alors la fonction $F : \mathbb{R}^d \to [0,1]$ définie par

$$F(x_1,\ldots,x_d) := C(F_1(x_1),\ldots,F_d(x_d)),$$

est la fonction de répartition d'une loi en dimension d avec marginales $F_1,\ldots,F_d$.

La démonstration de ce résultat se trouve dans [Nelsen 1999].

Définition. Soit $\mathbf{X}$ un vecteur aléatoire de fonction de répartition F dont les marginales sont continues. On dira que la copule C qui satisfait (1.1) est *la copule de* $\mathbf{X}$ ou aussi que C *est la copule de* F.

Par exemple, d'après ce résultat, lorsque dans le cas bivarié on a la représentation

$$F(x,y) = C(F_1(x), F_2(y))$$

la loi jointe est partagée en deux effets, celui des marginales et la copule, de manière que cette dernière représente seulement l'association entre les deux marginales. Les copules séparent donc le comportement des marginales de la loi jointe. Pour cette raison, les copules sont connues aussi comme fonctions de dépendance (voir [Deheuvels 1979]). Cette particularité permet donc aussi d'établir que tout problème de modélisation multivarié peut être décomposé en deux étapes : la première sur les distributions marginales et la seconde sur celle de la copule.

Quelques exemples de copules utilisées dans ce travail sont présentés en Annexe A. Dans le Chapitre 3 une classe de copules dite invariantes par agrégation sera introduite. Aussi, une copule que nous appellerons copule échiquier empirique sera proposée pour estimer la dépendance.

Mesure Spectrale

La mesure spectrale a été introduite, dans le cadre de la variation régulière, par [de Haan & Resnick 1977]. Elle permet de caractériser la dépendance dans les extrêmes d'un vecteur de risques dit à variation régulière. Plusieurs définitions équivalentes de la mesure spectrale sont disponibles dans la littérature, voir par exemple [Beirlant *et al.* 2006], [De Haan & Ferreira 2007] et [Resnick 2007]. On présente brièvement ici la définition de variation régulière multivariée et de mesure spectrale qu'on utilise dans le Chapitre 4 où l'on propose deux estimateurs, l'un pour le cas absolument continu et l'autre pour le cas discret. Auparavant, on présente la notion de variation régulière dans le cas univarié.

Définition. Soit f une fonction mesurable et positive, f est dite à variation régulière à l'infini avec indice $\rho \in \mathbb{R}$ si

$$\lim_{t\to\infty} \frac{f(xt)}{f(t)} = x^\rho, \text{ pour tout } x > 0.$$

Si $\rho = 0$, f est dite à variation lente.

Remarque. De manière équivalente, une fonction f est à variation régulière avec indice ρ, si et seulement si

$$f(x) = x^\rho L(x)$$

où L est une fonction à variation lente.

Si X est une variable aléatoire réelle avec fonction de répartition F, on dit que X est à variation régulière si $\bar{F}(x) := 1 - F(x)$ est à variation régulière à l'infini. Cette notion est intéressante dans les applications car la vitesse de convergence de $\bar{F}(x)$ vers 0 nous permet de mesurer la probabilité qu'il survienne un événement extrême.

La notion de variation régulière peut se généraliser au cadre multivarié de la manière suivante. Soit $\|\cdot\|$ une norme sur $\mathbb{R}^d$ et $\mathbf{X}$ un vecteur aléatoire en $\mathbb{R}^d$ que l'on suppose presque sûrement différent du vecteur 0. Les coordonnées polaires du vecteur sont décrites par

$$(R, S) := \left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|} \right) \in \mathbb{R}_+ \times \mathbb{S}_{d-1}$$

où $\mathbb{S}_{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$ est la sphère de rayon unitaire pour la norme $\|\cdot\|$.

Définition. Soit $\|\cdot\|$ une norme sur $\mathbb{R}^d$. Un vecteur aléatoire $\mathbf{X}$ avec coordonnées polaires (R, S) est dit à variation régulière d'indice α s'il existe une mesure σ sur $\mathbb{S}_{d-1}$ (qui dépend de la norme $\|\cdot\|$) telle que

$$\Pr((r^{-1}R, S) \in B | R > r) \to (v_\alpha \times \sigma)[B], \quad \text{quand } r \to \infty,$$

pour tout $B \in \mathcal{B}((1,\infty) \times \mathbb{S}_1)$ avec $(v_\alpha \times \sigma)[\partial B] = 0$, où v_α est une mesure dans $(1,\infty)$ telle que $v_\alpha[(x,\infty)] = x^{-\alpha}$ pour tout $x \geq 1$. La mesure σ est connue sous le nom de *mesure spectrale* ou *mesure angulaire* de $\mathbf{X}$.

La mesure spectrale nous donne des informations sur les directions dans lesquelles se trouvent les événement extrêmes d'un vecteur aléatoire $\mathbf{X}$ et l'indice α donne la vitesse à laquelle le rayon décroît en distribution. Dans le Chapitre 4 un estimateur de la mesure spectrale pour un vecteur à variation régulière de dimension 2 sera proposé.

Calcul du risque agrégé

D'autre part, dans ce travail on s'intéresse à l'estimation du risque agrégé. Plus précisément on s'intéresse à l'estimation de la distribution de $\Psi(\mathbf{X})$. Même quand le modèle du vecteur de risques est supposé connu, le calcul de la distribution agrégée n'est pas trivial. Si on dispose d'un portefeuille composé de risques $X_1, X_2, \ldots, X_d$, dont on connaît la fonction de répartition jointe F ou, de manière équivalente, un portefeuille dont on connaît les fonctions de répartition marginales F_i, $i = 1, \ldots, d$, et la copule C, alors la fonction de répartition de $\Psi(\mathbf{X})$, qu'on notera par F_Ψ, est donnée par

$$F_\Psi(s) = P(\Psi(\mathbf{X}) \leq s) = \int_{\mathcal{S}} dF(x_1, \ldots, x_d). \tag{1.2}$$

où $\mathcal{S}$ est l'ensemble $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^d \ : \ \Psi(\mathbf{x}) \leq s\}$.

On peut rarement trouver une expression explicite pour la fonction de répartition ou pour la densité du risque agrégé. Des expressions explicites existent quand Ψ est une fonction d'agrégation simple comme le maximum, le minimum, ou la projection sur une coordonnée. Par contre, si Ψ est la somme, dans peu d'exemples on connait une expression explicite pour le risque agrégé. Deux exemples connus sont le vecteur de risques Gaussien et le vecteur de risques marginaux exponentiels et indépendants. Un autre exemple où l'on connait la distribution de la somme est le cas du modèle Pareto-Clayton qu'on

présentera dans la Section 2.5. Néanmoins, en général, il faut recourir à des approximations numériques ou à des simulations. Différentes méthodes numériques peuvent s'appliquer pour approcher l'intégrale (1.2), voir par exemple [Arbenz *et al.* 2011]. Cependant lorsque le nombre de risques d'un portefeuille est élevé, il est en général nécessaire de passer par des méthodes de Monte Carlo. Toutefois, l'estimation par la méthode de Monte Carlo peut s'avérer limitée si, par exemple, on doit calculer la valeur d'un quantile élevé, comme la VaR à 99,5%, car dans ce cas le nombre de simulations nécessaires pour une bonne estimation peut être très élevé. Dans le Chapitre 2 on propose une méthode d'approximation des quantiles du risque agrégé pour des niveaux proches de 1. La méthode peut être utilisée en grande dimension.

1.1 Résumé des résultats présentés dans le Chapitre 2

Dans le Chapitre 2, issu de l'article : Cuberos, A., Masiello, E., & Maume-Deschamps, V. (2015). High level quantile approximations of sums of risks. *Dependence Modeling*, *3*(1), on s'intéressera au calcul de la Value-at-Risk (VaR) et la Tail Value-at-Risk (TVaR) de la somme $X_1 + \cdots + X_d$ définies respectivement par

$$\mathrm{VaR}_p(S) = F_S^{\leftarrow}(p) \quad \text{et} \quad \mathrm{TVaR}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_u(S)\,\mathrm{d}u,$$

pour des niveaux de confiance $p \in]0,1[$ proches de 1. F_S est la fonction de répartition de S et $F_S^{\leftarrow}$ son inverse généralisée.

La théorie des valeurs extrêmes permet d'obtenir des estimations de la VaR pour des niveaux de confiance proches de 1 ([Embrechts *et al.* 1997, Weissman 1978]). Cependant les méthodes basées sur cette théorie nécessitent l'estimation des paramètres qui parfois ne sont pas faciles à estimer. Aussi, il est possible d'appliquer des méthodes de Monte Carlo pour estimer la VaR mais, lorsque p est proche de 1, le nombre de simulations nécessaires pour donner des estimations précises peut être très élevé, et de nouvelles méthodes sont toujours les bienvenues.

Soit M le risque maximum dans le portefeuille de la compagnie d'assurance, $M = \max\{X_1, \ldots, X_d\}$. La fonction de répartition de M, notée F_M, se calcule par

$$F_M(x) = F(x, \ldots, x).$$

Ainsi F_M est directement déterminée par la fonction de répartition F du portefeuille, de sorte que l'intégration numérique ou des méthodes de Monte Carlo ne sont pas nécessaires pour son calcul. Cela signifie également que $\mathrm{VaR}_p(M)$ peut être facilement calculée pour un niveau de confiance p donné, car au pire une simple inversion numérique est nécessaire.

Estimation de la VaR de la somme en utilisant le maximum

La Proposition 2.2.3, nous donne quelques conditions sur $\mathbf{X}$ pour que la VaR et la TVaR de la somme et du maximum soient asymptotiquement équivalentes dans le sens où il existe un $\Delta \geq 1$ tel que

$$\mathrm{Var}_{1-p}(S) \sim \mathrm{VaR}_{1-\Delta^{-1}p}(M) \quad \text{et} \quad \mathrm{TVaR}_{1-p}(S) \sim \mathrm{TVaR}_{1-\Delta^{-1}p}(M),$$

lorsque $p \to 0$.[1] Ce résultat est intéressant car il permet d'estimer la VaR (ou la TVaR) de la somme en utilisant la VaR (ou la TVaR) de la valeur maximale, qui est plus facile à calculer à condition de disposer d'un estimateur de Δ. Dans la Section 2.4 on propose un méthode pour estimer Δ.

En particulier on démontrera que notre méthode est utilisable pour des vecteurs à variation régulière multivariés (voir Section 2.3.2). Dans ce cadre, des résultats similaires mais en utilisant X_1 au lieu de $\max(\mathbf{X})$ ont été démontrés dans [Barbe *et al.* 2006] et [Alink *et al.* 2007]. Dans la section 2.6.4 nous montrons empiriquement que notre méthode donne de meilleurs résultats pour l'estimation de la VaR : en effet la convergence de $\overline{F}_S(t)/\overline{F}_M(t)$ semble en général bien plus rapide que celle de $\overline{F}_S(t)/\overline{F}_1(t)$.

Dans la Proposition 2.3.5 on démontre qu'un vecteur $\mathbf{X}$ dont les marginales sont à variations régulières, mais pas forcement identiques, et dont la dépendance est décrite par une copule archimédienne (ou de survie archimédienne) vérifie aussi les conditions de la Proposition 2.2.3. Notre méthode s'applique donc encore une fois.

On se concentrera sur le cas où les marginales sont Pareto et la copule est une survie de Clayton car, dans ce cas spécifique, la VaR de la somme peut se calculer explicitement. Ceci nous permettra d'effectuer une étude numérique à la fin du Chapitre 2.

1. $a(t) \sim b(t)$ lorsque $t \to l$, pour $l \in [-\infty, \infty]$ signifie que $\lim_{t \to l} \frac{a(t)}{b(t)} = 1$.

Nous soulignons que notre méthode donne de bons résultats lorsque la dépendance entre les risques est importante et qu'au moins une des distributions marginales a une queue lourde. Les résultats sont notamment meilleurs que lorsqu'un seul risque, par exemple X_1 est utilisé au lieu du maximum du vecteur. Nous allons comparer notre méthode aux méthodes de la théorie des valeurs extrêmes et aussi à la méthode de Monte Carlo, en particulier pour les quantiles de très haut niveau de confiance et pour des dimensions de portefeuille grande et très grande (10 et 150).

Finalement, on présente quelques résultats du chapitre dans le contexte des fonctions à variation consistante. La classe des fonctions à variation consistante constitue une classe plus générale que les fonctions à variation régulière (voir Section 2.8).

1.2 Résumé des résultats présentés dans le Chapitre 3

Dans le Chapitre 3, issu d'un article co-écrit avec Esterina Masiello et Véronique Maume-Deschamps, soumis pour publication et dont une version courte a été publiée dans les actes de l'Actuarial and Financial Mathematical Conference 2015, on s'intéressera à l'estimation de la distribution d'une agrégation de risques lorsque les distributions des risques marginaux sont connues et lorsque une certaine information sur la dépendance entre les risques est disponible. Habituellement, cette information est disponible via quelques observations de la distribution conjointe ou également par un avis d'experts.

Considérons un vecteur de risques $\mathbf{X} = (X_1, \ldots, X_d)$ et son agrégation $\Psi(\mathbf{X})$. Dans ce chapitre, nous nous concentrons essentiellement sur $\Psi = \sum$ et nous nous intéressons à l'estimation de la VaR de $\Psi(\mathbf{X})$. Pour ce faire, nous supposerons que les distributions $F_1, \ldots, F_d$ des risques marginaux $X_1, \ldots, X_d$ sont connues et que de l'information sur la dépendance entre elles est donnée. On supposera que cette information sur la dépendance provient de l'une des trois possibilités suivantes : échantillon de la copule, connaissance de la dépendance dans la queue de la distribution ou connaissance de la dépendance d'un sous-vecteur.

En général, ni les risques marginaux, ni la dépendance du vecteur de risque $\mathbf{X}$ ne sont connus. Cependant, dans de nombreux cas, la connaissance des distributions marginales est beaucoup plus importante que celle de

la dépendance. Par exemple, lorsque certaines observations du vecteur $\mathbf{X}$ sont disponibles, l'inférence statistique qu'on peut faire sur les distributions marginales est de meilleure qualité que l'inférence effectuée sur la distribution multivariée. En outre, les échantillons disponibles pour les lois marginales peuvent être beaucoup plus grands que ceux disponibles pour la distribution conjointe. De plus, sur chaque risque marginal, il est plus habituel de trouver des informations supplémentaires, comme par exemple des avis d'experts ou des informations préalables. Donc, même si l'hypothèse de la connaissance de la distribution marginale peut sembler peu réaliste, il y a, cependant, dans la pratique beaucoup plus d'information sur les distributions marginales que sur la structure de dépendance du vecteur de risque.

Dans le contexte de l'agrégation des risques avec incertitude sur la structure de dépendance, la classe suivante a été introduite dans [Bernard *et al.* 2014]. Un risque agrégé S est dit admissible pour les distributions marginales $F_1, \dots, F_d$ si S peut s'exprimer comme $S = \Psi(X_1, \dots, X_d)$ où $X_i \sim F_i$ pour $i = 1, \dots, d$. La classe de risques admissibles avec distributions marginales est défini comme suit :

$$\mathfrak{S}_d(F_1, \dots, F_d, \Psi) = \{\Psi(X_1, \dots, X_d) : X_i \sim F_i, i = 1, \dots, d\}.$$

Certaines propriétés intéressantes de cette classe ont été présentées dans [Bernard *et al.* 2014] quand Ψ est la somme. Une classe associée à celle-ci, mais du point de vue des copules est définie dans la suite.
La classe de copules $\mathcal{C}(\mathbf{X}, \Psi)$ qu'on appellera invariantes par agrégation pour $\mathbf{X}$ et Ψ, est définie comme suit

$$\mathcal{C}(\mathbf{X}, \Psi) = \{C \in \mathcal{C} : \Psi(\mathbf{X}^C) \overset{\mathcal{L}}{=} \Psi(\mathbf{X})\}$$

où $\mathbf{X}^C$ est un vecteur avec les mêmes risques marginaux que X mais avec copule C. La classe de copules invariantes par agrégation est liée à la classe de risques admissibles de la même manière que les copules sont liées à la classe de Fréchet (voir l'Annexe A)

$$F \in \mathfrak{F}_d(F_1, \dots, F_d) \Leftrightarrow \exists C : F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_n(x_d))$$
$$S \in \mathfrak{S}_d(F_1, \dots, F_d, \Psi) \Leftrightarrow \exists \mathcal{C}(\mathbf{X}, \Psi) : \forall C \in \mathcal{C}(\mathbf{X}, \Psi) \quad S \overset{\mathcal{L}}{=} \Psi(\mathbf{X}^C),$$

Dans la Section 3.2, nous présentons quelques exemples et des résultats qui montrent explicitement que cette classe n'est pas triviale.

Par définition, toute copule dans la classe $\mathcal{C}(X, \Psi)$ caractérise $\Psi(\mathbf{X})$. On démontrera que dans certains cas la classe $\mathcal{C}(X, \Psi)$ contient toujours une copule

symétrique, i.e telle que $C(x_1, \ldots, x_d) = C(x_{\sigma(1)}, \ldots, x_{\sigma(d)})$ (voir Proposition 3.2.5).

Ces exemples démontrent que sous certaines conditions la pleine connaissance de la distribution de la copule est inutile lorsque l'on étudie une agrégation, i.e. la copule contient beaucoup plus d'information que nécessaire pour l'étude de la distribution d'un risque agrégé. L'information donnée par la copule de $\mathbf{X}$, pour l'étude de $\Psi(\mathbf{X})$ est la même que celle donnée par toute copule de la classe $\mathcal{C}(\mathbf{X}, \Psi)$. Cela peut être considéré comme une justification du fait que lorsque les lois marginales sont connues, il peut y avoir une certaine souplesse dans l'estimation de la copule afin d'estimer la distribution agrégée.

Dans ce qui suit, nous allons définir la version empirique des copules échiquier, en utilisant la copule empirique. Le but principal de cette copule n'est pas de décrire précisément la structure de dépendance qu'on suppose non connue, mais de l'utiliser spécifiquement pour estimer la distribution de $\Psi(\mathbf{X})$.

Copule échiquier empirique

La copule empirique, introduite dans [Deheuvels 1979], peut être utilisée pour estimer non paramétriquement une copule (voir la définition dans l'Annexe A).

Si un échantillon de $\mathbf{X}$ est disponible, nous allons utiliser la copule empirique $\widehat{C}_n$ associée à l'échantillon et sa mesure de probabilité $\widehat{\mu}$, pour définir les versions empiriques des copules échiquier (*checkerboard copulas* en anglais) définies par [Mikusinski & Taylor 2010]. Les copules échiquier ont la particularité d'être uniformes sur chaque élément d'une partition régulière $\mathcal{I}_m$ de $[0,1]^d$. Trois types de copules échiquier empirique seront introduites, la première sans information supplémentaire, la deuxième avec information dans un sous vecteur et la troisième avec information dans la queue. On démontrera que ces éléments définis sont bien des copules à condition que m, la taille de la longueur de la partition, divise n, la taille de l'échantillon (voir Proposition 3.3.8).

Dans la Section 3.4, nous proposons une approche non paramétrique, pour estimer la distribution de $\Psi(\mathbf{X})$ lorsque les lois marginales sont connues, et on la compare avec l'estimation empirique classique de la fonction quantile $F^{-1}_{\Psi(\mathbf{X})}$. Pour ce faire, nous utilisons la fonction de distribution de $\hat{X}_1 + \ldots + \hat{X}_d$ où $\hat{\mathbf{X}}$ est un vecteur avec les mêmes marginales que $\mathbf{X}$ mais avec la dépendance donnée par une des copules échiquiers empiriques présentées ci-dessus, pour estimer la $\mathrm{VaR}_p(S)$ pour $S = X_1 + \cdots + X_d$ à différents niveaux de confiance p, $0 < p < 1$.

Comme dans le Chapitre 2, nous examinons le modèle de Pareto-Clayton (présenté dans la Section 3.4.1) parce que, dans ce cas, la valeur exacte de $\mathrm{VaR}_p(S)$ est connue, de sorte que nous pouvons comparer nos résultats de simulation avec la valeur exacte.

Les résultats obtenus sont dans tous les cas meilleurs que ceux obtenus par une estimation empirique classique. Les estimations des quantiles élevés s'améliorent lorsque l'on introduit de l'information sur la queue de la copule ou sur un sous vecteur. Finalement la méthode proposée permet d'intégrer différents types d'information et ainsi d'améliorer l'estimation d'une agrégation lorsque cette information est disponible.

1.3 Résumé des résultats présentés dans le Chapitre 4

Dans le Chapitre 4, on s'intéressera à l'estimation de la mesure spectrale d'un vecteur bivarié et plus particulièrement à son estimation lorsque celle-ci est discrète. L'estimation de la mesure spectrale a été abordée dans plusieurs articles, voir par exemple [Coles & Tawn 1991], [Joe *et al.* 1992] pour des estimateurs paramétriques, et [Einmahl *et al.* 1997], [Einmahl *et al.* 2001], [Einmahl & Segers 2009] pour des estimateurs non-paramétriques. Cependant, à notre connaissance, ces résultats ne sont pas applicables si la mesure spectrale est discrète, seul l'estimateur proposé par [Einmahl & Segers 2009] peut s'utiliser dans les cas particulier d'indépendance ou comonotonicité mais pas dans le cas discret général.

La mesure spectrale d'un vecteur de risques est discrète lorsque, par exemple, ses marges subissent des chocs communs à queue lourde. Dans la Section 4.3, on montre que la probabilité de ruine multivariée pour un modèle des chocs communs de Poisson peut s'approximer en fonction d'une mesure spectrale discrète. Cet exemple montre l'intérêt de proposer un estimateur de la mesure spectrale applicable au cas discret.

On considère la représentation en coordonnées polaires d'un vecteur $\mathbf{X} = (X_1, X_2)$ différent de zéro donnée par

$$(R, \Theta) := \left(\|\mathbf{X}\|, \arctan\left(\frac{\mathbf{X}}{\|\mathbf{X}\|} \right) \right).$$

Soit $F^{(r)}$ la fonction de distribution de Θ conditionné par $(R > r)$, i.e. $F^{(r)}(\theta) = \Pr(\Theta < \theta | R > r)$. Alors par définition de la mesure spectrale

de $\mathbf{X}$, notée σ, on a que pour chaque $\theta \in [0, 2\pi)$, lorsque $r \to \infty$,

$$F^{(r)}(\theta) \to \sigma([0, \theta]).$$

On dénotera par $\Theta^{(r)}$ une variable aléatoire dont la fonction de distribution est $F^{(r)}$. Si (k_n) et (r_n) sont deux suites telle que

$$k_n \to \infty, \quad k_n/n \to 0, \text{ et } r_n \to \infty,$$

on supposera que pour chaque n, on dispose d'échantillons i.i.d. pour $\Theta^{(r_n)}$ de taille k_n

$$\{\Theta_1^{(r_n)}, \ldots, \Theta_{k_n}^{(r_n)}\} \qquad n = 1, 2, \ldots.$$

Dans la pratique pour un échantillon de taille n du vecteur bivarié $\mathbf{X}$ on extrait les angles Θ qui accompagnent les vecteurs avec les k normes les plus élevées. On utilisera cet échantillon pour l'estimation.

Estimation de la mesure spectrale

Dans ce cadre on propose d'abord un estimateur de la densité de la mesure spectrale lorsque celle-ci existe. C'est à dire que cet estimateur n'est pas utilisable directement pour estimer une mesure spectrale discrète car celle-ci n'admet pas de densité. Puis dans un deuxième temps, on proposera une modification de cet estimateur afin d'estimer une mesure spectrale discrète.

Estimation de la densité d'une mesure spectrale

On suppose d'abord que la mesure spectrale σ admet une densité f_σ et que de plus :

(i) $F^{(r)}$ est différentiable pour chaque $r > 0$. On dénotera par f_r sa dérivée.

(ii) $f_r(\theta) \to f_\sigma(\theta)$ pour chaque $\theta \in I$, lorsque $r \to \infty$.

On propose une méthode d'estimation basée sur l'estimation des densités $f^{(r)}$ de $F^{(r)}$ par la méthode à noyau.

Soit $(k_n) \in \mathbb{N}$ et $(r_n) \in \mathbb{R}$ des suites telles que

$$k_n \to \infty, \quad k_n/n \to 0, \text{ et } r_n \to \infty.$$

Alors, si pour chaque fenêtre h est la même et $h_n = h(k_n)$ satisfait $h_n \to 0$ et $h_n k_n \to \infty$ lorsque $n \to \infty$ alors pour chaque $\theta \in I$ on définit

$$\hat{f}_n(\theta) = \frac{1}{h_n k_n} \sum_{j=1}^{k_n} K\left(\frac{\theta - \Theta_j^{(r_n)}}{h_n}\right),$$

avec K un noyau. Sous certaines conditions, par exemple satisfaites si la densité spectrale f_σ et sa dérivée f'_σ sont bornées on démontre que (voir Proposition 4.4.1) pour chaque $\theta \in I$

$$\hat{f}_n(\theta) \xrightarrow[n\to\infty]{\Pr} f_\sigma(\theta).$$

Estimation de la mesure spectrale discrète

On suppose maintenant que la mesure spectrale est discrète et qu'elle peut s'écrire comme

$$\sigma = \sum_{i=1}^{m} \alpha_i \delta_{\theta_i}$$

où m est le nombre (fini) de masses, α_i et θ_i le poids et la localisation de la masse i et δ la fonction de Dirac. Si $m = 1$ alors l'estimation de la mesure spectrale consiste à estimer simplement sa localisation qu'on notera θ_0.

On suppose que, bien que la mesure spectrale soit discrète, les densités $f^{(r)}$ existent pour chaque r, c'est à dire que le caractère discret de σ est seulement atteint à l'infini. Dans ce cas on propose comme estimateur de θ_0, un point $\hat{\theta}_n$ qui satisfait

$$\hat{f}_n(\hat{\theta}_n) = \sup_\theta \hat{f}_n(\theta).$$

Sous certaines conditions on montrera la consistance faible de cet estimateur (voir Proposition 4.4.2).

Si la mesure spectrale a plusieurs masses on propose des estimateurs pour les poids α_i dans la Proposition 4.4.3. Cependant cet estimateur est consistant seulement si la localisation des masses est connue.

Chapitre 2

High level quantile approximations of sums of risks

Sommaire

Abstract

The approximation of a high level quantile or of the expectation over a high quantile (Value at Risk (VaR) or Tail Value at Risk (TVaR) in risk management) is crucial for the insurance industry. We propose a new method

to estimate high level quantiles of sums of risks. It is based on the estimation of the ratio between the VaR (or TVaR) of the sum and the VaR (or TVaR) of the maximum of the risks. We show that using the distribution of the maximum to approximate the VaR is much better than using the marginal. Our method seems to work well in high dimension (100 and higher) and gives good results when approximating the VaR (or TVaR) in high levels on strongly dependent risks where at least one of the risks is heavy tailed.

2.1 Introduction

Because of regulatory rules (such as Solvency 2 in Europe) or for internal risk management purposes, the estimation of high level quantiles of a sum of risks is of major interest both in finance and insurance industry.
Consider an insurance company that has a portfolio of $d \geq 2$ (possibly) dependent risks which is represented as a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with cumulative distribution function (c.d.f.) $\mathbf{F}(x_1, \ldots, x_d)$. We assume that all the risks are almost surely positive but we do not assume that they are identically distributed. Let S denote the aggregated risk

$$S = X_1 + \cdots + X_d.$$

We are interested here in the computation of the Value-at-Risk (VaR) and the Tail Value-at-Risk (TVaR) of the sum,

$$\mathrm{VaR}_p(S) = F_S^{\leftarrow}(p) \quad \text{and} \quad \mathrm{TVaR}_p(S) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_u(S)\,\mathrm{d}u,$$

for confidence levels $p \in (0,1)$ near 1, where F_S is the c.d.f. of S and $F^{\leftarrow}$ is its generalized inverse. Problems like this arise for insurance companies, for example, which are required to maintain a minimum capital requirement which is typically calculated as the VaR for the distribution of the sum at some high level of probability. Even when the distribution function $\mathbf{F}$ is known, good estimations for $\mathrm{VaR}_p(S)$ are not trivial since they require a precise calculation of F_S, which is given by the following integral

$$F_S(x) = \int_{\{x_1+\cdots+x_d \leq x\}} \mathrm{d}\mathbf{F}(x_1, \ldots, x_d).$$

This integral is more difficult to approximate when d is large and it is usually more efficient to apply Monte Carlo methods to estimate it (for a comprehensive introduction to Monte Carlo methods see [Weinzierl 2000]). Nevertheless, when p is near 1, the number of replications required to give precise

estimations is also large, so new methods are always well received. Classical Extreme Value Theory (EVT) allows one to get some estimation of the VaR ([Embrechts *et al.* 1997, Weissman 1978]), but EVT based methods requires an estimation of the EVT parameters, which is known to be not an easy task. Recently, in [Bernard *et al.* 2013, Embrechts *et al.* 2013a, Bernard & Vanduffel 2015], some approximations on the VaR are obtained for some specific models; see also [Fougères & Mercadier 2012] where theoretical results on the asymptotic behavior, when $p \to 1$, of the ratio

$$\frac{\mathrm{VaR}_p(S)}{\sum_{i=1}^{d} \mathrm{VaR}_p(X_i)}$$

are given. Results for the tail distribution of the sum of dependent subexponential risks are obtained in [Geluk & Tang 2009] and also in [Kortschak & Albrecher 2009] when risks are non-identically distributed and not necessarily positive. In [Arbenz *et al.* 2011], an algorithm to compute the distribution function of S is proposed and in [Cossette *et al.* 2014], bounds are obtained. Nevertheless, these results may be used to estimate $\mathrm{VaR}_p(S)$ for small dimensions ($d < 4$) and give ranges in dimension 4 or 5.

We present a method which seems to be quite accurate even for a large number of summands, in the order of several hundreds for instance (see Sections 2.6.2 and 2.6.3 for simulations in dimension 10 and dimension 150). Our method will be compared to the EVT driven ones as well as to the Monte Carlo method, especially for very high level quantiles and in dimension greater than 4.

Let us denote by M the maximum risk in the portfolio of the company, $M = \max\{X_1, \dots, X_d\}$. The c.d.f. of M, denoted F_M, is given by

$$F_M(x) = \mathbf{F}(x, \dots, x).$$

F_M is directly determined by the c.d.f. $\mathbf{F}$ of the portfolio, so that numerical integration or Monte Carlo methods are not necessary. This also means that $\mathrm{VaR}_p(M)$ can be easily calculated for any given level of confidence p, at most a simple numerical inversion is needed.

In this chapter we give some conditions on $\mathbf{X}$ under which the Value-at-Risk and the Tail Value-at-Risk of the sum and maximum are asymptotically equivalent in the sense that there exists some $\Delta \geq 1$ such that

$$\mathrm{VaR}_{1-p}(S) \sim \mathrm{VaR}_{1-\Delta^{-1}p}(M) \quad \text{and} \quad \mathrm{TVaR}_{1-p}(S) \sim \mathrm{TVaR}_{1-\Delta^{-1}p}(M),$$

when $p \to 0$ and where $a(t) \sim b(t)$ when $t \to l$, for $l \in [-\infty, \infty]$ means throughout this chapter that $\lim_{t \to l} \frac{a(t)}{b(t)} = 1$. This result is interesting because it allows to estimate the VaR (or TVaR) of the sum by using the VaR (or TVaR) of the maximum, which is easier to calculate, and the estimation of Δ.

For random vectors with common marginals (Fréchet, Gumbel, Weibull) and an Archimedean copula dependence structure [Alink *et al.* 2005] and [Alink *et al.* 2004] get an asymptotic approximation of the tail of S. These results are generalized in [Alink *et al.* 2007] to other dependence structures. In [Barbe *et al.* 2006], the same results are obtained in the multivariate regularly varying framework. Examples in which the limiting constant Δ can be computed explicitly are also given in [Embrechts *et al.* 2009]. Finally, in an independent framework, we would like to mention [Nguyen & Robert 2014] which obtain asymptotic approximations for the tail of a sum of Pareto marginals and also [Kratz 2014] which introduce a method called *Normex* to approximate the distribution of a sum of heavy tail marginals.

In this chapter, we consider the more general framework with non common marginals and regularly varying tails. We emphasize that our method applies when there are dependences between risks as well as the presence of heavy tailed marginal distributions (see Section 2.4 for more details). This may be a typical context for risk management applications in insurance and finance. Moreover, the proposed method is tractable, even in high dimension (dimension 150 tested).

The chapter is organized as follows. In Section 2.2, we recall the definition of regularly varying function and then present conditions under which the VaR and TVaR of the sum and the maximum are asymptotically equivalent. In Section 2.3, we give classes of random vectors satisfying our hypothesis. Section 2.4 is devoted to a methodology for the estimation of Δ. In Section 2.5, we give explicit expressions of the VaR on some specific models (introduced in [Marshall & Olkin 1988, Oakes 1989] and also considered in [Dacorogna *et al.* 2014] where the expression of the VaR is derived). In Section 2.6, we compare our method with classical ones on several models. Conclusions are given in Section 2.7.

2.2 Asymptotic results on the VaR and the TVaR of the sum and the maximum

In this section, we will first recall the definition of regularly varying functions.

Definition 2.2.1. Let f be a positive measurable function on $\mathbb{R}_+$.

We say that f is regularly varying at infinity of index $\rho \in \mathbb{R}$ if

$$\lim_{t\to\infty} \frac{f(xt)}{f(t)} = x^\rho,$$

for any $x > 0$. Similarly, we say that $f : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at 0 if we replace $t \to \infty$ by $t \to 0$.

Regularly variation of f at $a > 0$ is defined as regularly variation at infinity for the function $f(a - 1/t)$.

If $\rho = 0$ then f is said to be slowly varying.

Definition 2.2.2. A random variable X with distribution function F is said to have a regularly varying upper tail if its survival function $\overline{F} = 1 - F$ is regularly varying at infinity.

Examples of regularly varying distributions are Pareto, Cauchy, Burr and stable with exponent $\alpha < 2$.

Let δ be the real valued function defined by $\delta(t) = \overline{F}_S(t)/\overline{F}_M(t)$. Throughout this chapter we will consider the condition

$$\Delta = \lim_{t\to\infty} \frac{\overline{F}_S(t)}{\overline{F}_M(t)} \quad \text{exists.} \tag{2.1}$$

The following result is somewhat a *folklore theorem*, it links the Value-at-Risk of the sum and the maximum in case where the survival function of the maximum, $\overline{F}_M$, is regularly varying. The result still holds for the TVaR . Recall that we do not assume that the marginal distributions are either identically distributed or independent.

Proposition 2.2.3. *Let $X = (X_1, \ldots, X_d)$ be a vector of positive random variables (r.v.s). Suppose that assumption (2.1) holds and that $\overline{F}_M$ is regularly varying with index $-\rho$. Then,*

(i) $1 \le \Delta \le d^\rho$;

(ii) $\mathrm{VaR}_{1-p}(S) \sim \mathrm{VaR}_{1-\Delta^{-1}p}(M)$ *as* p *tends to* 0;

(iii) if $\mathrm{TVaR}_p(M)$ *exists for all* p, *then*

$$\mathrm{TVaR}_{1-p}(S) \sim \mathrm{TVaR}_{1-\Delta^{-1}p}(M)$$

as p *tends to* 0.

Proof. Since $\overline{F}_M$ is regularly varying, (ii) follows from properties of regularly varying functions and (iii) follows from Karamata's Theorem.

Remark that as we always assume that marginal risks are almost surely positive we have

$$\{\max\{X_1,\dots,X_d\} > t\} \subset \{X_1 + \cdots + X_d > t\} \subset \{\max\{X_1,\dots,X_d\} > t/d\}$$

In particular

$$\overline{F}_M(t) \leq \overline{F}_S(t) \leq \overline{F}_M(t/d) \tag{2.2}$$

and thus $\delta(t) \leq \overline{F}_M(t/d)/\overline{F}_M(t)$. So that if $\overline{F}_M$ is regularly varying with index $-\rho$ then $\Delta \leq d^\rho$ and (i) follows. □

Classes of random vectors that satisfy the assumptions of Proposition 2.2.3 will be given in Section 2.3 while in Section 2.4 we will provide a method to estimate Δ.

2.3 Random vectors where the limit Δ exists

In this section we explore several situations in which the limit Δ exists and Proposition 2.2.3 applies.

2.3.1 Multivariate regular framework

Alink et al. ([Alink *et al.* 2005], [Alink *et al.* 2004] and [Alink *et al.* 2007]) studied the asymptotic behavior of the tail of the sum when the marginals of the vector $\mathbf{X} = (X_1,\dots,X_d)$ are identically distributed as one of the three extreme value families: Gumbel, Fréchet or Weibull and when the dependence within the vector is given by an Archimedean copula. Then Barbe et al. ([Barbe *et al.* 2006]) generalized these results under the framework of the multivariate regular variation distributions. Their main contribution is the explicit calculation of the limit

$$\lim_{t\to\infty} \frac{\overline{F}_S(t)}{\overline{F}_1(t)},$$

where F_1 is the common distribution function of the marginal risks $X_1,\dots,X_d$.

This kind of results suggest that we may approximate the VaR (and TVaR) of the sum simply by the VaR (and TVaR) of X_1. This point will be detailed in Section 2.6.4 where it will be shown empirically that maximum based estimation gives indeed better results than F_1 based one.

Let us recall the definition of multivariate regularly varying random vectors.

Definition 2.3.1 (Multivariate Regular Variation)**.** A random vector $\mathbf{X}$ is said to be multivariate regularly varying of index $-\beta$, $\beta > 0$ if there exists a finite measure $\mu_{\|\ \|}$ (which depends on the chosen norm $\|\ \|$) on $\Gamma_d = \left\{ \frac{\mathbf{x}}{\|\mathbf{x}\|} : \mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\} \right\}$ and a function $b : (0,\infty) \to (0,\infty)$, such that for all $x > 0$ and all $A \subset \Gamma_d$,

$$\lim_{t\to\infty} t \Pr\left(\|\mathbf{X}\| > xb(t), \frac{\mathbf{X}}{\|\mathbf{X}\|} \in A \right) = \frac{\mu_{\|\ \|}(A)}{x^\beta}.$$

Using the L^1 norm, $\|\mathbf{X}\|_1 = |X_1| + \cdots + |X_d|$, the L^∞ norm, $\|\mathbf{X}\|_\infty = \max\{[X_1|, \ldots, |X_d|\}$ and $b(t) = F_1^{\leftarrow}(1 - 1/t)$, one finds

$$\Delta = \lim_{t\to\infty} \frac{\overline{F}_S(t)}{\overline{F}_M(t)} = \frac{|\mu_{\|\cdot\|_1}|}{|\mu_{\|\cdot\|_\infty}|}, \tag{2.3}$$

where $|\mu|$ is the total mass of the measure μ. So that, when $\mathbf{X}$ is multivariate regularly varying Proposition 2.2.3 applies.

We are also interested in random vectors whose coordinates are not identically distributed. Results for identically distributed marginals will not lead to results for arbitrary marginals. This is the purpose of the next section where different kinds of dependence structure are also considered.

2.3.2 Examples where condition (2.1) holds.

In this section we show that condition (2.1) holds for three classes of multivariate distributions, namely those for which

- a regularly varying marginal clearly dominates the other marginals,
- the dependence structure is the survival of a regularly varying Archimedean and the marginals are regularly varying,
- the dependence structure is regularly varying Archimedean and the marginals are regularly varying.

We now state our result when one marginal is regularly varying and dominates the others.

Proposition 2.3.2. *Let $\boldsymbol{X}$ be a random vector in $\mathbb{R}_+^d$ with marginal distributions $F_i, 1 \le i \le d$. If $\overline{F}_1$ is regularly varying and*

$$\lim_{t\to\infty} \overline{F}_j(t)/\overline{F}_1(t) = 0$$

for any $2 \le j \le d$, then (2.1) holds with $\Delta = 1$.

Proof. The proof splits into two parts, one showing that $\overline{F}_M \sim \overline{F}_1$ at infinity, one showing that $\overline{F}_S \sim \overline{F}_1$ at infinity. The result follows in combining these two asymptotic equivalences.

(i) If X_1 exceeds t, so is M, and if M exceeds t, at least one of the X_i does. Therefore, we have

$$\overline{F}_1(t) \leq \overline{F}_M(t) \leq \sum_{1\leq i\leq d} \overline{F}_i(t).$$

This ensures that under the assumptions of Proposition 2.3.2, $\overline{F}_M \sim \overline{F}_1$ at infinity.

(ii) Since the inequality $X_1 \geq t$ implies $S \geq t$, we have

$$\overline{F}_1(t) \leq \overline{F}_S(t). \tag{2.4}$$

Furthermore, for any positive ε, decomposing the event $\{S > t\}$ according to whether $\max_{2\leq i\leq d} X_i \leq t\varepsilon$ or not, we have

$$P\{S > t\} \leq P\{X_1 > t(1-d\varepsilon)\} + P\Big\{\max_{2\leq i\leq d} X_i > t\varepsilon\Big\}.$$

Applying Bonferroni's inequality, we obtain

$$\overline{F}_S(t) \leq \overline{F}_1(t(1-d\varepsilon)) + \sum_{2\leq i\leq d} \overline{F}_i(t\varepsilon).$$

Since $\overline{F}_1$ is regularly varying of index ρ say, and dominates the other $\overline{F}_i$, we obtain that

$$\limsup_{t\to\infty} \overline{F}_S(t)/\overline{F}_1(t) \leq (1-d\varepsilon)^\rho.$$

Since ε is arbitrary, we have

$$\limsup_{t\to\infty} \overline{F}_S(t)/\overline{F}_1(t) \leq 1$$

Combined with (2.4), this yields that $\overline{F}_S \sim \overline{F}_1$ at infinity. □

We now consider a dependence structure between the components of the random vector given by an Archimedean copula or the survival copula of an Archimedean copula, and give a sufficient condition for condition (2.1) to hold. We first recall the definitions of Archimedean copulas and survival copulas.

Definition 2.3.3. (Archimedean Copulas) A generator is a function ψ from $[0,1]$ to $[0,\infty]$ such that

(i) ψ is decreasing with $\psi(1) = 0$,
(ii) the first d derivatives of $\psi^{\leftarrow}$ exists,
(iii) for any $k = 0, 1, \ldots, d$ and any t positive, $(-1)^k \frac{d^k}{dt^k}\psi^{\leftarrow}(t) \geq 0$,
where $\psi^{\leftarrow}$ denotes the pseudo-inverse of ψ defined by

$$\psi^{\leftarrow}(s) = \begin{cases} \psi^{-1}(s) & \text{if } 0 \leq s \leq \psi(0) \\ 0 & \text{if } \psi(0) \leq s \leq +\infty \ . \end{cases}$$

The Archimedean copula C with generator ψ is the distribution function on $[0,1]^d$ defined by

$$C(u_1, \ldots, u_d) = \psi^{\leftarrow}(\psi(u_1) + \ldots \psi(u_d)).$$

Definition 2.3.4. (Survival copula) Given a copula C, we define:

$$C^*(u_1, \ldots, u_d) = P(U_1 > 1 - u_1, \ldots, U_d > 1 - u_d)$$

with $(U_1, \ldots, U_d)$ having C as distribution function. C^* is a copula known as the survival copula of C.

We can now extend the result of Alink et al. [2,3,4] and Barbe et al. [6], to a situation where the marginal distributions are not identical.

Proposition 2.3.5. *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector with nonnegative components and marginal distributions F_i, $1 \leq i \leq d$. Suppose that for some regularly varying functions h there exists some a_i, not all 0, such that*

$$\lim_{t\to\infty} \overline{F}_i(t)/h(t) = a_i, \qquad 1 \leq i \leq d. \tag{2.5}$$

Let the dependence structure of $\boldsymbol{X}$ be given by one of the following:
(i) a survival copula of an Archimedean copula with generator ψ which is regularly varying at 0 with negative index,
(ii) an Archimedean copula with generator ψ which is regularly varying at 1 with negative index.
Then $\boldsymbol{X}$ is multivariate regularly varying, and condition (2.1) holds.

Note that since a regularly varying function is ultimately positive, the a_i are nonnegative. If a_i is positive, then (2.5) implies that $\overline{F}_i$ is regularly varying with the same index of regular variation as h. In particular, if one marginal tail is regularly varying and dominates the others, then h could be the corresponding survival function.

Proof. (i) First we assume that the dependence of $\mathbf{X}$ is the survival copula of an Archimedean copula with generator ψ which is regularly varying at 0 with negative index.

Using Bonferroni's identity and agreeing that a sum over an empty set is 0, we have, for any $x_1, \ldots, x_d$ positive,

$$P(\cup_{1\leq i\leq d}\{X_i \geq tx_i\}) = \sum_{I\subset\{1,2,\ldots,d\}} (-1)^{1+\sharp I} P\{X_i > tx_i : i \in I\}. \tag{2.6}$$

If I is such that $a_i = 0$ for some i in I, we define

$$I_0 = \{i \in I : a_i = 0\}.$$

We then have

$$\begin{aligned} P\{X_i > tx_i : i \in I\} &\leq P\{X_i > tx_i : i \in I_0\} \\ &\leq \sum_{i\in I_0} P\{X_i > tx_i\} \\ &= o(h(t)) \end{aligned} \tag{2.7}$$

as t tends to infinity.
If I is such that all a_i, $i \in I$, are positive, we have

$$P\{X_i > tx_i : i \in I\} = \psi^{\leftarrow}\left(\sum_{i\in I} \psi\left(h(t)\frac{\overline{F}_i(tx_i)}{h(t)}\right)\right). \tag{2.8}$$

Since h is regularly varying with index $-\rho$ say, we have

$$\lim_{t\to\infty} \frac{\overline{F}_i(tx_i)}{h(t)} = \lim_{t\to\infty} \frac{\overline{F}_i(tx_i)}{h(tx_i)} \frac{h(txi)}{h(t)} = a_i x_i^{-\rho}. \tag{2.9}$$

Since ψ is regularly varying at 0 with index $-\theta$ say, the uniform convergence theorem (Theorem 1.2.1 in [Bingham *et al.* 1989]) and (2.9) ensure that

$$\psi\Big(h(t)\frac{\overline{F}_i(tx_i)}{h(t)}\Big) \sim \psi \circ h(t)(a_i x_i^{-\rho})^{-\theta}$$

as t tends to infinity. Since $\psi^{\leftarrow}$ is regularly varying with index $-1/\theta$, we then have, using (2.8) and the uniform convergence theorem,

$$P\{X_i > tx_i : i \in I\} \sim h(t)\Big(\sum_{i\in I}(a_i x_i^{-\rho})^{-\theta}\Big)^{-1/\theta} \tag{2.10}$$

as t tends to infinity.
Note that if we take the limit of the right hand side of (2.10) as one of the a_i tends to 0, we obtain 0. Therefore, as long as we agree that $1/0 = \infty$ and

$1/\infty = 0$, we may capture (2.7) in (2.10). Then, considering (2.6), and using that at least one of the a_i does not vanish, we obtain

$$P(\cup_{1\le i\le d}\{X_i \ge tx_i\}) \sim h(t) \sum_{I\subset\{1,2,\dots,d\}} (-1)^{1+\sharp I}\Big(\sum_{i\in I}(a_i x_i^{-\rho})^{-\theta}\Big)^{-1/\theta}$$

as t tends to infinity.
It then follows from Theorem 6.1 in [Resnick 2007] that the distribution of $\mathbf{X}$ is multivariate regularly varying. Condition (2.1) then follows from (2.3).
(ii) We now assume that the dependence of $\mathbf{X}$ is an Archimedean copula with generator ψ which is regularly varying at 1 with negative index.
By definition, for any $x_1, \dots, x_d$ positive,

$$P(\cup_{1\le i\le d}\{X_i \ge tx_i\}) = 1 - \psi^{\leftarrow}\left(\sum_{i=1}^{d}\psi\Big(F_i(tx_i)\Big)\right).$$

Since ψ is regularly varying at 1 with index $-\theta$ say, and h is regularly varying with index $-\rho$ say, then $1-\psi^{\leftarrow}$ is regularly varying at 0 with index θ^{-1} and $\psi\circ h$ is regularly varying at infinity with index $-\rho\theta$. Then, by using the same arguments as above we can conclude that

$$P(\cup_{1\le i\le d}\{X_i \ge tx_i\}) \sim h(t)\left(\sum_{i=1}^{d}(a_i x_i^{-\rho})^{\theta}\right)^{-1/\theta}$$

as t tends to infinity and $\mathbf{X}$ is multivariate regularly varying. □

Notice that Proposition 2.3.5 implies that a random vector $\mathbf{X}$ with regularly varying marginals, for example Pareto distributed marginals not necessarily with same scale or shape parameters, and dependence structure given by one of the copulas listed below, satisfies the assumption of Proposition 2.2.3. Possible dependence structures are:

- independence (recall that the independent copula is an Archimedean copula with generator $\psi(t) = -\ln(t)$, and thus regularly varying at 1 with index -1),
- Gumbel copula with parameter $\theta \ge 1$ (which is an Archimedean copula with generator $\psi(t) = -\ln(t)^{\theta}$ and thus regularly varying at 1 with index $-\theta$),
- survival copula of a Clayton copula with parameter $\theta > 0$ (which is an Archimedean copula with generator $\psi(t) = (t^{-\theta}-1)/\theta$ and thus regularly varying at 0 with index $-\theta$).

2.4 Approximation of the limit Δ

In this section, we assume that the limit Δ exists and we show how to estimate it using samples of $\mathbf{X}$. We will use this estimation to approximate $\mathrm{VaR}_{1-p}(S)$, for different values of p close to 0 using Proposition 2.2.3.

Recall that δ is the real valued function defined by $\delta(t) = \overline{F}_S(t)/\overline{F}_M(t)$ and continue to denote by Δ its limit at infinity if it exists.

If a sample of $\mathbf{X}$ is available, the function δ can be estimated using the empirical cumulative distribution function (e.c.d.f.) of S and M. As we assume that F_M can be easily calculated by the c.d.f. $\mathbf{F}$ of the portfolio, at least two versions of the empirical delta should interest us:

$$\widehat{\delta}(t) = \frac{1 - \widehat{F}_S(t)}{1 - F_M(t)} \qquad \text{and} \qquad \tilde{\delta}(t) = \frac{1 - \widehat{F}_S(t)}{1 - \widehat{F}_M(t)}$$

where $\widehat{F}_S$ and $\widehat{F}_M$ are the e.c.d.f.s of S and M respectively, based on the sample of $\mathbf{X}$. The first version $\widehat{\delta}$ may be more tractable statistically, while the second $\tilde{\delta}$ has the nice property that $\tilde{\delta} \geq 1$. In order to obtain some insight on the convergence of δ to its limit Δ, we plot, in Figure 2.1, functions $\widehat{\delta}$ and $\tilde{\delta}$ for four different models which are multivariate regularly varying.

In the first model (sum of 10 independent Pareto distributions with tail index 1) we notice that the limit $\delta(t)$ seems to be 1 but the convergence is not fast enough to consider using this limit to approximate $\mathrm{VaR}_p(S)$ even for higher confidence levels p. For the second model (sum of 10 Pareto distributions with tail index 1 and dependence structure given by a Gumbel copula of parameter 1.5) the convergence is a lot faster, $\delta(t)$ seems to be close to its limit for t greater than the VaR at the 95% confidence level. The two models in the lower side (sum of 10 Pareto distributions: five with tail index 1 and five with tail index 3 both in the independent and Gumbel copula dependent case) behave the same as the ones in the upper side.
The models on the right side of Figure 2.1 correspond to cases where our method will be applicable: the limit Δ is reached by $\widehat{\delta}(t)$ for t near the $\mathrm{VaR}_{0.95}$. These models exhibit a strong dependence combined with at least one of the marginal risks with a very heavy tail. Even if this is a limitation of our method we should remark that this kind of models are also those where Monte Carlo methods are less efficient to approximate the VaR or the TVaR, so that it may be interesting to have an alternative method of approximation.

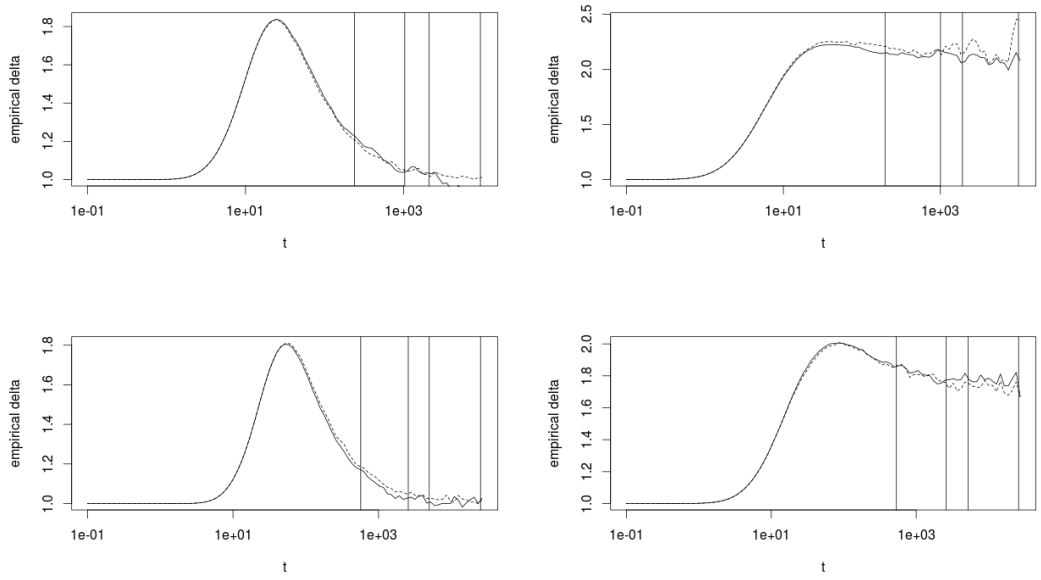

Figure 2.1 – Four plots of $\widehat{\delta}$ (solid) and $\tilde{\delta}$ (dashed) for different models, based on samples with size 10^4. Vertical lines are displayed at the empirical VaR of the sum at confidence levels 95%, 99%, 99.5%, 99.9%. Each model is a sum of 10 Pareto distributions with different tail indexes and different dependence structures. From top-left to bottom-right, we find: 1) independent Pareto distributions with tail index one; 2) the tail index is still one but dependence is given by a Gumbel copula of parameter 1.5; 3) independent Pareto distributions: five with tail index one and the other five with tail index 3; 4) the same as 3) but dependence is given by a Gumbel copula of parameter 1.5.

On a possible estimator of Δ

Let $(S_1, \ldots, S_n)$ be an i.i.d. sample of S. According to Donsker's Theorem, the empirical process

$$\sqrt{n}(\widehat{F}_S(t) - F_S(t))$$

converges in distribution to a Gaussian process with zero mean and covariance given by

$$F_S(t_1) - F_S(t_1)F_S(t_2)$$

for $t_1 \leq t_2$. Thus, given any sequence $0 < t_1 < \cdots < t_k$, the vector

$$\sqrt{n}\left(\widehat{\delta}(t_1) - \delta(t_1), \ldots, \widehat{\delta}(t_k) - \delta(t_k)\right)$$

converges in law to a centred Gaussian vector with covariances given by

$$\frac{F_S(t_i) - F_S(t_i)F_S(t_j)}{(1 - F_M(t_i))(1 - F_M(t_j))} = \frac{\delta(t_j)}{1 - F_M(t_i)} - \delta(t_i)\delta(t_j)$$

for any $i \leq j$. As a consequence

$$\sqrt{n}\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{\delta}(t_i) - \frac{1}{k}\sum_{i=1}^{k}\delta(t_i)\right)$$

converges to a normal distribution with zero mean and variance

$$\frac{1}{k^2}\sum_{1\leq i\leq j\leq k}\left\{\frac{\delta(t_j)}{1 - F_M(t_i)} - \delta(t_i)\delta(t_j)\right\}. \tag{2.11}$$

If we assume that the values t_i are large enough, the approximation $\delta(t_i) \approx \Delta$ holds for each $i = 1, \ldots, k$ and the variance (2.11) can be approximated by

$$\frac{\Delta}{k^2}\sum_{i=1}^{k}\left\{\frac{i}{1 - F_M(t_i)}\right\} - \frac{\Delta^2(k+1)}{2k}.$$

In practice we should plot points $\left(S_{(i)}, \widehat{\delta}(S_{(i)})\right)$ where $S_{(1)} < \cdots < S_{(n)}$ is the ordered sample of S and then choose a threshold in such a way that the approximation $\delta(S_{(n-i)}) \approx \Delta$ holds for any $0 \leq i \leq k$. The choice of the threshold is a recurrent and difficult problem in EVT, for which few theoretical results exist and are generally hardly applicable in practice. We propose then to estimate Δ by

$$\widehat{\Delta} = \frac{1}{k}\sum_{i=1}^{k}\widehat{\delta}(S_{(n-i)}). \tag{2.12}$$

As an example, the behavior of $\widehat{\delta}(x)$ for the Pareto-Clayton model, which will be described in Section 2.5, may be seen on Figure 2.2. The estimation $\widehat{\Delta}$ is represented by the solid line while dashed lines are for the estimated 95% confidence interval. See also Figure 2.3 for the shape of the δ function and the limit Δ.

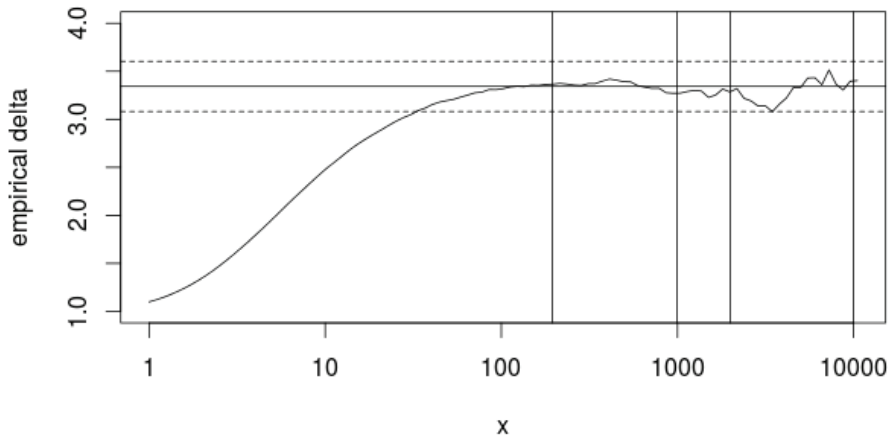

Figure 2.2 – Shape of the $\widehat{\delta}$ function of the Pareto-Clayton model with parameters $\alpha = 1, \beta = 1$ and $d = 10$ based on samples of size 10^4. Vertical lines are displayed at the empirical VaR of the sum at confidence levels 95%, 99%, 99.5%, 99.9%. The estimation $\widehat{\Delta}$ with its estimated 95% confidence interval is represented by the horizontal lines.

2.5 Some explicit calculations

In this section we will consider a simplified model in order to obtain explicit formulas for F_S and F_M and to better understand the scope and the limitations of our Δ estimation. The model is described by the following compound process: let Λ be a positive random variable and let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector such that

$$\Pr(X_1 > x_1, \ldots, X_d > x_d \,|\, \Lambda = \lambda) = \prod_{i=1}^{d} e^{-\lambda x_i},$$

for each $x_1, \ldots, x_d \geq 0$.

That means that conditionally on the value of Λ the marginals of $\mathbf{X}$ are independent and exponentially distributed. In general, the final distribution of $\mathbf{X}$ does not have independent marginals and they are not exponential either. Actually the dependence structure of $\mathbf{X}$ and its marginal distributions will depend on the distribution of Λ.

Some particular Λ distributions define some well-known models in which the explicit calculation of F_S and F_M is possible. For example when Λ is Gamma distributed, then the marginals of $\mathbf{X}$ are of Pareto type with dependence given by a survival Clayton copula. When Λ is Levy distributed

the marginals will be Weibull distributed with a Gumbel survival copula. These models have been introduced in [Marshall & Olkin 1988, Oakes 1989] and used in [Albrecher *et al.* 2011] to derive explicit formulas for ruin probabilities. In [Cénac *et al.* 2014, Maume-Deschamps *et al.* 2015], explicit results for the minimum of some risk indicators are obtained for this kind of models. We also would like to mention that the computation of the VaR for this model is given in [Dacorogna *et al.* 2014].

Let us consider the case where Λ is Gamma(α, β) distributed with density

$$f_\Lambda(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

In this case, the X_i's are Pareto(α, β) distributed with tail given by

$$\overline{F}_i(x) = \left(1 + \frac{x}{\beta}\right)^{-\alpha}$$

and the dependence structure is described by a survival Clayton copula with parameter $1/\alpha$. Through this chapter we will refer to this model as a Pareto-Clayton vector with parameters (α, β). This model is a particular Multivariate Pareto of type II with location parameters $\mu_i = 0$ and scale parameters $\sigma_i = \beta$ for $i = 1, \ldots, d$ (see for example [Yeh 2007]). As already noticed in Section 2.3.2, this model satisfies the hypothesis of Proposition 2.3.5 so that the limit Δ exists.

In the Pareto-Clayton model, the exact distribution function of $S = \sum_{i=1}^d X_i$ can be calculated. Conditionally on $\Lambda = \lambda$, the sum S is Gamma distributed with parameters $(1/\lambda, d)$, distribution also known as the Erlang distribution. Then, as here we are assuming that Λ is Gamma(α, β) distributed, the total distribution of S is the result of compounding two Gamma distributions, more precisely

$$S \sim \text{ Gamma}(1/\Lambda, d) \text{ where } \Lambda \sim \text{ Gamma}(\alpha, \beta).$$

It is well known that the result of this compound distribution is the so-called Beta prime distribution (see [Dubey 1970]). The c.d.f. of S can be expressed in terms of F_β, the c.d.f. of the Beta$(d\beta, \alpha)$ distribution, as

$$F_S(x) = F_\beta\left(\frac{x}{1+x}\right).$$

Naturally, the inverse of F_S can also be expressed in function of the inverse of the Beta distribution

$$F_S^{\leftarrow}(p) = \frac{F_\beta^{\leftarrow}(p)}{1 - F_\beta^{\leftarrow}(p)}.$$

In this example, the δ function is explicitly calculated (see Figure 2.3). Moreover, computer algebra softwares allow us to calculate explicitly the limit Δ for specified parameters.

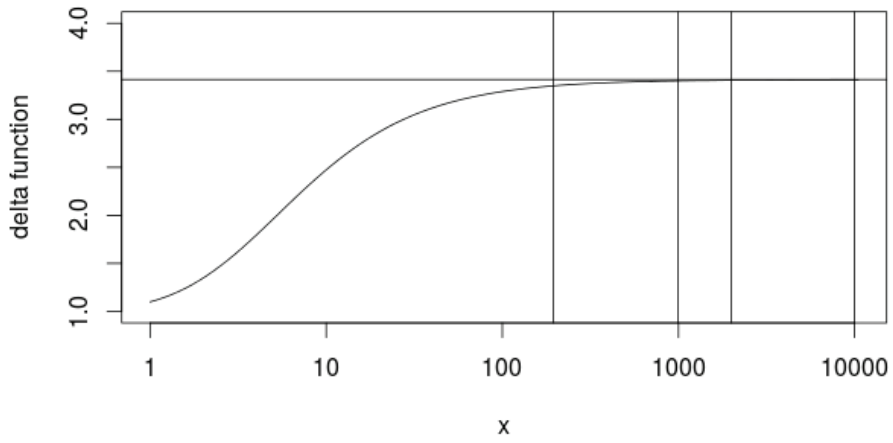

Figure 2.3 – Shape of the δ function of the Pareto-Clayton model, with parameters $\alpha = 1, \beta = 1$ and $d = 10$. Vertical lines are displayed at the VaR of the sum at confidence levels 95%, 99%, 99.5%, 99.9%. The limit $\Delta \approx 3.4142$ is represented by the horizontal line.

In order to see how fast the function δ converges to its limit Δ, we plot the function $p \mapsto \delta(\mathrm{VaR}_p(S))$ for different values of the parameter α and different dimensions d (see Figure 2.4). We remark that $\delta(x)$ is already very close to Δ when $x = \mathrm{VaR}_{0.95}(S)$, for $\alpha \leq 2.5$. The lower the value α, the flatter the tail of δ and thus the limit Δ is attained rapidly. Remark that the lower the levels of α, the heavier the tails of the Pareto marginals. Finally, this plot confirms the intuition that for heavier marginals the tail of the sum is better approximated by the tail of the maximum. A similar phenomenon in the i.i.d. case has been noted in [Barbe & McCormick 2005] when approximating $\overline{F}_S$ by $\overline{F}_1$.

2.6 Some numerical examples

In this section we show how the ideas presented in the above section can help to estimate in practice the VaR and the TVaR of a sum at confidence levels close to 1.
We compare the estimation done via the Δ-limit estimation (*New* in the tables

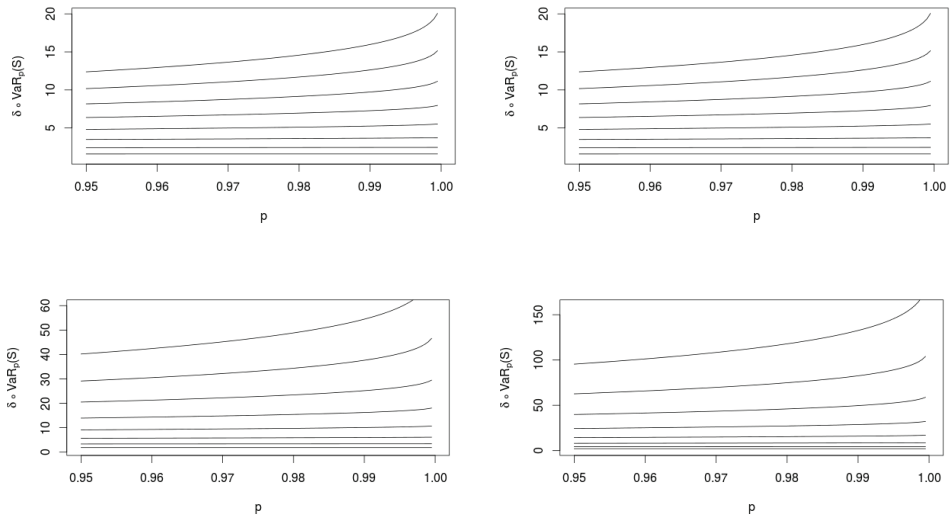

Figure 2.4 – Four plots of the $p \mapsto \delta(\mathrm{VaR}_p(S))$ function of the Pareto-Clayton model for dimensions $d = 2, 6, 10$ and 14 (from top-left to bottom-right) are represented. For each dimension, the curves with $\alpha = 0.5, 1, 1.5, 2, 2.5, 3, 3.5$ and 4 are plotted and they can be seen from bottom to top on each chart.

below) as described in Section 2.4 with other common quantile estimation methods, with the same sample size:

1. The direct Monte Carlo quantile estimation (MC).
2. The quantile estimation from a GPD fitted distribution where parameters are estimated using maximum likelihood method (GPD 1).
3. The quantile estimation from a GPD fitted distribution where parameters are estimated using the moment method (GPD 2).
4. The high quantile estimate based on a method by Weissman [Weissman 1978] (Weiss.).

We first consider the Pareto-Clayton model presented in Section 2.5 (dimension 2 and 10), where exact values for the Value-at-Risk are computable (see Section 2.6.1 and Section 2.6.2). Then, we test our method with a different model where exact values are not known.
In order to study the performance of our estimator and to compare it with the main competitors, we consider the root-mean-squared error (RMSE) loss function. When n estimations have been performed, it is defined by

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\mathrm{VaR}}_p(S^i) - \mathrm{VaR}_p(S)\right)^2},$$

where $\widehat{\mathrm{VaR}}_p(S^i)$ represents the estimate of $\mathrm{VaR}_p(S)$ for any of the different methods presented above, on the ith sample. In the case where the exact value is not known, in Section 2.6.3, we compare our results on a sample of size 10^5 with several methods (1-4 above) to a Monte Carlo quantile estimation based on a very large sample of size 3×10^8. This last estimation is considered as the exact VaR value in the RMSE computation.

2.6.1 Pareto-Clayton model dimension 2

Here we consider the model presented in Section 2.5. We first consider $d = 2$ and $\alpha = 1$ which corresponds to a model with Pareto marginals with $\alpha = 1$ and dependence given by a survival Clayton copula with parameter $\theta = 1$.

In Table 2.1, the exact VaR at different confidence levels (from 95% to 99.95%) is presented. In Table 2.2 and Table 2.3, we present the RMSE criterion in percentage of the real value based on 1000 simulations at different confidence levels. At each simulation a sample of size 10^4 in Table 2.2 and size 10^5 in Table 2.3 is used to estimate the VaR. On each method (New, GPD 1, GPD 2 and Weiss) the threshold used on each estimation corresponds to the empirical 95% quantile. Clearly, in term of RMSE, our method performs better than classical methods at each confidence level, even for very high levels. When increasing the size of the sample (10^5 instead of 10^4) classical methods improve but our method still produces the best results.

VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
194.5	994.5	1994.5	9994.5	19994.5

Table 2.1 – Exact Value-at-Risk at different confidence levels on the Pareto-Clayton model in dimension $d = 2$ with $\alpha = 1$.

2.6.2 Pareto-Clayton model dimension 10

We consider again the Pareto-Clayton model but here $d = 10$ and $\alpha = 1$ which corresponds to a model with Pareto marginals with $\alpha = 1$ and dependence given by a survival Clayton copula with parameter $\theta = 1$. Results are presented in Tables 2.4, 2.5 and 2.6. As above, on each method (New, GPD 1, GPD 2 and Weiss) the threshold used on each estimation corresponds to the empirical 95% quantile. We mention that even in dimension 10, the estimation remains efficient for high level quantiles.

Method	VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
New	1.9%	1.7%	1.7%	1.7%	1.7%
MC	4.4%	10.3%	14.1%	38.2%	76.2%
GPD 1	11.3%	8.5%	11.8%	23.8%	30.2%
GPD 2	4.4%	11.1%	15.1%	25.1%	29.9%
Weiss.	4.4%	11.2%	15.1%	25.0%	29.6%

Table 2.2 – RMSE in percentage of the real value based on 1000 simulations. At each simulation a sample of size 10^4 is used to estimate the VaR.

Method	VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
New	0.7%	0.5%	0.6%	0.6%	0.6%
MC	1.4%	3.1%	4.4%	9.7%	14.4%
GPD 1	5.2%	2.6%	3.6%	7.2%	8.9%
GPD 2	1.4%	3.7%	4.7%	7.7%	9.1%
Weiss.	1.4%	3.9%	4.9%	7.7%	9.0%

Table 2.3 – RMSE in percentage of the real value based on 1000 simulations. At each simulation a sample of size 10^5 is used to estimate the VaR.

We also remark that our method is more efficient than classical ones from level 0.99.

2.6.3 A model with 150 different Pareto marginals and Gumbel copula

We apply now our method to a model where the exact value of $\mathrm{VaR}_p(S)$ is not known. The model is composed of 150 marginals Pareto(α_i, β_i) distributed with parameters $\alpha_i = (3 - i \mod (3))/2$ and $\beta_i = 5 - i \mod (5)$ for $i = 1, \ldots, 150$, where $i \mod (j)$ denotes the reminder of i divided by j. The model is then composed of fifty Pareto marginals of tail index 0.5, fifty of tail index 1 and fifty with tail index 1.5, and different scale parameters within $1, 2, \ldots, 5$. The dependence structure is given by a Gumbel copula of parameter 1.5. Recall that for this model, by the comments that follow Proposition 2.3.5, the limit Δ exists.

Table 2.7 presents the VaR estimation based on a classical Monte Carlo quantile estimation with a sample of size 3×10^8. We assume this estimation is the "real VaR" in the computation of the RMSE presented in Table 2.8. On each method (New, GPD 1, GPD 2 and Weiss) the threshold used on

VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
194.5	994.5	1994.5	9994.5	19994.5

Table 2.4 – Exact Value-at-Risk at different confidence levels on the Pareto-Clayton model in dimension $d = 10$ with $\alpha = 1$.

Method	VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
New Method	8.4%	7.8%	7.7%	7.7%	7.7%
MC	4.5%	10.1%	14.5%	43.6%	85.5%
GPD 1	10.7%	8.5%	12.1%	25.0%	32.1%
GPD 2	4.5%	11.3%	15.6%	26.5%	31.8%
Weiss.	4.5%	11.4%	15.5%	26.1%	31.2%

Table 2.5 – RMSE in percentage of the real value based on 1000 simulations. At each simulation a sample of size 10^4 is used to estimate the VaR.

each estimation corresponds to the empirical 99% quantile. It is notable that our method is very stable with respect to others and is more efficient to approximate the VaR_p from $p = 0.99$.

2.6.4 Comparison of the method using $\max(X)$ vs X_1

The method of estimation of the Value-at-Risk of the sum proposed in this chapter relies on the convergence of the function $\delta(t) = \overline{F}_S(t)/\overline{F}_M(t)$. When the convergence is assured and it is fast enough, it has been shown that the proposed method gives accurate and stable estimations of the VaR at high levels. In theory, similar results could be obtained if the maximum M is replaced by X_1 where X_1 is assumed to have the heaviest tail in the vector X. In this section we compare numerically the estimation of the VaR using, on one side, $\delta(t) = \overline{F}_S(t)/\overline{F}_M(t)$ and, on the other side, $\delta'(t) = \overline{F}_S(t)/\overline{F}_{X_1}(t)$, i.e we compare the approximation of $\mathrm{VaR}_{1-p}(S)$ by $\mathrm{VaR}_{1-p/\Delta}(M)$ and $\mathrm{VaR}_{1-p/\Delta'}(X_1)$ where Δ and Δ' are the approximated limits of $\delta(t)$ and $\delta'(t)$ respectively estimated using (2.12).

We first consider the model $(X_1, \ldots, X_{10})$ where X_1 is Pareto distributed with $\alpha = 0.9$ and $X_2, \ldots, X_{10}$ are Pareto distributed with $\alpha = 1$. The dependence structure is given by a Gumbel copula with parameter 2. Empirical δ and δ' functions are displayed in Figure 2.5.

The δ function becomes almost horizontal before the VaR of the sum at the 95% confidence level whereas δ' does not seem to be close to the limit on

Method	VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
New	2.6%	2.2%	2.2%	2.3%	2.3%
MC	1.4%	3.2%	4.6%	10.1%	14.8%
GPD 1	4.3%	2.7%	3.8%	7.4%	9.2%
GPD 2	1.4%	3.6%	4.8%	7.8%	9.2%
Weiss.	1.4%	4.1%	5.2%	7.9%	9.1%

Table 2.6 – RMSE in percentage of the real value based on 1000 simulations. At each simulation a sample of size 10^5 is used to estimate the VaR.

VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
8.1981e06	3.2770e07	8.1545e08	3.2561e09

Table 2.7 – Estimated Value-at-Risk at different confidence levels for the model described in Section 2.6.3 estimated with a sample of size 3×10^8.

the displayed range. Then, the estimation of the VaR using δ' seems to be not accurate. This is confirmed by Table 2.9 where some VaR estimations are presented. From now on, the threshold used for the Δ and the Δ' approximations using formula (2.12) corresponds to the 95% empirical quantile and for each estimation a sample of size 10^5 is generated.

Even in the case where all the marginal risks are equal the use of the max seems to give better results. We consider the model $(X_1, \ldots, X_{10})$ where all the X_i's are Pareto distributed with the same index $\alpha = 1$. The dependence structure is given by a Gumbel copula with parameter 2. Empirical δ and δ' functions are displayed in Figure 2.6.

As above the δ function seems to converge faster than δ' but in this case the difference is not as important as in Figure 2.5. In Table 2.10 some VaR estimations are presented. Again, estimations provided by using the estimation of Δ are of better quality than the ones provided by using the estimation of Δ'.

Mathematically speaking, some work remains to be done to understand why the approximation of $\overline{F}_S$ by $\overline{F}_M$ is so much better than that by $\overline{F}_1$. This will be the object of further investigations.

2.7 Conclusion

In this chapter, we give some conditions under which the tail distribution of the sum can be approximated by using the tail of the maximum of a vector.

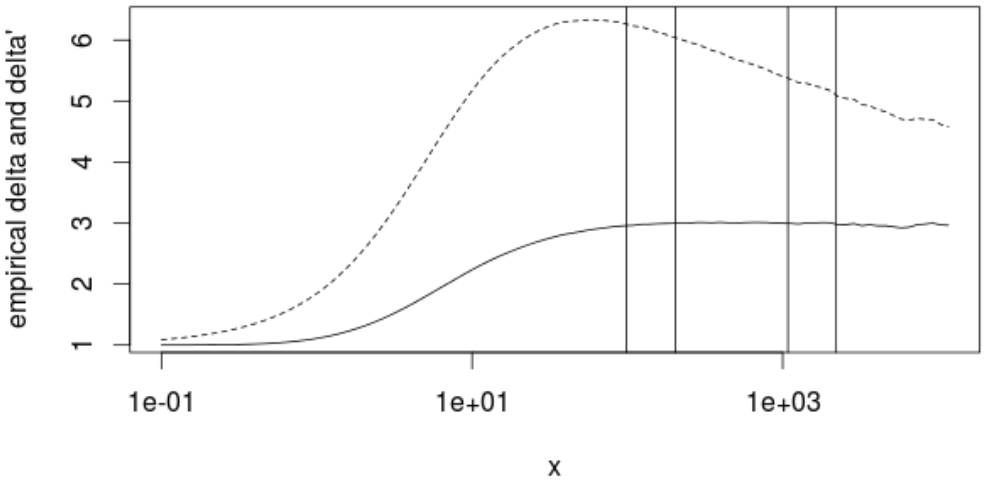

Figure 2.5 – Shape of an empirical $\delta(x)$ (solid) and $\delta'(x)$ (dashed) functions based on 10^5 simulations. Vertical lines are displayed at the empirical VaR of the sum at confidence levels 95%, 99%, 99.5%, 99.9%.

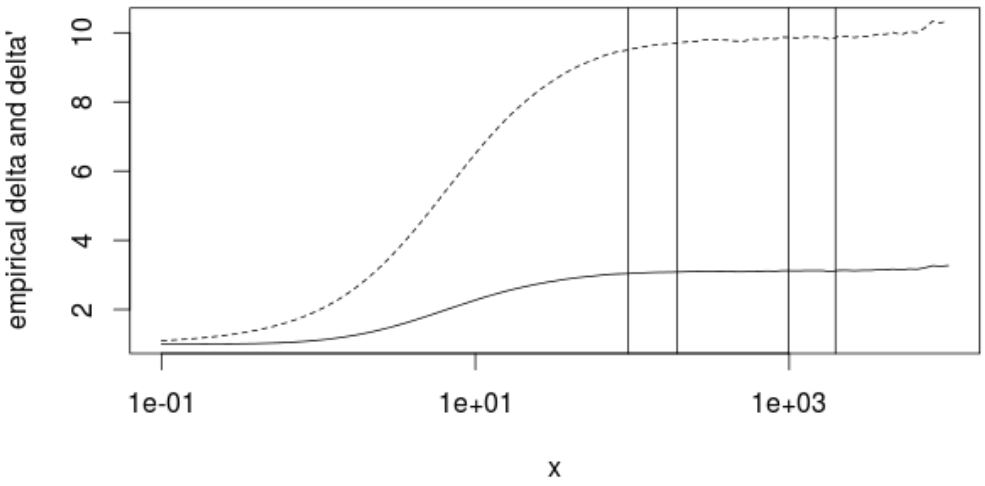

Figure 2.6 – Shape of an empirical $\delta(x)$ (solid) and $\delta'(x)$ (dashed) functions based in 10^5 simulations. Vertical lines are displayed at the empirical VaR of the sum at confidence levels 95%, 99%, 99.5%, 99.9%

Method	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
New	5.0%	4.9%	5.0%	5.0%
MC	6.2%	9.2%	21.2%	30.9%
GPD 1	5.9%	7.7%	12.4%	16.3%
GPD 2	5.9%	7.9%	13.1%	15.4%
Weiss.	5.9%	7.9%	13.0%	15.3%

Table 2.8 – RMSE in percentage of the estimated VaR presented in Table 2.7 based on 1000 simulations. At each simulation a sample of size 10^5 is used to estimate the VaR.

	VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
MC (3×10^8)	200	1058	2166	11201	22809
New method	203	1067	2188	11665	24083
using $\max(X)$	(2%)	(2%)	(2%)	(5%)	(6%)
New method	188	1126	2432	14549	31428
using X_1	(6%)	(7%)	(12%)	(30%)	(38%)

Table 2.9 – First line: Monte Carlo VaR estimation using 3×10^8 simulations. Second and third lines: mean and RMSE of 1000 VaR estimations using the max and the Δ' approximations. The RMSE is presented in % of the MC estimation.

We show how the VaR or the TVaR on high levels for the sum can be approximated, by first estimating a limiting constant Δ. The models in which our results can be applied include those where marginals are regularly varying and such that dependence is given by an Archimedean copula or survival copula. We do not require the marginals to be identically distributed and the method works for very high dimensions d ($d = 150$ for exemple). Our method gives a good approximation for the VaR and the TVaR when the convergence of $\delta(x)$ to Δ is fast enough. This generally happens when at least one of the marginal risks is strongly heavy tailed and when the dependence is strong. In particular, the method is not suitable e.g. for the case of two independent Pareto distributions. We also remark that the models for which our method applies correspond generally to those where Monte Carlo approximations are less efficient and there so is a real need for alternative methods.

	VaR 95%	VaR 99%	VaR 99.5%	VaR 99.9%	VaR 99.95%
MC (3×10^8)	196	1003	1996	9977	19931
New method	202	1068	2189	11671	24097
using $\max(X)$	(4%)	(7%)	(10%)	(17%)	(21%)
New method	188	1126	2434	14556	31444
using X_1	(5%)	(13%)	(22%)	(46%)	(58%)

Table 2.10 – First line: Monte Carlo VaR estimation using 3×10^8 simulations. Second and third lines: mean and RMSE of 1000 VaR estimations using the max and the Δ' approximations. The RMSE is presented in % of the MC estimation.

2.8 On a generalization to the consistently varying class of functions

Some of the results presented in Chapter 2 could be generalized under the consistently varying framework. We first give the definition and derive some general properties on the consistently varying functions and its generalized inverses. Finally we show how Propositions 2.2.3 and 2.3.5 can be generalized to hold for consistently varying functions.

Definition 2.8.1. We say that f is consistently varying at infinity, denoted as $f \in \mathcal{C}_\infty$, if

$$\lim_{t\downarrow 1} \liminf_{x\to\infty} \frac{f(tx)}{f(x)} = 1.$$

Similarly to the regularly varying definition, f is said to be consistently varying at $a > 0$, denoted by $f \in \mathcal{C}_a$, if $g(x) = f(a - 1/x)$ is consistently varying at infinity.

Notice that every regularly varying function is consistently varying. Examples of consistently varying functions which are not regularly varying can be found in [Cline & Samorodnitsky 1994].

2.8.1 Some results on consistently varying functions

It is well known that the sum and composition of regularly varying functions are again regularly varying: if f_i is regularly varying at infinity with index ρ_i, $i = 1, 2$, then $f_1 + f_2$ is regularly varying with index $\rho = \max\{\rho_1, \rho_2\}$ and if $f_2(\infty) = \infty$ then $f_1 \circ f_2$ is regularly varying with index $\rho = \rho_1\rho_2$ (see for example [Bingham *et al.* 1989]). Below we prove that functions which are consistently varying at infinity also satisfy these closure properties.

Proposition 2.8.2. *Let f and g be two non-increasing functions consistently varying at infinity, then the following is satisfied:*
(i) $g \circ (1/f)$ is consistently varying at infinity if $f(\infty) = 0 = g(\infty)$;
(ii) $f + g$ is consistently varying at infinity.

Proof. (i) Set $\varepsilon > 0$ and choose $s' > 1$ and $s > 1$ such that

$$\liminf_{x\to\infty} \frac{g(s'x)}{g(x)} > 1 - \varepsilon \text{ and } \liminf_{x\to\infty} \frac{f(sx)}{f(x)} > 1/s'.$$

Then

$$\liminf_{x\to\infty} \frac{g(1/f(sx))}{g(1/f(x))} \geq \liminf_{x\to\infty} \frac{g(s'/f(x))}{g(1/f(x))} > 1 - \varepsilon,$$

which proves the proposition.
(ii) Set $\varepsilon > 0$, and define for $t > 1$

$$L(t) := \liminf_{x\to\infty} \frac{f(tx) + g(tx)}{f(x) + g(x)} = \liminf_{x\to\infty} \left(\frac{f(tx)/f(x)}{1 + g(x)/f(x)} + \frac{g(tx)/g(x)}{f(x)/g(x) + 1} \right).$$

Then, as f and g are consistently varying at infinity, there exist reals $s > 1$ and $N > 0$ such that $f_i(sx)/f_i(x) > 1 - \varepsilon$ for all $x > N$, and $i = 1, 2$. Then

$$L(s) \geq (1 - \varepsilon) \liminf_{x\to\infty} \left(\frac{1}{1 + g(x)/f(x)} + \frac{1}{1 + f(x)/g(x)} \right) = (1 - \varepsilon).$$

As L is non-increasing, then also $L(s') \geq 1 - \varepsilon$ for all $1 < s' < s$. Finally, as $L(t) \leq 1$ for all $t > 1$ we have shown

$$\lim_{t\downarrow 1} L(t) = 1.$$

□

Now, we present some properties for consistently varying functions and their inverses. Before we should recall some well known facts on generalized inverse functions.

Definition 2.8.3. Let f be a non-decreasing function and h a non-increasing function. The generalized inverses of f and h are defined respectively as

$$f^{\leftarrow}(t) = \inf\{s : f(s) \geq t\} \quad \text{and} \quad h^{\leftarrow}(t) = \inf\{s : h(s) \leq t\}.$$

Remark that if h is positive we have then $h^{\leftarrow}(t) = (1/h)^{\leftarrow}(1/t)$.

Proposition 2.8.4. *If f and h are two right-continuous functions, respectively non-decreasing and non-increasing, then for any x and y the following is satisfied:*

(i) $f^{\leftarrow}(f(x)) \leq x$, $f(f^{\leftarrow}(x)) \geq x$ *and* $f^{\leftarrow}(y) \leq x \Leftrightarrow y \leq f(x)$
(ii) $h^{\leftarrow}(h(x)) \leq x$, $h(h^{\leftarrow}(x)) \leq x$ *and* $h^{\leftarrow}(y) \leq x \Leftrightarrow y \geq h(x)$.

Proposition 2.8.5. *Let f be a consistently varying and non-increasing measurable function $f : \mathbb{R}_+ \to \mathbb{R}_+$ and $a(x)$ and $b(x)$ two positive sequences such that $a(x) \sim b(x)$ as $t \to \infty$ and $a(\infty) = \infty$. Then the following is satisfied:*
(i) The function $x \mapsto f^{\leftarrow}(1/x)$ is consistently varying at infinity
(ii) If $f(\infty) = 0$, then $f^{\leftarrow} \circ f(x) \sim x$ and $f \circ f^{\leftarrow}(1/x) \sim 1/x$ when $x \to \infty$
(iii) $f(a(x)) \sim f(b(x))$ and $f^{\leftarrow}(1/a(x)) \sim f^{\leftarrow}(1/b(x))$ when $x \to \infty$

Proof. (i) Set $H_x(t) = f(tx)/f(x)$ and $H(t) = \liminf_{x\to\infty} H_x(t)$. We will first show that $\liminf_{x\to\infty} H_x^{\leftarrow}(\omega) \leq H^{\leftarrow}(\omega)$ for all $0 < \omega < 1$. Take $\omega \in (0, 1)$, for each $x > 0$ let us denote by $I_x(\omega)$ the value

$$I_x(\omega) = \inf\{H_s^{\leftarrow}(\omega) \,:\, s \geq x\}.$$

Then $H_x^{\leftarrow}(\omega) \geq I_x(\omega)$ and by the last equivalence of Proposition 2.8.4 (ii) $\omega \leq H_x(I_x(\omega))$, for all $x > 0$. We have then that $\omega \leq \inf\{H_s(I_s(\omega)) \,:\, s \geq x\}$ for any $x > 0$. Notice now that as $I_x(\omega)$ is non-decreasing on x and that each H_x is a non-increasing function then

$$\omega \leq \inf\{H_s(I_s(\omega)) \,:\, s \geq x\} \leq \inf\{H_s(I_x(\omega)) \,:\, s \geq x\} \leq H(I_x(\omega))$$

for all $x > 0$. By Proposition 2.8.4 (ii) we have

$$I_x(\omega) \leq H^{\leftarrow}(\omega)$$

for all $x > 0$. Thus by taking limits we find

$$\liminf_{x\to\infty} H_x^{\leftarrow}(\omega) \leq H^{\leftarrow}(\omega).$$

Now, we prove the announced result. For each $x > 0$ and $0 < \omega < 1$ we have $H_x{}^{\leftarrow}(\omega) = f^{\leftarrow}(\omega f(x))/x$. Proposition 2.8.4 (ii) gives, $f \circ f^{\leftarrow}(1/x) \leq 1/x$ for all $x > 0$ and then for $0 < \omega < 1$ we have $f^{\leftarrow}(\omega/x) \leq f^{\leftarrow}(\omega f \circ f^{\leftarrow}(1/x))$. Now, we have

$$1 \leq \liminf_{x\to\infty} \frac{f^{\leftarrow}(\omega/x)}{f^{\leftarrow}(1/x)} \leq \liminf_{x\to\infty} \frac{f^{\leftarrow}(\omega f \circ f^{\leftarrow}(1/x))}{f^{\leftarrow}(1/x)} = \liminf_{x\to\infty} H_x^{\leftarrow}(\omega) \leq H^{\leftarrow}(\omega).$$

As $f \in \mathcal{C}_\infty$ implies $H(t) \uparrow 1$ when $t \downarrow 1$, and thus, as H is non-increasing, then $H^{\leftarrow}(\omega) \downarrow 1$ when $\omega \uparrow 1$ and

$$\lim_{\omega\uparrow 1} \liminf_{x\to\infty} \frac{f^{\leftarrow}(\omega/x)}{f^{\leftarrow}(1/x)} = 1,$$

which proves the Proposition.
(ii) Proposition 2.8.4 (ii) gives $f^{\leftarrow}(f(x)) \leq x$ for all $x > 0$. By definition of the generalized inverses it follows that

$$0 \leq x - f^{\leftarrow} \circ f(x) \leq \lim_{s\uparrow 1} f^{\leftarrow}(sf(x)) - \lim_{t\downarrow 1} f^{\leftarrow}(tf(x))$$

where the last expression represents the size of the possible jump of $f^{\leftarrow}$ at $f(x)$. Take $\varepsilon > 0$, by (i) the mapping $x \mapsto f^{\leftarrow}(1/x)$ is consistently varying so we can choose $N > 0$, $0 < s < 1$ and $t > 1$ such that for all $x > N$

$$\frac{f^{\leftarrow}(s/x)}{f^{\leftarrow}(1/x)} \leq 1 + \frac{\varepsilon}{2} \quad \text{and} \quad \frac{f^{\leftarrow}(t/x)}{f^{\leftarrow}(1/x)} \geq 1 - \frac{\varepsilon}{2}.$$

Then

$$0 \leq 1 - \frac{f^{\leftarrow} \circ f(x)}{x} \leq \frac{f^{\leftarrow}(sf(x)) - f^{\leftarrow}(tf(x))}{f^{\leftarrow} \circ f(x)} \leq \varepsilon$$

where the last inequality holds for all $x > f^{\leftarrow}(1/N)$. As ε was arbitrary it had been proven that $f^{\leftarrow} \circ f(x) \sim x$. The proof that $f \circ f^{\leftarrow}(1/x) \sim 1/x$ is similar.
(iii) Set $\varepsilon > 0$, and let T be such that $1-\varepsilon \leq a(x)/b(x) \leq 1+\varepsilon$ for any $x \geq T$. Then as f is non-increasing, for $x \geq T$ we have

$$\frac{f((1+\varepsilon)b(x))}{f(b(x))} \leq \frac{f(a(x))}{f(b(x))} \leq \frac{f((1-\varepsilon)b(x))}{f(b(x))}.$$

Applying limits in the equation above we get,

$$\liminf_{x\to\infty} \frac{f((1+\varepsilon)b(x))}{f(b(x))} \leq \liminf_{x\to\infty} \frac{f(a(x))}{f(b(x))}$$

and

$$\limsup_{x\to\infty} \frac{f(a(x))}{f(b(x))} \leq \limsup_{x\to\infty} \frac{f((1-\varepsilon)b(x))}{f(b(x))}.$$

As $f \in \mathcal{C}_\infty$ then

$$\lim_{\varepsilon\downarrow 0} \liminf_{x\to\infty} \frac{f((1+\varepsilon)b(x))}{f(b(x))} = 1.$$

Similarly, as

$$\limsup_{x\to\infty} \frac{f((1-\varepsilon)b(x))}{f(b(x))} = \liminf_{x\to\infty} \frac{f(b(x))}{f((1-\varepsilon)b(x))}$$

then $f \in \mathcal{C}_\infty$ implies

$$\lim_{\varepsilon\downarrow 0} \limsup_{x\to\infty} \frac{f((1-\varepsilon)b(x))}{f(b(x))} = 1.$$

Thus we had proved

$$\lim_{x\to\infty} \frac{f(a(x))}{f(b(x))} = 1.$$

The proof of $f^{\leftarrow}(1/a(x)) \sim f^{\leftarrow}(1/b(x))$ follows from (i). □

2.8.2 Generalization of some results to the consistently varying case

Proposition 2.2.3 links the VaR of the sum and the maximum in case where the survival function of the maximum is regularly varying. We now generalize this result to the consistently varying case and give an example where the tail of the maximum is consistently varying.

Proposition 2.8.6. *Let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a vector of positive random variables (r.v.s). Suppose that $\overline{F}_M$ is consistently varying and that $\delta(x) := \frac{\overline{F}_S(x)}{\overline{F}_M(x)} \to \Delta$ as $x \to \infty$, for some $1 \leq \Delta < \infty$. Then*

$$\mathrm{VaR}_{1-p}(S) \sim \mathrm{VaR}_{1-\Delta^{-1}p}(M) \text{ as } p \to 0.$$

Proof. By Proposition 2.8.5 (ii) and (iii) we have

$$t \sim \overline{F}_M^{\leftarrow} \circ \overline{F}_M(t) \sim \overline{F}_M^{\leftarrow}\left(\Delta^{-1}\overline{F}_S(t)\right).$$

Again, by combining Proposition 2.8.5 (ii) and (iii) we have as $p \to 0$

$$\overline{F}_S^{\leftarrow}(p) \sim \overline{F}_M^{\leftarrow}\left(\Delta^{-1}\overline{F}_S \circ \overline{F}_S^{\leftarrow}(p)\right) \sim \overline{F}_M^{\leftarrow}(\Delta^{-1}p)$$

After rewriting the last equation in terms of the VaR function the result follows. □

Finally we show that under the framework of Archimedean dependence the maximum of consistently varying vectors is again consistently varying and we thus generalize partially Proposition 2.3.5.

Proposition 2.8.7. *Let $\boldsymbol{X} = (X_1, \dots, X_d)$ be a random vector with consistently varying marginal tails. Suppose that the copula C of $\boldsymbol{X}$ is Archimedean with generator ψ and that $x \mapsto \psi(1 - 1/x)$ is consistently varying. Then the tail of the maximum is consistently varying.*

Proof. The proof follows Proposition 2.3.5, but we use here Propositions 2.8.5 (i) and 2.8.2 (i). The closure by sum of the consistently varying functions stated in Proposition 2.8.2 (ii) is also required. □

References

[Albrecher *et al.* 2011] Hansjörg Albrecher, Corina Constantinescu and Stéphane Loisel. *Explicit ruin formulas for models with dependence among risks.* Insurance: Mathematics and Economics, vol. 48, pages 265–270, 2011.

[Alink *et al.* 2004] Stan Alink, Matthias Löwe and Mario V. Wüthrich. *Diversification of aggregate dependent risks.* Insurance: Mathematics and Economics, vol. 35, no. 1, pages 77–95, 2004.

[Alink *et al.* 2005] Stan Alink, Matthias Löwe and Mario V. Wüthrich. *Analysis of the expected shortfall of aggregate dependent risks.* ASTIN Bulletin, vol. 35, no. 1, pages 25–43, 2005.

[Alink *et al.* 2007] Stan Alink, Matthias Löwe and Mario V. Wüthrich. *Diversification for general copula dependence.* Statistica Neerlandica, vol. 61, no. 4, pages 446–465, Novembre 2007.

[Arbenz *et al.* 2011] P. Arbenz, P. Embrechts and G. Puccetti. *The* AEP *algorithm for the fast computation of the distribution of the sum of dependent random variables.* Bernoulli, vol. 17, no. 2, pages 562–591, 2011.

[Barbe & McCormick 2005] Philippe Barbe and William P McCormick. *Asymptotic expansions of convolutions of regularly varying distributions.* Journal of the Australian Mathematical Society, vol. 78, no. 03, pages 339–371, 2005.

[Barbe *et al.* 2006] Philippe Barbe, Anne-Laure Fougères and Christian Genest. *On the Tail Behavior of Sums of Dependent Risks.* ASTIN Bulletin, vol. 36, no. 2, pages 361–373, 2006.

[Bernard & Vanduffel 2015] Carole Bernard and Steven Vanduffel. *A new approach to assessing model risk in high dimensions.* Journal of Banking & Finance, vol. 58, pages 166–178, 2015.

[Bernard *et al.* 2013] Carole Bernard, Ludger Rüschendorf and Steven Vanduffel. *Value-at-Risk bounds with variance constraints.* Available at SSRN 2342068, 2013.

[Bingham *et al.* 1989] Nicholas H Bingham, Charles M Goldie and Jef L Teugels. Regular variation, volume 27. Cambridge university press, 1989.

[Cénac *et al.* 2014] Peggy Cénac, Stephane Loisel, Véronique Maume-Deschamps et Clémentine Prieur. *Risk indicators with several lines of business: comparison, asymptotic behavior and applications to optimal reserve allocation.* Annales de l'ISUP, vol. 58, no. 3, pages 3–26, 2014.

[Cline & Samorodnitsky 1994] Daren BH Cline and Gennady Samorodnitsky. *Subexponentiality of the product of independent random variables.* Stochastic Processes and their Applications, vol. 49, no. 1, pages 75–98, 1994.

[Cossette *et al.* 2014] Hélène Cossette, Marie-Pier Côté, Mélina Mailhot and Etienne Marceau. *A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks.* Journal of Multivariate Analysis, vol. 130, pages 1–20, 2014.

References

[Dacorogna *et al.* 2014] Michel Dacorogna, Leila El Bahtouri and Marie Kratz. Explicit diversification benefit for dependent risks. Preprint, 2014.

[Dubey 1970] Satya D Dubey. *Compound gamma, beta and F distributions*. Metrika, vol. 16, no. 1, pages 27–31, 1970.

[Embrechts *et al.* 1997] Paul Embrechts, Claudia Klüppelberg and Thomas Mikosch. Modelling extremal events: for insurance and finance, volume 33. Springer, 1997.

[Embrechts *et al.* 2009] Paul Embrechts, Johanna Neslehová and Mario V Wüthrich. *Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness*. Insurance: Mathematics and Economics, vol. 44, no. 2, pages 164–169, 2009.

[Embrechts *et al.* 2013] Paul Embrechts, Giovanni Puccetti and Ludger Rüschendorf. *Model uncertainty and VaR aggregation*. Journal of Banking & Finance, vol. 37, no. 8, pages 2750–2764, 2013.

[Fougères & Mercadier 2012] Anne-Laure Fougères and Cécile Mercadier. *Risk measures and multivariate extensions of Breiman's theorem*. Journal of Applied Probability, vol. 49, pages 364–384, 2012.

[Geluk & Tang 2009] Jaap Geluk and Qihe Tang. *Asymptotic tail probabilities of sums of dependent subexponential random variables*. Journal of Theoretical Probability, vol. 22, no. 4, pages 871–882, 2009.

[Kortschak & Albrecher 2009] Dominik Kortschak and Hansjörg Albrecher. *Asymptotic results for the sum of dependent non-identically distributed random variables*. Methodology and Computing in Applied Probability, vol. 11, no. 3, pages 279–306, 2009.

[Kratz 2014] Marie Kratz. *Normex, a new method for evaluating the distribution of aggregated heavy tailed risks*. Extremes, vol. 17, no. 4, pages 661–691, 2014.

[Marshall & Olkin 1988] Albert W Marshall and Ingram Olkin. *Families of multivariate distributions*. Journal of the American Statistical Association, vol. 83, no. 403, pages 834–841, 1988.

[Maume-Deschamps *et al.* 2015] Véronique Maume-Deschamps, Didier Rullière and Khalil Saïd. On capital allocation by minimizing multivariate risk indicators. Preprint, 2015.

[Nguyen & Robert 2014] Quang Huy Nguyen and Christian Y Robert. *New efficient estimators in rare event simulation with heavy tails*. Journal of Computational and Applied Mathematics, vol. 261, pages 39–47, 2014.

[Oakes 1989] David Oakes. *Bivariate survival models induced by frailties*. Journal of the American Statistical Association, vol. 84, no. 406, pages 487–493, 1989.

[Resnick 2007] Sidney I Resnick. Heavy-tail phenomena: probabilistic and statistical modeling. Springer, 2007.

[Weinzierl 2000] Stefan Weinzierl. *Introduction to monte carlo methods*. arXiv preprint hep-ph/0006269, 2000.

[Weissman 1978] Ishay Weissman. *Estimation of parameters and large quantiles based on the k largest observations*. Journal of the American Statistical Association, vol. 73, no. 364, pages 812–815, 1978.

[Yeh 2007] Hsiaw-Chan Yeh. *The frailty and the Archimedean structure of the general multivariate Pareto distributions*. Bulletin Institute of Mathematics Acedemia Sinica, vol. 2, no. 3, pages 713–729, 2007.

Chapitre 3

Copula's approximations : application to quantile estimation

Sommaire

Abstract

Estimating high level quantiles of aggregated variables (mainly sums or weighted sums) is crucial in risk management for many application fields such as finance, insurance, environment... This question has been widely treated but new efficient methods are always welcome; especially if they apply in (relatively) high dimension. We propose an estimation procedure based on the *checkerboard copula*. It allows to get good estimations from a (quite) small sample of the multivariate law and a full knowledge of the marginal laws. This situation is realistic for many applications. Estimations may be improved by including in the checkerboard copula some additional information (on the law

of a sub-vector or on extreme probabilities). Our approach is illustrated by numerical examples.

3.1 Introduction

Consider a random vector $\mathbf{X} = (X_1, \ldots, X_d)$ and a measurable function $\Psi : \mathbb{R}^d \to \mathbb{R}^+$, called the aggregation function. In the context of quantitative risk management $\mathbf{X}$ is known as a risk vector and generally represents the profit-losses of a portfolio at a given future date. $\Psi(\mathbf{X})$, the aggregated risk, represents its total future position. The main examples of aggregation functions are: the sum, max, weighted sums or a slightly more complex function that may include stop-loss reinsurance type function on each of the marginals. In this chapter, we will be essentially concerned with $\Psi = \sum$, which is the most commonly studied aggregation function. We are interested here in the estimation of p-quantiles, $0 < p < 1$, of $\Psi(\mathbf{X})$: $Q_p(\Psi(\mathbf{X})) = \inf\{x \in \mathbb{R}, F_\Psi(x) \geq p\}$. To this purpose, we will assume that the distributions $F_1, \ldots, F_d$ of the marginal variables $X_1, \ldots, X_d$ are known and that some information on the dependence between them is given. Usually this information is available via some observations of the joint distribution and also via expert opinion.

In practice, neither the marginals nor the dependence of the risk vector $\mathbf{X}$ will be known. However, in many cases, the information available on the marginal distributions is much more important than the one on the dependence structure. For example, when some observations of the vector $\mathbf{X}$ are available, inferences one can do on the marginal distributions give better results than inferences one can do on the multivariate distribution. Also, samples available for marginal laws may be much larger than those available for the joint distribution. So, even if the assumption of the knowledge of marginal distributions may seem not realistic, there is, however, in practice much more knowledge on the marginal distributions than on the dependence structure of the random vector. These situations arise e.g. for environmental data, in insurance context...

When the marginals are known but the dependence is unknown, the re-arrangement algorithm (introduced in special cases in [Rüschendorf 1983] and [Rüschendorf 1982]) allows to obtain bounds on the distribution of $\Psi(\mathbf{X})$ ([Puccetti & Rüschendorf 2012]) working well for $d \geq 30$. By improving the re-arrangement algorithm, bounds on the VaR are obtained in ([Embrechts *et al.* 2013b]) in high dimen-

sional ($d \geq 1000$) inhomogeneous portfolio. Cases in which some kind of dependence information is available lead to narrower bounds ([Bernard *et al.* 2013, Bernard & Vanduffel 2015]) for the risk measure at hand. Bounds are also derived in ([Cossette *et al.* 2014]) for dependence structures described by different copula models. A general mathematical framework which interpolates between marginal knowledge and full knowledge of the distribution function of $\mathbf{X}$ is considered in ([Embrechts & Puccetti 2010]).

In this chapter, we propose to use the checkerboard copula (introduced in [Mikusinski & Taylor 2010]) to merge the information given by a small sample of the distribution of $\mathbf{X}$ with the known marginal distributions. Moreover, we introduce the checkerboard copula with information on the tail and with information on a sub-vector, to take into account some additional informations which may improve the quantile estimation (see Section 3.3.1). Some simulations are provided in Section 3.4.
We begin (see Section 3.2) with a brief discussion on the admissible multivariate distribution with fixed marginals and aggregated laws: in other words, given marginal laws and a distribution for $\Psi(\mathbf{X})$, what are the possible multivariate distributions for $\mathbf{X}$? Conclusions are provided in Section 3.5.

3.2 The invariant aggregation copula class

Let $\mathbf{F}$ be the distribution function of $\mathbf{X} = (X_1, \ldots, X_d)$. By Sklar's Theorem, there exists a copula distribution C on $[0,1]^d$ such that

$$\mathbf{F}(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

When the marginal random variables of $\mathbf{X}$ are absolutely continuous this copula C is unique. We will assume that the marginals of $\mathbf{X}$ are absolutely continuous. The aggregation function $\Psi : \mathbb{R}^d \to \mathbb{R}^+$ is considered to be measurable and non-decreasing on each variable. Let us denote by $F_\Psi(x) = \mathbb{P}(\Psi(\mathbf{X}) \leq x)$. Of course, the copula of the vector $\mathbf{X}$ determines the distribution of $\Psi(\mathbf{X})$. Nevertheless, the copula specification may be redundant, as for any copula C there may exist an infinite set of copulas $\mathcal{C}_{\Psi,C}$ such that $\Psi(\mathbf{X}^C) \overset{\mathcal{L}}{=} \Psi(\mathbf{X}^{C'})$ for any $C' \in \mathcal{C}_{\Psi,C}$, where $\mathbf{X}^C$ denotes a random vector with same marginals as $\mathbf{X}$ with copula C.

The Fréchet class of the marginal distributions $F_1, \ldots, F_d$, denoted by $\mathfrak{F}_d(F_1, F_2, \ldots, F_d)$, consists of all d-multivariate distributions with $F_1, \ldots, F_d$ as marginals. This class is completely determined by the class of all d-copulas, i.e.:

$$\mathfrak{F}_d(F_1, \ldots, F_d) = \{F : F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right)\}.$$

Moreover, when the marginals are absolutely continuous there is a bijective correspondence between both classes.

In the context of risk aggregation the following more useful class has been introduced in [Bernard *et al.* 2014].

Definition 3.2.1. An aggregate risk S is called an admissible risk of marginal distributions $F_1, \dots, F_d$ if it can be written as $S = \Psi(X_1, \dots, X_d)$ where $X_i \sim F_i$ for $i = 1, \dots, d$. The admissible risk class is defined by the set of admissible risks of given marginal distributions but unknown dependence structure:

$$\mathfrak{S}_d(F_1, \dots, F_d, \Psi) = \{\Psi(X_1, \dots, X_d) : X_i \sim F_i, i = 1, \dots, d\}.$$

Some interesting properties of this class have been presented in [Bernard *et al.* 2014] when Ψ is the sum. Here we present a related class from the copula point of view.

Definition 3.2.2. Let $\mathbf{X}$ be a random vector and Ψ an aggregation function. The class of copulas

$$\mathcal{C}(\mathbf{X}, \Psi) = \{C \in \mathcal{C} : \Psi(\mathbf{X}^C) \overset{\mathcal{L}}{=} \Psi(\mathbf{X})\}$$

is the invariant aggregation copula class of $\mathbf{X}$ and Ψ.

The invariant aggregation copula class is related to the set of admissible risks, in a similar way as the copulas are related to the Fréchet class:

$$F \in \mathfrak{F}_d(F_1, \dots, F_d) \Leftrightarrow \exists C : F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$
$$S \in \mathfrak{S}_d(F_1, \dots, F_d, \Psi) \Leftrightarrow \exists \mathcal{C}(\mathbf{X}, \Psi) : \forall C \in \mathcal{C}(\mathbf{X}, \Psi) \quad S \overset{\mathcal{L}}{=} \Psi(\mathbf{X}^C),$$

In what follows, we present an example and some results that show explicitly that this class is not trivial.

Example of non trivial classes $\mathcal{C}(\mathbf{X}, \Psi)$. For $\Psi = \sum$ and $\Psi = \max$, it is easy to prove that the classes $\mathcal{C}(\mathbf{X}, \Psi)$ are not trivial.

Example 3.2.3. We construct explicitly two different random vectors (X, Y) and (X', Y') such that $X \overset{\mathcal{L}}{=} X'$, $Y \overset{\mathcal{L}}{=} Y'$ and $X + Y \overset{\mathcal{L}}{=} X' + Y'$. Let (X, Y) be any random vector in $\mathbb{R}^2$ with density f. Suppose that for some $\varepsilon > 0$ and some $a < b$ and $c < d$ with $b - a = d - c$ we have that $f(x, y) > \varepsilon$ for any $(x, y) \in [a, b] \times [c, d]$. The equality

$$\int_0^{2\pi} \int_0^{2\pi} \sin(x - y)\, dxdy = 0$$

implies that

$$g(x,y) = f(x,y) + \varepsilon \sin\left(2\pi\frac{x-a}{b-a} - 2\pi\frac{y-c}{d-c}\right) I_{[a,b]\times[c,d]}(x,y)$$

is a density function different to f. Moreover, as for any t the following equations hold,

$$\int_0^t \int_0^{2\pi} \sin(x-y)\,dxdy = 0 \quad \text{and} \quad \int_0^{2\pi} \int_0^t \sin(x-y)\,dxdy = 0.$$

Thus, the marginal densities of f and g are identical. Furthermore, it can also be checked easily that

$$\int_0^{2\pi} \int_0^{2\pi} \sin(x-y) I_{\{0\le x+y\le t\}}(x,y)\,dxdy = 0$$

for any $t > 0$, thus if (X', Y') is a random vector with density g, it satisfies that $X' + Y' \overset{\mathcal{L}}{=} X + Y$.

The example above may be generalized in any dimension.

Proposition 3.2.4. *If $\boldsymbol{X}$ admits a density f then $\mathcal{C}(\boldsymbol{X}, +)$ has infinite elements.*

Proof. Let $h(x_1, \ldots, x_d) = \sum_{1\le i<j\le d} \sin(x_i - x_j)$. Then h satisfies the following properties:

(i) $\int_0^t \int_{[0,2\pi]^{d-1}} h(x_1, \ldots, x_d)\,dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d dx_i = 0$ for all $i = 1, \ldots, d$ and all $t > 0$,

(ii) $\int_{[0,2\pi]^d} h(x_1, \ldots, x_d) I_{\{x_1+\cdots x_d \le t\}}\,dx_1 \cdots dx_d = 0$ for all $t > 0$,

(iii) $h \neq 0$ on $[0, 2\pi]^d$.

Let f be the density of $\mathbf{X}$, and let $\varepsilon > 0$ a point $a = a_1, \ldots a_d$ in $\mathbb{R}^d$ and a positive number δ, such that $f(x_1, \ldots, x_d) > \varepsilon$ for all $a_i < x_i < a_i + \delta$ for $i = 1, \ldots, d$. then

$$g(x_1, \ldots, x_d) = f(x_1, \ldots, x_d) + \varepsilon_0 h\left(2\pi\frac{x_1 - a_1}{\delta}, \ldots, 2\pi\frac{x_d - a_d}{\delta}\right) I_{[a_1, a_1+\delta]\times\cdots\times[a_d, a_d+\delta]}(x_1, \ldots, x_d)$$

is a density function for any $0 < \varepsilon_0 < \varepsilon$. If $\mathbf{X}'$ is a random vector with density g then by the properties of h, $\mathbf{X}' \overset{\mathcal{L}}{\neq} \mathbf{X}$ and $\mathbf{X}' \in \mathcal{C}(\mathbf{X}, +)$. As the same is true for any $0 < \varepsilon_0 < \varepsilon$, we have shown that there are infinite elements in $\mathcal{C}(\mathbf{X}, +)$. □

By definition, any element of the class $\mathcal{C}(\mathbf{X}, \Psi)$ characterizes $\Psi(\mathbf{X})$. The following result shows that in some cases we can always find a symmetrical copula in $\mathcal{C}(\mathbf{X}, \Psi)$.

Proposition 3.2.5. *If $\boldsymbol{X}$ admits a density with identical marginals and Ψ is a symmetrical aggregation function then there exists a symmetrical copula C such that $\Psi(\boldsymbol{X}) \overset{\mathcal{L}}{=} \Psi(\boldsymbol{X}^C)$.*

Proof. Let $f(x_1, \ldots, x_d)$ be the density of $\mathbf{X}$. Define $g(x_1, \ldots, x_d)$ as

$$g(x_1, \ldots, x_d) = \frac{1}{d!} \sum_{\sigma \in S_d} f(x_{\sigma(1)}, \ldots, x_{\sigma(d)}),$$

where S_d is the set of all the permutations of $\{1 \ldots, d\}$. Let $\mathbf{X}'$ be a random vector with density g. Then it is easy to check that the marginals of $\mathbf{X}'$ are distributed as the marginals of $\mathbf{X}$. It follows equally, from the symmetry of Ψ, that $\Psi(\mathbf{X}) \overset{\mathcal{L}}{=} \Psi(\mathbf{X}')$. As the density of $\mathbf{X}'$ is completely symmetrical so is its copula. □

Remark. In the case of d dimensional Archimedean copulas, it is known that the copula C is uniquely determined by its diagonal δ, $\delta(t) = C(t, \ldots, t)$ if $\delta'(1-) = d$ (see [Frank 1996, Sungur & Yang 1996] in dimension 2 and [Erdely *et al.* 2014] in higher dimension). This means that if $\Psi = \max$ or $\Psi = \min$, given a fixed common law for $X_1, \ldots, X_d$ and a fixed law for $\Psi(\mathbf{X})$, then there is only one Archimedean copula which leaves $\Psi(\mathbf{X})$ and the marginal laws invariant. Nevertheless, using constructions in [Nelsen *et al.* 2008], infinitely many copulas with a fixed diagonal may be constructed.

Below we provide a construction of infinitely many laws of random vectors with a fixed distribution for their max and fixed marginal distributions (remark that if the marginal laws are not the same, then the law of the max is not determined by the diagonal of the copula).

Proposition 3.2.6. *Assume that $\boldsymbol{X}$ is absolutely continuous with density f such that $\inf_K f > 0$ with $K = \prod_{i=1}^{d} [a_i, b_i]$. If K is symmetric with respect to the diagonal, then there exists a density function φ such that $f \equiv \varphi$ outside K, $f \neq \varphi$ on K and the random vector $\widetilde{\boldsymbol{X}}$ whose density function is φ is such that*

- *for $i = 1, \ldots, d$, $\widetilde{X}_i \overset{\mathcal{L}}{=} X_i$,*
- *$\max(\boldsymbol{X}) \overset{\mathcal{L}}{=} \max(\widetilde{\boldsymbol{X}})$.*

Proof. We sketch the proof in dimension 2. Let $f > \varepsilon$ on K and $\varphi = f + \varepsilon\gamma$ where γ has its support in K as shown in Figure 3.1. It is easy to verify that

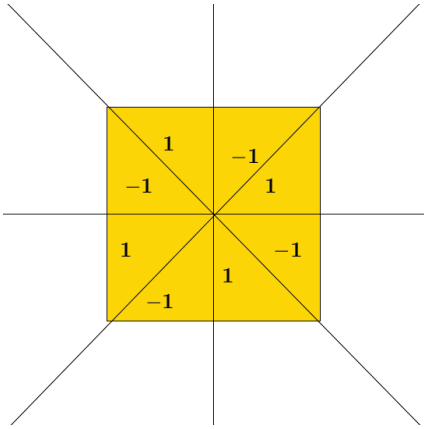

Figure 3.1 – Values of γ in K

the random vector $\widetilde{\mathbf{X}}$ whose density is φ has the same marginal laws as $\mathbf{X}$ and $\max(\widetilde{\mathbf{X}}) \overset{\mathcal{L}}{=} \max(\mathbf{X})$. □

Even if the example above seems trivial it shows that the full knowledge of the copula distribution is unnecessary when studying an aggregation: there is some redundant information.

3.3 Non-parametric estimation of the aggregation distribution when the marginals are known

We have seen in the above section that the exact copula estimation can be considered as a redundant exercise when estimating the distribution of an aggregation of $\mathbf{X}$. The information given by the copula of $\mathbf{X}$, for the study of $\Psi(\mathbf{X})$ is the same as any copula from the class $\mathcal{C}(\mathbf{X}, \Psi)$. This may be seen as a justification of the fact that when the marginals are known, there is some flexibility in the copula estimation in order to estimate the aggregated distribution.

In this section we propose a non-parametric estimator of the distribution of $\Psi(\mathbf{X})$ when marginals $F_1, \ldots, F_d$ are known and an independent and identically distributed (i.i.d.) sample $\mathbf{X}^1, \ldots, \mathbf{X}^n$ is given. The sample size n is quite small. The estimation of F_Ψ will allow us to obtain an estimation of the p-quantile $Q_p(\Psi(\mathbf{X}))$. This estimation will be compared to the one obtained from the empirical cumulative distribution function

$$\widehat{F_\Psi}(t) = \frac{1}{n} \sum_{i=1}^{n} I_{\{\Psi(\mathbf{X}^i) \leq t\}}.$$

Before illustrating the estimation procedure (Section 3.3.3), we need to define three kinds of checkerboard copulas (Section 3.3.1) as well as an empirical version of each of them (Section 3.3.2).

3.3.1 Checkerboard copulas

As in the above sections, let F denote the cumulative distribution function (c.d.f.) of $\mathbf{X}$, C its copula function and F_i the c.d.f of X_i, $i = 1, \ldots, d$.
Let μ_C be the probability measure associated to C, i.e such that:

$$\mu_C \left(\prod_{i=1}^{d} [0, u_i] \right) = C(u_1, \ldots, u_d)$$

for any $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$.
By a μ-decomposition of a set $A \subset \mathbb{R}^d$ we mean a finite family of measurable sets $\{A_i \subset A\}$ such that

1. $\mu(A_i \cap A_j) = 0$ whenever $i \neq j$
2. $\sum_i \mu(A_i) = \mu(A)$.

We now give the definitions of the checkerboard approximations and checkerboard copulas which has been introduced by [Li *et al.* 1998].

Definition 3.3.1. A measure μ^* is a checkerboard approximation for a copula C if there exists a λ-decomposition $\mathcal{A} = \{(a_i, b_i)\}$ of I^d, the d-dimensional unit cube, made out of d-intervals such that for all i,

1. μ^* is uniform on (a_i, b_i);
2. $\mu^*(A) = \mu_C(A)$ for any $A \in \mathcal{A}$,

where λ is the d-dimensional Lebesgue measure.

For $m \in \mathbb{N}$, let us consider the regular λ-decomposition of the unite cube $[0,1]^d$ denoted as $\mathcal{I}_m$ and consisting of m^d d-cubes with side length $1/m$:

$$I_{i,m} = \prod_{j=1}^{d} \left[\frac{i_j - 1}{m}, \frac{i_j}{m} \right], \ i = (i_1, \ldots, i_d), \ i_j \in \{1, \ldots, m\}.$$

μ_m^* is the checkerboard approximation associated to the regular decomposition $\mathcal{I}_m$.

We shall denote by C_m^* the checkerboard copula associated to the measure μ_m^*. The definition of the checkerboard copula may then be rewritten as:

$$C_m^*(\mathbf{x}) = \sum_i m^d \mu_C(I_{i,m}) \lambda([\mathbf{0}, \mathbf{x}] \cap I_{i,m})$$

where $[\mathbf{0},\mathbf{x}] = \prod_{i=1}^{d}[0,x_i]$, for $\mathbf{x} = (x_1,\ldots,x_d) \in [0,1]^d$. In [Mikusinski & Taylor 2010], it is proved that C_m^* is a copula and that it approximates C. The following proposition provides a more precise bound on the approximation of C by C_m^*, by a factor of 2, that the one presented in [Li *et al.* 1998], pag 613.

Proposition 3.3.2. *Let C_m^* be the checkerboard copula defined above. We have:*

$$\sup_{\boldsymbol{x}\in[0,1]^d} |C_m^*(\boldsymbol{x}) - C(\boldsymbol{x})| \leq \frac{d}{2m}.$$

Proof. This is clear that $C_m^*(\mathbf{x}) = C(\mathbf{x})$ for any $\mathbf{x} \in [0,1]^d$ with $x_k = \frac{i}{m}$, $i \in \{1,\ldots,m\}^d$, $k = 1,\ldots,d$. For $i \in \{1,\ldots,m\}^d$ and $k \in \{1,\ldots,d\}$, we denote by B_i^{k+} and B_i^{k-} the (half)-strips:

$$B_i^{k+} = \left\{\mathbf{x} \in [0,1]^d,\ \frac{i_k}{m} - \frac{1}{2m} < x_k \leq \frac{i_k}{m}\right\} \text{ and}$$

$$B_i^{k-} = \left\{\mathbf{x} \in [0,1]^d,\ \frac{i_k - 1}{m} < x_k \leq \frac{i_k}{m} - \frac{1}{2m}\right\}.$$

If $\mathbf{x} \in I_{i,m}$ then,

$$\begin{aligned}
|C_m^*(\mathbf{x}) - C(\mathbf{x})| &\leq \sum_{k=1}^{d} |\mu_m^*(B_i^{k-}) - \lambda(B_i^{k-})| \mathbf{1}_{B_i^{k-}}(\mathbf{x}) + \\
&\quad \sum_{k=1}^{d} |\mu_m^*(B_i^{k+}) - \lambda(B_i^{k+})| \mathbf{1}_{B_i^{k+}}(\mathbf{x}) \\
&\leq \sum_{k=1}^{d} \min(\mu_m^*(B_i^{k-}), \lambda(B_i^{k-})) \mathbf{1}_{B_i^{k-}}(\mathbf{x}) + \\
&\quad \sum_{k=1}^{d} \min(\mu_m^*(B_i^{k+}), \lambda(B_i^{k+})) \mathbf{1}_{B_i^{k+}}(\mathbf{x}) \\
&= \frac{d}{2m}
\end{aligned}$$

since μ_m^* and λ are both associated to a copula,

$$\mu_m^*(B_i^{k-}) = \lambda(B_i^{k-}) = \mu_m^*(B_i^{k+}) = \lambda(B_i^{k+}) = \frac{1}{2m}.$$

The announced result follows. □

We also define two kinds of *checkerboard copula with additional information*. First of all, we consider the case where the distribution of a sub-vector $\mathbf{X}^J = (X_i)_{i\in J}$, $J \subset \{1,\dots,d\}$, is known, $|J| = k < d$. Denote C^J the copula of $\mathbf{X}^J$. Let μ^J be the probability measure on $[0,1]^k$ associated to C^J. For $i = (i_1,\dots,i_d)$, let $\mathbf{x} = (x_1,\dots,x_d) \in [0,1]^d$, $\mathbf{x}^J = (x_j)_{j\in J}$, $\mathbf{x}^{-J} = (x_j)_{j\notin J}$ and

$$I_{i,m}^J = \left\{ \mathbf{x} \in [0,1]^d \ / \ x_j \in \left[\frac{i_j - 1}{m}, \frac{i_j}{m} \right], \ j \in J \right\},$$

$$I_{i,m}^{-J} = \left\{ \mathbf{x} \in [0,1]^d \ / \ x_j \in \left[\frac{i_j - 1}{m}, \frac{i_j}{m} \right], \ j \notin J \right\}.$$

The checkerboard copula with information on $\mathbf{X}^J$ is defined below.

Definition 3.3.3. Consider the probability measure on $[0,1]^d$ defined by

$$\mu_m^J([0,\mathbf{x}]) = \sum_{i\subset\{1,\dots,d\}} \frac{m^{d-k}}{\mu^J(I_{i,m}^J)} \mu_C(I_{i,m}) \lambda([\mathbf{0},\mathbf{x}^{-J}] \cap I_{i,m}^{-J}) \mu^J([0,\mathbf{x}^J] \cap I_{i,m}^J).$$

Let C_m^J, the checkerboard copula with additional information on $\mathbf{X}^J$, be defined by $C_m^J(\mathbf{x}) = \mu_m^J([0,\mathbf{x}])$.

Proposition 3.3.4. *C_m^J is a copula that approximates C:*

$$\sup_{x\in[0,1]^d} |C_m^J(\boldsymbol{x}) - C(\boldsymbol{x})| \leq \frac{d}{2m}.$$

If $\boldsymbol{X}^J$ and $\boldsymbol{X}^{-J}$ are independent then,

$$\sup_{x\in[0,1]^d} |C_m^J(\boldsymbol{x}) - C(\boldsymbol{x})| \leq \frac{d-k}{2m}.$$

Proof. The definition of C_m^J insures that it is a cumulative distribution function on $[0,1]^d$. The fact that C_m^J is a copula then follows from an easy computation to get that $C_m^J(\mathbf{x}) = x_k$ whenever $\mathbf{x} = (x_j)_{j=1,\dots,d}$, with $x_j = 1$ for $j \neq k$.
The rest of the proof is done as in that of Proposition 3.3.2. □

We may also add information on the tail and so define the following particular checkerboard copula.

Definition 3.3.5. For $p \in]0,1[$, let $E_p = \left(\prod_{i=1}^d [0,p]^d\right)^c$ and $\mathcal{E}_p$ the λ-decomposition of E_p consisting of the hyper rectangles $[a_1,b_1] \times \cdots \times [a_d,b_d]$ where $[a_i,b_i] = [0,p]$ or $[a_i,b_i] = [p,1]$ for all $i = 1,\dots d$ with at least one of $[a_i,b_i] = [p,1]$. We assume that $\mu_C(E)$ is known for each $E \in \mathcal{E}_p$. Consider the

λ-decomposition of the d-cube $[0,p]^d$ given by $\mathcal{J}_m$ consisting of the elements $J_{i,m} = p \cdot I^d_{m,i}$ for d-tuple $i = (i_1, \ldots, i_d)$ in $\{0, 1/m, \ldots, (m-1)/m\}^d$. Define $C_m^{\mathcal{E}_p}$ as the checkerboard copula associated to the λ-decomposition of the unit d-cube $\mathcal{J}_m \cup \mathcal{E}_p$, that is

$$C_m^{\mathcal{E}_p}(\mathbf{x}) = (1 - \mu_C(E_p))C_m^*(\mathbf{x}_p) + \sum_{E \in \mathcal{E}_p} \frac{\mu_C(E)}{\lambda(E)} \lambda([\mathbf{0}, \mathbf{x}] \cap E),$$

where $\mathbf{x}_p = (\min\{x_1/p, 1\}, \ldots, \min\{x_d/p, 1\})$.

This is the checkerboard copula with extra information on the tail. The idea of this copula is to combine two types of information, one on the "center" of the distribution $[0,p]^d$, provided by a data sample (see below the empirical version of this checkerboard copula in Definition 3.3.7) and other on the "tail" of the distribution E_p provided in some cases with the help of expert judgment.

In what follows, we will define an empirical version of the checkerboard copulas defined above, by using the empirical copula.

3.3.2 Empirical checkerboard copulas

The empirical copula, introduced by [Deheuvels 1979], may be used to estimate non parametrically the copula.

Definition 3.3.6. Let $\mathbf{X}^1, \ldots, \mathbf{X}^n$ be n independent copies of $\mathbf{X}$. Each of them writes $\mathbf{X}^j = (X_1^j, \ldots, X_d^j)$. Let $R_i^1, \ldots, R_i^n$, $i = 1, \ldots, d$ be their marginals ranks, i.e.,

$$R_i^j = \sum_{k=1}^{n} 1\{X_i^{(j)} \geq X_i^{(k)}\},\ i = 1, \ldots, d,\ j = 1, \ldots, n$$

where $X_i^{(1)} < \cdots < X_i^{(n)}$ are the order statistics associated to the ith coordinate sample $(X_i^1, \ldots X_i^n)$. The empirical copula $\widehat{C}_n$ of $\mathbf{X}^1, \ldots \mathbf{X}^n$ is defined as

$$\widehat{C}_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{k=1}^{n} 1\left\{\frac{1}{n} R_1^k \leq u_1, \ldots, \frac{1}{n} R_d^k \leq u_d\right\}.$$

It is well known (see [Fermanian *et al.* 2004] e.g.) that the empirical copula may be used to estimate C. Nevertheless, it is not a copula as its marginal laws are discrete. We shall use the empirical copula $\widehat{C}_n$ and the empirical probability measure $\widehat{\mu}$ associated to $\widehat{C}_n$, to define an empirical version of the checkerboard copulas defined above. We introduce these definitions below.

Definition 3.3.7. Let $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ be n independent copies of $\mathbf{X}$.

— The empirical checkerboard copula (ECBC) $\widehat{C}_m^*$ is defined by

$$\widehat{C}_m^*(\mathbf{x}) = \sum_i m^d \widehat{\mu}(I_{i,m}) \lambda([\mathbf{0}, \mathbf{x}] \cap I_{i,m}).$$

— The ECBC with information on a sub-vector $\mathbf{X}^J$ is defined by

$$\widehat{C}_m^J(\mathbf{x}) = \sum_{i \subset \{1,\dots,d\}} \frac{m^{d-k}}{\mu^J(I_{i,m}^J)} \widehat{\mu}(I_{i,m}) \lambda([\mathbf{0}, \mathbf{x}^{-J}] \cap I_{i,m}^{-J}) \mu^J([\mathbf{0}, \mathbf{x}^J] \cap I_{i,m}^J).$$

— The ECBC with information on the tail is defined by:

$$\widehat{C}_m^{\mathcal{E}_p}(\mathbf{x}) = (1 - \mu_C(E_p)) \widehat{C}_m^*(\mathbf{x}_p) + \sum_{E \in \mathcal{E}_p} \frac{\mu_C(E)}{\lambda(E)} \lambda([\mathbf{0}, \mathbf{x}] \cap E),$$

where $\mathbf{x}_p = (\min\{x_1/p, 1\}, \dots, \min\{x_d/p, 1\})$.

In [Li *et al.* 1998] it has been shown that when C is a copula its checkerboard approximation on a regular partition is always a copula. This is not true in general for the empirical copulas which are not copulas. In the next proposition we show that $\widehat{C}_m^*$, defined above, is a copula whenever the integer m from the length size of the partition $\mathcal{I}_m$ divides n, the sample size.

Proposition 3.3.8. *Let m, n be integers such that m divides n. Then, the empirical chekerboard copula $\widehat{C}_m^*$ defined on the regular partition $\mathcal{I}_m$ and based on an i.i.d sample of size n is a copula.*

Proof. Suppose that $m \leq n$ and that m divides n. By definition the empirical checkerboard copula is a distribution function, we should simply check that the marginals are uniform. Without losing generality we show only that the projection on the first coordinate of the measure induced by $\widehat{C}_m^*$ is uniform, or equivalently that $\widehat{C}_m^*(\mathbf{x}) = x_1$ for any $\mathbf{x} \in [0,1]^d$ with $x_j = 1$ for $j \neq 1$.
For $\ell \in \{1, \dots, m\}$, consider the strip B_ℓ^1:

$$B_\ell^1 = \left\{ \mathbf{x} \in [0,1]^d, \ \frac{\ell - 1}{m} < x_1 \leq \frac{\ell}{m} \right\} = \left] \frac{\ell - 1}{m}, \frac{\ell}{m} \right] \times [0,1]^{d-1}.$$

The empirical copula is concentrated on n points of $[0,1]^d$ whose coordinates are of the form $\frac{j}{n}$, $j = 1, \dots, n$. Moreover, there is exactly one mass on each strip B_j^1, $j = 1, \dots, n$. So that if $k = n/m$, then the number of masses of $\widehat{C}_n$ on each strip B_ℓ^1, $\ell = 1, \dots, m$ is exactly k, which means that $\widehat{\mu}(B_\ell^1) = \frac{k}{n} = \frac{1}{m}$.

Let $\mathbf{x} = (x_1, 1, \ldots, 1)$, $x_1 \in [0,1]$, $x \in B_\ell^1$ with $\frac{\ell-1}{m} < x_1 \leq \frac{\ell}{m}$.

$$\begin{aligned}
\widehat{C}_m^*(\mathbf{x}) &= \sum_{i \in \{1,\ldots,m\}^d} m^d \widehat{\mu}(I_{i,m}) \lambda([\mathbf{0}, \mathbf{x}] \cap I_{i,m}) \\
&= \sum_{j<\ell} \widehat{\mu}(B_j^1) + \sum_{I_{i,m} \subset B_\ell^1} m^d \widehat{\mu}(I_{i,m}) \frac{(x_1 - \frac{\ell-1}{m})}{m^{d-1}} \\
&= \sum_{j<\ell} \widehat{\mu}(B_j^1) + m \left(x_1 - \frac{\ell-1}{m} \right) \widehat{\mu}(B_\ell^1) \\
&= \frac{\ell-1}{m} + \left(x_1 - \frac{\ell-1}{m} \right) = x_1,
\end{aligned}$$

which concludes the proof. □

Remarks. (i) The same calculation shows that the empirical checkerboard copulas with additional information $\widehat{C}_m^J$ and $\widehat{C}_m^{\mathcal{E}}$ are copulas provided that m divides n.

(ii) In dimension 2, and when $n = m$, this result follows from Theorem 2.2 of [Li *et al.* 1998] and the fact that $(n\Delta_{i,j}(\widehat{C}_n^*))_{i,j=1}^n$ is an $n \times n$ doubly stochastic matrix, where

$$\Delta_{i,j}(C) = C\left(\frac{i}{n}, \frac{j}{n}\right) - C\left(\frac{i-1}{n}, \frac{j}{n}\right) - C\left(\frac{i}{n}, \frac{j-1}{n}\right) + C\left(\frac{i-1}{n}, \frac{j-1}{n}\right).$$

Figure 3.2 shows simulations of different empirical checkerboard copulas with and without information on the tail and two different decompositions. Simulating a sample of size N from $\widehat{C}_m^*$ given a sample of size n of $\mathbf{X}$ is very easy because it suffices to

1. get the sample rank,
2. determine the number $n_{i,m}$ of points of the rank sample in each $I_{i,m}$,
3. choose one element of $\mathcal{I}_m$ such that the probability to chose $I_{i,m}$ is $\frac{n_{i,m}}{n}$,
4. take a point in the chosen $I_{i,m}$ at step (3) uniformly,

repeat steps (3) and (4) N times.

3.3.3 Estimation procedure

In what follows $\widehat{C}_m^{cb}$ denotes any of the three ECBCs defined above. We propose the following estimation procedure.
Assume the marginal laws are known and a (quite small) sample of size n of $\mathbf{X}$ is available.

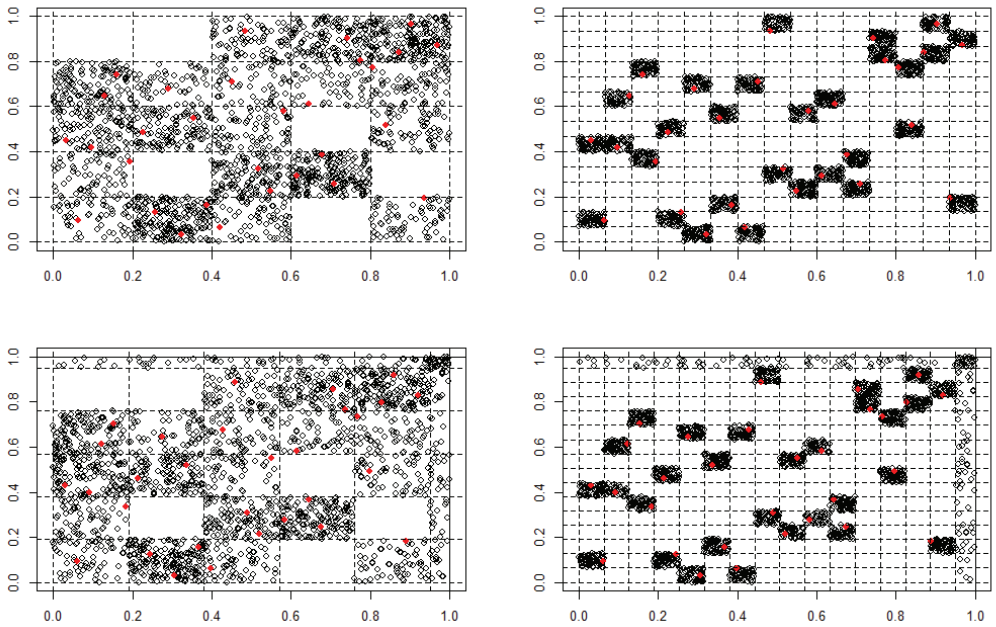

Figure 3.2 – Simulation of four checkerboard empirical copulas based on the same sample of size $n = 30$. At the top there is no information on the tail whereas on the bottom information on $\mathcal{E}_p$, with $p = 0.95$ is assumed. On the left the regular decomposition is $\mathcal{I}_5$, on the right $\mathcal{I}_{15}$. The red points are the sample rank points.

1. Estimate μ by $\widehat{\mu}$ using the empirical copula.
2. Obtain the ECBC $\widehat{C}_m^{cb}$ (depending if some additional information is known).
3. Simulate a sample of size N from the copula $\widehat{C}_m^{cb}$ for N large enough:
$$(u_1^{(1)}, \ldots, u_d^{(1)}), \ldots, (u_1^{(N)}, \ldots, u_d^{(N)})$$
4. Get a sample of $\Psi(\mathbf{X})$ using the marginals to transform the above sample:
$$\Psi\left(F_1^{\leftarrow}(u_1^{(1)}), \ldots, F_d^{\leftarrow}(u_d^{(1)})\right), \ldots, \Psi\left(F_1^{\leftarrow}(u_1^{(N)}), \ldots, F_d^{\leftarrow}(u_d^{(N)})\right)$$
5. Estimate the distribution function F_Ψ of $\Psi(\mathbf{X})$ empirically using the above sample. It will be denoted by $\widehat{F}(\Psi)$.

Below, we state a convergence result which may have some interest from a theoretical point of view although it will not be used since we are working with (relatively) small samples of size n.

Proposition 3.3.9. *For some $A > 0$, let $A\sqrt{n} \leq m \leq n$. Assume that $\Psi(\boldsymbol{X})$ is absolutely continuous and C has continuous partial derivatives. Then,*

$$\|F_\Psi - \widehat{F}(\Psi)\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Use the convergence result by ([Fermanian *et al.* 2004]). □

3.4 Numerical Application

In this section we use the estimator of the distribution function F_Ψ to estimate the quantiles $Q_p(S)$ for $S = X_1 + \cdots + X_d$ at different confidence levels $0 < p < 1$. We will consider the Pareto-Clayton model, defined in Section 3.4.1, because, in that case, the exact value of $Q_p(S)$ can be calculated, so that we may compare our simulation results to the exact one.

3.4.1 The Pareto-Clayton model

We consider $\mathbf{X} = (X_1, \ldots, X_d)$ such that:

$$\mathbb{P}(X_1 > x_1, \ldots, X_d > x_d \,|\, \Lambda = \lambda) = \prod_{i=1}^{d} e^{-\lambda x_i},$$

that is, conditionally to the value of Λ the marginals of $\mathbf{X}$ are independent and exponentially distributed.
If Λ is Gamma distributed, then the X_i's are Pareto distributed with dependence given by a survival Clayton copula.
If Λ is Levy distributed, then the X_i's are Weibull distributed with a Gumbel survival copula.
These models have been introduced in [Marshall & Olkin 1988, Oakes 1989]. In the context of multivariate risk theory, they have been used e.g. in [Maume-Deschamps *et al.* 2015] and [Dacorogna *et al.* 2014].
In what follows, we consider that $\Lambda \rightsquigarrow \Gamma(\alpha, \beta)$, so that the X_i's are Pareto (α, β) distributed and the dependence structure is described by a survival Clayton copula with parameter $1/\alpha$. In [Dubey 1970], it is shown that, in this case, S follows the so-called Beta prime distribution:

$$F_S(x) = F_\beta\left(\frac{x}{1+x}\right)$$

where F_β is the distribution function of the Beta$(d\beta, \alpha)$ distribution. The inverse of F_S (or quantile function of S) can also be expressed in terms of the

inverse of the Beta distribution

$$Q_p(S) = F_S^{\leftarrow}(p) = \frac{F_\beta^{\leftarrow}(p)}{1 - F_\beta^{\leftarrow}(p)}.$$

From these results, we may compute $Q_p(S)$.

3.4.2 Simulation study in dimension 2 (with and without information on the tail).

We consider a Pareto-Clayton model in dimension 2, with $\beta = 1$ and $\alpha = 2$. The multivariate sample is of size $n = 30$ and for each presented method we performed $N = 1000$ estimations of the p-quantile at different confidence levels. Table 3.1 presents the mean and the root mean squared error of the $N = 1000$ estimations. The estimations were calculated using the ECBC with and without information on the tail on different λ-regular decompositions $\mathcal{I}_m$ for $m = 6, 15, 30$. The information on the tail is introduced on $\mathcal{E}_p$, for $p = 0.95, 0.99$, by giving to each E in $\mathcal{E}_p$ the measure $\mu_C(E)$ where C is the survival Clayton copula with parameter $1/2$. For comparison a direct estimation from the empirical distribution of S is given. All the methods we proposed perform much better than the empirical estimation based on the multivariate sample alone. The estimations based on the ECBC with λ-decomposition $\mathcal{I}_m$, perform better when $m = 6$ and $m = 15$ than when $m = 30$. ECBC with $m = 6$ performs slightly better than ECBC with $m = 15$ for the estimation of the quantiles with confidence levels lower than 99.5% and slightly worst on the higher levels. When the information on the tail is introduced on $\mathcal{E}_p$ with $p = 0.95$ the estimation on the quantile with confidence level 90%, 95% and 99% is considerably improved. When it is introduced on $\mathcal{E}_p$ with $p = 0.95$ the estimations improve on the higher confidence levels 99.5% and 99.9%.

3.4.3 Simulation study introducing information on a sub-vector

In order to assess the gain that the knowledge of the information on a sub-vector may give to the estimation, we performed here the following simulation study. Let $\mathbf{X} = (X_1, X_2, X_3)$ be the model where $X_1 = X_2 = Y/2$, and $X_3 \sim Y$ where Y is Pareto distributed with $\alpha = 2$. We assume that (Y, X_3) is a Pareto-Clayton model as in section 3.4.2. That is, X_1 and X_2 are comonotonic (or fully dependent) and the dependence between X_1 and X_3 is given by a survival Clayton of parameter $1/2$. Clearly the distribution of the sum

Quantile	80%	90%	95%	99%	99.5%	99.9%
Exact value	2.5	4.1	6.4	16.0	23.2	53.4
Empirical	2.5	4.0	6.1	12.2	13.2	14.0
	(26%)	(31%)	(39%)	(72%)	(70%)	(78%)
ECBC (m=6)						
No tail information	2.6	4.4	6.6	14.8	20.8	45.7
	(9%)	(8%)	(6%)	(8%)	(11%)	(15%)
Information on $\mathcal{E}_p$	2.6	4.4	6.4	14.2	22.7	49.5
p=0.99	(9%)	(8%)	(5%)	(11%)	(3%)	(8%)
Information on $\mathcal{E}_p$	2.7	4.1	6.1	15.6	21.8	46.8
p=0.95	(10%)	(5%)	(4%)	(3%)	(6%)	(13%)
ECBC (m=15)						
No tail information	2.5	4.2	6.8	15.5	21.5	46.4
	(12%)	(13%)	(11%)	(9%)	(10%)	(14%)
Information on $\mathcal{E}_p$	2.5	4.3	6.8	14.3	22.7	49.5
p=0.99	(12%)	(12%)	(12%)	(11%)	(3%)	(8%)
Information on $\mathcal{E}_p$	2.6	4.3	6.2	15.6	21.8	46.8
p=0.95	(11%)	(10%)	(4%)	(3%)	(6%)	(13%)
ECBC (m=30)						
No tail information	2.5	4.2	6.6	15.8	22.0	47.0
	(13%)	(15%)	(17%)	(13%)	(12%)	(14%)
Information on $\mathcal{E}_p$	2.5	4.2	6.7	14.3	22.7	49.5
p=0.99	(13%)	(16%)	(16%)	(11%)	(3%)	(8%)
Information on $\mathcal{E}_p$	2.6	4.4	6.2	15.6	21.8	46.8
p=0.95	(13%)	(11%)	(4%)	(3%)	(6%)	(13%)

Table 3.1 – The mean and the RMSE in % of the exact value for 1000 estimations of the Quantile for a Pareto-Clayton sum in dimension 2.

$S = X_1 + X_2 + X_3$ is equal to the distribution of the sum of the Pareto-Clayton model in dimension 2, with parameters $\alpha = 2$ and $\beta = 1$ and thus the exact value of the quantiles can be easily computed. We compare the results on the quantiles estimation using the ECBC method without and with information on the sub-vector (X_1, X_2) and $\lambda-$decompositions $\mathcal{I}_m$ for $m = 6, 15, 30$. As before the multivariate sample is of size $n = 30$ and for each method we performed $N = 1000$ estimations of the quantile at different confidence levels. The results are presented in Table 3.2.

It can be noticed that the RSME of the quantile estimation is lower when the information on (X_1, X_2) is introduced in the ECBC of dimension 3 and the gap is more important on higher confidence levels.

Quantile	80%	90%	95%	99%	99.5%	99.9%
Exact Value	2.5	4.1	6.4	16.0	23.2	53.4
ECBC (m=6)						
No information	2.7	4.6	6.6	14.0	19.1	40.7
	(13%)	(13%)	(7%)	(13%)	(18%)	(24%)
Information on	2.6	4.4	6.6	14.8	20.8	45.7
(X_1, X_2)	(9%)	(8%)	(6%)	(8%)	(11%)	(15%)
ECBC (m=10)						
No information	2.5	4.6	7.0	14.5	19.8	41.3
	(12%)	(13%)	(12%)	(11%)	(15%)	(23%)
Information on	2.5	4.3	6.7	15.2	21.2	46.1
(X_1, X_2)	(11%)	(9%)	(9%)	(8%)	(10%)	(15%)
ECBC (m=30)						
No information	2.5	4.2	6.8	15.9	21.4	43.3
	(14%)	(16%)	(19%)	(14%)	(14%)	(21%)
Information on	2.5	4.2	6.6	15.8	21.9	47.1
(X_1, X_2)	(13%)	(16%)	(17%)	(13%)	(13%)	(14%)

Table 3.2 – The mean and the RMSE in % of the exact value for 1000 estimations of the Quantiles in dimension 3, with or without using the knowledge of the comonotonic dependence between X_1 and X_2.

Remark. Simulating with respect to the empirical checkerboard copula with information on a sub-vector may be a difficult task because one has to simulate with respect to a given copula conditionally to belonging to a given set $I_{i,m}$. In the case of a comonotonic sub-vector, this becomes trivial because we only need to simulate one coordinate uniformly.

Simulation results with the same kind of model in dimension 6 are presented in Table 3.3. We assumed $\mathbf{X} = (X_1, \ldots, X_6)$ with $X_1 = X_2 = Y/2$ and X_3, X_4, X_5 and X_6 distributed as Y, a Pareto r.v. with parameter $\alpha = 2$. The copula of $\mathbf{X}$ is assumed to be a survival Clayton of parameter $1/2$. As above, the size of the multivariate sample is $n = 30$ and for each method we performed $N = 1000$ estimations of the quantile at different confidence levels. Again, by introducing the information on the sub-vector (X_1, X_2) we get a smaller RMSE than in the case where no information is added. On the other hand, we also remark that by increasing the dimension (from $d = 3$ to $d = 6$) we get higher RMSE, for the same sample size, which is an expected behavior.

Quantile	80%	90%	95%	99%	99.5%	99.9%
Exact Value	6.1	9.8	14.9	36.4	52.4	120.1
ECBC (m=6)						
No information	7.2	11.0	15.0	28.2	37.3	74.3
	(19%)	(14%)	(7%)	(23%)	(29%)	(38%)
Information on	7.1	10.9	15.0	28.9	38.4	77.5
(X_1, X_2)	(17%)	(13%)	(7%)	(21%)	(27%)	(36%)
ECBC (m=10)						
No information	6.8	11.3	16.1	30.0	39.2	76.3
	(16%)	(18%)	(13%)	(19%)	(26%)	(37%)
Information on	6.7	11.1	16.0	30.6	40.3	79.7
(X_1, X_2)	(15%)	(16%)	(12%)	(17%)	(24%)	(34%)
ECBC (m=30)						
No information	6.3	10.4	16.3	33.7	43.4	81.3
	(15%)	(19%)	(20%)	(18%)	(22%)	(33%)
Information on	6.3	10.4	16.2	34.0	44.1	84.2
(X_1, X_2)	(15%)	(18%)	(19%)	(18%)	(20%)	(31%)

Table 3.3 – The mean and the RMSE in % of the exact value for 1000 estimations of the Quantiles in dimension 6.

3.4.4 A simulation in higher dimension

We conclude this simulation section with a simulation in dimension 10. We consider a Pareto-Clayton model with $\beta = 1$, $\alpha = \frac{1}{2}$. The multivariate sample size is $n = 75$, then $n = 150$. We perform $N = 1000$ runs for the checkerboard method without any kind of information. In this case, we have simply taken $m = n$. Results are presented in Table 3.4. It is notable that the checkerboard method performs well even in dimension 10. Let us emphasize that $n = 150$ in dimension 10 is not much.

3.5 Conclusion

In this chapter, we have constructed empirical checkerboard copulas with and without additional information on the joint law. We have used them to get efficient estimations of the quantiles of the sum when using a (relatively) small sample of the joint law and the knowledge of the marginal laws. We have also proved (in the case of max and $\sum$) that there exist infinitely many copulas for given marginal laws and given aggregation law. This theoretical result indicates that when the marginal laws are known, if we are only interested in the aggregated law, the full knowledge of the joint law is not necessary.

Quantile	80%	90%	95%	99%	99.5%	99.9%
Exact value	12.2	19.2	29	70.1	100.8	230.5
Empirical	12.6	20	29.9	62.2	75.8	86.7
$n = 75$	(12%)	(15%)	(19%)	(41%)	(67%)	(181%)
ECBC	12.5	20.1	31.2	74.8	92.4	152.6
$n = 75, m = 75$	(10%)	(13%)	(16%)	(21%)	(22%)	(53%)
Empirical	12.4	19.6	30.3	67.3	89.9	121
$n = 150$	(9%)	(11%)	(14%)	(27%)	(40%)	(108%)
ECBC	12.4	19.6	29.8	75.4	107.6	173.9
$n = 150, m = 150$	(7%)	(9%)	(12%)	(18%)	(22%)	(38%)

Table 3.4 – Mean and RMSE in % of the exact value for the Pareto-Clayton sum in dimension 10.

We are aware that many theoretical and practical questions have to be studied further, among which:

— the optimal choice of m with respect to the sample size n,

— the quantification of the impact of plugging additional information in the empirical checkerboard copula,

— the number N of simulations necessary to reach a certain precision level,

— developing efficient algorithms to simulate with respect to the empirical checkerboard copula with information on a sub-vector, for other copulas than the comonotonic one,

— developing efficient algorithms to simulate with respect to the empirical checkerboard copula with information on the tail in dimension larger than 2.

References

[Bernard & Vanduffel 2015] Carole Bernard and Steven Vanduffel. *A new approach to assessing model risk in high dimensions.* Journal of Banking & Finance, vol. 58, pages 166–178, 2015.

[Bernard *et al.* 2013] Carole Bernard, Ludger Rüschendorf and Steven Vanduffel. *Value-at-Risk bounds with variance constraints.* Available at SSRN 2342068, 2013.

[Bernard *et al.* 2014] Carole Bernard, Xiao Jiang and Ruodu Wang. *Risk Aggregation with Dependence Uncertainty.* Insurance: Mathematics and Economics, vol. 54, no. 1, pages 93–108, 2014.

References

[Cossette *et al.* 2014] Hélène Cossette, Marie-Pier Côté, Mélina Mailhot and Etienne Marceau. *A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks.* Journal of Multivariate Analysis, vol. 130, pages 1–20, 2014.

[Dacorogna *et al.* 2014] Michel Dacorogna, Leila El Bahtouri and Marie Kratz. Explicit diversification benefit for dependent risks. Preprint, 2014.

[Deheuvels 1979] Paul Deheuvels. *La fonction de dépendance empirique et ses propriétés.* Acad. Roy. Belg. Bull. Cl. Sci., vol. 65, no. 5, pages 274 – 292, 1979.

[Dubey 1970] Satya D Dubey. *Compound gamma, beta and F distributions.* Metrika, vol. 16, no. 1, pages 27–31, 1970.

[Embrechts & Puccetti 2010] Paul Embrechts and Giovanni Puccetti. *Risk aggregation.* Copula Theory and Its Applications, 2010.

[Embrechts *et al.* 2013] Paul Embrechts, Giovanni Puccetti and Ludger Rüschendorf. *Model uncertainty and VaR aggregation.* Journal of Banking & Finance, vol. 37, no. 8, pages 2750–2764, 2013.

[Erdely *et al.* 2014] Arturo Erdely, José M González-Barrios and María M Hernández-Cedillo. *Frank's condition for multivariate Archimedean copulas.* Fuzzy Sets and Systems, vol. 240, pages 131–136, 2014.

[Fermanian *et al.* 2004] Jean-David Fermanian, Dragan Radulovic and Marten Wegkamp. *Weak convergence of empirical copula processes.* Bernoulli, vol. 10, no. 5, pages 847–960, 2004.

[Frank 1996] M.J. Frank. *Diagonals of copulas and Schröeder's equation.* Procedings Aequ. Math., no. 150, 1996.

[Li *et al.* 1998] Xin Li, P Mikusiński and Michael D Taylor. *Strong approximation of copulas.* Journal of Mathematical Analysis and Applications, vol. 225, no. 2, pages 608–623, 1998.

[Marshall & Olkin 1988] Albert W Marshall and Ingram Olkin. *Families of multivariate distributions.* Journal of the American Statistical Association, vol. 83, no. 403, pages 834–841, 1988.

[Maume-Deschamps *et al.* 2015] Véronique Maume-Deschamps, Didier Rullière and Khalil Saïd. On capital allocation by minimizing multivariate risk indicators. Preprint, 2015.

[Mikusinski & Taylor 2010] Piotr Mikusinski and Michael D Taylor. *Some approximations of n-copulas.* Metrika, vol. 72, no. 3, pages 385–414, 2010.

[Nelsen *et al.* 2008] Roger B Nelsen, José Juan Quesada-Molina, José Antonio Rodríguez-Lallena and Manuel Úbeda-Flores. *On the construction of copulas and quasi-copulas with given diagonal sections.* Insurance: Mathematics and Economics, vol. 42, no. 2, pages 473–483, 2008.

[Oakes 1989] David Oakes. *Bivariate survival models induced by frailties.* Journal of the American Statistical Association, vol. 84, no. 406, pages 487–493, 1989.

[Puccetti & Rüschendorf 2012] Giovanni Puccetti and Ludger Rüschendorf. *Computation of sharp bounds on the distribution of a function of dependent risks.* Journal of Computational and Applied Mathematics, vol. 236, no. 7, pages 1833–1840, 2012.

[Rüschendorf 1982] Ludger Rüschendorf. *Random variables with maximum sums.* Advances in Applied Probability, pages 623–632, 1982.

[Rüschendorf 1983] Ludger Rüschendorf. *Solution of a statistical optimization problem by rearrangement methods.* Metrika, vol. 30, no. 1, pages 55–61, 1983.

[Sungur & Yang 1996] Engin A Sungur and Yimin Yang. *Diagonal copulas of Archimedean class.* Communications in Statistics-Theory and Methods, vol. 25, no. 7, pages 1659–1676, 1996.

Chapitre 4

Kernel based estimation of the spectral measure

Sommaire

Abstract

In this chapter we propose a kernel based estimator for the density of the spectral measure of a bivariate distribution with regular variation. An extension of our method allows to estimate discrete spectral measures. Some convergence properties are obtained. A simulation study is proposed to investigate the performance of the estimator. As a motivation to our study, we present the approximation of the ruin probability of a multi-branch insurance company where the dependence between claims in different branches is given by a common shock model with discrete spectral measure.

4.1 Introduction

Statistics of multivariate extremes have important applications in fields like insurance, finance and environmental sciences. A fundamental issue when dealing with bivariate or multivariate extremes is the estimation of the dependence structure. The dependence structure on the extremes can be described by different alternatives, for example the spectral measure, the tail dependence function and the exponent measure.

Our interest in this chapter is principally to estimate the spectral measure when it is discrete, or more precisely when it can be expressed as

$$\sigma = \sum_{i=1}^{m} \alpha_i \delta_{\theta_i}, \tag{4.1}$$

where δ_θ is the Dirac function which is equal to 1 at θ and 0 otherwise.

That is, we want to estimate the locations θ_i and the weights α_i of its masses. Discrete spectral measures often occur in common shock multivariate models. For example it occurs when calculating the ruin probability of a multi-branch insurance company where the dependence between claims in different branches is given by a common shock model (see [Hult & Lindskog 2006]).

Estimation of the spectral measure has been treated in many papers. For example some parametric estimators of the spectral measure have been proposed by [Coles & Tawn 1991] and [Joe *et al.* 1992], and some nonparametric estimators by [Einmahl *et al.* 1997], [Einmahl *et al.* 2001] and [Einmahl & Segers 2009].

However, most of the estimations of the spectral measure proposed in the literature do not apply to the discrete case. The estimator proposed by [Einmahl & Segers 2009] includes only the special discrete cases of full dependency and independence. Only recently, in the framework of tail dependence function, [Einmahl *et al.* 2012] proposed an estimation that includes the general discrete case. In this chapter we explore this problem and propose an estimator that can estimate discrete spectral measures under some conditions.

The rest of the chapter is organised as follows. In Section 4.2 we define bivariate regular variation and spectral measure. In Section 4.3 we present the asymptotic approximation of the multivariate ruin probability proposed by [Hult & Lindskog 2006]. We show how, when the claim process has a common shock, its spectral measure is discrete and thus a link can be established

between ruin probabilities and discrete spectral measures. In Section 4.4 we introduce a simple estimator of a spectral measure using the kernel function. We present first, in subsection 4.4.1, an estimator that only applies to spectral measures which are continuous. This estimator is shown to be weakly consistent under some conditions. Then, in subsection 4.4.2, using basic results on mode estimation the estimator is modified in order to estimate discrete spectral measures. In Section 4.5 some results on simulated data are presented in order to investigate the performance of the proposed estimator. Finally some conclusions are provided in Section 4.6.

4.2 Bivariate regular variation

Let $\|\cdot\|$ denote any norm on $\mathbb{R}^2$. The polar coordinates of a vector $\mathbf{X} \neq 0$ are defined by

$$(R, S) := \left(\|\mathbf{X}\|, \frac{\mathbf{X}}{\|\mathbf{X}\|} \right) \in \mathbb{R}_+ \times \mathbb{S}_1$$

where $\mathbb{S}_1 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$ denotes the unit sphere for this norm.

A bivariate random vector $\mathbf{X}$ is said to have regular variation of index α if there exists a probability measure σ on $\mathbb{S}_1$ such that

$$\Pr((r^{-1}R, S) \in B | R > r) \to (v_\alpha \times \sigma)[B], \quad \text{as } r \to \infty, \tag{4.2}$$

for any $B \in \mathcal{B}((1,\infty) \times \mathbb{S}_1)$ with $(v_\alpha \times \sigma)[\partial B] = 0$, where v_α is the measure in $(1,\infty)$ such that $v_\alpha[(x,\infty)] = x^{-\alpha}$ for any $x \geq 1$ and ∂B is the set of boundary points of B.

The measure σ is the so-called *spectral* or *angular measure* of $\mathbf{X}$. Notice that letting $B = (1,\infty) \times B'$ in the definition above, we have that if $\mathbf{X}$ has regular variation then

$$\Pr(S \in B' | R > r) \to \sigma[B'], \quad \text{as } r \to \infty,$$

for any $B' \in \mathcal{B}(\mathbb{S}_1)$ such that $\sigma[\delta B'] = 0$. The spectral measure was introduced by [de Haan & Resnick 1977]. For a general background see, for instance, the monographs by [Beirlant *et al.* 2006] and [De Haan & Ferreira 2007].

Notation. Usually we don't work directly with $\mathbb{S}_1$ but with some of its parametrisations. When $\|\cdot\|$ is the euclidean norm, a standard parametrisation is for example $\theta \mapsto (\cos(\theta), \sin(\theta))$ for $\theta \in [0, 2\pi)$. In what follows we show that we may work directly with a measure on the domain of the parametrisation.

Proposition 4.2.1. *Consider a parametrization $s : I \subset \mathbb{R} \to \mathbb{S}_1$ that is bijective and such that the points of discontinuity of s^{-1} have null measure under*

σ. Let Θ be the random vector in I defined by $\Theta = s^{-1}(S) = s^{-1}(\boldsymbol{X}/\|\boldsymbol{X}\|)$. If $\boldsymbol{X}$ has regular variation then there exists a measure $\tilde{\sigma}$ in I such that

$$\Pr(\Theta \in B | R > r) \to \tilde{\sigma}[B], \quad \text{as } r \to \infty,$$

for any $B \in \mathcal{B}(I)$ with $\tilde{\sigma}[\partial B] = 0$.

Proof. Define the measure $\tilde{\sigma}$ on I by $\tilde{\sigma}(B) = \sigma[s(B)]$ for any $B \in \mathcal{B}(I)$. Let $B \in \mathcal{B}$ such that $\tilde{\sigma}[\partial B] = 0$. Now because the points of discontinuity of s^{-1} have measure 0, then $\sigma[\partial s(B) \setminus s(\partial B)] = 0$. This and the assumption that $\tilde{\sigma}[\delta B] = \sigma[s(\delta B)] = 0$ imply $\sigma[\partial s(B)] = 0$. We have then by definition of regular variation and the bijectivity of s that

$$\Pr(\Theta \in B | R > r) = \Pr(S \in s(B) | R > r) \to \sigma[s(B)] = \tilde{\sigma}[B],$$

when $r \to \infty$, as desired. □

From now on by spectral measure we will always mean the measure $\tilde{\sigma}$ in I of Proposition 4.2.1. and by abuse of notation it will be denoted by σ. Similarly the polar coordinates or polar transformation of a bivariate vector $\mathbf{X} \neq 0$ will be the pair $(R, \Theta) = (R, s^{-1}(S))$ for a parametrisation s satisfying the hypothesis of Proposition 4.2.1.

4.3 Motivation: spectral measure and ruin probabilities

In the classical ruin problem, an insurance company gains capital from premium incomes and loses capital as a result of claims. Estimates are then given for the probability that the company ever incurs ruin, that is the probability that company's total capital becomes negative.

More precisely, the risk reserve process is modeled by $R(t) = u + tp - S(t)$, where u denotes the initial capital, p is the premium rate per unit time and $S(t)$ is a stochastic process modeling the amount of cumulative claims up to time t. The ruin probability is then given by

$$\psi(u) = \Pr(R(t) < 0 \text{ for some } t \geq 0).$$

Ruin theory for the univariate risk model has been discussed extensively in the literature; for a summary of main results, see, for exemple, [Asmussen & Albrecher 2010]. An extension of classical risk theory towards a multidimensional model is possible while allowing $R(t)$, u, p and $S(t)$ to be vectors, with possible dependence between the components of $\mathbf{S}(t)$.

We are interested in insurance claims which occur in different lines of business (for example, driving insurance, house insurance, health, etc.) of the same company or in different countries. The mathematical framework is as follows. Let us consider an insurance company with d lines of business. Suppose $\{\mathbf{Z}_k, k \geq 1\}$ are $i.i.d.$ $\mathbb{R}^d$-valued random vectors representing claim sizes and Z is regularly varying, i.e. satisfies (4.2), with $\alpha > 1$. Let $N_t = \#\{n \geq 1 : T_n \leq t\}$ denote the total number of claims up to time t and $T_0 = 0$, $T_n = W_1 + \cdots + W_n$ where $\{W_k, k \geq 1\}$ is a sequence of $i.i.d.$ random variables representing the interarrival times.

We consider a d-dimensional risk process $\{\mathbf{R}_t, t \geq 0\}$ given by

$$\mathbf{R}_t = u\mathbf{b} + t\mathbf{p} - \sum_{k=1}^{N_t} \mathbf{Z}_k$$

where u is the initial capital and capital allocation to the different lines of business is determined according to $u\mathbf{b}$ with $\mathbf{b} \in (0,1]^d$ and $b^{(1)} + \cdots + b^{(d)} = 1$. $\mathbf{p} \in (0,\infty)^d$ is the premium income rate which is assumed to be constant. The company is said to be insolvent at t if the risk reserve process $\mathbf{R}_t$ lies in some predefined region $\mathcal{F} \subset \mathbb{R}^d$ called the insolvency region or ruin set. For example a possible ruin set is given by

$$\mathcal{F}_\beta = \left\{ \mathbf{x} : \beta \sum_{k=1}^{d} (x^{(k)} \vee 0) < - \sum_{k=1}^{d} (x^{(k)} \wedge 0) \right\}$$

where $\beta \in [0,1]$ is the fraction of capital which can be transferred from one business line with positive reserve to cover losses in other business lines with negative positions.

The multivariate ruin probability over an infinite horizon is defined as

$$\psi_{d,\mathcal{F}}(u) = \Pr\left(\mathbf{R}_t \in \mathcal{F} \text{ for some } t \geq 0\right)$$

for the given ruin set $\mathcal{F} \subset \mathbb{R}^d$. For given $T > 0$, the ruin probability in finite time is

$$\psi_{d,\mathcal{F}}(u,T) = \Pr(\mathbf{R}_t \in \mathcal{F} \text{ for some } t \in [0,T]).$$

Let us define $\mathbf{c} = E(W)\mathbf{p} - E(\mathbf{Z})$ and assume that the net profit condition, $\mathbf{c} \in (0,\infty)^d$, holds, that is, each business line has a positive safety loading. Assume further that $E(W^\gamma) < \infty$ for some $\gamma > \alpha$. Under these hypothesis, [Hult & Lindskog 2006] state that it is possible to approximate infinite and finite-time ruin probabilities of an insurance company for large initial capital u by

$$\psi_{d,\mathcal{F}}(u) \approx \int_0^\infty \mu(v\mathbf{c} + \mathbf{b} - \mathcal{F})\, dv\, u \Pr(|\mathbf{Z}| > u) \tag{4.3}$$

$$\psi_{d,\mathcal{F}}(u,T) \approx \mathrm{E}\,(N_T)\mu(\mathbf{b} - \mathcal{F}) \Pr(|\mathbf{Z}| > u) \tag{4.4}$$

where μ is a measure on $\mathbb{R}^d \backslash \{0\}$ related to the spectral measure σ of $\mathbf{Z}$ through

$$\mu(\mathbf{z} : |\mathbf{z}| > r, \mathbf{z}/|\mathbf{z}| \in S) = r^{-\alpha}\sigma(S) \tag{4.5}$$

for $r > 0$ and Borel sets $S \subset \mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$.

Proposition 4.3.1. *When the spectral measure σ is discrete and has a representation as in (4.1), and A is any subset of $\mathbb{R}^d$ which satisfies that $\lambda \boldsymbol{a} \in A$ for any $\lambda > 1$ and any $\boldsymbol{a} \in A$ then*

$$\mu(A) = \sum_{i=1}^{m} \alpha_i r_i(A)^{-\alpha} \tag{4.6}$$

where $r_i(A)$ is the distance from the origin to the closest intersection point between the set A and the line l_i passing through the origin in direction of θ_i, that is

$$r_i(A) = \inf\{\|\boldsymbol{x}\| : \boldsymbol{x} \in A \cap l_i\}.$$

Proof. Suppose that $\sigma = \delta_\theta$, that $A \subset \mathbb{R}^d$ satisfies the hypothesis given in the proposition and that l is the line passing through the origin and in direction of θ. Then $\mu(A) = \mu(A \cap l) = \mu(\mathbf{z} : |\mathbf{z}| > r, \mathbf{z}/|\mathbf{z}| \in \{\theta\})$, where $r = \inf\{\|\mathbf{x}\| : \mathbf{x} \in A \cap l\}$. And thus $\mu(A) = r^{-\alpha}\sigma(\{\theta\})$ as desired. The same procedure can be generalized easily to the case of multiple (but finite) masses. □

Let us suppose that we have some events that cause claims in the different lines of business and some other events that are specific to each line of business. In insurance, events that hit different lines of business could be, for example, natural catastrophes like hurricanes or earthquakes.

Losses generated by business line specific claims occur according to a Poisson process $(N_{j,t})_{t\geq 0}$ with intensity λ_j, $j \in 1, \ldots, d$. The claim sizes $(Z_{j,k})_{k\geq 1}$ are considered to be *i.i.d.*, with $Z_{j,k} \overset{d}{=} \nu^{(j)} Z$ and $\nu^{(j)} \in (0, \infty)$ for $j = 1, \ldots, d$. Assume further that all the counting processes $(N_{i,t})_{t\geq 0}$ and claim sizes $(Z_{i,k})_{k\geq 1}$, $i \in \{0, \ldots, d\}$, are independent.

Losses which are not line specific occur according to a multivariate Poisson process $(N_{0,t})_{t\geq 0}$ with intensity λ_0 and claim sizes $\{\mathbf{Z}_{0,k}\}$ where $(\mathbf{Z}_{0,k})_{k\geq 1}$ is a sequence of i.i.d. r.v.s with $\mathbf{Z}_{0,k} \overset{\mathcal{L}}{=} \mathbf{Z}_0$.

The claim amount process caused by losses of all types can be written as

$$\mathbf{C}_t = \sum_{k=1}^{N_{0,t}} \mathbf{Z}_{0,k} + \sum_{k=1}^{N_{1,t}} Z_{1,k}\mathbf{e}_1 + \cdots + \sum_{k=1}^{N_{d,t}} Z_{d,k}\mathbf{e}_d = \sum_{k=1}^{N_t} \mathbf{Z}_k$$

where $N_t = N_{0,t} + \cdots + N_{d,t}$ is a Poisson process with intensity $\overline{\lambda} = \sum_{i=0}^{d} \lambda_i$ counting all shocks and business line specific claims. Further, $\mathbf{Z}_k$ has the stochastic representation

$$\mathbf{Z}_k \stackrel{d}{=} \mathbf{Z}_{0,1}\delta_0(\xi) + Z_{1,1}\mathbf{e}_1\delta_1(\xi) + \cdots + Z_{d,1}\mathbf{e}_d\delta_d(\xi)$$

where ξ is a random vector satisfying $\Pr(\xi = k) = \lambda_k/\overline{\lambda}$ for $k \in \{0, \ldots, d\}$ and independent of $\mathbf{Z}_{0,1}, Z_{1,1}, \cdots, Z_{d,1}$ and $(\mathbf{e}_1, \ldots, \mathbf{e}_d) = ((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1))$ are the basis vectors of $\mathbb{R}^d$.

According to (4.3), ruin probabilities for large initial capital u depends on the measure μ of $\mathbf{Z}$ that is related to the spectral measure σ through (4.5).

If σ_0 is the spectral density of $\mathbf{Z}_0$, then the spectral density of $\mathbf{Z}$ is given by

$$\sigma = \frac{\lambda_0}{\overline{\lambda}}\sigma_0 + \sum_{i=1}^{d} \frac{\lambda_0}{\overline{\lambda}}\delta_{\mathbf{e}_i}.$$

When we specialize to processes where the common claims $\mathbf{Z}_0$ have discrete spectral measure then the spectral measure of $\mathbf{Z}$ is also discrete. Proposition 4.3.1 allows to give another representation to the approximations of the ruin probabilities (4.3):

$$\lim_{u\to\infty} \frac{\psi_{d,\mathcal{F}}(u)}{u \Pr(Z > u)} = \sum_{j=1}^{m} \alpha_i \int_0^{\infty} r_i(v\mathbf{c} + \mathbf{b} - \mathcal{F})dv$$
$$\lim_{u\to\infty} \frac{\psi_{d,\mathcal{F}}(u)}{E(N_T) \Pr(Z > u)} = \sum_{j=1}^{m} \alpha_i r_i(\mathbf{b} - \mathcal{F}),$$

where r_i has been defined in Proposition 4.3.1.

Then, if the quantities α_i and θ_i are estimated, an estimator of the multivariate ruin probabilities could be obtained by a plug-in method. The estimation of these quantities are the subject of the following section.

4.4 Kernel based estimators of the spectral measure

In this section we propose an estimator of the density of the spectral measure of a bivariate random vector, then a modification of this estimator that applies to discrete spectral measures.

4.4.1 Estimation of the spectral measure density

In this subsection we assume that the spectral measure admits a density function. Let $\mathbf{X}$ be a bivariate regularly varying random vector with (R, Θ) its polar transformation.

We will denote by $F^{(r)}$ the distribution function of Θ conditionally to the event $R > r$:

$$F^{(r)}(\theta) = \Pr(\Theta < \theta | R > r).$$

Then, by definition of the spectral measure of $\mathbf{X}$, we have that for each $\theta \in I$ when $r \to \infty$

$$F^{(r)}(\theta) \to F_\sigma(\theta),$$

where $F_\sigma(\theta) = \sigma([-\infty, \theta] \cap I)$. In this section we will make the following assumptions:

(i) $F^{(r)}$ is differentiable for each $r > 0$, we denote by f_r its derivative.
(ii) For each $\theta \in I$ when $r \to \infty$ we have $f_r(\theta) \to f_\sigma(\theta)$, with $F'_\sigma = f_\sigma$ the density of the spectral measure.

For any real $r > 0$ and integer $k > 0$, let $\hat{f}_{r,k}$ be a kernel estimator for f_r based on k observations. For a sequence $\Theta_1^{(r)}, \ldots, \Theta_k^{(r)}$ of i.i.d. random variables distributed as $F^{(r)}$, the kernel estimator $\hat{f}_{r,k}$ can be written as

$$\hat{f}_{r,k}(\theta) = \frac{1}{hk} \sum_{j=1}^{k} K\left(\frac{\theta - \Theta_j^{(r)}}{h}\right),$$

for any $\theta \in I$, where K is a kernel function and $h = h(k)$ is the window function. We assume that the kernel K is a non-negative measurable function such that:

— $\int K(x)\,dx = 1$
— $\int K^2(x)\,dx < \infty$,
— $\int |x| K^2(x)\,dx < \infty$.

and that the window h satisfies: $h \to 0$ and $hk \to \infty$ as $k \to \infty$.

Notice that both the expectation and variance of $\hat{f}_{r,k}(\theta)$ exist by the integrability of K^2. It is straightforward to check that they can be written as

$$\mathrm{E}\left[\hat{f}_{r,k}(\theta)\right] = \int f_r(\theta - hx) K(x)\,dx, \tag{4.7}$$

and

$$\mathrm{Var}\left[\hat{f}_{r,k}(\theta)\right] = \frac{1}{kh} \int f_r(\theta - hx) K^2(x)\,dx - \frac{1}{k}\left(E[\hat{f}_{r,k}(\theta)]\right)^2. \tag{4.8}$$

Let $(k_n) \in \mathbb{N}$ and $(r_n) \in \mathbb{R}$ be two sequences such that

$$k_n \to \infty, \quad k_n/n \to 0, \text{ and } r_n \to \infty, \tag{4.9}$$

then for each n we have a kernel estimator $\hat{f}_{r_n,k_n}$. If we have a common window function h and $h_n = h(k_n)$ satisfies $h_n \to 0$ and $h_n k_n \to \infty$ as $n \to \infty$ then, for each $\theta \in I$, we will denote by

$$\hat{f}_n(\theta) = \frac{1}{h_n k_n} \sum_{j=1}^{k_n} K\left(\frac{\theta - \Theta_j^{(r_n)}}{h_n}\right). \tag{4.10}$$

Proposition 4.4.1. *Under the assumptions and notations of this section, if for each $r > 0$ the derivative of f_r is bounded by some $M_r < \infty$, and if*

$$(i)\ \ f_{r_n}(\theta)/(k_n h_n) \to 0, \quad \textit{and} \quad (ii)\ \ M_{r_n} h_n \to 0$$

as $n \to \infty$, then for all $\theta \in I$

$$\hat{f}_n(\theta) \xrightarrow[n\to\infty]{\Pr} f_\sigma(\theta).$$

Proof. The triangle inequality allows us to write

$$|\hat{f}_n(\theta) - f_\sigma(\theta)| \le |\hat{f}_n(\theta) - f_{r_n}(\theta)| + |f_{r_n}(\theta) - f_\sigma(\theta)|.$$

By assumption $f_{r_n}(\theta) \to f_\sigma(\theta)$ so it is sufficient to show that $|\hat{f}_n(\theta) - f_{r_n}(\theta)|$ converges to 0 in probability. Applying again the triangle inequality gives

$$|\hat{f}_n(\theta) - f_{r_n}(\theta)| \le |\hat{f}_n(\theta) - E[\hat{f}_n(\theta)]| + |E[\hat{f}_n(\theta)] - f_{r_n}(\theta)|.$$

The proof is completed by showing that $|E[\hat{f}_n(\theta)] - f_{r_n}(\theta)| \to 0$ and that $|\hat{f}_n(\theta) - E[\hat{f}_n(\theta)]| \to 0$ in probability.
For each $n > 0$ and x let $\xi_n(x)$ by a real between θ and $\theta - h_n x$ such that

$$f_{r_n}(\theta - h_n x) = f_{r_n}(x) + h_n x f'_{r_n}(\xi_n(x)). \tag{4.11}$$

Substituting (4.11) into (4.7) yields to

$$\begin{aligned}\left|\mathrm{E}\left[\hat{f}_n(\theta)\right] - f_{r_n}(\theta)\right| &= h_n \left|\int x K(x) f'_{r_n}(\xi_n(x))\, dx\right| \\ &\le M_{r_n} h_n \int |x| K(x)\, dx\end{aligned}$$

and by assumption (ii) of the proposition the convergence to 0 follows. Similarly substituting (4.11) into (4.8) yields to

$$\mathrm{Var}\left[\hat{f}_n(\theta)\right] \le \frac{1}{k_n h_n} f_{r_n}(\theta) \int K^2(x)\, dx + \frac{M_{r_n}}{k_n} \int |x| K^2(x)\, dx - \frac{1}{k_n}\left(E[\hat{f}_n(\theta)]\right)^2$$

Assumption (i) of the proposition and the window function properties imply that the right terms of the inequality above go to 0. Finally Chebyshev's inequality and the convergence to 0 of the $\mathrm{Var}\,[\hat{f}_n(\theta)]$ imply

$$\left| E[\hat{f}_n(\theta)] - \hat{f}_n(\theta) \right| \xrightarrow[n\to\infty]{\Pr} 0$$

which completes the proof.

□

Remarks. Notice that assumption (i) of Proposition 4.4.1 is true if the density of the spectral measure f_σ is bounded in I. Similarly, assumption (iii) is achieved if f'_σ exists and is bounded in I.

In practice, given a sample of size n, it is usual to assume either k_n or r_n being fixed and the other being random. For example, we can take r_n as given and put k_n as the (random) number of exceedances of r_n. So instead of assuming we have a fixed sequence $(k_n) \in \mathbb{Z}$ we will suppose we have a random sequence $(K_n) \in \mathbb{Z}$ that satisfies:

$$K_n \xrightarrow[n\to\infty]{\Pr} \infty \text{ and } K_n/n \xrightarrow[n\to\infty]{\Pr} 0. \tag{4.12}$$

4.4.2 Estimation of a discrete spectral measure

In Section 4.4.1 we proposed an estimator for the density of the spectral measure. Here we assume that the spectral measure is discrete and thus such a density does not exist. Nevertheless we will assume that densities f_r exist for each $r > 0$, i.e., the discrete nature of the spectral measure is only reached at infinity. We will begin by proposing an estimator when the discrete measure consists of one single mass, we will move then to the more general case of a finite number of masses.

4.4.2.1 One mass estimation

Let us assume that the spectral measure of $\mathbf{X}$ is concentrated on a point $\theta_0 \in I$, i.e the spectral measure σ is a dirac measure δ_{θ_0} for some θ_0 in I. This means that $\Theta^{(r)} \xrightarrow[r\to\infty]{d} \theta_0$.
As before, given sequences (k_n) and (r_n) satisfying (4.9), we assume that for each n we have, families of i.i.d. random variables distributed as $\Theta^{(r_n)}$ of size k_n

$$\{\Theta_1^{(r_n)}, \ldots, \Theta_{k_n}^{(r_n)}\} \qquad n = 1, 2, \ldots.$$

Moreover, given that also $\Theta^{(r_n)} \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0$, because θ_0 is a point, we have that $\Theta_1^{(r_n)}$ is a consistent estimator of θ_0. We can also take $\Theta_{j_n}^{(r_n)}$ as the estimator

of θ_0 for any sequence (j_n), $1 \leq j_n \leq k_n$. However this may not be the more efficient estimator, especially when the last convergence is slow. The estimator can be improved using all the information contained in each sample of size k_n as we explain below.
For each n, let $\hat{f}_n$ be the kernel estimator of f_{r_n} defined by (4.10) and drawn from a sample $\Theta_1^{(r_n)}, \ldots, \Theta_{k_n}^{(r_n)}$. The sample estimate of the mode $\hat{\theta}_n$ is then defined as a point such that

$$\hat{f}_n(\hat{\theta}_n) = \sup_{\theta} \hat{f}_n(\theta). \tag{4.13}$$

We propose $\hat{\theta}_n$ as an estimator of θ_0 and show its consistency under some conditions. The problem of estimating the location of the mode has been widely studied in the context of the kernel density estimation, see for example [Parzen 1962], [Romano 1988] and [Vieu 1996].

We will assume that for each r, $0 < r < \infty$, the random variable $\Theta^{(r)}$ is absolutely continuous so that a density f_r exists. If the spectral measure is discrete with only one mass it is natural to think that eventually the densities f_r have only one mode, noted as $\theta_0^{(r)}$, and that this mode converges to θ_0, the mass of the spectral measure, as $r \to \infty$.

For a couple (r, k), let $\hat{\theta}_k^{(r)}$ denote the sample mode of the distribution $\Theta^{(r)}$ estimated from a sample of size k . Notice that with our notation $\hat{\theta}_n \stackrel{d}{=} \hat{\theta}_{k_n}^{(r_n)}$. We assume that conditions in f_r are given such that $\hat{\theta}_k^{(r)} \xrightarrow[k\to\infty]{\Pr} \theta_0^{(r)}$. Sufficient conditions are for example (see [Romano 1988]):

1. f_r is continuous on a neighbourhood of the mode $\theta_0^{(r)}$.
2. For every $\zeta > 0$, $\sup\{f_r(t) : |\theta_0^{(r)} - t| > \zeta\} < f_r(\theta_0^{(r)})$

Given $\varepsilon > 0$ let us denote

$$\zeta_{r,k}(\varepsilon) = \Pr\left(|\theta_0^{(r)} - \hat{\theta}_k^{(r)}| > \varepsilon\right). \tag{4.14}$$

By [Romano 1988], if eventually for all r conditions 1. and 2. are satisfied for each f_r, then $\hat{\theta}_k^{(r)}$ is a consistent estimator of $\theta_0^{(r)}$ so that we have $\zeta_{r,k}(\varepsilon) \xrightarrow[k\to\infty]{} 0$. We will indeed assume that $\zeta_{r,k}(\varepsilon)$ is (eventually) non increasing in k.

Proposition 4.4.2. *Suppose that $\boldsymbol{X}$ has a spectral measure concentrated in $\theta_0 \in I$ such that the densities f_r exist and satisfy conditions 1. and 2. of this section, and such that $\theta_0^{(r)}$, the mode of f_r, converges to θ_0. If for each $\varepsilon > 0$ the sequence $\zeta_{r,k}(\varepsilon)$, given by (4.14), is eventually decreasing, both in r and k, then $\hat{\theta}_n$ is a consistent estimator of θ_0.*

Proof. The proof is easy because by assumption $\theta_0^{(r_n)} \to \theta_0$ so it is sufficient to show that $\Pr(|\theta_0^{(r_n)} - \hat{\theta}_n| > \varepsilon)$ converges to 0, as $n \to \infty$. The last is true because from our assumption in $\zeta_{r,k}(\varepsilon)$ it follows that $\zeta_{r_n,k_n}(\varepsilon) \to 0$, as $n \to \infty$, for any sequence r_n, k_n satisfying (4.9). □

Checking the validity of the hypothesis on $\zeta_{r,k}$ seems not to be easy because the exact distribution of the sample mode is not generally known. In fact we do not know non trivial examples, when we can give explicitly its distribution. To sum up, the assumption on $\zeta_{r,k}(\varepsilon)$ can be described roughly as:

— For given r, the more data we have then better is the estimation of the mode of f_r.

— Samples of $\Theta^{(r)}$ of fixed size k gives better estimation of the mode when r increases.

Both of them look natural given that the convergences $\zeta_{r,k}(\varepsilon) \to 0$ as $k \to \infty$ and $\Pr(|\theta_0^{(r)} - \Theta^{(r)}| > \varepsilon) \to 0$ as $r \to \infty$, for any $\varepsilon > 0$, are both satisfied under our framework.

4.4.2.2 Multiple Masses

In this section we will generalize our method in order to obtain an estimator of the discrete spectral measure which applies in case of multiple, but finite, number of masses. In this case we need to estimate not only the location of the masses as before but also their weights. We assume in this section that the spectral measure σ can be written as

$$\sigma = \sum_{i=1}^{m} \alpha_i \delta_{\theta_i} \tag{4.15}$$

where m is the number of masses of the spectral measure, θ_i their locations and α_i their weights.
We begin by showing that if we have consistent estimators for the locations then the weights can be estimated consistently using the empirical distribution function.

Proposition 4.4.3. *If $\hat{\theta}_{i,k}$ is a consistent estimator of θ_i then $\widehat{F}_{k_n}^{(r_n)}(\hat{\theta}_{i,k}+\zeta)$ is a consistent estimator of $\sigma((-\infty, \theta_i])$, where $\widehat{F}_{k_n}^{(r_n)}$ is the empirical distribution function of the sample $\{\Theta_1^{(r_n)}, \ldots, \Theta_{k_n}^{(r_n)}\}$ and $\zeta > 0$ is sufficiently small such that σ has no masses between θ_i and $\theta_i + \zeta$.*

Proof. The consistency is proven by showing that the following three elements:

$$(i) \quad |\widehat{F}_{k_n}^{(r_n)}(\hat{\theta}_{i,k_n}+\zeta) - F^{(r_n)}(\hat{\theta}_{i,k_n}+\zeta)|, \qquad (ii) \quad |F^{(r_n)}(\hat{\theta}_{i,k_n}+\zeta) - F^{(r_n)}(\theta_i+\zeta)|$$

$$\text{and} \quad (iii) \quad |F^{(r_n)}(\theta_i + \zeta) - F_\sigma(\theta_i)|,$$

converge to 0 in probability as $n \to \infty$, where $F_\sigma(\theta) = \sigma([-\infty, \theta] \cap I)$.

The convergence of (i) follows from the Dvoretzk-Kiefer-Wolfowitz inequality (see [Dvoretzky *et al.* 1956]),

$$\Pr(\sup_\theta |\widehat{F}_{k_n}^{(r_n)}(\theta) - F^{(r_n)}(\theta)| > \varepsilon) \leq 2e^{-2k_n\varepsilon^2}.$$

Convergence of (iii) is implied directly by the definition of the spectral probability measure. To prove the convergence of (ii) notice that

$$\begin{aligned} |F^{(r_n)}(\hat{\theta}_{i,k_n} + \zeta) - F^{(r_n)}(\theta_i + \zeta)| \leq |F^{(r_n)}(\hat{\theta}_{i,k_n} + \zeta) - F_\sigma(\hat{\theta}_{i,k_n} + \zeta)| \\ + |F_\sigma(\hat{\theta}_{i,k_n} + \zeta) - F_\sigma(\theta_i + \zeta)| + |F_\sigma(\theta_i + \zeta) - F^{(r_n)}(\theta_i + \zeta)|. \end{aligned}$$

The convergence of the last two terms on the right-hand side of this inequality follows from the convergences $\hat{\theta}_{i,k_n} \xrightarrow{\Pr} \theta_i$ and $F^{(r_n)} \xrightarrow{F} \sigma$, respectively, and the continuity of F_σ in $\theta_i + \zeta$. To show that

$$|F^{(r_n)}(\hat{\theta}_{i,k_n} + \zeta) - F_\sigma(\hat{\theta}_{i,k_n} + \zeta)|$$

converges to 0 in probability notice that $F^{(r_n)}(\cdot) \to F_\sigma(\cdot)$ in all the continuity points of F_σ. For given $\varepsilon > 0$ let $V_n = \{\theta \in I : |F^{(r_n)}(\theta) - F_\sigma(\theta)| > \varepsilon\}$. Then $\cap_n V_n$ contains exactly the points of discontinuity of F_σ and so, again by consistency of $\hat{\theta}_{i,k_n}$, we have that $\Pr(\hat{\theta}_{i,k_n} + \zeta \in V_n) \to 0$, which proves the proposition. □

Unfortunately, when the number of masses is multiple and unknown the estimation of their location is not trivial and requires in depth exploration of the densities $\hat{f}_n$. For example we should apply bump-hunting techniques like in [Eddy 1980] to estimate the location of multiple modes. As for the number of masses, it could be approximated using the results given in [Silverman 1981].

If the number of masses of the discrete spectral measures is known, say d, the estimation could be simplified by assuming that the location of the masses occur on the d highest bumps of the density. We do not present here theoretical results on the convergence of the locations. However some numerical results are presented using this simplified approach in Section 4.5.2.

4.5 Application to simulated data

The following simple model has been used in order to simulate data on which the proposed estimator could be applied.

Let A be a $2 \times d$ real matrix, $\mathbf{Z} = (Z_1, \ldots, Z_d)$ a random vector of d independent r.v.s with survival function $\overline{F}$ regular varying with index $-\alpha$, and $\mathbf{W} = (W_1, W_2)$ a vector of 2 independent r.v.s with survival function $\overline{G}$ regular varying with index $-\beta$ with $\beta > \alpha > 0$. Then, a direct calculation shows that

$$\mathbf{X} = A\mathbf{Z} + \mathbf{W}$$

has a discrete spectral measure with masses

$$\sigma(\{\theta_i\}) = \alpha_i = \frac{(a_{1i}^2 + a_{2i}^2)^{\alpha/2}}{\sum_{j=1}^{d}(a_{1j}^2 + a_{2j}^2)^{\alpha/2}}, \tag{4.16}$$

at the points $\theta_i = \arctan(a_{1i}/a_{2i})$ for $i = 1, \ldots d$.

4.5.1 One mass example

Using the notation above, in the following example we will take $d = 1$ and $A = (1, 1)^t$. We will assume that Z_1 is Pareto distributed with index α, and that W_1 and W_2 are Pareto distributed with index β. Thus the location of the mass is in $\pi/4$. For the simulation study we fix $\alpha = 1$ and we let β, the size of the sample n and the subsample k_n vary. For each simulation the kernel density of the k_n values with corresponding largest norm has been used to estimate the mode according to (4.13). In order to study the performance of our estimator we present the mean and root-mean-squared error (RMSE) based in $m = 1000$. The RMSE is given in percentage of the real location θ_0,

$$\text{RMSE} = \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\widehat{\theta}_{n,i} - \theta_0\right)^2}}{\theta_0} \times 100,$$

where $\widehat{\theta}_{n,i}$ is an estimation of θ_0 based on the ith sample simulated of size n from which a subsample of k_n is extracted. The results are presented in Table 4.1.

We notice that the estimated location are close to $\pi/4 \approx 0.785$ and that the RMSE values are very low. When β gets closer to α the results deteriorate as expected.

4.5.2 Multiple masses

In the example below we assume that the number of masses is $d = 3$, and that Z_i, is Pareto distributed with index α for $i = 1, 2, 3$ and that W_1 and W_2

	$n = 1000$ $k_n = 200$	$n = 1000$ $k_n = 100$	$n = 1000$ $k_n = 50$	$n = 500$ $k_n = 100$	$n = 500$ $k_n = 50$	$n = 500$ $k_n = 25$
$\beta = 2$	0.787	0.783	0.766	0.777	0.801	0.750
	(1.40%)	(1.43%)	(1.56%)	(1.90%)	(1.85%)	(1.84%)
$\beta = 1.5$	0.768	0.798	0.802	0.812	0.782	0.804
	(1.95%)	(1.98%)	(1.97%)	(3.12%)	(3.85%)	(3.25%)
$\beta = 1.2$	0.801	0.835	0.765	0.771	0.810	0.818
	(4.81%)	(4.78%)	(5.97%)	(7.12%)	(7.85%)	(6.84%)

Table 4.1 – Mean and RMSE in percentage of the real location based on 1000 simulations.

are Pareto distributed with index β. The matrix A is such that the masses are located in $\theta_1 = 0.7$, $\theta_2 = 0.8$ and $\theta_3 = 1.3$. More precisely we assume

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0.61 & 0.67 & 0.92 \end{pmatrix}.$$

Thus, by equation (4.16) we have $\alpha_1 \approx 0.31, \alpha_2 \approx 0.32$ and $\alpha_3 \approx 0.36$. For the simulation study we let the index β and the subsample k_n vary. For each simulation the kernel density of the k_n values with corresponding largest norm has been used and the three highest bumps of the density $\hat{f}_n$ used as mass locations. Then the weights of the masses in these points were estimated using Proposition 4.4.3. The results are presented in Tables 4.2 and 4.3.

	θ_1	θ_2	θ_3	α_1	α_2	α_3
Exact Value	0.700	0.800	1.300	0.314	0.323	0.363
$k_n = 200$	0.724	0.791	1.345	0.305	0.313	0.359
	(3.9%)	(4.2%)	(1.5%)	(5.6%)	(6.1%)	(0.9%)
$k_n = 100$	0.733	0.765	1.334	0.295	0.299	0.359
	(4.7%)	(4.5%)	(1.6%)	(7.2%)	(7.1%)	(1.4%)
$k_n = 50$	0.745	0.781	1.354	0.281	0.283	0.371
	(7.3%)	(8.1%)	(1.9%)	(10.2%)	(11.1%)	(3.4%)

Table 4.2 – Mean and RMSE in percentage of the exact value of the location and masses based on 1000 simulations with $\alpha = 1$ and $\beta = 2$.

It can be seen that the estimation of the location and the mass of the third point gives good results with very low RMSE. On the other hand the estimation of θ_1 and θ_2 gives higher RMSE. In fact due to its closeness, the estimation of the density around θ_1 and θ_2 may present in some cases only one bump biasing the estimation.

	θ_1	θ_2	θ_3	α_1	α_2	α_3
Exact Value	0.700	0.800	1.300	0.314	0.323	0.363
$k_n = 200$	0.731	0.787	1.334	0.301	0.311	0.360
	(4.2%)	(4.1%)	(1.7%)	(5.9%)	(7.1%)	(1.3%)
$k_n = 100$	0.733	0.765	1.330	0.289	0.295	0.357
	(5.0%)	(5.1%)	(1.7%)	(7.7%)	(7.8%)	(1.4%)
$k_n = 50$	0.745	0.761	1.350	0.279	0.286	0.375
	(12.3%)	(14.1%)	(3.1%)	(16.3%)	(16.8%)	(4.0%)

Table 4.3 – Mean and RMSE in percentage of the exact value of the location and masses based on 1000 simulations with $\alpha = 1$ and $\beta = 1.5$.

4.6 Conclusion

The estimation of the spectral measure, and in particular of a discrete spectral measure could be important in order to, for example, approximate the ruin probability of a multivariate risk process. In this chapter an estimator of the spectral measure based on the kernel estimator of the densities of the conditioned distributions

$$F^{(r)}(\theta) = \Pr(\Theta < \theta | R > r),$$

was presented. Under some hypothesis, satisfied for example if the densities of $F^{(r)}$ and their derivatives are bounded, the consistency of the estimator was demonstrated. Then, it has been shown that the proposed method can be modified to estimate the location and masses of a discrete spectral measure, which to our knowledge, is quite new in the literature, as few results are known and only recently some estimators were proposed.

However there are multiple limitations in our procedure. When the discrete spectral measure has one mass, the consistency of the estimator of the location of the mass has been obtained assuming an hypothesis on the distribution of the sample mode, which, even if natural, cannot be verified. Moreover, when the spectral measure has multiple masses, we only prove consistency of the estimator of the masses weights when the locations are known. Estimating the location of the masses requires to use bump-hunting techniques like in [Eddy 1980]. However when assuming that the number of masses is known the location can be estimated by looking for the highest bumps of the estimated densities $\hat{f}_n$ and the simulation study shows good results. In any case further research is required.

References

[Asmussen & Albrecher 2010] Søren Asmussen and Hansjörg Albrecher. Ruin probabilities, volume 14. World scientific, 2010.

[Beirlant *et al.* 2006] Jan Beirlant, Yuri Goegebeur, Johan Segers and Jozef Teugels. Statistics of extremes: theory and applications. John Wiley & Sons, 2006.

[Coles & Tawn 1991] Stuart G Coles and Jonathan A Tawn. *Modelling extreme multivariate events.* Journal of the Royal Statistical Society. Series B (Methodological), pages 377–392, 1991.

[De Haan & Ferreira 2007] Laurens De Haan and Ana Ferreira. Extreme value theory: an introduction. Springer Science & Business Media, 2007.

[de Haan & Resnick 1977] Laurens de Haan and Sidney I Resnick. *Limit theory for multivariate sample extremes.* Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 40, no. 4, pages 317–337, 1977.

[Dvoretzky *et al.* 1956] Aryeh Dvoretzky, Jack Kiefer and Jacob Wolfowitz. *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator.* The Annals of Mathematical Statistics, pages 642–669, 1956.

[Eddy 1980] William F. Eddy. *Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data: Comment.* Journal of the American Statistical Association, vol. 75, no. 369, pages pp. 59–61, 1980.

[Einmahl & Segers 2009] John HJ Einmahl and Johan Segers. *Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution.* The Annals of Statistics, pages 2953–2989, 2009.

[Einmahl *et al.* 1997] John HJ Einmahl, Laurens de Haan and Ashoke Kumar Sinha. *Estimating the spectral measure of an extreme value distribution.* Stochastic Processes and their Applications, vol. 70, no. 2, pages 143–171, 1997.

[Einmahl *et al.* 2001] John HJ Einmahl, Laurens de Haan and Vladimir I Piterbarg. *Nonparametric estimation of the spectral measure of an extreme value distribution.* Annals of Statistics, pages 1401–1423, 2001.

[Einmahl *et al.* 2012] John HJ Einmahl, Andrea Krajina, Johan Segers*et al. An M-estimator for tail dependence in arbitrary dimensions.* The Annals of Statistics, vol. 40, no. 3, pages 1764–1793, 2012.

[Hult & Lindskog 2006] Henrik Hult and Filip Lindskog. *Heavy-tailed insurance portfolios: buffer capital and ruin probabilities.* Rapport technique, vol. 1441, 2006.

[Joe *et al.* 1992] Harry Joe, Richard L Smith and Ishay Weissman. *Bivariate threshold methods for extremes.* Journal of the Royal Statistical Society. Series B (Methodological), pages 171–183, 1992.

[Parzen 1962] Emanuel Parzen. *On estimation of a probability density function and mode.* The annals of mathematical statistics, pages 1065–1076, 1962.

[Romano 1988] Joseph P Romano. *On weak convergence and optimality of kernel density estimates of the mode.* The Annals of Statistics, pages 629–647, 1988.

[Silverman 1981] Bernard W Silverman. *Using kernel density estimates to investigate multimodality.* Journal of the Royal Statistical Society. Series B (Methodological), pages 97–99, 1981.

[Vieu 1996] Philippe Vieu. *A note on density mode estimation.* Statistics & probability letters, vol. 26, no. 4, pages 297–307, 1996.

Chapitre 5

Conclusions et perspectives

Dans cette thèse, plusieurs façons d'estimer la dépendance et l'agrégation d'un vecteur de risques ont été proposées. Tout d'abord, on présente une méthode pour estimer le quantile (VaR) d'une somme de risques dépendants à des niveaux de confiance extrêmes. Toujours dans le contexte de l'estimation du risque agrégé, on introduit des nouvelles copules, basées sur la copule empirique et la copule échiquier, et on propose d'estimer le risque agrégé lorsque les risques marginaux sont connus mais la structure de dépendance est inconnue. Finalement, deux estimateurs pour la mesure spectrale ont été proposés afin d'estimer la dépendance dans les extrêmes.

La méthode présentée dans le deuxième chapitre propose d'approximer la VaR de la somme en utilisant la VaR du maximum, qui est généralement facile et simple à calculer si la fonction de répartition du vecteur de risques est connue. On a constaté que dans les modèles à queue épaisse et fortement dépendants cette approximation donne de bons résultats. Plus précisément on montre que cette approximation peut s'utiliser si ces risques individuels sont à variation régulière avec une copule (ou copule de survie) Archimédienne. Cependant une caractérisation précise des modèles pour lesquels la méthode est applicable n'a pas été trouvée. De même, nous avons présenté une méthode pour estimer le paramètre Δ mais ses propriétés statistiques n'ont pas été véritablement explorées. Donc, proposer des estimateurs pour la valeur Δ et en étudier les propriétés fait partie des perspectives possibles pour ce travail.

D'autre part, nous avons comparé notre estimateur avec un autre estimateur basé sur le rapport de la somme et l'un des risques, et qui n'utilise donc pas l'information sur la structure du vecteur de dépendance. Notre méthode, qui utilise, elle, l'information sur la dépendance du vecteur de risques donnée par la fonction de répartition du maximum, donne de meilleurs résultats dans les exemples étudiés. Ceci s'explique parce que la convergence du rapport $\overline{F}_S(t)/\overline{F}_M(t)$ semble bien plus rapide que celle du rapport $\overline{F}_S(t)/\overline{F}_1(t)$. Mesurer plus précisément le gain obtenu par notre méthode doit encore être exploré et peut faire l'objet d'un travail futur.

Dans le troisième chapitre, le concept de classe de copules invariantes par agrégation a été introduit et certaines propriétés pour cette classe ont été démontrées. En particulier, il a été montré que cette classe est infinie lorsque la fonction d'agrégation est le maximum ou la somme (sous certaines conditions du vecteur de risques). Nous pensons que ces résultats peuvent être généralisés, par exemple pour des fonctions d'agrégation plus générales que la somme ou le maximum. De toute façon, les résultats obtenus montrent que, si l'objectif principal est d'estimer le risque agrégé d'un vecteur, au moins pour la somme et le maximum, toute l'information contenue dans la structure de dépendance, matérialisée par la copule n'est pas nécessaire.

Dans ce contexte on introduit une copule que nous appelons copule échiquier empirique. Le but principal de cette copule n'est pas de décrire précisément la structure de dépendance qu'on suppose dans ce chapitre non connue, mais de l'utiliser spécifiquement pour estimer le risque agrégé. Les propriétés asymptotiques de la copule empirique et de la copule échiquier empirique sont similaires. Cependant cette dernière est plus appropriée pour estimer le risque agrégé et d'ailleurs elle peut intégrer différents types d'informations sur la dépendance comme l'information contenue dans un échantillon, dans un sous-vecteur ou dans les extrêmes.

Les résultats numériques présentés nous montrent que le gain obtenu lorsque on introduit de l'information dans la copule échiquier empirique est important. Cependant, théoriquement ce gain doit encore être plus précisément mesuré. Aussi, la taille de la partition (notée $\mathcal{I}_m$ dans le chapitre) est un paramètre qui joue un rôle important et son choix optimal peut être le sujet d'une étude future.

Dans le quatrième chapitre un estimateur de la mesure spectrale pour un vecteur à variation régulière de dimension 2 est proposé. La méthode d'estimation repose sur l'approximation par noyau des densités et sur des résultats sur l'estimation du mode. La consistance des estimateurs proposés a été démontrée sous certaines conditions. Il reste à étudier la vitesse de convergence ainsi que la normalité asymptotique. De même, la généralisation des résultats pour des vecteurs de dimension plus grande que 2 peut être envisagée.

Annexe A

Exemples de copules

Dans cet annexe on présentera quelques exemples de copules.

Exemple (**Copule indépendante**). *La copule indépendante (ou copule produit) est la copule des vecteurs aléatoires dont les marginales sont indépendantes,*

$$C(F_1(x_1), \ldots, F_d(x_d)) = F_1(x_1) \cdots F_d(x_d).$$

La copule indépendante est donc définie par la formule

$$C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i. \tag{A.1}$$

Cette copule caractérise l'indépendance entre les variables : les marginales d'un vecteur aléatoire continu $\boldsymbol{X}$ sont indépendantes si et seulement si la copule de $\boldsymbol{X}$ est la copule indépendante.

Exemple (**Copule de Gauss et copule de Student**). *Soit Φ la fonction de répartition d'une variable aléatoire normale centrée et réduite et Φ_R^d la fonction de répartition d'un vecteur normal standard de dimension d avec matrice de corrélation R. Alors*

$$C(u_1, \ldots, u_d) = \Phi_R^d(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)) \tag{A.2}$$

est la copule de Gauss ou copule normale de dimension d.

De manière identique, si Φ_ν es la fonction de répartition d'une variable Student à ν degrés de liberté et $\Phi_{R,\nu}^d$ est la fonction de répartition d'une Student multivariée avec matrice de corrélation R et ν degrés de liberté, alors

$$C(u_1, \ldots, u_d) = \Phi_{R,\nu}^d(\Phi_\nu^{-1}(u_1), \ldots, \Phi_\nu^{-1}(u_d)) \tag{A.3}$$

est la copule de Student en dimension d.

Copule archimédienne

Définition. Une fonction $\psi : [0,1] \to [0,\infty]$ est un *générateur d'ordre d* si elle satisfait les conditions :

(i) ψ est décroissante et $\psi(1) = 0$;

(ii) L'inverse généralisé de ψ, défini par

$$\psi^{\leftarrow}(t) = \inf\{u \in [0,1] \mid \psi(u) \leq t\} \text{ pour tout } t \in [0,\infty]$$

est d fois continûment dérivable.

(iii) Pour $k = 0, 1, \ldots d$,

$$(-1)^k \frac{d^k \psi^{\leftarrow}(t)}{dt^k} \geq 0, \text{ pour tout } t > 0.$$

Définition. (Copule archimédienne)
Une copule C est dite *archimédienne* s'il existe un générateur ψ d'ordre d tel que

$$C(u_1, \ldots, u_d) = \psi^{\leftarrow}\{\psi(u_1) + \cdots + \psi(u_d)\},$$

pour tout $(u_1, \ldots, u_d) \in [0,1]^d$. C est appelée la copule archimédienne de générateur ψ.

Exemple. *Les fonctions suivantes sont des générateurs de copules archimédienne.*

— $\psi(t) = (-\ln t)^\theta$, $\theta \geq 1$. *La copule engendrée,*

$$C_\theta(u_1, \ldots, u_d) = \exp\left(-\left([-\ln u_1]^\theta + \cdots + [-\ln u_d]^\theta\right)^{1/\theta}\right),$$

est connue sous le nom de copule de Gumbel.

— $\psi(t) = (t^{-1/\theta} - 1)$, $\theta > 0$. *La copule engendrée,*

$$C_\theta(u_1, \ldots, u_d) = \left(u_1^{-1/\theta} + \cdots + u_d^{-1/\theta} - (d-1)\right)^{-\theta},$$

est connue sous le nom de copule de Clayton

Copule empirique

Nous présentons maintenant l'estimation empirique de la copule. Elle porte le nom de copule empirique.

Définition. Considérons n copies, $\mathbf{X}^1, \ldots, \mathbf{X}^n$, indépendantes de $\mathbf{X}$. Chacune d'elles se décompose par $\mathbf{X}^j = (X_1^j, \ldots, X_d^j)$. Soit $R_i^1, \ldots, R_i^n$, $i = 1, \ldots, d$, les rang marginaux, i.e.,

$$R_i^j = \sum_{k=1}^{n} 1\{X_i^{(j)} \geq X_i^{(k)}\},\ i = 1, \ldots, d,\ j = 1, \ldots, n$$

où $X_i^{(1)} < \cdots < X_i^{(n)}$ sont les statistiques d'ordre associées à l' i-ème échantillon $X_i^1, \ldots X_i^n$. La copule empirique $\widehat{C}_n$ de $\mathbf{X}^1, \ldots \mathbf{X}^n$ est définie par

$$\widehat{C}_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{k=1}^{n} 1 \left\{ \frac{1}{n} R_1^k \leq u_1, \ldots, \frac{1}{n} R_d^k \leq u_d \right\}.$$

Il est bien connu (voir par exemple [Fermanian *et al.* 2004]) que la copule empirique peut être utilisée pour estimer C. Néanmoins, elle a le désavantage de ne pas être une copule, malgré son nom, car ses lois marginales ne sont pas uniformes en $[0, 1]$.

Copule échiquier

Soit μ_C la mesure de probabilité associée à C, c'est à dire, telle que :

$$\mu_C(\Pi_{i=1}^d [0, u_i]) = C(u_1, \ldots, u_d)$$

pour tout $u = (u_1, \ldots, u_d) \in [0, 1]^d$.
Par une μ-décomposition d'un ensemble $A \subset \mathbb{R}^d$, nous entendons une famille finie d'ensembles mesurables $\{A_i \subset A\}$ telle que

1. $\mu(A_i \cap A_j) = 0$ si $i \neq j$
2. $\sum_i \mu(A_i) = \mu(A)$.

Définition. Une mesure μ^* est une approximation échiquier d'une copule C s'il existe une λ-décomposition $\mathcal{A} = \{(a_i, b_i)\}$ de l'hypercube I^d, composée de d-intervalles tel que :

1. μ^* est uniforme sur (a_i, b_i).
2. $\mu^*(A) = \mu_C(A)$ pour tout $A \in \mathcal{A}$,

où λ est la mesure de Lebesgue en dimension d. Pour $m \in \mathbb{N}$, nous considérons la λ -décomposition régulière de l'hypercube $[0, 1]^d$ notée $\mathcal{I}_m$ et consistant en m^d d-cubes d'une longueur de côté $1/m$:

$$I_{i,m} = \prod_{j=1}^{d} \left[\frac{i_j - 1}{m}, \frac{i_j}{m} \right], \ i = (i_1, \ldots, i_d), \ i_j \in \{1, \ldots, m\}.$$

μ_m^* est l'approximation échiquier associé à la décomposition régulière $\mathcal{I}_m$.

Nous noterons C_m^* la copule échiquier associée à la mesure μ_m^*. La définition de la copule échiquier peut alors s'écrire comme :

$$C_m^*(x) = \sum_i m^d \mu_C(I_{i,m}) \lambda([0, x] \cap I_{i,m})$$

où $[0, x] = \prod_{i=1}^d [0, x_i]$, pour $x = (x_1 \ldots, x_d) \in [0, 1]^d$. Dans [Mikusinski & Taylor 2010] il est prouvé que C_m^* approxime C.

Les bornes de Fréchet

Définition. La *classe de Fréchet* de marginales $F_1, \ldots, F_d$, notée $\mathfrak{F}(F_1, \ldots, F_d)$, est l'ensemble de toutes les fonctions de répartitions de $\mathbb{R}^d$ dont la i-ème marge est F_i.

Grâce au Théorème de Sklar, $F \in \mathfrak{F}(F_1, \ldots, F_d)$ si et seulement s'il existe une copule C tel que $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$ pour tout $\mathbf{x} \in \mathbb{R}^d$.

Théorème. *Soient $C_W, C_M : [0,1]^d \to [0,1]$ les fonctions définies par :*

$$C_W(\boldsymbol{u}) = \max\{u_1 + \cdots + u_d - (d-1), 0\} \quad et \quad C_M(\boldsymbol{u}) = \min\{u_i, i = 1, \ldots, d\}.$$

Alors pour toute $F \in \mathfrak{F}(F_1, \ldots, F_d)$ et tout $\boldsymbol{x} \in \mathbb{R}^d$

$$C_W(F_1(x_1), \ldots, F_d(x_d)) \leq F(\boldsymbol{x}) \leq C_M(F_1(x_1), \ldots, F_d(x_d)).$$

Les fonctions C_W et C_M s'appellent les bornes supérieure et inférieure de Fréchet-Hoeffding.

Remarque. Il est possible de vérifier que C_M est une d-copule pour tout d, par contre C_W est une copule seulement quand $d = 2$.
Cela veut dire que la borne supérieure appartient toujours à la classe de Fréchet $\mathfrak{F}$. La borne inférieure fait partie de $\mathfrak{F}$ seulement si $d = 2$. Néanmoins cette borne est précise en toutes les dimensions dans le sens suivant :
Pour tout d et tout $u \in [0,1]^d$ il existe une copule C_u (qui dépend de $\mathbf{u}$) telle que $C_u(u) = C_W(u)$.

Remarque. La copule C_M représente la structure de dépendance parfaite positive entre les variables : C_M est la copule de $\mathbf{X}$ si on peut trouver une variable aléatoire réelle Z et d fonctions croissantes $h_i, i = 1, \ldots, d$, telles que $\mathbf{X} \overset{\mathcal{L}}{=} (h_1(Z), \ldots, h_d(Z))$.
Au contraire, la copule C_W en dimension $d = 2$ représente la dépendance parfaite négative : C_W est la copule de $\mathbf{X}$ s'il existe une variable aléatoire Z et une fonction h décroissante tel que $X \overset{\mathcal{L}}{=} (Z, h(Z))$.

Bibliographie

[Albrecher *et al.* 2011] Hansjörg Albrecher, Corina Constantinescu et Stéphane Loisel. *Explicit ruin formulas for models with dependence among risks*. Insurance : Mathematics and Economics, vol. 48, pages 265–270, 2011. (Cité en page 32.)

[Alink *et al.* 2004] Stan Alink, Matthias Löwe et Mario V. Wüthrich. *Diversification of aggregate dependent risks*. Insurance : Mathematics and Economics, vol. 35, no. 1, pages 77–95, 2004. (Cité en pages 20 et 22.)

[Alink *et al.* 2005] Stan Alink, Matthias Löwe et Mario V. Wüthrich. *Analysis of the expected shortfall of aggregate dependent risks*. ASTIN Bulletin, vol. 35, no. 1, pages 25–43, 2005. (Cité en pages 20 et 22.)

[Alink *et al.* 2007] Stan Alink, Matthias Löwe et Mario V. Wüthrich. *Diversification for general copula dependence*. Statistica Neerlandica, vol. 61, no. 4, pages 446–465, Novembre 2007. (Cité en pages 10, 20 et 22.)

[Arbenz *et al.* 2011] P. Arbenz, P. Embrechts et G. Puccetti. *The* AEP *algorithm for the fast computation of the distribution of the sum of dependent random variables.* Bernoulli, vol. 17, no. 2, pages 562–591, 2011. (Cité en pages 9 et 19.)

[Asmussen & Albrecher 2010] Søren Asmussen et Hansjörg Albrecher. Ruin probabilities, volume 14. World scientific, 2010. (Cité en page 74.)

[Barbe & McCormick 2005] Philippe Barbe et William P McCormick. *Asymptotic expansions of convolutions of regularly varying distributions*. Journal of the Australian Mathematical Society, vol. 78, no. 03, pages 339–371, 2005. (Cité en page 33.)

[Barbe *et al.* 2006] Philippe Barbe, Anne-Laure Fougères et Christian Genest. *On the Tail Behavior of Sums of Dependent Risks*. ASTIN Bulletin, vol. 36, no. 2, pages 361–373, 2006. (Cité en pages 10, 20 et 22.)

[Beirlant *et al.* 2006] Jan Beirlant, Yuri Goegebeur, Johan Segers et Jozef Teugels. Statistics of extremes : theory and applications. John Wiley & Sons, 2006. (Cité en pages 7 et 73.)

[Bernard & Vanduffel 2015] Carole Bernard et Steven Vanduffel. *A new approach to assessing model risk in high dimensions*. Journal of Banking & Finance, vol. 58, pages 166–178, 2015. (Cité en pages 19 et 51.)

[Bernard *et al.* 2013] Carole Bernard, Ludger Rüschendorf et Steven Vanduffel. *Value-at-Risk bounds with variance constraints*. Available at SSRN 2342068, 2013. (Cité en pages 19 et 51.)

[Bernard *et al.* 2014] Carole Bernard, Xiao Jiang et Ruodu Wang. *Risk Aggregation with Dependence Uncertainty*. Insurance : Mathematics and Economics, vol. 54, no. 1, pages 93–108, 2014. (Cité en pages 12 et 52.)

[Bingham *et al.* 1989] Nicholas H Bingham, Charles M Goldie et Jef L Teugels. Regular variation, volume 27. Cambridge university press, 1989. (Cité en pages 26 et 41.)

[Cénac *et al.* 2014] Peggy Cénac, Stephane Loisel, Véronique Maume-Deschamps et Clémentine Prieur. *Risk indicators with several lines of business : comparison, asymptotic behavior and applications to optimal reserve allocation*. Annales de l'ISUP, vol. 58, no. 3, pages 3–26, 2014. (Cité en page 32.)

[Cline & Samorodnitsky 1994] Daren BH Cline et Gennady Samorodnitsky. *Subexponentiality of the product of independent random variables*. Stochastic Processes and their Applications, vol. 49, no. 1, pages 75–98, 1994. (Cité en page 41.)

[Coles & Tawn 1991] Stuart G Coles et Jonathan A Tawn. *Modelling extreme multivariate events*. Journal of the Royal Statistical Society. Series B (Methodological), pages 377–392, 1991. (Cité en pages 14 et 72.)

[Cossette *et al.* 2014] Hélène Cossette, Marie-Pier Côté, Mélina Mailhot et Etienne Marceau. *A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks*. Journal of Multivariate Analysis, vol. 130, pages 1–20, 2014. (Cité en pages 19 et 51.)

[Dacorogna *et al.* 2014] Michel Dacorogna, Leila El Bahtouri et Marie Kratz. Explicit diversification benefit for dependent risks. Preprint, 2014. (Cité en pages 20, 32 et 63.)

[De Haan & Ferreira 2007] Laurens De Haan et Ana Ferreira. Extreme value theory : an introduction. Springer Science & Business Media, 2007. (Cité en pages 7 et 73.)

[de Haan & Resnick 1977] Laurens de Haan et Sidney I Resnick. *Limit theory for multivariate sample extremes*. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 40, no. 4, pages 317–337, 1977. (Cité en pages 7 et 73.)

[Deheuvels 1979] Paul Deheuvels. *La fonction de dépendance empirique et ses propriétés*. Acad. Roy. Belg. Bull. Cl. Sci., vol. 65, no. 5, pages 274 – 292, 1979. (Cité en pages 6, 13 et 59.)

[Dubey 1970] Satya D Dubey. *Compound gamma, beta and F distributions*. Metrika, vol. 16, no. 1, pages 27–31, 1970. (Cité en pages 32 et 63.)

[Dvoretzky *et al.* 1956] Aryeh Dvoretzky, Jack Kiefer et Jacob Wolfowitz. *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*. The Annals of Mathematical Statistics, pages 642–669, 1956. (Cité en page 83.)

[Eddy 1980] William F. Eddy. *Density Estimation and Bump-Hunting by the Penalized Likelihood Method Exemplified by Scattering and Meteorite Data : Comment*. Journal of the American Statistical Association, vol. 75, no. 369, pages pp. 59–61, 1980. (Cité en pages 83 et 86.)

[Einmahl & Segers 2009] John HJ Einmahl et Johan Segers. *Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution*. The Annals of Statistics, pages 2953–2989, 2009. (Cité en pages 14 et 72.)

[Einmahl *et al.* 1997] John HJ Einmahl, Laurens de Haan et Ashoke Kumar Sinha. *Estimating the spectral measure of an extreme value distribution*. Stochastic Processes and their Applications, vol. 70, no. 2, pages 143–171, 1997. (Cité en pages 14 et 72.)

[Einmahl *et al.* 2001] John HJ Einmahl, Laurens de Haan et Vladimir I Piterbarg. *Nonparametric estimation of the spectral measure of an extreme value distribution*. Annals of Statistics, pages 1401–1423, 2001. (Cité en pages 14 et 72.)

[Einmahl *et al.* 2012] John HJ Einmahl, Andrea Krajina, Johan Segers*et al. An M-estimator for tail dependence in arbitrary dimensions*. The Annals of Statistics, vol. 40, no. 3, pages 1764–1793, 2012. (Cité en page 72.)

[Embrechts & Puccetti 2010] Paul Embrechts et Giovanni Puccetti. *Risk aggregation*. Copula Theory and Its Applications, 2010. (Cité en page 51.)

[Embrechts *et al.* 1997] Paul Embrechts, Claudia Klüppelberg et Thomas Mikosch. Modelling extremal events : for insurance and finance, volume 33. Springer, 1997. (Cité en pages 9 et 19.)

[Embrechts *et al.* 1999] Paul Embrechts, Alexander McNeil et Daniel Straumann. *Correlation : pitfalls and alternatives*. RISK, vol. 12, no. 5, pages 69–71, 1999. (Cité en page 5.)

[Embrechts *et al.* 2009] Paul Embrechts, Johanna Neslehová et Mario V Wüthrich. *Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness*. Insurance : Mathematics and Economics, vol. 44, no. 2, pages 164–169, 2009. (Cité en page 20.)

[Embrechts *et al.* 2013a] Paul Embrechts, Giovanni Puccetti et Ludger Rüschendorf. *Model uncertainty and VaR aggregation*. Journal of Banking & Finance, vol. 37, no. 8, pages 2750–2764, 2013. (Cité en page 19.)

[Embrechts *et al.* 2013b] Paul Embrechts, Giovanni Puccetti et Ludger Rüschendorf. *Model uncertainty and VaR aggregation*. Journal of Banking & Finance, vol. 37, no. 8, pages 2750–2764, 2013. (Cité en page 50.)

[Erdely *et al.* 2014] Arturo Erdely, José M González-Barrios et María M Hernández-Cedillo. *Frank's condition for multivariate Archimedean copulas*. Fuzzy Sets and Systems, vol. 240, pages 131–136, 2014. (Cité en page 54.)

[Fermanian *et al.* 2004] Jean-David Fermanian, Dragan Radulovic et Marten Wegkamp. *Weak convergence of empirical copula processes*. Bernoulli, vol. 10, no. 5, pages 847–960, 2004. (Cité en pages 59, 63 et 93.)

[Fougères & Mercadier 2012] Anne-Laure Fougères et Cécile Mercadier. *Risk measures and multivariate extensions of Breiman's theorem*. Journal of Applied Probability, vol. 49, pages 364–384, 2012. (Cité en page 19.)

[Frank 1996] M.J. Frank. *Diagonals of copulas and Schröeder's equation*. Procedings Aequ. Math., no. 150, 1996. (Cité en page 54.)

[Frees & Valdez 1998] Edward W Frees et Emiliano A Valdez. *Understanding relationships using copulas*. North American actuarial journal, vol. 2, no. 1, pages 1–25, 1998. (Cité en page 5.)

[Geluk & Tang 2009] Jaap Geluk et Qihe Tang. *Asymptotic tail probabilities of sums of dependent subexponential random variables*. Journal of Theoretical Probability, vol. 22, no. 4, pages 871–882, 2009. (Cité en page 19.)

[Genest & MacKay 1986] Christian Genest et R Jock MacKay. *Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données*. Canadian Journal of Statistics, vol. 14, no. 2, pages 145–159, 1986. (Cité en page 5.)

[Hoeffding 1940] Wassily Hoeffding. *ScaleâInvariant Correlation Theory*. In The collected works of Wassily Hoeffding, pages 57–107. Springer, 1940. (Cité en page 5.)

[Hult & Lindskog 2006] Henrik Hult et Filip Lindskog. *Heavy-tailed insurance portfolios : buffer capital and ruin probabilities*. Rapport technique, vol. 1441, 2006. (Cité en pages 72 et 75.)

[Joe *et al.* 1992] Harry Joe, Richard L Smith et Ishay Weissman. *Bivariate threshold methods for extremes*. Journal of the Royal Statistical Society. Series B (Methodological), pages 171–183, 1992. (Cité en pages 14 et 72.)

[Joe 1997] Harry Joe. Multivariate models and multivariate dependence concepts. CRC Press, 1997. (Cité en page 5.)

[Kortschak & Albrecher 2009] Dominik Kortschak et Hansjörg Albrecher. *Asymptotic results for the sum of dependent non-identically distributed random variables*. Methodology and Computing in Applied Probability, vol. 11, no. 3, pages 279–306, 2009. (Cité en page 19.)

[Kratz 2014] Marie Kratz. *Normex, a new method for evaluating the distribution of aggregated heavy tailed risks*. Extremes, vol. 17, no. 4, pages 661–691, 2014. (Cité en page 20.)

[Li *et al.* 1998] Xin Li, P Mikusiński et Michael D Taylor. *Strong approximation of copulas*. Journal of Mathematical Analysis and Applications, vol. 225, no. 2, pages 608–623, 1998. (Cité en pages 56, 57, 60 et 61.)

[Marshall & Olkin 1988] Albert W Marshall et Ingram Olkin. *Families of multivariate distributions*. Journal of the American Statistical Association, vol. 83, no. 403, pages 834–841, 1988. (Cité en pages 20, 32 et 63.)

[Maume-Deschamps *et al.* 2015] Véronique Maume-Deschamps, Didier Rullière et Khalil Saïd. On capital allocation by minimizing multivariate risk indicators. Preprint, 2015. (Cité en pages 32 et 63.)

[McNeil *et al.* 2005] Alexander J McNeil, Rüdiger Frey et Paul Embrechts. Quantitative risk management : Concepts, techniques and tools. Princeton university press, 2005. (Cité en page 5.)

[Mikusinski & Taylor 2010] Piotr Mikusinski et Michael D Taylor. *Some approximations of n-copulas*. Metrika, vol. 72, no. 3, pages 385–414, 2010. (Cité en pages 13, 51, 57 et 93.)

[Nelsen *et al.* 2008] Roger B Nelsen, José Juan Quesada-Molina, José Antonio Rodríguez-Lallena et Manuel Úbeda-Flores. *On the construction of copulas and quasi-copulas with given diagonal sections*. Insurance : Mathematics and Economics, vol. 42, no. 2, pages 473–483, 2008. (Cité en page 54.)

[Nelsen 1999] Roger B Nelsen. An introduction to copulas. Springer, 1999. (Cité en pages 5 et 6.)

[Nguyen & Robert 2014] Quang Huy Nguyen et Christian Y Robert. *New efficient estimators in rare event simulation with heavy tails*. Journal of Computational and Applied Mathematics, vol. 261, pages 39–47, 2014. (Cité en page 20.)

[Oakes 1989] David Oakes. *Bivariate survival models induced by frailties*. Journal of the American Statistical Association, vol. 84, no. 406, pages 487–493, 1989. (Cité en pages 20, 32 et 63.)

[Parzen 1962] Emanuel Parzen. *On estimation of a probability density function and mode*. The annals of mathematical statistics, pages 1065–1076, 1962. (Cité en page 81.)

[Puccetti & Rüschendorf 2012] Giovanni Puccetti et Ludger Rüschendorf. *Computation of sharp bounds on the distribution of a function of dependent risks*. Journal of Computational and Applied Mathematics, vol. 236, no. 7, pages 1833–1840, 2012. (Cité en page 50.)

[Resnick 2007] Sidney I Resnick. Heavy-tail phenomena : probabilistic and statistical modeling. Springer, 2007. (Cité en pages 7 et 27.)

[Romano 1988] Joseph P Romano. *On weak convergence and optimality of kernel density estimates of the mode*. The Annals of Statistics, pages 629–647, 1988. (Cité en page 81.)

[Rüschendorf 1982] Ludger Rüschendorf. *Random variables with maximum sums*. Advances in Applied Probability, pages 623–632, 1982. (Cité en page 50.)

[Rüschendorf 1983] Ludger Rüschendorf. *Solution of a statistical optimization problem by rearrangement methods*. Metrika, vol. 30, no. 1, pages 55–61, 1983. (Cité en page 50.)

[Silverman 1981] Bernard W Silverman. *Using kernel density estimates to investigate multimodality*. Journal of the Royal Statistical Society. Series B (Methodological), pages 97–99, 1981. (Cité en page 83.)

[Sklar 1959] M Sklar. Fonctions de répartition à n dimensions et leurs marges. Université Paris 8, 1959. (Cité en page 5.)

[Sungur & Yang 1996] Engin A Sungur et Yimin Yang. *Diagonal copulas of Archimedean class*. Communications in Statistics-Theory and Methods, vol. 25, no. 7, pages 1659–1676, 1996. (Cité en page 54.)

[Vieu 1996] Philippe Vieu. *A note on density mode estimation*. Statistics & probability letters, vol. 26, no. 4, pages 297–307, 1996. (Cité en page 81.)

[Weinzierl 2000] Stefan Weinzierl. *Introduction to monte carlo methods*. arXiv preprint hep-ph/0006269, 2000. (Cité en page 18.)

[Weissman 1978] Ishay Weissman. *Estimation of parameters and large quantiles based on the k largest observations*. Journal of the American Statistical Association, vol. 73, no. 364, pages 812–815, 1978. (Cité en pages 9, 19 et 34.)

[Yeh 2007] Hsiaw-Chan Yeh. *The frailty and the Archimedean structure of the general multivariate Pareto distributions*. Bulletin Institute of Mathematics Acedemia Sinica, vol. 2, no. 3, pages 713–729, 2007. (Cité en page 32.)

Modélisation de la dépendance et estimation du risque agrégé

Résumé : Cette thèse porte sur l'étude de la modélisation et estimation de la dépendance des portefeuilles de risques et l'estimation du risque agrégé.

Dans le Chapitre 2, nous proposons une nouvelle méthode pour estimer les quantiles de haut niveau pour une somme de risques. Elle est basée sur l'estimation du rapport entre la VaR de la somme et la VaR du maximum des risques. Nous utilisons des résultats sur les fonctions à variation régulière. Nous comparons l'efficacité de notre méthode avec quelques estimations basées sur la théorie des valeurs extrêmes, sur plusieurs modèles. Notre méthode donne de bons résultats lors de l'approximation de la VaR à des niveaux élevés lorsque les risques sont fortement dépendants et au moins l'un des risques est à queue épaisse.

Dans le Chapitre 3, nous proposons une procédure d'estimation pour la distribution d'un risque agrégé basée sur la copule échiquier. Elle permet d'obtenir de bonnes estimations à partir d'un petit échantillon de la loi multivariée et une connaissance complète des lois marginales. Cette situation est réaliste pour de nombreuses applications. Les estimations peuvent être améliorées en incluant dans la copule échiquier des informations supplémentaires (sur la loi d'un sous-vecteur ou sur des probabilités extrêmes). Notre approche est illustrée par des exemples numériques.

Finalement, dans le Chapitre 4, nous proposons un estimateur de la mesure spectrale basé sur l'estimation à noyau de la densité de la mesure spectrale d'une distribution à variation régulière bivariée. Une extension de notre méthode permet d'estimer la mesure spectrale discrète. Certaines propriétés de convergence sont obtenues.

Mots clés : Agrégation des risques, copules, copules empiriques, copules échiquier, estimation de la VaR, fonctions à variation régulier , mesure spectrale

Dependence modelling and risk aggregation estimation

Abstract : This thesis comprises three essays on estimation methods for the dependence between risks and its aggregation.

In the first essay we propose a new method to estimate high level quantiles of sums of risks. It is based on the estimation of the ratio between the VaR (or TVaR) of the sum and the VaR (or TVaR) of the maximum of the risks. We use results on regularly varying functions. We compare the efficiency of our method with classical ones, on several models. Our method gives good results when approximating the VaR or TVaR in high levels on strongly dependent risks where at least one of the risks is heavy tailed.

In the second essay we propose an estimation procedure for the distribution of an aggregated risk based on the *checkerboard copula*. It allows to get good estimations from a (quite) small sample of the multivariate law and a full knowledge of the marginal laws. This situation is realistic for many applications. Estimations may be improved by including in the checkerboard copula some additional information (on the law of a sub-vector or on extreme probabilities). Our approach is illustrated by numerical examples.

In the third essay we propose a kernel based estimator for the spectral measure density of a bivariate distribution with regular variation. An extension of our method allows to estimate discrete spectral measures. Some convergence properties are obtained.

Keywords: risk aggregation, copulas, empirical copulas, checkerboard copulas, value at risk estimation, regularly varying functions, spectral measure