Katholieke Universiteit Leuven

Faculteit Wetenschappen
Departement Wiskunde

Dependence structures and limiting results, with applications in finance and insurance

Arthur Charpentier

Promotors:
Prof. Dr. J. Beirlant
Prof. Dr. M. Denuit
Jury:
Prof. Dr. A.L. Fougères
Prof. Dr. I. Gijbels
Prof. Dr. J. Dhaene
Prof. Dr. C. Gouriéroux
Prof. Dr. W. Schoutens

Proefschrift ingediend tot het behalen van de graad van Doctor in de Wetenschappen

Leuven 2006

Remerciements

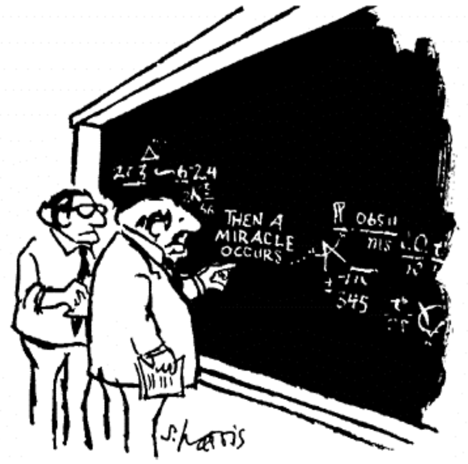
Une saine tradition veut que l'on remercie ceux qui ont, de près ou de loin, contribué à la finalisation de cette thèse, qui m'ont apporté en particulier la confiance nécessaire pour cette aventure. Mon enthousiasme (et peut être aussi le temps qu'il m'aura fallu pour boucler cette thèse) me conduit sûrement à faire une liste un peu longue, mais je tenais vraiment à remercier certaines personnes.

Avant tout, je tenais à remercier Jan Beirlant et Michel Denuit, qui ont accepté d'encadrer cette thèse. Merci à Jan pour sa disponibilité, lui qui (malgré toutes ses activités prenantes) trouvait toujours le temps pour passer me voir dès que j'étais à Leuven. Ses commentaires et réflexions m'ont ainsi permis de trouver des applications connexes, vastes et intéressantes, aux travaux que je menais. Merci aussi pour l'aide apportée pour l'inscription à la Katholieke Universiteit Leuven. Merci à Michel pour ses encouragements constants, et pour les nombreuses discussions que nous avons pu avoir. Je lui dois une partie de l'orientation de cette thèse. Je n'oublie pas qu'il a réussi à me faire croire que j'avais du temps libre pour écrire des livres (et que cette thèse, ainsi que les quelques cours que j'ai pu donner, ne suffiraient pas à occuper ces quelques années). Merci également de m'avoir fait découvrir quelques bières belges, qui malheureusement rendent délicates les interventions tôt le lendemain matin.

Je souhaitais également remercier Anne-Laure Fougères, Irène Gijbels, Jan Dhaene, Christian Gouriéroux et Wim Schoutens de m'avoir fait l'honneur d'accepter de faire partie du jury de cette thèse. Compte tenu de leurs emplois du temps chargés, je suis d'autant plus honoré qu'ils aient accepté de lire de façon aussi minutieuse ce travail.

Merci aussi à Marco Scarsini, Roger Nelsen, Moshe Shaked, Mark Meerschaert, Edward Omey, Victor Rivero, Johan Segers et Christian Genest, pour avoir pris le temps de répondre à mes nombreuses questions, y compris les plus naïves, et qui m'ont très souvent orienté vers des articles que je n'avais pas vu passer. J'aurais aussi une pensée pour Bruno Bassan avec qui nous avions longuement discuté pour voir si nos deux approches était complémentaires, sur la dynamique dans les modèles de durées. J'ai infiniment de regrets que notre collaboration n'ait pas pu aboutir.

Je souhaitais aussi remercier également mes co-auteurs. Molti grazie al Alessandro Juri (e Mario Wüthrich) ! Many thanks to Johan Segers who tried to help me to write down equations meticulously, or scrupulously.



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Parmi les autres coauteurs, un grand merci à Jean-David Fermanian et Olivier Scaillet, dont les intuitions sur les estimateurs à noyaux m'impressionneront toujours...

Je tenais aussi à remercier Jean-Luc Besson et Alain Trognon, qui m'auront tous deux permis de trouver un financement pour cette thèse, mais surtout, au delà de cet aspect vénal, pour les remarques et conseils qu'ils m'ont prodigués.

Merci à Jan Beirlant, Jan Dhaene, Marc Goovaerts et Christian Gouriéroux qui m'ont accueilli dans leurs labos, à Leuven ou à Paris, me permettant de cotoyer des personnes qui m'ont beaucoup aidé et appris, et qui, matériellement, m'ont permis d'avoir ainsi accès à un certain nombre d'ouvrages et d'articles.

J'aurais aussi une petite pensée pour tous ceux que j'aurais eu plaisir à croiser continuellement dans les diverses conférences, Emiliano Valdez, Jan Dhaene, Etienne Marceau, Paul Embrechts, Anne-Catherine Favre, Stéphane Loisel... Les discussions que j'ai pu avoir avec eux ont très largement nourri (et inspiré) cette thèse, ou divers travaux menés en parrallèle. Merci aussi à un certain nombre de professionels (essentiellement de l'assurance) qui m'ont constament soumis des problèmes intéressants et stimulants, François Bucchini, Gilles Bénéplanc, Michel Laparra ou encore Jean Berthon.

Parmi les us et coutumes relatifs au doctorat, il y a l'incontournable soutenance. J'aurais souhaité remercier Philippe Berthet, Philippe Briand, et Florent Malrieu qui ont accepté de perdre un peu de leur temps pour me permettre de répéter sous leur yeux. Merci à toutes leurs critiques et conseils.

Ma rencontre avec Thierry Roncalli en 1998 m'avait poussé à me mettre à Gauss, puis j'ai redécouvert SAS pour des raisons professionelles, avant que Michel Denuit ne me pousse à me mettre à SPlus, et qu'enfin David Sibaï ne me pousse à franchir le pas de ne plus travailler qu'avec R... Ceci explique en partie que les graphiques dans cette thèse manquent parfois de cohérence puisqu'ils ont été fait avec ces trois logiciels. J'aurais

d'ailleurs souhaité remercier en plus de David, André Calmels et Philippe Donnay qui m'ont dévoilé quelques mystères de l'informatique. A moins que ce ne soit le contraire, et que depuis que je les entends discuter, l'informatique me paraît de plus en plus mystérieux... je ne sais plus.

Ma reconnaissance va aussi vers tous les membres des laboratoires de Finance et Assurance, et de Statistique, du CREST (en particulier Patrice Bertail, Christian Robert, Jean-David Fermanian, Benoît Menoni, mais surtout Romuald Elie dont le bureau est aussi resté longtemps éclairé tard le soir, et pour les soirées chez *Papa*), de l'entresol de l'ENSAE (Laure Turner, Vincent Loonis, Emmanuel Jessua, Xavier Mary , Stephan Morley-Pedge, Emmanuel Masse, Xavier d'Haultfeuille, Philippe Février et Sylviane Gastaldo (pour leur immense soutien)), de la Fédération Française des Sociétés d'Asurance (et plus particulièrement Julie Dauriol et Benoît Lebrun qui m'auront supporté plus d'un an, et qui ont accepté mon vélo dans le bureau sans (trop) râler), à Hong Kong (en particulier Evariste Yeung qui aura réussi à me faire croire que j'étais capable de me couper les cheveux tout seul), du labo d'actuariat à Leuven (Katrien Antonio et Tom Hoedemakers), du labo de probas de Paris VI (y especialmente mi familia mejicana, Victor Rivero (y Maeka Rozada), Joaquin Fontbana, Karine Bertin), et le clan rennais (Jean-Baptiste et Magali Bardet, Florent et Lise Malrieu). J'aurais aussi une pensée pour plusieurs personnes croisées à Dauphine (Martine Bellec, Christian Hess, Pierre Cazes, Nizar Touzi) qui m'ont donné l'opportunité de découvrir les joies de l'enseignement à Dauphine. Leurs conseils en la matière m'ont été très précieux ! Merci aussi à Alain Chateauneuf, Dimitrios Konstantinides, Jean Claude Augros et Jean-Jacques Lilti qui m'ont solicité pour donner encore plus de cours. Merci d'ailleurs à toutes les autres personnes que j'ai pu cotoyer durant ces trois ans, qui m'auront communiqué leur passion pour l'enseignement et pour la recherche. Il y a aussi toutes les autres personnes que j'ai sûrement oubliées ici, mais que je n'oublie pas.

Pour finir, merci aux amis et à la famille pour leur soutien et leurs encouragements. Un grand merci à ma mère pour son sampiternel "*et ta thèse, ça en est où ?* ". C'est sûrement en grande partie pour faire cesser ce questionnement que je me suis décidé à la finir un jour. Sinon un immense remerciement à Hélène qui aura su m'encourager et me soutenir. Et j'aurais enfin une petite pensée pour Maël et Romane - et leurs "*perturbations aléatoires*" - qui auront su me faire croire que leurs petits problèmes étaient autrement plus importants que tout ce qui a pu être abordé dans cette thèse. Et au delà des remerciements, je tenais à leur présenter des excuses pour avoir été aussi peu présent pendant une trop longue partie de ces quelques années...

Table of contents

Introduction et pricipaux résultats

La littérature des articles en mathématiques financières ou en sciences actuarielles de ces dernières décennies permet de voir que parmi les sujets fondamentaux figurent l'impact des queues de distributions et des événements extrêmes, ainsi que la dépendance entre risques. Dans le cadre de l'analyse du portefeuille, Markowitz (1952) a ainsi montré que la diversification optimale du risque d'un portefeuille devait dépendre des corrélations entre les différents actifs. Plus récemment, un grand nombre de faits stylisés ont montré qu'il était nécessaire, dans des problématiques de risque de crédit, de prendre en compte les liaisons ou interactions qui peuvent exister entre différents emprunteurs, pouvant engendrer des contagions entre les défauts, ou faillites.

Dans un tout autre contexte, comme le notait *Le Monde* en septembre 2002, "*la destruction des tours de New York a provoqué une secousse sans précédent chez les assureurs, mettant à mal tout leur système de fonctionnement. Aucun d'entre eux n'avait prévu qu'une telle corrélation entre les branches « vie » et « non vie » de l'assurance soit possible, mêlant les dommages aux biens, les pertes d'exploitation, les accidents du travail et les décès. Ce cumul - et le coût qui en découle estimé entre* 36 *et* 54 *milliards de dollars (entre* $36,7$ *milliards et* $55,1$ *milliards d'euros) - oblige la profession à reconsidérer sa manière d'appréhender les catastrophes*". Alors que l'hypothèse d'indépendance était sous-jacente, et fondamentale, dans la plupart des modèles actuariels, ce risque de cumul a montré qu'il n'était plus possible de modéliser les risques indépendamment les uns des autres.

C'est dans ce contexte que s'inscrit cette thèse, afin d'étudier les aspects nonlinéaires des structures de dépendance, dans un cadre de gestion des risques multiples. Nous nous intéresserons plus particulièrement aux changements quant à la structure de dépendance: la dépendance qui peut exister entre événements extrêmes n'est pas nécessairement du même type que celle qui lie l'ensemble des risques. Nous insisterons en particulier sur les résultats limites, quand toutes les composantes (e.g. les coûts associés) sont relativement grandes.

1 Modéliser les risques multiples

Cette partie est développée dans le Chapitre 1.

Le chapitre introductif de cette thèse propose une revue de la littérature sur la modélisation des risques multiples, en introduisant les principales notions qui seront développées dans cette thèse. La notion fondamentale dans toute cette thèse est celle de *classe de Fréchet*: soit $F_1, ..., F_d$ des fonctions de répartition (univariée), que l'on supposera continues dans l'intégralité de la thèse, on note $\mathcal{F}(F_1, ..., F_d)$ l'ensemble des fonctions de répartition $\mathbb{R}^d \to \mathbb{R}$ dont les lois marginales sont respectivement $F_1, ..., F_d$. Un cas particulier est celui de la classe des fonctions *copules* $\mathcal{C}$, où les lois marginales sont uniformes,

Définition 1.1. *Une copule C est une fonction de répartition sur $[0,1]^d$ d-dimensionnelle, dont les lois marginales sont uniformes sur $[0,1]$, où d est un entier strictement positif.*

On parlera aussi de copule associée à une fonction de répartition ou, par abus de langage, à un vecteur aléatoire:

Définition 1.2. *Soit $\boldsymbol{X} = (X_1, ..., X_d)$ un vecteur aléatoire de $\mathbb{R}^d$ dont les marges sont continues (les fonctions de répartitions marginales $F_i(x) = \mathbb{P}(X_i \leq x)$ sont continues). La copule associée à $\boldsymbol{X}$ est la fonction de répartition du vecteur $(F_1(X_1), ..., F_d(X_d))$, i.e.*

$$C(u_1, ..., u_n) = \mathbb{P}(F_1(X_1) \leq u_1, ..., F_d(X_d) \leq u_d), (u_1, ..., u_d) \in [0,1]^d. \tag{1}$$

On notera que la définition 1.2 est bien dans le cadre de la définition 1.1 puisque pour tout i, $F_i(X_i)$ suit une loi uniforme sur $[0,1]$. Les copules (selon la terminologie retenue par Sklar (1959)) sont alors les "*fonctions de dépendance*" définie par Deheuvels (1979): ces fonctions ne dépendent que de la *structure de dépendance* des composantes du vecteur $\boldsymbol{X}$, et elles sont en particulier invariantes par transformation croissante des marges. Ainsi, le vecteur $(X_1, ..., X_d)$ a la même copule que $(\phi_1(X_1), ..., \phi_d(X_d))$ pour des fonctions ϕ_i croissantes. De plus, ce sont des fonctions qui "*couplent*" les lois marginales, au sens où

$$F_{\mathbf{X}}(\boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, ..., X_d \leq x_d) = C(F_1(x_1), ..., F_d(x_d)). \tag{2}$$

Remarque 1.3. *Cette invariance par transformation croissante des marges est un point essentiel qui légitime dans toute cette thèse l'utilisation des copules comme outils central afin de modéliser la "dépendance". Transformer les marges permet en effet de les rendre comparables. Dans la théorie des extrêmes multivariés (Resnick* (1987)*), les marges sont généralement ramenées à une même loi (e.g. Fréchet de paramètre* 1*), car sinon, une composante dont la queue serait beaucoup plus épaisse que les autres écraserait toutes les autres. En transformant les marges (par des transformations croissantes), les extrêmes sont préservés. Notons que Spearman (1906) et Hoeffding (1940) notaient déjà l'importance de la transformation des marges afin de rendre les composantes "comparables".*

Une illustration de l'utilité des copules peut être le cas de l'indépendance. Les variables aléatoires $X_1, ..., X_d$ sont indépendantes si et seulement si la copule du vecteur $\boldsymbol{X} = (X_1, ..., X_d)$ est la copule indépendante $C^{\perp}(u_1, ..., u_d) = u_1 \times ... \times u_d$ (parfois appelée aussi copule produit). On parlera de *comonotonie* (ou de *dépendance parfaite positive*) si toute variable X_i s'écrit comme une fonction croissante de n'importe quelle autre variable X_j. Cette notion se caractérise en terme de copule par le fait que la copule associée est nécessairement $C^+(u_1, ..., u_d) = \min\{u_1, ..., u_d\}$.

Un exemple important de copules, qui sera traité dans plusieurs chapitres de cette thèse, est celui des copules Archimédiennes:

Définition 1.4. *Une copule C est dite Archimédienne s'il existe une fonction $\phi : [0,1] \mapsto \mathbb{R}^+$, telle que l'inverse $\phi^{\leftarrow}(\cdot) = \inf\{u \in [0,1], \phi(u) \leq \cdot\}$ soit complètement monotone à l'ordre d, i.e.*

$$(-1)^k \frac{d^k \phi^{\leftarrow}(t)}{dt^k} \geq 0, \text{ pour } k = 0, 1, ..., d,$$

avec $\phi(1) = 0$*, telle que* $C(u_1, ..., u_d) = \phi^{\leftarrow}(\phi(u_1) + ... + \phi(u_d))$*. La fonction* ϕ *est appelée générateur de la copule.*

Ces copules Archimédiennes correspondent à un cas d'indépendance conditionnelle, la loi du facteur d'exogénéité pouvant être reliée au générateur via sa transformée de Laplace (on parlera d'approche dite par "*frailty*"): soit $\boldsymbol{X} = (X_1, ..., X_d)$ un vecteur aléatoire, tel que les composantes soient indépendantes, conditionnellement à un facteur d'exogénéité aléatoire Θ. On suppose de plus que

$$\mathbb{P}(X_i \leq x_i | \Theta = \theta) = G_i(x_i)^\theta, \text{ où } G_i \text{ est une fonction de répartition.} \quad (3)$$

La copule de $\boldsymbol{X}$ est une copule Archimédienne, de générateur $\phi = \psi^{\leftarrow}$, où ψ est la transformée de Laplace de Θ. En fait les copules Archimédiennes sont un cas particulier des $\mathcal{H}$-copules, définies comme $h^{\leftarrow}(C(h(u_1), ..., h(u_n)))$ (où C est une copule, $h \in \mathcal{H}$ des fonctions de distorsion, c'est à dire des bijections croissantes de $[0, 1]$, et détaillées dans la Section 1.3.2), en posant $\phi = -\log h$, et $C = C^\perp$. Une propriété fondamentale de ces copules Archimédiennes est étudiée dans le Chapitre 3: la stabilité de cette famille par troncature, ou par déformation dynamique.

Si le générateur est un outil naturel pour caractériser les copules, d'autres fonctions peuvent aussi être utilisées, en particulier la *fonction de distribution de Kendall* K, définie comme la fonction de répartition de la variable $C(X_1, ..., X_d)$, où $(X_1, ..., X_d)$ est un vecteur aléatoire de fonction de répartition C, i.e.

$$K(t) = \mathbb{P}(C(X_1, ..., X_d) \leq t), t \in [0, 1].$$

Si C est une copule Archimédienne, de générateur ϕ, alors $K(t) = t - \lambda(t)$ où $\lambda(t) = \phi(t)/\phi'(t)$ où ϕ' désigne la dérivée à droite de ϕ sur $[0, 1)$. Réciproquement, à partir de K ou λ il est possible d'obtenir le générateur ϕ en notant que

$$\phi(u) = \phi(u_0) \exp\left(\int_{u_0}^{u} \frac{1}{\lambda(t)} dt\right)$$

pour $0 < u_0 < 1$ et $0 \leq u \leq 1$.

Définition 1.5. *Soit* $(U_1, ..., U_d)$ *un vecteur aléatoire de fonction de répartition la copule* C*, alors la fonction de répartition de* $(1 - U_1, ..., 1 - U_d)$ *est une copule appelée copule de survie associée à* C*, et est notée* C^**.*

En particulier, si C est la copule de $\boldsymbol{X}$, C^* est la copule de $-\boldsymbol{X}$ (ou de n'importe quelle transformation décroissante). Cette propriété sera utile dans le chapitre 6, par exemple. Les *extrêmes* y sont caractérisés par l'appartenance à une région de l'espace où toutes les composantes du vecteur $\boldsymbol{X}$ sont importantes, ce qui conduit à étudier la copule dans la région proche du coin $\mathbf{1} = (1, ..., 1)$. Cette notion de copule de survie permettra d'alléger les notations, en notant que l'étude de C est équivalente à l'étude de C^* en $\mathbf{0} = (0, ..., 0)$.

Comme le notait Patton (2005) dans la revue Risk, "*copula theory should be of interest to anyone who has to deal with multiple sources of risk*". C'est également ce que recommande l'Association Actuarielle Internationale dans son rapport sur la solvabilité (IAA (2004)). Mais si les copules sont utiles pour mieux comprendre la structure de la dépendance, il convient de noter comme Mikosch (2005) qu'il existe "*many flexible classes of multivariate distributions in the literature, it is not forbidden to fit those to the data*", et qu'il est parfois difficile de savoir si l'approche par les copules est la plus simple, difficile de distinguer entre "*use or abuse of copulas*".

Parmi les interrogations, certains semblent douter de la représentation "*uniforme*" des marges. Hoeffding (1940) justifiait cette approche de la façon suivante: "*in order to be able to investigate the relationships between the variables X and Y independently of their scales [...] first, we imagine ourselves given a well-determined univariate distribution . Although from a purely theoretical point of view the form of the distribution does not matter*". Seules les considérations pratiques semblent légitimer le choix de la distribution. En finance la référence étant le modèle Gaussien, il semble parfois plus légitime de s'intéresser à une version des copules dont les marges seraient des lois $\mathcal{N}(0,1)$ (comme cela avait été proposé par Cook et Johnson (1981)). Et plus généralement, un certain nombre de lois ont souvent été proposées pour étudier la dépendance en enlevant des effets marginaux, en particulier la loi de Fréchet ou exponentielle pour les extrêmes multivariés (Renisck (1987)). Hoeffding (1940) suggérait d'utiliser la loi uniforme, mais sur $[-1/2,+1/2]$ "*in order to have something definite before our eyes from the outset*". La loi uniforme sur $[0,1]$ présente, elle, l'avantage de correspondre à une probabilité, ou un rang. Ainsi, s'intéresser au quadrant $[0.95,1]\times[0.95,1]$ a une interprétation directe et simple: on s'intéresse aux événements extrêmes, où chacune des composantes dépasse son quantile au seuil de 95%.

Cette transformation des marges, proposée pour passer des lois jointes aux copules, bien qu'intéressante en théorie, pose des problèmes de mise en oeuvre. A partir d'un échantillon $(X_1,Y_1),...,(X_d,Y_d)$, il est rare que les distributions marginales F_X et F_Y soient connues. Une idée "*naturelle*" (Rüschendorf (1976), Deheuvels (1979)) est de remplacer les lois marginales par leur version empirique $\widehat{F}_X$ et $\widehat{F}_Y$ (ce qui revient à remplacer les observations par leurs rangs respectifs). Une autre approche consiste à utiliser une approche paramétrique sur les lois marginales. Ghoudi et Rémillard (1998, 2004) quantifient ainsi la perte d'information dû à l'utilisation de pseudo-observations. On retiendra aussi les travaux de Deheuvels (1979, 1981), et plus généralement tous ceux basés sur les rangs. En ce qui concerne l'approche paramétrique, on notera que cette approche peut être particulièrement dangereuse puisqu'une mauvais adéquation entraîne des estimations (fortement) biaisées.

L'estimation, discutée dans le Chapitre 7, et le choix de la copule reste des problèmes complexes. En pratique, certaines familles sont ajustées à cause de leur facilité d'usage, et non pas parce qu'elles seraient adéquate pour la modélisation. Il convient néanmoins de garder en mémoire certaines propriétés, qui permettent d'intuiter une forme naturelle de modèle. En particulier, les copules Archimédiennes traduisent généralement de l'indépendance conditionnelle entre les composantes (approche "*frailty*" présentée dans la Section 1.3). Un autre souci est celui de la dimension. La plupart des copules paramétriques sont symétriques, et ne permettent pas de prendre en compte la complexité de la dépendance en dimension très grande (on pensera aux indices boursiers basés sur 40 ou 500 titres). La dimension $d=2$ reste généralement privilégiée dans l'étude multivariée (comme dans les Chapitres 2 et 7): la distribution est alors visuellement représentable par une surface dans l'espace.

De façon générale, l'analyse et la modélisation de la dépendance reste un exercice périeux, mais essentiel en gestion des risques. Les dépendances entre risques correspondent dans cette thèse à celles qui existent au sein d'un vecteur aléatoire. Dans un cadre financier, de valorisation de portefeuille ou d'options multisupport (sur plusieurs actions sous-jacentes, par exemple), il peut s'agir des dépendances qui existent entre les prix (ou les rendements) de plusieurs titres à une date fixée. En assurance décès, il s'agit de la dépendance entre les durées de vie de deux époux, dans un contrat d'assurance sur deux têtes. En assurance non-vie, on peut s'intéresser à la dépendance entre l'indemnité versé à un assuré par un assureur et les frais d'expertise qui reste à la charge de l'assureur, pour valoriser un contrat de réassurance. Il s'agit toujours de dépendance entre composantes d'un vecteur aléatoire. Un autre aspect essentiel est celui de la dépendance sérielle. Mais bien qu'il y ait un lien entre les copules et les probabilités de transition d'un processus Markovien (Olsen, Darsow et Nguyen (1996), Section 1.3), cette relation est difficile d'utilisation. En effet, la plupart des processus usuels en temps discret (GARCH, ARFIMA) comme en temps continu (processus de Levy, processus de diffusion) n'ont pas de copule simple, voire connue explicitement. Pour ce type de dépendance, les copules ne semble pas être l'outil privilégié.

Néanmoins, il convient de noter que si les copules sont un outil puissant et intéressant pour décrire et comprendre la structure de la dépendance, dans un certain nombre de cas, des modèles possédant une interprétation naturelle peuvent être utilisés, plutôt qu'une modélisation par copules. Par exemple, dans les contrats d'assurance sur deux têtes, offrant une rente au décès du père et/ou de la mère, Norberg (1989) avait proposé un modèle à 4 états, où le décès d'un conjoint affectait le risque de mortalité du survivant, en ajoutant une surmortalité. Cette approche est beaucoup plus simple à comprendre que l'ajustement d'une copule sur les probabilités jointes de décès. Mais si la modélisation est plus simple, l'étude de la copule associée apporte en revanche un très grand nombre d'informations, afin de comprendre où se situe la dépendance la plus forte, par exemple. La dépendance est-elle plus forte pour les âges très élevés, ou au contraire pour les âges les plus faibles ? Est-on plus sensible au décès du conjoint à 40 ans, ou à 80 ?

2 Déformation des structures de dépendance et risque de crédit

Cette partie est développée dans le Chapitre 2.

Le second chapitre présente les principaux résultats obtenus dans Charpentier et Juri (2004), ainsi que quelques résultats de Charpentier (2003). En risque de crédit, comme en fiabilité, on s'intéresse à des temps de faillite d'émetteurs de titres, ou de défaillance de composants, c'est à dire un vecteur aléatoire $\boldsymbol{X} = (X_1, ..., X_n)$, où X_i désigne le temps résiduel avant la faillite du $i$ème émetteur. Si C désigne la copule reflétant la structure de dépendance entre les défauts des émetteurs à la date 0, notons que si à la date t aucun émetteur n'a fait défaut, la structure de dépendance aura changé. Formellement, il n'y a en effet aucune raison pour que la copule de $\boldsymbol{X}$ et celle de $\boldsymbol{X}|\boldsymbol{X} > t$ coïncident. Ce chapitre permet de formaliser ce que les praticiens de la finance appellent la *dépendance à court, moyen ou long terme*, exprimant ainsi le fait que la dépendance change au cours du temps.

Soit C^* la copule de survie du vecteur $\boldsymbol{X} = (X_1, ..., X_d)$ à la date 0, c'est à dire la fonction de répartition du vecteur $\boldsymbol{U} = (U_1, ..., U_d) = (1 - F_1(X_1), ..., 1 - F_d(X_d))$. Si à la date $t > 0$ aucun émetteur n'a fait défaut, la structure de dépendance est caractérisée par la copule du vecteur $(X_1, ..., X_d)$ sachant $\{X_1 > t, ..., X_d > t\}$. De façon équivalente, il s'agit d'étudier la copule associée au vecteur

$$(U_1, ..., U_d) \text{ sachant } \{U_1 \leq 1 - F_1(t), ..., U_d \leq 1 - F_d(t)\}, \tag{4}$$

quand t varie. On notera que la fonction de répartition associée n'est pas une copule, puisque les marges ne sont plus uniformes sur $[0, 1]$.

Définition 2.1. *Soit $\boldsymbol{U} = (U_1, ..., U_n)$ un vecteur aléatoire dont les marges sont uniformément distribuées sur $[0, 1]$, de fonction de répartition C. On note $\Phi(C, \boldsymbol{r})$, ou $C_{\boldsymbol{r}}$, la copule du vecteur conditionnel*

$$(U_1, ..., U_n)|U_1 \leq r_1, ..., U_d \leq r_d, \tag{5}$$

où $r_1, ..., r_d \in (0, 1]$. En particulier, en notant $F_{i|\boldsymbol{r}}(\cdot)$ la fonction de répartion marginale de U_i sachant $\{U_1 \leq r_1, ..., U_i \leq r_i, ..., U_d \leq r_d\}$,

$$F_{i|\boldsymbol{r}}(x_i) = \frac{C(r_1, ..., r_{i-1}, x_i, r_{i+1}, ..., r_d)}{C(r_1, ..., r_{i-1}, r_i, r_{i+1}, ..., r_d)},$$

et, en notant l'inverse généralisée par l'exposant $^{\leftarrow}$, au sens où $F^{\leftarrow}(u) = \sup\{x, F(x) \leq u\}$, la copule conditionnelle s'écrit

$$\Phi(C, \boldsymbol{r})(\boldsymbol{u}) = C_{\boldsymbol{r}}(\boldsymbol{u}) = \frac{C(F^{\leftarrow}_{1|\boldsymbol{r}}(u_1), ..., F^{\leftarrow}_{d|\boldsymbol{r}}(u_d))}{C(r_1, ..., r_d)}. \tag{6}$$

Proposition 2.2. *En particulier, si $\boldsymbol{X} = (X_1, ..., X_d)$ un vecteur aléatoire de copule de survie C^*, la copule du vecteur conditionnel $(X_1, ..., X_d)$ sachant $\{X_1 > t, ..., X_d > t\}$ est $\Phi(C^*, 1 - F_1(t), ..., 1 - F_d(t))$.*

Definition 2.1. *Une copule C sera dite invariante par troncature si et seulement si $\boldsymbol{U}$ sachant $\boldsymbol{U} \leq \boldsymbol{r}$ a la même copule que $\boldsymbol{U}$, quel que soit $\boldsymbol{r} \in (0, 1]^d$, i.e. $\Phi(C, \boldsymbol{r}) = C_{\boldsymbol{r}} = C$.*

Théorème 2.3. *Les seules copules absolument continues invariantes par troncatures, i.e. $\Phi(C, \boldsymbol{r}) = C$ pour tout $\boldsymbol{r} \in (0, 1]^d$ sont les copules de Clayton*

$$C(u_1, ..., u_d) = (u_1^{-\alpha} + ... + u_d^{-\alpha} - (d - 1))^{-1/\alpha}, \alpha > 0,$$

avec le cas limite de la copule indépendante ($\alpha \to 0$). Notons que la copule comonotone ($\alpha \to \infty$) est également invariante (mais elle n'est pas absolument continue).

Il est en fait possible de conjecturer que la copule comonotone est la seule copule non absolument continue à être invariante. Juri et Wüthrich (2003) avaient noté que la copule de Clayton était invariante, mais elle est en réalité la seule à vérifier cette propriété.

En fait, Juri et Wüthrich (2003, 2004) s'était intéressé à ces copules conditionnelles, dans des cas particuliers très restrictifs (copules Archimédiennes pour Juri et Wüthrich (2003) et pour des copules symétriques pour Juri et Wüthrich (2004), avec dans les deux cas des lois marginales identiques, c'est dire $\boldsymbol{r} = r\boldsymbol{1}$). Leur objectif était d'étudier des comportements limites quand $t \to \infty$. Comme l'avait noté Juri et Wüthrich (2004), les copules obtenus présentaient la particularité d'être des copules invariants, sur la diagonale, au sens où on a $\Phi(C, r\boldsymbol{1}) = C$ pour tout $r \in (0, 1]$. En effet, dans le cas des modèles de durées, la notion d'invariance nécessaire est beaucoup plus faible que la notion d'invariance globale de la Définition 2.1,

Définition 2.4. *Soient $r_1, ..., r_d$ des fonctions continues $\mathbb{R}^+ \to (0,1]$, telles que $r_i(t) \to 0$ quand $t \to \infty$ et $r_i(0) = 1$. Soit $\mathcal{D}$ le graphe associé, $\mathcal{D} = \{(r_1(t), ..., r_d(t)), t \geq 0\}$. On dira que C est une copule invariante (par troncature) suivant la direction $\mathcal{D}$ si*

$$\Phi(C, r_1(t), ..., r_d(t)) = C_{\boldsymbol{r}(t)} = C \text{ pour tout } t \geq 0. \quad (7)$$

Dans l'approche que j'essayais d'avoir, la caractérisation des copules invariantes aboutissait alors à une impasse, puisque cette Equation (7) (même uniquement en dimension 2) aboutissait à une équation fonctionnelle peu usuelle. En revanche, la caractérisation des copules limites, c'est à dire des copules C_0 non dégénérées telles que $\Phi(C, r\mathbf{1}) \to C_0$ quand $r \to 0$ s'obtenait dans Juri et Wüthrich (2004) en considérant des résultats de *variation régulière* et d'équations fonctionnelles classiques (fonctions homogènes).

Le lien entre ces deux approches a été rendues possibles à l'aide d'un résultat de continuité et surtout d'un *théorème de point fixe*,

Proposition 2.5. *L'application qui à une copule C donnée associe à $\boldsymbol{r} \in (0,1]^d$ la copule du vecteur conditionnel, $\boldsymbol{r} \mapsto \Phi(C, \boldsymbol{r})$ est continue (pour la norme infinie $\|\cdot\|_\infty$).*

Le théorème de point fixe, et d'invariance des copules limites ne peut toutefois s'obtenir que sur des conditions fortes: il faut que la direction $\mathcal{D}$ soit invariante par changement d'échelle. Notons que cela est le cas pour les fonctions puissances. On a alors le résultat suivant,

Proposition 2.6. *Une copule C_0 est une copule invariante sous la direction*

$$\mathcal{D} = \{(t^{-\alpha_1}, ..., t^{-\alpha_d}), t \geq 0\}, \text{ où } \alpha_i > 0, i = 1, ..., d,$$

si et seulement si il existe une copule C telle que

$$\lim_{t\to\infty} \|\Phi(C, t^{-\alpha_1}, ..., t^{-\alpha_d}) - C_0\|_\infty = \lim_{t\to\infty} \|C_{t^{-\boldsymbol{\alpha}}} - C_0\|_\infty = 0,$$

où $\|\cdot\|_\infty$ désigne la norme sup.

En fait, un résultat plus général peut être énoncé pour les directions *proches*. Pour cela, on introduit la notion de *variation régulière*. Une fonction r sera à variation régulière d'indice $\theta \in \mathbb{R}$ en l'infini si $r(tx)/r(t) \to x^{-\theta}$ quand $t \to \infty$, que l'on notera $r(\cdot) \in \mathcal{R}_{-\theta}$ (cette notion sera largement étudiée par la suite).

Théorème 2.7. *Une copule C_0 est une copule limite suivant la une direction $\mathcal{D} = \{(r_1(t), ..., r_d(t)), t \geq 0\}$, où $r_i(\cdot)$ est à variation régulière d'indice $\alpha_i > 0$, si et seulement si C_0 est une copule invariante sous la direction $\mathcal{D} = \{(t^{-\alpha_1}, ..., t^{-\alpha_d}), t \geq 0\}$.*

Ainsi, la recherche de copules invariantes est *équivalente* à celle des copules limites (pour laquelle Juri et Wüthrich (2004) avait apporté une première piste). L'idée était néanmoins d'avoir un résultat beaucoup plus général, s'affranchissant de l'hypothèse de lois identiques pour chacune des composantes, mais aussi de symétrie de la copule, nécessaire dans l'approche de Juri et Wüthrich (2004).

Pour caractériser ces comportements limites, une définition relativement générale de la variation régulière en dimension 2 est nécessaire (plus forte que celle utilisée en théorie *usuelle* des extrêmes multivariées, développée dans le Chapitre 6). Rappelons que dans un cadre univarié, une fonction f est dite à variation régulière s'il existe une fonction $\lambda(\cdot)$ telle que $f(tx)/f(t) \to \lambda(x)$ quand $t \to \infty$. En utilisant des résultats d'équations fonctionnelles (sur les fonctions homogènes), on montre qu'il existe un réel θ tel que $\lambda(x) = x^{-\theta}$, appelé indice de variation régulière. de Haan, Omey et Resnick (1984) avaient cherché, dans un cadre bivarié, à caractériser les fonctions $\lambda(\cdot, \cdot)$ telles que

$$\lim_{t\to\infty} \frac{f(r(t)x, s(t)y)}{f(r(t), s(t))} = \lambda(x, y). \tag{8}$$

On parle alors de variation régulière *suivant la direction* (r, s).

Proposition 2.8. *Soit $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ une fonction monotone en chacune des composantes, au delà d'un certain seuil, et r et s deux fonctions à variation régulière en $+\infty$, respectivement de paramètres α et β strictement positifs. S'il existe une fonction positive h et une fonction non nulle λ telles qu'aux points de continuité (x, y) de λ,*

$$\lim_{t\to\infty} \frac{f\left(r\left(t\right)x, s\left(t\right)y\right)}{h\left(t\right)} = \lambda\left(x, y\right), \tag{9}$$

alors λ vérifie l'équation fonctionnelle $\lambda\left(t^\alpha x, t^\beta y\right) = t^\theta \lambda\left(x, y\right)$ pour tout $x, y, t > 0$, dont la solution générale est

$$\lambda(x, y) = \begin{cases} x^{\theta/\alpha}\kappa(yx^{-\beta/\alpha}) & \text{si } x \neq 0 \\ cy^{\theta/\beta} & \text{si } x = 0 \text{ et } y \neq 0 \\ 0 & \text{si } x=0 \text{ et } y = 0, \end{cases}$$

où c est une constante, et κ une fonction $\mathbb{R}^+ \to \mathbb{R}^+$.

Les propriétés de cette notion de variation régulière sont présentées dans le Chapitre 6, en particulier le lien avec celle plus restreinte de Resnick (1987), ou celle *a priori* plus générale de Meerschaert et Scheffler (2001). A partir de cette notion, il est possible de caractériser la forme des copules limites par troncature.

Soient α, β, θ des constantes positives et P, Q des fonctions de répartitions sur $[0, 1]$. Soit $\mathcal{H}(\alpha, \beta, \theta)$ l'ensemble des fonctions de répartitions H définies sur $[0, 1]^2$ de la forme

$$H(x, y) = x^{\theta/\alpha}h(yx^{-\beta/\alpha}), \text{ où } h(t) = \begin{cases} Q(t) & \text{si } t \in [0, 1] \\ t^{\theta/\beta}P(t^{-\alpha/\beta}) & \text{si } t \in (1, \infty) \end{cases}$$

On notera $\Gamma(P, Q, \alpha, \beta, \theta)$ la copule associée, définie par

$$\Gamma(P, Q, \alpha, \beta, \theta)(u, v) = \begin{cases} Q^{\leftarrow}(v)^{\theta/\beta}P(P^{\leftarrow}(u)Q^{\leftarrow}(v)^{-\alpha/\beta}), \text{ si } P^{\leftarrow}(u)^\beta \leq Q^{\leftarrow}(v)^\alpha \\ P^{\leftarrow}(u)^{\theta/\alpha}Q(P^{\leftarrow}(u)^{-\beta/\alpha}Q^{\leftarrow}(v)), \text{ si } P^{\leftarrow}(u)^\beta > Q^{\leftarrow}(v)^\alpha \end{cases}$$

Théorème 2.9. *Dans le cas bivarié, si les fonctions de sruvie $1 - F_X$ et $1 - F_Y$ sont à variation régulière de paramètres $\alpha, \beta \geq 0$ respectivement, de telle sorte que C^* soit à variation régulière en $(0, 0)$ suivant la direction $(1 - F_X(\cdot), 1 - F_Y(\cdot))$ alors il existe $\theta > 0$, P et Q deux fonctions de répartitions sur $[0, 1]$ telle que la copule limite soit $\Gamma(P, Q, \alpha, \beta, \theta)$.*

Cette caractérisation des copules limites nous donne également une caractérisation des copules invariantes par troncatures quand la direction est fixée (c'est à dire quand les lois marginales F_X et F_Y des durées sont données), parmi lesquelles on retrouve la copule de Clayton (Section 1.6), mais aussi la copule de Marshall-Olkin, par exemple (Section 1.7). On notera en particulier que ces copules sont invariantes suivant la direction $\mathcal{D} = \{(P(t^\alpha), Q(t^\beta)), t \in (0,1]\}$.

Notons que plusieurs pistes d'ouvertures sont envisageables. Tout d'abord la famille des lois limites est présentée sous cette forme car les paramètres apparaissent naturellement, mais elle est non identifiable (en particulier seul le rapport α/β doit intervenir, donc un paramètre est redondant). Aussi, la caractérisation des copules appartenant à cette famille n'est pas triviale. Enfin, parmi les applications, seul le cas du risque de crédit est présenté (validant théoriquement certains résultats intuités par Schönbücher et Schubert (2001), par exemple), mais de nombreuses applications seraient possibles, en particulier sur les structures par terme des annuités sur deux têtes par exemple, où la structure de dépendance doit se modifier au fur et à mesure que le temps avance: l'impact de l'effet "*coeur brisé*" (Denuit, Dhaene, le Bailly de Tilleghem et Teghem (2001)) serait alors peut être moins important que prévu.

3 Dépendance de queue inférieure pour les copules Archimédiennes

Cette partie est développée dans le Chapitre 3.

Le troisième chapitre présente les principaux résultats obtenus dans Charpentier et Segers (2006*a*, 2006*b*). Dans la continuité du chapitre précédant, nous allons nous restreindre ici à l'étude des copules conditionnelles dans le cas où C est une copule Archimédienne. En effet, les copules Archimédiennes définissent une famille *stable* par troncature (résultat publiée dans Juri et Wüthrich (2003) en dimension 2, mais qui se généralise en dimension d quelconque),

Proposition 3.1. *Si $\boldsymbol{X} = (X_1, ..., X_d)$ admet pour copule une copule Archimédienne C, de générateur ϕ, $(X_1, ..., X_d)$ sachant $\{X_1 < x_1, ..., X_d < x_d\}$, pour tout $x_1, ..., x_d \in \mathbb{R}$, admettra également pour copule une copule Archimédienne, mais de générateur différent, noté $\phi_{\boldsymbol{F}(\boldsymbol{x})} = \phi(tc) - \phi(c)$ où $c = C(F_1(x_1), ..., F_d(x_d))$.*

Le générateur de la copule conditionnelle est obtenu par une simple transformation géométrique, en appliquant une homothétie à la restriction de ϕ à $[0, c]$, de telle sorte que $\phi_{\boldsymbol{F}(\boldsymbol{x})}$ soit un générateur (en particulier $\phi_{\boldsymbol{F}(\boldsymbol{x})}(1) = 0$).

L'étude de la convergence des copules Archimédiennes nécessite au préalable des résultats liant la convergence des générateurs, et celle des copules associées. Pour cela, il fallait généraliser deux résultats de Genest et Rivest (1986), le premier donnant des conditions pour que la limite des copules Archimédienne soit Archimédienne,

Proposition 3.2. *Les cinq résultats suivants sont équivalents,*

(i) $\lim_{n\to\infty} C_n(u, v) = C(u, v)$ *pour tout* $(u, v) \in [0, 1]^2$*,*

(ii) $\lim_{n\to\infty} \phi_n(x)/\phi_n'(y) = \phi(x)/\phi'(y)$ *pour tout* $x \in (0, 1]$ *et* $y \in (0, 1)$ *tel que* ϕ' *soit continue en* y*,*

(iii) $\lim_{n\to\infty} \lambda_n(x) = \lambda(x)$ *pour tout* $x \in (0,1)$ *tel que* λ *soit continue en* x*,*

(iv) il existe des constantes positives κ_n *telles que* $\lim_{n\to\infty} \kappa_n \phi_n(x) = \phi(x)$ *pour tout* $x \in [0,1]$*,*

(v) $\lim_{n\to\infty} K_n(x) = K(x)$ *pour tout* $x \in (0,1)$ *tel que* K *soit continue en* x*.*

Le second résultat donne des conditions pour que la limite des copules Archimédienne soit comonotone,

Proposition 3.3. *Les quatre résultats suivants sont équivalents,*

(i) $\lim_{n\to\infty} C_n(u,v) = C^+(u,v) = \min(u,v)$ *pour tout* $(u,v) \in [0,1]^2$*,*

(ii) $\lim_{n\to\infty} \lambda_n(x) = 0$ *pour tout* $x \in (0,1)$*,*

(iii) $\lim_{n\to\infty} \phi_n(y)/\phi_n(x) = 0$ *pour tout* $0 \leq x < y \leq 1$*,*

(iv) $\lim_{n\to\infty} K_n(x) = x$ *pour tout* $x \in (0,1)$*.*

Des deux propositions 3.2 et 3.3, on pourrait croire que la limite de n'importe quelle suite de copules Archimédienne est nécessairement soit Archimédienne, soit comonotone. Mais ce n'est pas le cas, comme le montre le contre exemple donné dans la section 3.3.5.

A partir de ces résultats, il a été possible de reprendre le résultat de Juri et Wüthrich (2003), en le généralisant en dimension quelconque, mais aussi, en corrigeant le cas particulier d'indépendance asymptotique,

Proposition 3.4. *Soit* C *une copule Archimédienne de générateur* ϕ*, et* $0 \leq \alpha \leq \infty$*. Notons* $C(\cdot,\cdot;\alpha)$ *la copule de Clayton de paramètre* α*. Considérons les quatre résultats suivants:*

(i) $\lim_{u\to 0} C_u(x,y) = C(x,y;\alpha)$ *pour tout* $(x,y) \in [0,1]^2$*;*

(ii) $-\phi' \in \mathcal{R}_{-\alpha-1}$*.*

(iii) $\phi \in \mathcal{R}_{-\alpha}$*.*

(iv) $\lim_{u\to 0} u\phi'(u)/\phi(u) = -\alpha$*.*

Si $\alpha = 0$ *(indépendance asymptotique),*

$$\textit{(i)} \Longleftrightarrow \textit{(ii)} \Longrightarrow \textit{(iii)} \Longleftrightarrow \textit{(iv)},$$

et si $\alpha \in (0,\infty]$ *(dépendance asymptotique),*

$$\textit{(i)} \Longleftrightarrow \textit{(ii)} \Longleftrightarrow \textit{(iii)} \Longleftrightarrow \textit{(iv)}.$$

Plus précisémément, dans le cas général $\alpha \in [0,\infty]$, les conditions *(i)* et *(ii)* impliquent *(iii)* et *(iv)*, la réciproque étant vrai seulement dans le cas $\alpha > 0$. En particulier, un contre exemple permettra de contredire le Théorème 3.5 de Juri et Wüthrich (2003),

Proposition 3.5. *Il existe des copules Archimédiennes* C *de générateur* ϕ *dont la dérivée est continue, à variation lente à l'origine, telles que la copule conditionnelle ne converge pas vers la copule indépendante.*

Ce résultat est en fait très lié à la notion de *variation régulière au second ordre*, liée à la classe de de Haan (Chapitre 3 de Bingham, Goldie et Teugels (1987)). En effet, le Lemme 3.4. de Juri et Wüthrich (2003) prétendait que si ϕ est un générateur différentiable et à variation lente à l'origine, alors il existe une fonction positive g sur $(0,1)$ telle

$$\lim_{u\to 0}\frac{\psi(ux)-\psi(u)}{g(u)} = -\log(x) \tag{10}$$

pour tout $0<x<\infty$. Cette condition est la définition d'appartenance à la classe de de Haan Π, avec pour fonction auxiliaire g, notée $\phi\in\Pi_g$, qui s'interprète de manière équivalente comme $-\phi'\in\mathcal{R}_{-1}$, et dans ce cas, $g(s)\sim -s\phi'(s)$ quand $s\to 0$. Si le théorème de Karamata implique qu'alors $\phi\in\mathcal{R}_0$, la réciproque n'est en général pas vraie.

Pour conclure ce chapitre certains résultats plus généraux sont donnés. Afin de mieux comprendre leur intérêt, revenons à l'interprétation pour les modèles de durées, ou de risque de crédit. En dimension $d=2$, la seule information qui pouvait être disponible à la date t est qu'il n'y avait eu aucun défaut. En dimension $d\geq 2$, l'information peut être plus complexe, car certains titres seulement peuvent être encore en vie. Le conditionnement peut alors se faire sur une information partielle. Plus formellement, on a le résultat suivant,

Proposition 3.6. *Soit C une copule Archimédienne en dimension d, de générateur ψ. Si ψ est à variation régulière d'indice $-\theta\in[-\infty,0]$, et J est un ensemble non-vide de $\{1,\dots,d\}$ et si $0<y_j<\infty$ pour tout $j\in J$, alors, pour tout $(x_1,\dots,x_d)\in(0,\infty)^d$,*

$$\begin{aligned}&\lim_{s\to 0}\mathbb{P}(\forall i=1,\dots,d: U_i\leq sx_i \mid \forall j\in J: U_j\leq sy_j) \qquad (11)\\ &= \left(\frac{\sum_{j\in J^c}x_j^{-\theta}+\sum_{j\in J}\{\min(x_j,y_j)\}^{-\theta}}{\sum_{j\in J}y_j^{-\theta}}\right)^{-1/\theta}.\end{aligned}$$

La copule limite associée est alors la copule de Clayton.

Ces deux résultats ne traitent que du cas de dépendance asymptotique. Le cas d'indépendance asymptotique, obtenu lorsque $C(s,\dots,s)=o(s)$ en 0, est plus délicat.

Proposition 3.7. *Soit C une copule Archimédienne en dimension d, de générateur ψ. Si ψ est à variation lente en 0, alors $C(s,\dots,s)=o(s)$, qui implique que $\log\{\psi(s)\}/\log(s)\to 0$ lorsque $s\to 0$. Notons toutefois que pour chacune des implications, aucune réciproque n'est vraie.*

Une idée est alors d'utiliser l'idée qu'avait développée Ledford et Tawn (1996, 1997), en dimension 2.

Proposition 3.8. *Soit C une copule Archimédienne en dimension d, de générateur ψ. Si*

$$\lim_{t\to\infty}\frac{D(\log\psi^{\leftarrow})(dt)}{D(\log\psi^{\leftarrow})(t)}=\frac{1}{d\eta},$$

alors $s\mapsto C(s,\dots,s)$ est à variation régulière en 0 d'indice $1/\eta$.

On a alors les deux résultats suivants,

Proposition 3.9. *Soit C une copule Archimédienne en dimension d, de générateur ψ. Si la fonction $\phi = -1/D(\log \psi^{\leftarrow})$ est à variation régulière d'indice $-\infty < \tau \leq 1$ et si $\phi(t) = o(t)$ lorsque $t \to \infty$, alors pour tout ensemble J non-vide de $\{1, \ldots, d\}$ et pour tout $\boldsymbol{x} \in (0, \infty)^d$,*

$$\begin{aligned} &\mathbb{P}(\forall j \in J : U_j \leq s x_j ; \forall j \in J^c : U_j \leq \psi^{\leftarrow}\{x_j^{-1}\phi(\psi(s))\}) \\ \sim \quad &\psi^{\leftarrow}\{|J|\psi(s)\} \prod_{j \in J} x_j^{|J|^{-\tau}} \prod_{j \in J^c} \exp\left(-|J|^{-\tau} x_j^{-1}\right) \text{ lorsque } s \to 0. \end{aligned} \tag{12}$$

Proposition 3.10. *Soit C une copule Archimédienne en dimension d, de générateur ψ dont le générateur vérifie les conditions précédantes, alors, pour tout $\boldsymbol{x} \in (0, \infty)^d$ et tout $(u_j)_{j \in J} \in (0, 1]^{|J|}$,*

$$\begin{aligned} &\mathbb{P}(\forall j \in J : U_j \leq s u_j x_j ; \forall j \in J^c : U_j \leq \psi^{\leftarrow}\{x_j^{-1}\phi(\psi(s))\} \mid \forall j \in J : U_j \leq s x_j) \\ \to \quad &\prod_{j \in J} u_j^{|J|^{-\tau}} \prod_{j \in J^c} \exp\left(-|J|^{-\tau} x_j^{-1}\right), \text{ lorsque } s \to 0. \end{aligned}$$

4 Dépendence et relation d'ordre: comparaison de la dépendance

Cette partie est développée dans le Chapitre 4.

Le quatrième chapitre présente les principaux résultats obtenus dans présentés dans Charpentier (2003, 2004). L'étude des relations d'ordre entre vecteurs aléatoires a permis de définir un grand nombre de relations de comparaisons (Shaked et Shantikumar (1994), Müller et Stoyan (2001) ou Denuit, Dhaene, Goovaerts et Kaas (2005) pour des applications en sciences actuarielles). Dans un premier temps, nous verrons comment construire des "*relation d'ordre de dépendance*" à partir de relations d'ordre définies sur le même espace de Fréchet. Une approche axiomatique permet de définir des propriétés de relation de dépendence.

Définition 4.1. *Une relation binaire $\preceq$ définie sur l'ensemble des copules $\mathcal{C}$ est appelée relation de dépendance si $\preceq$ vérifie les propriétés de transitivité, de réflexivité et d'antisymmétrie, la relation de concordance (si $C_X \preceq C_Y$ alors $C_{\boldsymbol{X}}(\cdot) \leq C_{\boldsymbol{Y}}(\cdot)$), de bornes ($C^- \preceq C \preceq C^+$ pour toute copule C), de transposition, et de fermeture par convergence faible.*

On a alors une bijection entre ordonancement sur l'ensemble d'une classe de Fréchet et sur l'ensemble des copules,

Proposition 4.2. *Soit $\mathcal{C}$ la classe des copules, et $\mathcal{F}$ une classe de Fréchet ($\mathcal{F} = \mathcal{F}(F_1, ..., F_d)$) dont les marges sont absolument continues. Soit $\preceq$ une relation de dépendance sur $\mathcal{C}$, et définissons $\leq$ par $F_{\boldsymbol{X}} \leq F_{\boldsymbol{Y}}$, pour $F_{\boldsymbol{X}}, F_{\boldsymbol{Y}} \in \mathcal{F}(F_1, ..., F_d)$, si et seulement si $C_{\boldsymbol{X}} \preceq C_{\boldsymbol{Y}}$ où $C_{\boldsymbol{X}}$ et $C_{\boldsymbol{Y}}$ sont les copules induites respectivement par $F_{\boldsymbol{X}}$ et $F_{\boldsymbol{Y}}$, alors $\leq$ est une relation d'ordre sur $\mathcal{F}$. Réciproquement, si $\leq$ est une relation d'ordre sur $\mathcal{F}$, définissons $\preceq$ par $C_{\boldsymbol{X}} \preceq C_{\boldsymbol{Y}}$ si et seulement si $F_{\boldsymbol{X}} \preceq F_{\boldsymbol{Y}}$ où $F_{\boldsymbol{X}}$ et $F_{\boldsymbol{Y}}$ sont les fonctions de la classe de Fréchet $\mathcal{F}$ dont les copules induites sont $C_{\boldsymbol{X}}$ et $C_{\boldsymbol{Y}}$ respectivement, alors $\leq$ est une relation de dépendance (sur $\mathcal{C}$).*

Définition 4.3. *Soit $\preceq$ une relation de dépendance, alors on dira que $\boldsymbol{Y}$ présente plus de dépendance que $\boldsymbol{X}$ si $F_{\boldsymbol{X}}(\boldsymbol{X}) \preceq F_{\boldsymbol{Y}}(\boldsymbol{Y})$, que l'on notera $\boldsymbol{X} \preceq_d \boldsymbol{Y}$.*

Cette notion permet de légitimer ce qui fera fait dans ce chapitre, à savoir comparer des copules, sans se préoccuper des marges. En particulier, nous avons ici une notion qui permet de comparer la dépendance entre $\boldsymbol{X}$ et $\boldsymbol{X}|\boldsymbol{X} > t$ (étude de la dépendance au fur et à mesure que le temps avance). En particulier, par la suite, nous nous restreindrons au cas particulier des structures de dépendance Archimédiennes, et plus spécifiquement, à la sous-classe des copules Archimédiennes obtenues par "*frailty*", correspondant à de l'indépendance conditionnelle, et où la transformée de Laplace du facteur d'exogénéité caractérise le générateur de la copule Archimédienne sous-jacente. Plus précisément, en reprenant les notations de l'équation (3), on a le résultat suivant

Proposition 4.4. *Si $\boldsymbol{X} = (X_1, ..., X_n)$ admet pour copule une copule Archimédienne avec une représentation factorielle, où Θ est le facteur d'hétérogénéité Θ, il en sera de même pour $(X_1, ..., X_n)$ sachant $\{X_1 > x_1, ..., X_n > x_n\}$, pour tout $x_1, ..., x_n \in \mathbb{R}$, mais la transformée de Laplace du facteur aura changée.*

Et plus généralement, si la copule de $\boldsymbol{X} = (X_1, ..., X_n)$ est une $\mathcal{H}$-copule (de la forme $h^{\leftarrow}(C(h(u_1), ..., h(u_n)))$ où C est une copule, et h une fonction de distorsion), avec des conditions sur h et C, la copule conditionnelle sera également une $\mathcal{H}$-copule.

Enfin, des résultats peuvent être induits en terme de relation de dépendance usuelle ($\boldsymbol{X} \preceq \boldsymbol{Y}$ si $C_{\boldsymbol{X}}(\boldsymbol{u}) \leq C_{\boldsymbol{Y}}(\boldsymbol{u})$ pour tout $\boldsymbol{u}$):

Proposition 4.5. *Soit C une copule Archimédienne en dimension 2, de générateur ϕ deux dois dérivable, et notons $\psi(\cdot) = \log -D\phi(\cdot)$. Si ψ est concave alors*

$$C_{t_2}(\boldsymbol{u}) \leq C_{t_1}(\boldsymbol{u}) \text{ pour tout } \boldsymbol{u},$$

pour tout $t_1 > t_2$, où C_{t_i} est la copule conditionnelle à la date t_i: la copule présente de moins en moins de dépendance au fur et à mesure que le temps s'écoule.

Le cas des copules Archimédiennes est à mettre en rapport avec les travaux Bassan et Spizzichino (2001, 2004 et 2005) sur la déformation de la structure de dépendance dans les modèles de vieillissement (notion de "*aging*") pour des risques échangeables. Nous présenterons ainsi des propriétés de dépendance positive sur les copules Archimédiennes en fonction de propriétés vérifiées par le générateur (propriétés de vieillissement).

5 Dépendance de queue supérieure pour les copules Archimédiennes

Cette partie est développée dans le Chapitre 5.

Le troisième chapitre présente les principaux résultats obtenus dans Charpentier et Segers (2006*c*). Cette partie continue l'approche initiée à la fin du Chapitre 3, dans le cas de la queue inférieure, mais cette fois-ci en s'intéressant à la queue supérieure, c'est à dire aux orthants supérieurs du carré unité.

Un certain nombre d'articles traitant d'extrêmes mutivariés ont souligné la difficulté de la modélisation dans le cas d'indépendance asymptotique (Ledford et Tawn (1996, 1997), ou Draisma, Drees, Ferreira et de Haan (2004)). Ceci va se confirmer dans ce chapitre, où l'analyse va être plus compliquée que celui présentée dans le chapitre 3. En particulier, le début du chapitre présente plusieurs de lemmes techniques permettant d'obtenir par la suite des simplifications, et d'alléger les preuves, autant que faire se peut.

L'étude dans les orthants inférieurs faisaient intervenir des propriétés de variations régulières du générateur en 0. Dans le cas des orthants supérieurs, il est alors légitime d'étudier le comportement du générateur à l'autre borne du support, c'est à dire en 1, ou, de manière équivalente, le comportement de $\psi(1-\cdot)$ en 0. Or, par définition $\psi(1)=0$, aussi, un développement de Taylor donne $\psi(1-s)=-sD\psi(1)+o(s)$ quand $s\to 0$. En particulier, compte tenu de la convexité de ψ, si $\psi(1-\cdot)$ est à variation régulière d'indice θ, alors nécessairement $\theta\in[1,\infty]$. Et plus précisément, si $(-D)\psi(1)>0$, alors $\theta=1$ (la réciproque n'étant pas forcément vraie, ce qui va compliquer ici l'étude).

Proposition 5.1. *Soit $\boldsymbol{U}$ un vecteur aléatoire en dimension d, de distribution C, une copule Archimédienne de générateur ψ. Si $s\mapsto\psi(1-s)$ est à variation régulière d'indice $\theta\in(1,\infty]$ en 0, alors, pour tout ensemble non-vide J de $\{1,\dots,d\}$, pour tout $(x_1,\dots,x_d)\in(0,\infty)^d$ et tout $(y_j)_{j\in J}\in(0,\infty)^{|J|}$,*

$$\begin{aligned}&\mathbb{P}(\forall j=1,\dots,d:U_j\geq 1-sx_j\mid\forall j\in J:U_j\geq 1-sy_j)\\ &\quad\to\ \frac{r_d(z_1,\dots,z_d)}{r_{|J|}\left((y_j)_{j\in J}\right)}\ \text{lorsque } s\to 0,\end{aligned}\tag{13}$$

où $z_j=min(x_j,y_j)$ pour $j\in J$ et $z_j=x_j$ pour $j\in J^c$, en posant

$$r_k(u_1,\dots,u_k)=\begin{cases}\displaystyle\sum_{I\subset\{1,\dots,k\}:|I|\geq 1}(-1)^{|I|-1}\left(\sum_{i\in I}u_i^\theta\right)^{1/\theta} & \text{si } 1<\theta<\infty,\\ \min(u_1,\dots,u_d) & \text{si } \theta=\infty,\end{cases}$$

pour tout entier k et tout $(u_1,\dots,u_k)\in(0,\infty)^k$.

Notons que l'on retrouve ici un cas particulier de copule max-stable: la copule de Gumbel (ou logistique). Dans le cas où $\theta=1$, il convient de distinguer deux cas: soit $(-D)\psi(1)>0$ (correspondant au cas d'"*indépendance dans l'indépendance*") ou $(-D)\psi(1)=0$ ("*dépendance dans l'indépendance*").

Proposition 5.2. *Soit ψ un générateur en dimension d tel que $\psi^{\leftarrow}$ soit d fois continument dérivable, et $\boldsymbol{U}$ un vecteur en dimension d dont la fonction de répartition est la copule Archimédienne induite par ψ. Si $(-D)^d\psi^{\leftarrow}(0)<\infty$ alors $(-D)\psi(1)>0$. Soit J un sous-ensemble non-vide $\{1,\dots,d\}$ tel que J^c soit non-vide. Pour tout $\boldsymbol{v}\in(0,1]^d$,*

$$\begin{aligned}&\mathbb{P}(\forall j\in J:U_j\geq 1-sv_jx_j;\forall j\in J^c:U_j\leq v_j\mid\forall j\in J:U_j\geq 1-sx_j)\\ &\quad\to\ \frac{(-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J^c}\psi(v_j)\right)}{(-D)^{|J|}\psi^{\leftarrow}(0)}\prod_{j\in J}v_j\ \text{lorsque } s\downarrow 0.\end{aligned}\tag{14}$$

En particulier, on notera que la copule limite de la distribution conditionnelle de $(U_j)_{j\in J^c}$ sachant $U_j\geq 1-sx_j$ pour tout $j\in J$ admet pour copule une copule Archimédienne de générateur

$$\psi_{|J|}(\cdot)=\left(\frac{(-D)^{|J|}\psi^{\leftarrow}(\,\cdot\,)}{(-D)^{|J|}\psi^{\leftarrow}(0)}\right)^{\leftarrow}.\tag{15}$$

Dans le cas de dépendance dans l'indépendance, des conditions variation régulière au second ordre sont nécessaires,

Proposition 5.3. *Soit ψ un générateur en dimension d tel que $\psi^{\leftarrow}$ soit d fois continument dérivable, et $\boldsymbol{U}$ un vecteur en dimension d dont la fonction de répartition est la copule Archimédienne induite par ψ. Posons $f(s) = \psi(1-s)$. Si $s^{-1}f(s) \to 0$ lorsque $s \to 0$ et si $s \mapsto \mathcal{L}(s) = s(\mathrm{d}/\mathrm{d}s)\{s^{-1}f(s)\}$ est positive et à variation lente en 0, alors la fonction $g(s) = sf'(s)/f(s) - 1$ est positive et également à variation lente, avec $g(s) \to 0$ lorsque $s \to 0$. Si J est un sous-ensemble $\{1, \ldots, d\}$ contenant au moins deux éléments, alors, pour tout $\boldsymbol{x} \in (0, \infty)^d$ et $(y_j)_{j\in J} \in (0, \infty)^{|J|}$,*

$$\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - sx_i \mid \forall j \in J : U_j \geq 1 - sy_j) = \frac{r(z_1, \ldots, z_d)}{r((y_j)_{j\in J})}$$

lorsque $s \downarrow 0$, où $z_j = \min(x_j, y_j)$ pour $j \in J$ et $z_j = x_j$ pour $j \in J^c$, et où

$$\begin{aligned} r(x_1, \ldots, x_d) &= \sum_{I\subset\{1,\ldots,d\}:|I|\geq 1} (-1)^{|I|}(\textstyle\sum_I x_i) \log(\textstyle\sum_I x_i) \\ &= (d-2)! \int_0^{x_1} \cdots \int_0^{x_d} \left(\sum_{i=1}^d t_i\right)^{-(d-1)} dt_1 \cdots dt_d. \end{aligned}$$

En particulier, en dimension $d = 2$, si $J = \{1, 2\}$, notons que pour tout $(x, y) \in (0, \infty)^2$,

$$\begin{aligned} &\mathbb{P}(U \geq 1 - sx, V \geq 1 - sy) \\ &\quad \sim \alpha s(-\log s)^{-1}\{(x+y)\log(x+y) - x\log(x) - y\log(y)\}, \text{ lorsque } s \to 0. \end{aligned}$$

La fin de ce chapitre pousuit cette étude dans le cas des autres coins du carré unité.

6 Théorème de Pickands, Balkema et de Haan multivarié, et dépendance entre évènements extrêmes

Cette partie est développée dans le Chapitre 6.

Le sixième chapitre s'intéresse à une nouvelle application des résultats des chapitres précédents, et plus particulièrement ceux présentés dans le chapitre 2. On ne s'intéresse plus à des modèles de durée, mais aux comportements limites dans les queues de distributions, et donc à l'interprétation en terme de valeurs extrêmes. Rappelons que dans un cadre univarié, il y a trois approches *équivalentes* des extrêmes:

- étude de la loi limite normalisée du maximum $X_{n:n} = \max\{X_1, ..., X_n\}$ d'une suite de variables réelles i.i.d. (théorème dit de Fisher-Tippett),
- étude de la loi limite des excès, c'est à dire de $X - u$ sachant que $X > u$ quand $u \to \infty$ (théorème de Pickands-Balkema-de Haan),

- étude du comportement de la fonction de survie $\overline{F} = 1 - F$ en terme de variation régulière, en cherchant une fonction g telle que

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = g(x). \tag{16}$$

L'approche par la loi des excès est d'autant plus intéressante qu'elle apparaît naturelle dans un grand nombre d'applications. Si X correspond à une durée de vie, $X - u|X > u$ peut être interprétée comme une durée de vie résiduelle au delà de la date u. Si X est une perte (potentielle) financière et si u correspond à un quantile (aussi appelé Value-at-Risk, VaR), alors $X - u|X > u$ est la perte conditionnelle au delà de ce seuil, et son espérance s'appelle "*expected shortfall*". En réassurance, le seuil u est une franchise, appelée aussi priori et $X - u|X > u$ est alors l'indemnité versée par le réassureur. En fait, un grand nombre de mesures de risques extrêmes sont fonction de la distribution de X sachant $X > u$, pour un seuil u souvent grand. En pratique, le théorème de Pickands-Balkema-de Haan permet d'approcher la loi conditionnelle par une loi simple (la loi de Pareto généralisée), et permet d'obtenir des estimations robustes de ces mesures de risques. C'est pour ces raisons qu'il semble naturel d'étendre ce résultat limite au cadre multivarié.

Comme le notait Resnick (1987), dans un cadre multivarié, l'étude des *extrêmes* n'est pas évidente puisqu'il n'existe pas de relation d'ordre *naturelle* dans $\mathbb{R}^d$ (et donc il n'y a pas de manière *naturelle* et unique de définir les extrêmes). L'approche la plus étudiée (e.g. Tiago de Olivera (1958), Geoffroy (1958), Sibuya (1961) ou Resnick (1987)) consiste à étudier la loi limite du maximum par composante, c'est à dire, dans un cadre bivarié, la loi limite du couple

$$(X_{n:n}, Y_{n:n}) = (\max\{X_1, ..., X_n\}, \max\{Y_1, ..., Y_n\}),$$

Cependant, comme le notait Tawn (1988) en évoquant l'interprétation, "*a difficulty with this approach is that in some applications it may be impossible for* $(X_{n:n}, Y_{n:n})$ *to occur as a vector observation*". Néanmoins, cette approche peut être interprétée en utilisant une définition (relativement restrictive) de la variation régulière en dimension 2, généralisant l'Equation (16) sous la forme

$$\lim_{t\to\infty} \frac{\overline{F}(tx, ty)}{\overline{F}(t, t)} = g(x, y), \tag{17}$$

où g est alors nécessairement une fonction homogène $g(tx, ty) = t^\theta g(x, y)$ pour tout $x, y, t > 0$ et où θ est un paramètre réel (Resnick (2004)).

Le but de ce chapitre est de proposer une généralisation du théorème de Pickands-Balkema-de Haan, en étudiant la copule limite de (X, Y) sachant $\{X > x \text{ et } Y > y\}$, quand $x, y \to \infty$, sans supposer comme dans le chapitre précédant, ou dans Juri et Wüthrich (2004) et Wüthrich (2004), que la copule (ou la copule de survie) soit Archimédienne. La principale difficulté réside dans la définition de ce comportement limite "$x, y \to \infty$". Notons qu'en pratique deux types de modélisations peuvent être retenues: les convergences "*en niveau*" ($X > t, Y > t$ pour un même niveau t, où $t \to \infty$) ou "*en probabilité*" ($X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p)$ pour un même niveau p, où $p \to 0$). Afin de pouvoir avoir un cadre qui englobe ces deux cas particuliers, on retient l'approche directionnelle (présentée dans la section précédente, et l'équation (16) se généralise sous la forme

$$\lim_{t\to\infty} \frac{\overline{F}(r(t)x, s(t)y)}{\overline{F}(r(t), s(t))} = g(x, y), \tag{18}$$

si l'on se place dans la direction (r, s). On obtient ainsi deux versions bidimensionnelles du Théorème de Pickands, Balkema et de Haan. Dans un premier temps, on s'intéresse à une approche par seuils en probabilités, ou en quantiles,

$$(X, Y) \text{ sachant } \{X > F_X^{\leftarrow}(p) \text{ et } Y > F_Y^{\leftarrow}(p)\} \text{ quand } p \to 1,$$

et dans un second temps à une approche en niveau,

$$(X, Y) \text{ sachant } \{X > z \text{ et } Y > z\} \text{ quand } z \to \infty.$$

Pour énoncer les versions bivariées du théorème de Pickands, Balkema et de Haan, on se place sous les hypothèses suivantes. Soient X et Y, dans le max-domaine d'attraction de la loi de Fréchet, de paramètres respectifs α et β, strictement positifs, c'est à dire qu'il existe $a(\cdot)$ et $b(\cdot)$ telles que

$$\lim_{u\to\infty} 1 - \frac{1 - F_X(u + xa(u))}{1 - F_X(u)} = \lim_{u\to\infty} \mathbb{P}(X \le u + a(u) | X > u) = G_\alpha(x),$$

et de manière analogue,

$$\lim_{v\to\infty} 1 - \frac{1 - F_Y(v + yb(v))}{1 - F_Y(v)} = \lim_{v\to\infty} \mathbb{P}(Y \le v + b(v) | Y > v) = G_\beta(y),$$

où

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \xi x/\sigma)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-x/\sigma) & \xi = 0, \end{cases} \text{, avec } G_\xi = G_{\xi,1}.$$

Le premier théorème est obtenu en considérant des seuils définis par quantiles,

Théorème 6.1. *Supposons que C^*, copule de survie de (X, Y), soit à variation régulière sur la diagonale, c'est à dire qu'il existe une fonction continue $h : \mathbb{R}^+ \to \mathbb{R}^+$ avec $h(x) > 0$ pour $x > 0$, et telle que*

$$\lim_{u\to 0} \frac{C^*(xu, u)}{C^*(u, u)} = h(x) \text{ pour tout } x \ge 0. \tag{19}$$

Alors $h(0) = 0$, $h(1) = 1$, et il existe $\theta \in \mathbb{R}$ tel que

$$h(x) = x^\theta h\left(\frac{1}{x}\right) \text{ pour tout } x > 0,$$

On a la version bivariée du Theorème de Pickands-Balkema-de Haan, à savoir

$$\begin{aligned}&\lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a(F_X^{\leftarrow}(1-p))} > x, \frac{Y - F_Y^{\leftarrow}(1-p)}{b(F_Y^{\leftarrow}(1-p))} > y \middle| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p)\right) \\ = \; &(1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\beta}}\right),\end{aligned}$$

en posant $\gamma = \theta/\beta$. La convergence étant uniforme, notons que l'on peut aussi écrire

$$\lim_{p\to 0}\sup_{x,y} |\mathbb{P}\left(X - F_X^{\leftarrow}(1-p) > x, Y - F_Y^{\leftarrow}(1-p) > y \right| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p))$$
$$-(1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\beta}}\right)| = 0.$$

Notons que la copule associée est alors la copule duale de $C_h(x,y) = H(h^{\leftarrow}(x), h^{\leftarrow}(y))$, où $H(x,y) = y^\theta h(x/y)$. Cette version est relativement proche de celle obtenue par Juri et Wüthrich (2004), en supprimant l'hypothèse de lois marginales identiques. La seconde version repose sur la notion de variation régulière directionnelle.

Théorème 6.2. *Supposons que C^*, copule de survie de (X,Y), soit à variation régulière dans la direction $(1 - F_X(\cdot), 1 - F_Y(\cdot))$, de telle sorte qu'il existe une fonction λ telle que*

$$\lim_{z\to\infty} \frac{C^*((1 - F_X(z))x, (1 - F_Y(z))y)}{C^*(1 - F_X(z), 1 - F_Y(z))} = \lambda(x,y).$$

Les deux fonctions de survie marginales étant à variation régulière, d'après la Proposition 2.8, il existe un réel γ et une fonction $h : \mathbb{R} \to \mathbb{R}$ tels que $\lambda(x,y) = x^{\gamma/\alpha} h(yx^{-\beta/\alpha})$ si $x \neq 0$ et $\lambda(0,y) = cy^{\gamma/\beta}$ où c est une constante positive. On a la version bivariée du théorème de Pickands-Balkema-de Haan, à savoir

$$\lim_{z\to\infty} \mathbb{P}\left(\frac{X-z}{a(z)} > x, \frac{Y-z}{b(z)} > y \,\middle|\, X > z, Y > z\right) = (1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\alpha}}\right).$$

La convergence étant uniforme, notons que l'on peut aussi écrire

$$\lim_{z\to\infty}\sup_{x,y} |\mathbb{P}\left(X - z > x, Y - z\right| X > z, Y > z) = (1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\alpha}}\right)| = 0.$$

Ce résultat peut aussi se généraliser pour n'importe quel type de conditionnement, à condition que les seuils puissent s'écrire sous la forme $(r(t), s(t))$, où r et s sont des fonctions à variation régulière. Les deux exemples simples traités ici généralisent le théorème de Juri et Wüthrich (2004), en particulier X et Y ne sont pas nécessairement de même loi, et le premier résultat peut d'ailleurs se généraliser en supposant que X ou Y (voire les deux) sont à queue fine (Section 3.5). Dans le cas où X est à queue fine, par exemple, il suffit de substituer $\exp(x)$ à $(1+x)^{-1/\alpha}$.

Nous nous intéressserons alors à la généralisation des résultats des chapitres 3 et 6 en otant l'hypothèse de copule Archimédienne, en reprenant les résultats initiaux de Charptier et Segers (2006d), permettant de formaliser les résultats obtenus en terme de copules conditionnelles, et les approches usuelles en théorie des valeurs extrêmes multivariés. Pour cela, notons $\boldsymbol{Y} = 1/(1 - F_{\boldsymbol{X}}(\boldsymbol{X}))$, et supposons que $\boldsymbol{Y}$ soit à variation régulière de mesure exponent ν, i.e.

$$t\mathbb{P}[t^{-1}\boldsymbol{Y} \in \cdot\,] \xrightarrow{v} \nu(\,\cdot\,), \text{ as } t \to \infty.$$

Rappelons que ν est nécessairement homogène à l'ordre $\theta = 1$, $\nu(s\,\cdot\,) = s^{-1}\nu(\,\cdot\,)$ pour tout $s > 0$. En particulier, notons que

$$\mathbb{P}(\boldsymbol{Y} \le n\boldsymbol{y})^n \to \exp\{-\nu([\boldsymbol{0}, \boldsymbol{y}]^c)\}, \text{ lorsque } n \to \infty,$$

On peut alors définir la *fonction de dépendance de queue stable* par

$$l(\boldsymbol{x}) = \nu([\boldsymbol{0}, \boldsymbol{x}^{-1}]^c) = \nu(\{\boldsymbol{y} \in \mathbb{E} \mid \max(x_1 y_1, \ldots, x_d y_d) \geq 1\})$$

pour tout $\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\infty})$. De manière duale, on peut aussi définir

$$r(\boldsymbol{x}) = \nu([\boldsymbol{x}^{-1}, \boldsymbol{\infty}]) = \nu(\{\boldsymbol{y} \in \mathbb{E} \mid \min(x_1 y_1, \ldots, x_d y_d) \geq 1\}),$$

pour tout $\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\infty})$. Notons ainsi que

$$\lim_{s \downarrow 0} s^{-1} \mathbb{P}[U_1 \leq s x_1, \ldots, U_d \leq s x_d] = r(\boldsymbol{x})$$

On définie enfin la fonction H comme

$$\lim_{s \downarrow 0} \mathbb{P}(\boldsymbol{U} \leq s\boldsymbol{x}\boldsymbol{u} \mid \boldsymbol{U} \leq s\boldsymbol{x}] = H(\boldsymbol{u}; \boldsymbol{x}) = \frac{r(\boldsymbol{u}\boldsymbol{x})}{r(\boldsymbol{x})},$$

pour tout $\boldsymbol{x} \in (\boldsymbol{0}, \boldsymbol{\infty})$ et $\boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}]$.

Proposition 6.3. *Soit $\boldsymbol{U}$ un vecteur aléatoire de fonction de répartition C, une copule. Soit $C_{\boldsymbol{u}}$ la copule conditionnelle de $\boldsymbol{U}$ sachant $\{\boldsymbol{U} \leq \boldsymbol{u}\}$, alors*

$$\lim_{s \downarrow 0} C_{s\boldsymbol{x}}(\boldsymbol{p}) = C(\boldsymbol{p}; \boldsymbol{x})$$

pour tout $\boldsymbol{x} \in (\boldsymbol{0}, \boldsymbol{\infty})$ et $\boldsymbol{p} \in [\boldsymbol{0}, \boldsymbol{1}]$, où $C(\cdot, \boldsymbol{x})$ est la copule associée à $H(\cdot; \boldsymbol{x})$

A partir de ces notions, il sera étudié plusieurs exemples de lois limites usuelles (logistique négative de Coles et Tawn (1991), ou mélange asymétrique) afin d'étudier la copule limite associée.

Enfin, la dernière section du chapitre présente une partie des travaux de Charpentier (2003) sur les mesures de dépendances de queues conditionnelles. Rappelons que Patton (2004) avait proposé d'étudier les "*exceedences correlations*", définies par

$$r(u) = \begin{cases} corr(X, Y | X \leq F_X^{\leftarrow}(u) \text{ et } Y \leq F_Y^{\leftarrow}(u)) & \text{si } u \leq 0.5, \\ corr(X, Y | X > F_X^{\leftarrow}(u) \text{ et } Y > F_Y^{\leftarrow}(u)) & \text{si } u > 0.5. \end{cases}$$

La séparation entre quantiles élevés et faibles est motivée ici par les applications financières qui en découlent (Longin et Solnik (2001) ou Ang et Chen (2002)): les fortes baisses comme les fortes hausses jouent un rôle en finance. Néanmoins, en assurance, on s'intéresse soit aux hauts quantiles (si X et Y désignent des coûts de sinistres, par exemple), soit aux bas quantiles (si X et Y désignent des résultats). Or cette corrélation dépend très fortement des comportements maginaux, et ne correspondent pas à des mesures de concordance. L'idée naturelle est alors d'utiliser des mesures de corrélation de rang (par exemple le rho de Spearman, $\rho(X, Y) = corr(F_X(X), F_Y(Y))$) ou des mesures basées sur les probabilités de concordances (par exemple le tau de Kendall).

Définition 6.4. *Soient X et Y deux variables aléatoires. On définie les corrélations de rang de queue supérieure par*

$$\rho(u) = \rho((X, Y) | X > F_X^{\leftarrow}(u) \text{ et } Y > F_Y^{\leftarrow}(u)), u \in [0, 1),$$

et le taux de Kendall de queue supérieure par

$$\tau(u) = \tau((X, Y) | X > F_X^{\leftarrow}(u) \text{ et } Y > F_Y^{\leftarrow}(u)), u \in [0, 1).$$

De façon analogue on peut définir une corrélation de rang de queue inférieure ou un taux de Kendall de queue inférieure.

Ces mesures peuvent ainsi être utilisées pour construire des tests d'indépendance de queue, en testant si $\rho(u) \to 0$ quand $u \to 1$, comme le proposent Hájek et Sidák (1967), ou Behnen et Neuhaus (1989). Notons toutefois que si l'étude de la loi théorique des estimateurs "*naturels*" de $\rho(u)$ ou de $\tau(u)$ n'a donné que peu de résultats analytiques dans la littérature, dans le cas particulier de l'indépendance, des résultats sur les U statistiques permettent d'obtenir une normalité asymptotique. Aussi, en utilisant une approximation Gaussienne du nombre d'observations appartenant à la région $[F_X^{\leftarrow}(u), \infty) \times [F_Y^{\leftarrow}(u), \infty)$, on peut obtenir que la loi asymptotique de $\widehat{\rho}_n(u)$, comme celle de $\widehat{\tau}_n(u)$ est un mélange de lois normales,

$$\mathbb{P}(\widehat{\rho}_n(u) \leq r) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} \phi\left(\sqrt{z-1}y\right) \phi\left(\frac{z - n(1-u)^2}{\sqrt{nu(1-u)^2}}\right) dzdy,$$

et

$$\mathbb{P}(\widehat{\tau}_n(u) \leq r) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} \phi\left(\sqrt{z-1}y\right) \phi\left(\frac{z - n(1-u)^2}{\sqrt{nu(1-u)^2}}\right) dzdy,$$

où ϕ désigne la densité de la loi $\mathcal{N}(0,1)$. Une utilisation une approximation supplémentaire, il est alors possible de considérer un test d'indépendance de queue de la forme suivante. X et Y seront indépendants dans la queue supérieure si et seulement si

$$-u_{1-\alpha/2}\sqrt{n(1-u)^2} \leq \widehat{\rho}_n(u) \leq +u_{1-\alpha/2}\sqrt{n(1-u)^2}, \text{ pour } u \text{ proche de } 1,$$

et

$$-u_{1-\alpha/2}\frac{3}{2}\sqrt{n(1-u)^2} \leq \widehat{\rho}_n(u) \leq +u_{1-\alpha/2}\frac{3}{2}\sqrt{n(1-u)^2}, \text{ pour } u \text{ proche de } 1,$$

où $u_{1-\alpha/2}$ correspond au quantile d'ordre $1-\alpha/2$ de la loi $\mathcal{N}(0,1)$.

Ces mesures peuvent également servir de méthode graphique de validation d'ajustement de lois. Si les lois théoriques de $\widehat{\rho}_n(u)$ et de $\widehat{\tau}_n(u)$ ne sont pas connues, des simulations permettent néanmoins de les construire. Aussi des intervalles de confiance permettent de tester d'autres hypothèses que celle d'indépendance de queue. Ceci a été présenté dans Charpentier (2003), sur des données assurantielles.

Notons que là aussi, un grand nombre d'ouvertures sont possibles. Tout d'abord sur la caractérisation de l'indépendance de queue, liée à la recherche des caractéristiques des copules dont la copule limite par troncature est la copule indépendante. L'étude de la puissance du test d'indépendance proposé est également à étudier en détails, puisqu'il est relativement délicat de conclure, y compris pour un vecteur Gaussien (mais le problème se pose également pour les estimateurs classiques de λ qui ne permettent que difficilement de tester l'indépendance de queue, comme le souligne Draisma, Drees, Ferreira et de Haan (2004)). Enfin, là aussi des applications nombreuses sont envisageables. Dias et Embrechts (2004) avaient montré que pour la modélisation de la dépendance dans les quadrants inférieurs et supérieurs, la copule de Clayton (et sa version duale) étaient celles qui s'ajustaient le mieux sur des séries financières. Mais comme le notait Joe, Smith et Weissmann (1992) ou Smith (1994), dans de nombreuses applications de risques environnementaux, les risques extrêmes peuvent être caractérisés par le dépassement de deux seuils conjointement par deux variables (en particulier un fort niveau d'ozone est caractérisé par une vitesse de vent très faible et une température élevée). Les applications en hydrologie sont elles-aussi nombreuses, en particulier sur l'étude des "*low-flow*".

7 Estimation non-paramétrique de densités de copules

Cette partie est développée dans le Chapitre 7.

Ce chapitre présente les résultats de Charpentier, Fermanian et Scaillet (2005, 2006). En dimension 2, les copules sont des fonctions de répartition définies sur le carré $[0,1] \times [0,1]$. Sous l'hypothèse où cette fonction est deux fois continument dérivable, on peut définir la densité associée.

L'estimation de la densité de copule a été introduite dans Behnen, Husková et Neuhaus (1985), ou Gijbels et Mielniczuk (1990), avec des approches non-paramétriques. Or estimer une densité à support compact à l'aide de noyaux s'avère difficile, à cause des problèmes de bord (comme le notait déjà ce papier). Si l'utilisation d'estimateurs à noyaux reste satisfaisante asymptotiquement (Fermanian, Radulović et Wegkanmp (2003)), le biais à distance finie est d'autant plus gênant si l'on s'intéresse aux phénomènes extrêmes (c'est à dire au comportement au voisinage de $(0,0)$ ou de $(1,1)$). Ainsi, il est souvent délicat d'estimer de façon précise $\mathbb{P}(X > F_X^{\leftarrow}(u), Y > F_Y^{\leftarrow}(v))$ pour u et v proches de 1, où X et Y sont deux variables aléatoires de fonction de répartition F_X et F_Y respectivement.

Par exemple, il est naturel de considérer l'estimateur par noyau de la densité des $(\widehat{U}_i, \widehat{V}_i)$ où $\widehat{U}_i = \widehat{F}_X(X_i)$ et $\widehat{V}_i = \widehat{F}_Y(Y_i)$, où $\widehat{F}_X(\cdot)$ et $\widehat{F}_Y(\cdot)$ désignent les fonctions de répartition marginales empiriques (éventuellement en divisées par $(n+1)$ pour éviter la valeur 1),

$$\widehat{F}_X(\cdot) = \frac{1}{n+1}\sum_{i=1}^n \mathbf{1}(X_i \leq \cdot) \text{ et } \widehat{F}_Y(\cdot) = \frac{1}{n+1}\sum_{i=1}^n \mathbf{1}(Y_i \leq \cdot).$$

Behnen, Husková et Neuhaus (1985), puis Gijbeks et Mielniczuk (1990) proposent alors d'estimer la densité de la copule du vecteur (X,Y) par

$$\widehat{c}_h(u,v) = \frac{1}{nh^2}\sum_{i=1}^n K\left(\frac{u-\widehat{F}_X(X_i)}{h}, \frac{v-\widehat{F}_Y(Y_i)}{h}\right), (u,v) \in [0,1]^2$$

qui est l'estimateur par noyau de la densité (Rosenblatt (1956), Silverman (1986)). Mais comme l'avait noté Behnen, Husková et Neuhaus (1985), cet estimateur n'est pas consistent sur la bordure du carré unitaire. Plus particulièrement, cet estimateur a un biais multiplicatif sur les bords:

Proposition 7.1. *Soit $(U_1,V_1),...,(U_n,V_n)$ un échantillon dont le support est $[0,1]^2$, de densité $c(u,v)$, supposée deux fois continûment dérivable sur $(0,1)^2$. Si K est un noyau symétrique de support $[-1,+1]$, alors dans le coin*

$$\mathbb{E}(\widehat{c}_h(0,0)) = \frac{1}{4}\cdot c(0,0) - \frac{1}{2}[c_1(0,0)+c_2(0,0)]\int_0^1 \omega K(\omega)d\omega \cdot h + o(h).$$

sur la bordure

$$\mathbb{E}(\widehat{c}_h(u,0) = \frac{1}{2}\cdot c(u,0) - [c_1(u,0)]\int_0^1 \omega K(\omega)d\omega \cdot h + o(h),$$

pour tout $u \in (0,1)$, et à l'intérieur

$$\mathbb{E}(\widehat{c}_h(u,v)) = c(u,v) + \frac{1}{2}[c_{1,1}(u,v)+c_{2,2}(u,v)]\int_{-1}^1 \omega^2 K(\omega)d\omega \cdot h^2 + o(h^2),$$

pour tout $(u,v) \in (0,1)^2$.

Un biais multiplicatif de trois quarts quart apparaît dans les estimations dans les coins, ce qui conduit à une sous-estimation considérable des probabilités des événements associés (de très larges sinistres, dans les applications assurantielles).

Notons enfin que cet estimateurs, comme les estimateurs par noyaux usuels sont asymptotiquement Gaussiens,

Proposition 7.2. *Soit c une densité de copule deux fois dérivable sur $[0,1]\times[0,1]$. Pour tout $(u,v)\in[0,1]\times[0,1]$, $\widehat{c}_h(u,v)$ est asymptotiquement Gaussien, avec*

$$\sqrt{nh^2}[\widehat{c}_h(0,0)-c(0,0)]\xrightarrow{\mathcal{L}}\mathcal{N}(0,c(0,0)),\ \textit{lorsque } nh^2\to\infty \textit{ et } h\to 0,$$

$$\sqrt{nh^3}[\widehat{c}_h(u,0)-c(u,0)]\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{c(u,0)}{\sqrt{\pi u(1-u)}}\right),\ \textit{lorsque } nh^3\to\infty \textit{ eth}\to 0,$$

$$\sqrt{nh^4}[\widehat{c}_b(u,v)-c(u,v)]\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{c(u,v)}{\sqrt{\pi u(1-u)v(1-v)}}\right),\ \textit{lorsque } nh^4\to\infty \textit{ et } h\to 0.$$

Deux estimateurs sont proposés dans ce dernier chapitre, pour corriger ce biais à distance fini dû aux effets de bords. Le premier repose sur l'utilisation de produits de noyaux Betas, qui présentent l'avantage d'être définis sur le même support que la densité que l'on cherche à estimer:

Définition 7.3. *L'estimateur par noyaux Betas de la densité c est donné par*

$$\widehat{c}_b(u,v)=\frac{1}{n}\sum_{i=1}^{n}K\left(\widehat{F}_X(X_i),\frac{u}{b},\frac{1-u}{b}\right)K\left(\widehat{F}_Y(Y_i),\frac{v}{b},\frac{1-v}{b}\right),$$

où le réel b désigne une taille de fenêtre, et le noyau $K(\cdot,\alpha,\beta)$ est la densité de la loi Beta de paramètres α et β,

$$K\left(x,\alpha,\beta\right)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\mathbf{1}(x\in[0,1])=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}(x\in[0,1]).$$

Cet estimateur est alors nettement meilleur, puisqu'il permet d'éviter le biais multiplicatif au bord:

Proposition 7.4. *Si c est deux fois dérivable sur $[0,1]\times[0,1]$, le biais de $\widehat{c}(u,v)$ est de l'ordre de b, i.e.*

$$\mathbb{E}(\widehat{c}(u,v))=c(u,v)+\mathcal{Q}(u,v)\cdot b+o(b),\ \textit{pour tout } u,v\in[0,1],$$

où le biais $\mathcal{Q}(u,v)$ est

$$\mathcal{Q}(u,v)=(1-2u)c_1(u,v)+(1-2v)c_2(u,v)+\frac{1}{2}\left[u(1-u)c_{1,1}(u,v)+v(1-v)c_{2,2}(u,v)\right],$$

où $c_1(u,v)=\partial c(u,v)/\partial u$, $c_{1,1}(u,v)=\partial^2c(u,v)/\partial u^2$ et $c_{1,2}(u,v)=\partial^2c(u,v)/\partial u\partial v$.

Le biais est en $O(b)$ à l'intérieur du carré $[0,1]\times[0,1]$, alors qu'il est en $O(h^2)$ pour les noyaux symétriques. On a aussi les propriétés suivantes pour la variance de cet estimateur:

Proposition 7.5. *Si c est deux fois dérivable sur* $[0,1]\times[0,1]$. *Pour tout* $(u,v)\in[0,1]\times[0,1]$ *, la variance de* $\widehat{c}_b(u,v)$ *dans le coin* $(0,0)$*, est*

$$Var(\widehat{c}_b(0,0)) = \frac{1}{nb^2}[c(0,0)+o(n^{-1})],$$

dans l'intérieur de la bordure,

$$Var(\widehat{c}_b(0,v)) = \frac{1}{2nb^{3/2}\sqrt{\pi v(1-v)}}[c(u,0)+o(n^{-1})], \text{ pour tout } v\in(0,1),$$

et à l'intérieur,

$$Var(\widehat{c}_b(u,v)) = \frac{1}{4nb\pi\sqrt{v(1-v)u(1-u)}}[c(u,v)+o(n^{-1})], \text{ pour tout } u,v\in(0,1).$$

Une seconde approche peut permettre de corriger des effets de bords. L'idée du second estimateur de la densité repose sur une passage de $[0,1]^2$ à $\mathbb{R}^2$, afin de faire une estimation "*standard*", en utilisant un noyau Gaussien bivarié par exemple, puis de revenir à $[0,1]^2$ en faisant la transformation inverse. L'idée est ici de considérer "*à l'envers*" la transformation proposée par Devroye et Györfi (1981), qui proposait au contraire de ramener le problème de $\mathbb{R}$ à $[0,1]$, avant de revenir à $\mathbb{R}$.

Définition 7.6. *L'estimateur par* Φ*-transformation de la densité c est*

$$\widehat{c}_h(u,v) = \frac{\widehat{f}(\Phi^{\leftarrow}(u),\Phi^{\leftarrow}(v))}{\phi(\Phi^{\leftarrow}(u))\cdot\phi(\Phi^{\leftarrow}(v))},$$

où $\widehat{f}$ *est l'estimation par noyau de la densité du couple* $(\Phi^{\leftarrow}(U),\Phi^{\leftarrow}(V))$*, où* $\Phi:\mathbb{R}\mapsto[0,1]$ *est une fonction bijective, i.e.*

$$\widehat{f}(x,y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\cdot K\left(\frac{y-Y_i}{h}\right).$$

Par exemple, Φ peut être la fonction de répartition d'une loi elliptique (Gaussienne ou Student). Notons que cette procédure permet d'éviter le biais multiplicatif au bord.

Proposition 7.7. *Si* Φ *est continûment dérivable, de dérivée* ϕ*,*

$$\mathbb{E}(\widehat{c}_h(u,v)) = c(u,v) + \frac{o(h)}{\phi(\Phi^{\leftarrow}(u))\phi(\Phi^{\leftarrow}(v))}.$$

et de plus,

$$\begin{aligned} Var(\widehat{c}_h(u,v)) &= \frac{1}{\phi(\Phi^{\leftarrow}(u))\phi(\Phi^{\leftarrow}(v))}\left[\frac{c(u,v)}{nh^2}\left(\int K(\omega)^2 d\omega\right)^2\right] \\ &+ \frac{1}{\phi(\Phi^{\leftarrow}(u))^2\phi(\Phi^{\leftarrow}(v))^2} o\left(\frac{1}{nh^2}\right). \end{aligned}$$

Ces deux estimateurs seront alors utilisés pour estimer des densités de copules, à partir d'échantillons $(X_1, Y_1), ..., (X_n, Y_n)$ dont les lois marginales sont inconnues. Nous présenterons en particulier les résultats de Charpentier, Fermanian et Scaillet (2005) dans le cas où les Y_i sont des données censurées. On s'intéresse ainsi plus particulièrement à l'exemple traité par Frees et Valdez (1995), de tarification d'un contrat de réassurance, où les coûts des sinistres et les frais associés sont traités différemment. En effet, un biais est alors induit par la censure, puisque les données censurées sont généralement des données extrêmes (les coûts individuels des sinistres admettent une limite supérieure contractuelle, inconnue).

Efron et Tibshirani (1993) ont proposé d'utiliser des techniques de type bootstrap pour estimer ce biais. On considère alors B échantillons bootstrapés tirés en tirant suivant les estimateurs de Kaplan-Meier de la censure et de la variable non-censurée. Le biais de l'estimateur de la densité au point (u, v) est alors estimé par

$$\widehat{biais}\,(u,v) = \frac{1}{B}\sum_{k=1}^{B} \widehat{c}_h\,(u,v)^{*,k} - \widehat{c}_h\,(u,v)$$

où $\widehat{c}\,(u)^{*,b}$ est la densité obtenue sur le $k$ième échantillon bootstrapé. Comme nous le verrons sur des données simulées, cette technique permet de corriger de manière très efficace le biais de censure.

Là aussi, si les outils proposés ici permettent d'avoir une meilleure compréhension de la structure de dépendance dans les queues, un grand nombre de questions sont désormais posées. En particulier, la transformation proposée par Devroye et Györki (1981) donne de très bons résultats sur les simulations. Mais le poids des queues est sensiblement influencé par la queue de la distribution de la transformation considérée.

8 *Bijgevoegde stelling*, dépendance temporelle pour évènements climatiques

Cette partie est développée dans le Chapitre 8.

Ce dernier chapitre présente les résultats de Bouëtte et al. (2006), Charpentier et Sibaï (2006) et Charpentier (2006). Ce chapitre présente des résultats connexes à ceux présentés dans les chapitres précédants, sur les risques liés aux évènements climatiques.

La première section se focalise sur la modélisation des tempêtes. En effet, suite aux tempêtes de Décembre 1999 en France et en Belgique, il est apparu que le phénomène de persistance pouvait avoir des impacts collossaux (la première tempête *Lothar* étant survenue moins de 48 heures avant la seconde, *Martin*). Ce phénomène de persistance a été étudié et modélisé par Haslett et Raftery (1989). Malheureusement, en travaillant sur le même jeu de données, il est apparu que leurs motivations pour introduire des modèles à *mémoire longue* n'étaient pas fondées: la lente décroissance des autocorrélations était dû à un effet de saisonnalité, et non pas une racine unité fractionnaire. Néanmoins, en étudiant la série attentivement, il est apparu qu'en otant la composante saisonnière, il restait de la persistance dans les résidus. L'idée a alors été d'utiliser des processus GARMA (*Gegenbauer ARMA*) au lieu des modèles ARFIMA proposés par Haslett et Raftery (1989) (*ARIMA Fractionnaire*). A l'aide de simulations, nous verrons en particulier que si la vitesse du vent était modélisée à l'aide de phénomène à mémoire courte, les probabilités de survenance de tempêtes consécutives seraient fortement sous estimées.

La seconde section étend l'approche sur le vent à la modélisation de la température. L'idée est ici d'estimer correctement la période de retour de la canicule d'août 2003. De même que la canicule qui avait touché Chicago en 1995 (et qui tua plusieurs centaines de personnes), cette canicule n'était pas tant exceptionnelle par les températures atteintes (élevées certes, mais comme d'autres étés très chauds) que par la durée pendant laquelle les températures ont été élevées. En particulier, la température nocture (qui permet au corps de se refroidir) à été très élevée pendant plusieurs jours consécutifs. Là aussi, le risque principal est celui de persistance des températures à un niveau élevé. Toutefois, contrairement au vent qui est stationnaire, la température admet une tendance croissante facilement identificable sur une longue période (+3 degrés en un siècle). Après avoir oté cette tendance (linéaire), il restait, comme pour le vent, une composante saisonnière, et un bruit résiduel à mémoire longue. Mais comme le prétendait Dacunha-Castelle (2004) au lieu de se focaliser sur des processus fractionnaires, il est possible d'obtenir des résultats très proches en modélisant ce résidu par un processus à mémoire courte, avec des queues plus épaisses (que des variables Gaussiennes) . Aussi, dans cette section, nous comparerons deux modèles: un processus GARMA Gaussien (à mémoire longue et à queues fines), et un processus ARMA avec erreurs suivant une loi de Student (à mémoire courte et à queues épaisses). Comme nous le verrons sur des simulations, les périodes de retour dépendent alors fortement de la définition retenue pour la canicule d'août 2003 (11 jours consécutifs avec une température minimal excédant 19° C, ou 3 jours consécutifs avec une température minimal excédant 24° C). En particulier, dans le second cas, les modèles à mémoire longue donne une période de retour inférieure aux modèles à queue épaisse, et inversement pour le second cas.

Enfin, nous nous interesserons dans un troisième temps à l'étude des crues des fleuves, qui ont été étudiées abondamment entre 1940 et 1955. Deux approaches ont ainsi été proposées pour modéliser les maximas annuels. Gumbel (1941) notait que les maximas annuels suivaient une distribution particulière (appelée distribution de Gumbel), mais parrallèlement, sur des séries plus longues (700 ans, contre 100 dans les travaux de Gumbel) Hurst (1951) a noté que les maximas annuels n'étaient pas indépendants, mais au contraire très fortement dépendants: les processus fractionnaires étaient parfait pour modéliser la dynamique de telles séries. Entre ces deux approches a priori contradictoires (indépendance ou forte mémoire ?), il semble être délicat de trancher. Et les modèles retenus donnent de fortes différences si l'on cherche à calculer des périodes de retour de crues, par exemple. Afin de contourner cette difficulté, cette partie propose d'adapter des modèles de données haute fréquence, utilisés en finance afin de modéliser les prix transaction par transaction, à des niveaux de fleuves. Comme nous le verrons, cette modélisation donne des résultats différentes de ceux obtenus par Gumbel lors du calcul de la période de retour.

Par la suite, certaines notions peuvent avoir deux notations différentes entre les chapitres. Mais elles seront rappelées en début de chaque chapitre.

Chapter 1

Modeling dependent risks using copulae

1.1 Modeling multiple risks

In reinsurance, it could be interesting to split a claim cost into several components, e.g. the loss amount (paid to the insured) and the allocated expenses (lawyers, expertise...), see e.g. Frees and Valdez (1998) or Klugman and Parsa (1999). In that case, $\boldsymbol{X}_i = (X_{1,i}, X_{2,i})$'s are the amounts associated with i-th claim. In the context of reinsurance pricing, assuming that allocated expenses are prorata capita, the reimbursement of the reinsurer, with an excess-of-loss treaty (with infinite limit, and deductible d), when a claim expressed as $\boldsymbol{X} = (X_1, X_2)$ occurred, is

$$g(X_1, X_2) = \begin{cases} 0, \text{ if } X_1 \leq d, \\ X_1 - d + \dfrac{X_1 - d}{X_1} X_2, \text{ if } X_1 > d. \end{cases}$$

The pure premium per claim is then $\mathbb{E}(g(\boldsymbol{X}))$ which is based on the joint distribution of $\boldsymbol{X}$ since g is nonlinear.

Identically, for some financial derivatives, dependencies play an important rule. Let $X_{1,i}$ and $X_{2,i}$ denote the price at time i of two financial assets. Let $g_k(X_1, X_2)$ denote the payoff at maturity 1 of some European option, where g_k is, a priori, nonlinear, and where k is a strike.

- Quanto derivatives, with payoff $g_k(x_1, x_2) = x_2(x_1 - k)_+$. It is a call, expressed in domestic currency (X_2 denotes the exchange rate), and based on some overseas asset (with price X_1),
- Spreads derivatives, with payoff $g_k(x_1, x_2) = (x_1 - x_2 - k)_+$. It is a call on the spread between the prices of the two assets, X_1 and X_2,
- Basket derivatives, with payoff $g_k(x_1, x_2) = (\alpha x_1 + \beta x_2 - k)_+$. It is a call on the portfolio with two assets,
- Min-max derivatives, with respective payoffs $g_k(x_1, x_2) = (\min\{x_1, x_2\} - k)_+$ and $g_k(x_1, x_2) = (\max\{x_1, x_2\} - k)_+$. Those are call options respectively on the minimum and the maximum of two prices.

In life insurance, analogous of those financial derivatives can be considered. Consider a husband and his wife, and denote by T_x and T_y the survival life lengths, assuming that the man has age x and his wife y when they buy a life-insurance contract. Several contracts can be considered, where capital C_k is due each year k,

- as long as the spouses are both still alive, $g(T_x, T_y) = \sum_{k=1}^{\infty} v^k C_k \mathbf{1}(T_x > k \text{ and } T_y > k)$,
- as long as there is a survivor, $g(T_x, T_y) = \sum_{k=1}^{\infty} v^k C_k \mathbf{1}(T_x > k \text{ or } T_y > k)$.

Note that C_k can be stochastic if the capital is indexed on a financial asset, or if the income is indexed by some stochastic interest rate. The associated pure premium, called annuities when $C_k = 1$, can be written respectively (with standard actuarial notations)

$$a_{xy} = \sum_{k=1}^{\infty} v^k \mathbb{P}(T_x > k, T_y > k) \text{ and } a_{\overline{xy}} = \sum_{k=1}^{\infty} v^k \mathbb{P}(T_x > k \text{ or } T_y > k).$$

Those contracts are usually built for an husband and his wife, i.e. contracts with more risks can be considered if children are involved, or even higher when dealing with collective insurance contracts.

Applications with a high number of risks can also be considered, in credit risk for instance. Let $\boldsymbol{X} = (X_1, ..., X_d)$ denote the vector of indicator variables, indicating if the i-th contract defaulted during a given period of time. If a credit derivative is based on the occurrence of k defaults among d companies, and thus, the pricing is related to the distribution of the number of defaults, N, defined as $N = X_1 + ... + X_d$. Under the assumption of possible contagious risks, the distribution of N should integrate dependencies.

1.2 Distribution functions in $\mathbb{R}^d$, and copulae

A copula provides a uniform representation of a multivariate distribution. Copulae are important since they allow to separate the effect of the dependence from the effects of the marginal distributions (due to Sklar's Theorem, see Sklar (1959)).

Remark 1.2.1. *We shall keep in this thesis the word "copula" used first by Sklar (1959), which originates from the Latin noun for a "link or tie" that connects two different things. But actually, such a function appeared earlier in the literature, e.g. in Eyraud (1934) which considers a "fonction de corrélation" (which is the copula function), and latter on, in Hoeffding (1940) who considered a "standardized version of a random pair" (but which considered uniform distributions on* $[-1/2, +1/2]$ *instead of* $[0, 1]$*).*

Definition 1.2.2. *A d-dimensional copula is a d-dimensional distribution function restricted to* $[0, 1]^d$ *with standard uniform margins, for a non-negative integer d.*

In order to characterize those functions, observe that a proper definition of *increasing* function is needed. In dimension 1, recall that a distribution function F is increasing since a probability measure is positive: for $a < b$, $F(b) - F(a) = \mathbb{P}(X \in (a, b]) \geq 0$, where X is a random variable with distribution function F. The analogous in dimension d, is that whatever rectangle $[\boldsymbol{a}, \boldsymbol{b}]$, $\mathbb{P}(\boldsymbol{X} \in (\boldsymbol{a}, \boldsymbol{b}]) \geq 0$. This yields intuitively the notion of d-increasingness:

Definition 1.2.3. *A function $F : \Omega_1 \times ... \times \Omega_d \subset \mathbb{R}^d \to \mathbb{R}$ is said to be d-increasing if for all rectangle $(\boldsymbol{a}, \boldsymbol{b}]$ in $\Omega_1 \times ... \times \Omega_n$, $V_F((\boldsymbol{a}, \boldsymbol{b}]) \geq 0$, where*

$$V_F([\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{\boldsymbol{a}}^{\boldsymbol{b}} F(\boldsymbol{t}) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} ... \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} F(\boldsymbol{t}) \tag{1.1}$$

for all t, where

$$\Delta_{a_i}^{b_i} F(\boldsymbol{t}) = F(t_1, ..., t_{i-1}, b_i, t_{i+1}, ..., t_n) - F(t_1, ..., t_{i-1}, a_i, t_{i+1}, ..., t_n). \tag{1.2}$$

Note that in dimension 2, this can be written analogously

$$F(x_1, y_1) + F(x_2, y_2) \geq F(x_1, y_2) + F(x_2, y_1),$$

for all $x_1 \leq x_2$ and $y_1 \leq y_2$, and such a function will also be said to be supermodular.

Hence, copulae can be equivalently defined as functions $C : [0,1]^d \to [0,1]$ satisfying, for $u_1, ..., u_d, v_1, ..., v_d \in [0,1]$ with $x_i \leq y_i$ for all $i = 1, ..., d$, the conditions

$$C(1, ..., 1, u_i, 1, ..., 1) = u_i, \tag{1.3}$$

$$C(u_1, ..., u_{i-1}, 0, u_{i+1}, ..., u_d) = 0, \tag{1.4}$$

$$C \text{ is } d\text{-increasing.} \tag{1.5}$$

In fact, it is easily seen that Equations (1.3) and (1.4) translates into the uniformity of the margins. Moreover, Equations (1.3), (1.4), and (1.5) imply that C increases in each variable as well that C is Lipschitz-continuous with Lipschitz constant one. Note that in the case where C is differentiable, its density is

$$c(u_1, ..., u_d) = \frac{\partial^d C(u_1, ..., u_d)}{\partial u_1 ... \partial u_d}.$$

Example 1.2.4. *Figure (1.1) shows the shape of a copula (the so-called Gumbel copula), in dimension 2, defined as*

$$C(u, v) = \exp\left(-\left([-\log u]^\theta + [-\log v]^\theta\right)^{1/\theta}\right), u, v \in [0,1],$$

where $\theta \geq 1$ (here $\theta = 2$), with on top the surface of C and the associated level curves. Since this copula is twice differentiable, its density exists, and is plotted bellow. Note that when $\theta \to \infty$, the limiting copula is $C(u, v) = \min\{u, v\}$, also called upper Fréchet-Hoeffding copula (see Section 1.5.8). And finally, when $\theta \to 1$, the limiting copula is $C(u, v) = uv$, the independent copula.

1.3 Coupling marginal distributions, and Sklar theorem

One of the most important and useful result about copulae is Sklar's Theorem. A proof of Theorem 1.3.1 can be found e.g. in Nelsen (1999) or in Sklar (1959).

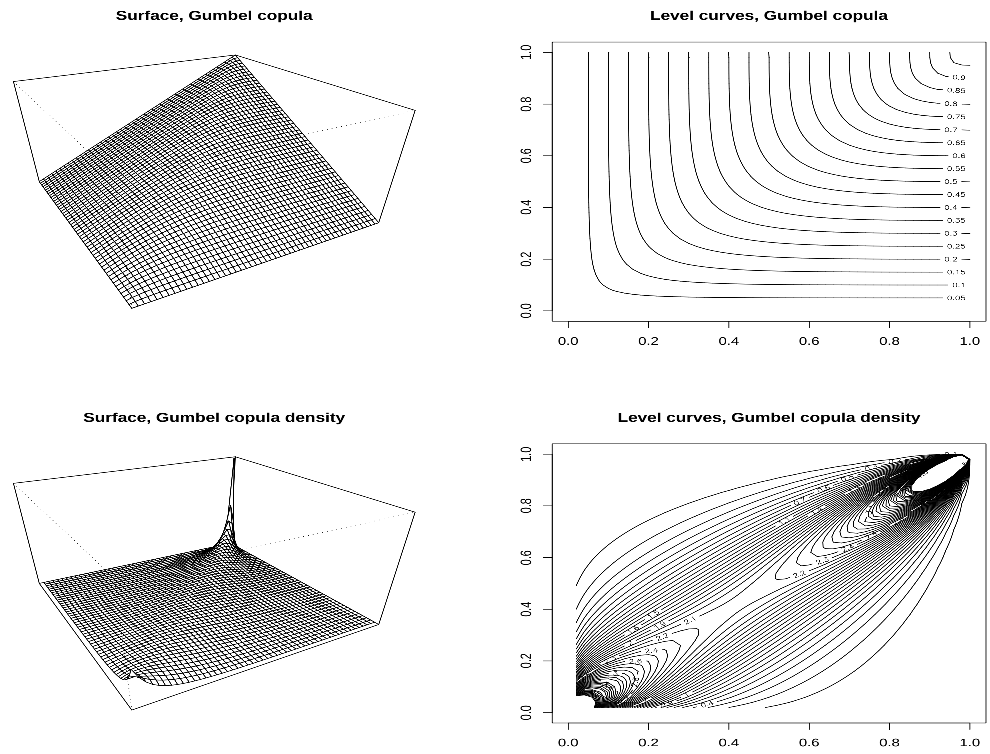


Figure 1.1: Copula on top (the distribution function), and the associated density below.

Theorem 1.3.1. *1. Let C be a d-dimensional copula and $F_1, ..., F_d$ be univariate distribution functions. Then, for $\boldsymbol{x} = (x_1, ..., x_d) \in \mathbb{R}^d$,*

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_d(x_d)) \tag{1.6}$$

defines a distribution function with marginal distribution functions $F_1, ..., F_d$.

2. Conversely, for a d-dimensional distribution function F with marginal distributions $F_1, ..., F_d$ there is a copula C satisfying Equation (1.6)*. This copula is not necessarily unique, but it is if $F_1, ..., F_d$ are continuous, given by*

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_n^{\leftarrow}(x_n)), \tag{1.7}$$

for any $\boldsymbol{u} = (u_1, , ..., u_d) \in [0,1]^d$, where $F_1^{\leftarrow}, ..., F_d^{\leftarrow}$ denote the generalized left continuous inverses of the F_i's, i.e. $F_i^{\leftarrow}(t) = \inf\{x \in \mathbb{R}, F_i(x) \geq t\}$ for all $0 \leq t \leq 1$.

Sklar's Theorem constitutes the motivation for calling copulae *dependence structures* that capture scale invariant dependence properties. In fact, we see from Equation (1.6) that C couples the marginals $F_1, ..., F_n$ to the joint distribution function F separating thus dependence and marginal behaviors.

Further, from the second part of the theorem, consider the following definition,

Definition 1.3.2. *Consider a random vector $\boldsymbol{X}$ with joint distribution function F and continuous marginal distributions $F_1, ..., F_d$. The copula C of $\boldsymbol{X}$ is the copula associated with F, i.e.*

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d)), \tag{1.8}$$

for any $\boldsymbol{u} = (u_1, , ..., u_d) \in [0,1]^d$.

Further, copulae satisfy an invariance property: the dependence structure between random variables does not changes under increasing and continuous transformations of the margins. More formally, let $\boldsymbol{X} = (X_1, ..., X_d)$ be a random vector with copula C. For any functions $\phi_i : \mathbb{R} \to \mathbb{R}$ continuous and strictly increasing functions, $i = 1, ..., d$, C is also the copula of $(\phi_1(X_1), ..., \phi_d(X_d))$ (see Joe (1997), or Embrechts, Hoeing and Juri (2002)). Note that the continuity assumptions of the ϕ_i's is not needed if the X_i's are continuous.

This invariance property means that copulae are the natural framework to study dependence properties which are invariant under increasing transformations of the margins (the "*scale invariance property*" in Hoeffding (1940)). For example, consider a portfolio of dependent insurance policies, and assume that the losses derive from a multivariate distribution. The joint distribution for the losses will have the same copula as the copula of the logarithm of the losses, or any integral-transforms of the losses.

Another example of copula is the following. Throughout the rest of the thesis we will encounter other examples of copulae.

Example 1.3.3. *If $d = 2$, the Marshall and Olkin copula with parameters $\alpha, \beta \in [0,1]$ is defined for $u, v \in [0,1]$ as*

$$C_{\alpha,\beta}(u,v) = uv \min\{u^{-\alpha}, v^{-\beta}\} = \min\{u^{1-\alpha}v, uv^{1-\beta}\}. \tag{1.9}$$

This copula can be extended in higher dimension as

$$C_{\alpha_1,...,\alpha_n}(u_1, ..., u_d) = u_1...u_d \min\{u_1^{-\alpha_1}, ..., u_d^{-\alpha_d}\},$$

where $\alpha_1, ..., \alpha_d \in [0,1]$ For some interpretations of this copula, see Section 1.7.

Definition 1.3.4. *Copula $C^{\perp}(u_1, ..., u_d) = x_1 \times ... \times x_d$ is called the independent copula.*

Note that a random vector $\boldsymbol{X}$ has independent components if and only if $C^{\perp}$ is a copula of $\boldsymbol{X}$.

Definition 1.3.5. *Given $F_1, ..., F_d$ some univariate distribution functions, the class of d-dimensional distribution functions F with marginal distributions $F_1, ..., F_d$ respectively, is called a Fréchet class, denoted $\mathcal{F}(F_1, ..., F_d)$.*

Note that those classes are bounded, i.e.

$$\max\{0, F_1(x_1) + ... + F_d(x_d) - (d-1)\} \leq F(x_1, ..., x_d) \leq \min\{F_i(x_i), i = 1, ..., d\}, \tag{1.10}$$

for all $\boldsymbol{x} = (x_1, ..., x_d)$, and any $F \in \mathcal{F}(F_1, ..., F_d)$. Those bounds are called Fréchet-Hoeffding bounds, and they will be studied in more details in section 1.4.

1.3.1 Survival distributions in $\mathbb{R}^d$

In the univariate case, the survival distribution function of random variable X is $\overline{F}_X(x) = \mathbb{P}(X > x) = 1 - F_X(x)$, $x \in \mathbb{R}$. The extension in dimension 2 yields

$$\overline{F}_{X,Y}(x,y) = \mathbb{P}(X > x, Y > y) = 1 - F_X(x) - F_Y(y) + F_{X,Y}(x,y) \neq 1 - F_{X,Y}(x,y),$$

where $F_X(x) = \mathbb{P}(X \leq x)$ and $F_Y(y) = \mathbb{P}(Y \leq y)$. If C denotes a copula of (X, Y), note that expression can be written

$$\overline{F}_{X,Y}(x, y) = \overline{F}_X(x) + \overline{F}_Y(y) - 1 + C(1 - \overline{F}_X(x), 1 - \overline{F}_Y(y)),$$

which may also be written

$$\overline{F}_{X,Y}(x, y) = C^*(\overline{F}_X(x), \overline{F}_Y(y)),$$

where

$$C^*(u, v) = u + v - 1 + C(1 - u, 1 - v), u, v \in [0, 1] \times [0, 1],$$

will be called survival copula (since it is a copula). Note that if C is the distribution function of $(F_X(X), F_Y(Y)$, C^* is the distribution function of $(\overline{F}_X(X), \overline{F}_Y(Y))$.

Example 1.3.6. *Figure 1.2 shows some random generations of copula C (here Gumbel copula), with the density of the copula on the right, and random simulations of copula C^*, and the associated density below. The survival copula is also called "rotated" copula since the density is simply rotated around the center of the unit square.*

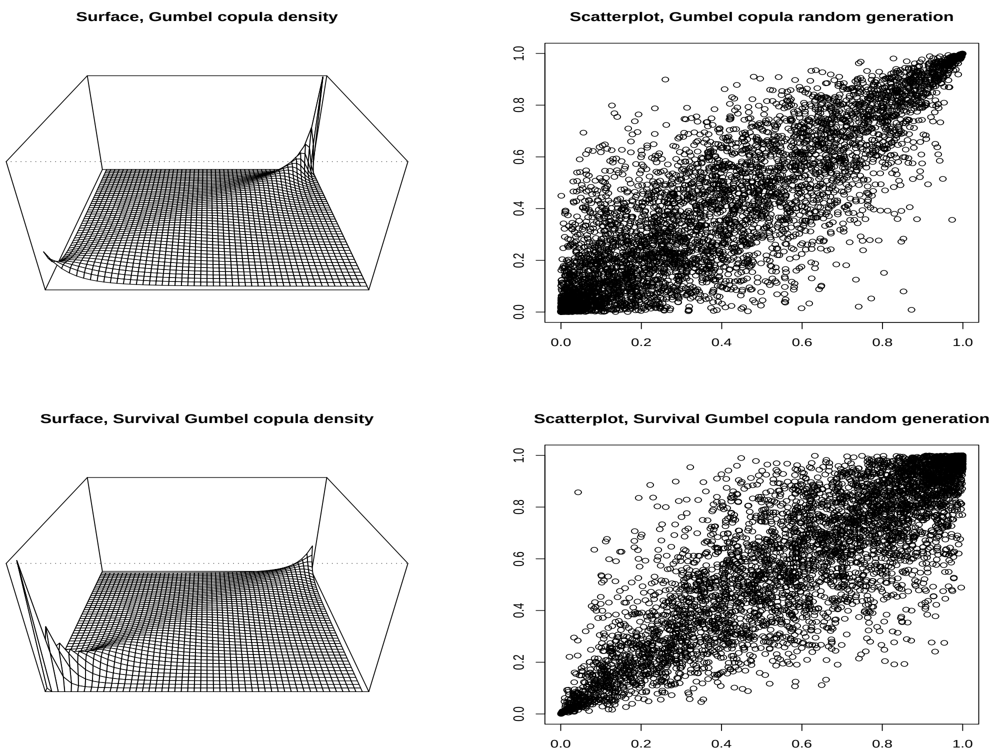


Figure 1.2: A copula on top, and the associated survival copula below.

This survival copula can be defined more generally in dimension d, using Poincarré's formula (see Feller (1971)).

Definition 1.3.7. *Let C denote a n-dimensional copula, the function C^* defined by*

$$C^*(u_1, ..., u_d) = \sum_{k=0}^{d} \left((-1)^k \sum_{i_1,...,i_d} C(1, ..., 1, 1 - u_{i_1}, 1, ...1, 1 - u_{i_k}, 1, ..., 1) \right), \quad (1.11)$$

for all $\boldsymbol{u} = (u_1, ..., u_d) \in [0,1] \times ... \times [0,1]$*, is a copula, called survival copula or dual copula, associated to* C*. Further, if* $\boldsymbol{X} = (X_1, ..., X_d)$ *has distribution function* C*, then* $(1 - X_1, ..., 1 - X_d)$ *has distribution function* C^**.*

1.3.2 Topological properties of the set of copulae and extensions

One of the most important property is that the class of copulae is convex, i.e. mixtures of copulae can be considered. More formally (see Nelsen (1999)), if $\{C_\theta, \theta \in \mathbb{R}\}$ is a collection of copulae, and H a distribution function on $\mathbb{R}$, the function

$$C(u_1, ..., u_d) = \int_{\mathbb{R}} C_\theta(u_1, ..., u_d)\, dH(\theta)$$

is a copula.

Furthermore, copulae are interesting when working on probabilistic arithmetic (see e.g. Williamson (1989)), which is the study of functions of risks $\psi(X, Y)$, e.g. $X + Y$. Recall that given two (univariate) distribution functions F_X and F_Y, the convolution $F_X \otimes F_Y$ is the function defined as

$$F_X \otimes F_Y(x) = \int_{\mathbb{R}} F_X(x - t)\, dF_Y(t) \text{ for all } x \in \mathbb{R}.$$

It is well known that $F_X \otimes F_Y$ is the distribution function of $X + Y$ when X and Y are independent, with respective distribution functions F_X and F_Y. Analogously, it might be interesting to see if there exists a function ψ such that the mixture $pF_X + (1-p)F_Y$ can be the distribution function of $\psi(X, Y)$. This question has arisen in Alsina and Schweizer (1988) and Alsina, Nelsen and Schweizer (1993).

Definition 1.3.8. *A binary operation* χ *on a set of distribution function is said to be derivable if there is a function* ψ*, Borel-measurable, such that, for all distribution functions* F_X *and* F_Y*, associated to random variables* X *and* Y *respectively,* $\chi(F_X, F_Y)$ *is the distribution function of* $\psi(X, Y)$*.*

Particular cases of binary operations are mixtures operations,

$$\chi(F_X, F_Y) = pF_X + (1-p)F_Y \text{ where } 0 < p < 1.$$

and geometric mixtures, $\chi(F_X, F_Y) = \sqrt{F_X F_Y}$. As shown in Alsina and Schweizer (1988), mixture and geometric mixtures operations are not derivable. In order to characterize operations on distribution functions that can be deduced from operations on random variables, quasi-copulae have been introduced (see Alsina, Nelsen and Schweizer (1993)).

In dimension 2, copulae were defined as functions $C : [0,1] \times [0,1] \to [0,1]$ such that

$$C(0, v) = C(u, 0) = 0, 0 \le u, v \le 1 \tag{1.12}$$

$$C(1, v) = v,\ C(u, 1) = u, 0 \le u, v \le 1 \tag{1.13}$$

$$C(u, v) + C(u', v') \ge C(u, v') + C(u', v), 0 \le u \le u' \le 1, 0 \le v \le v' \le 1. \tag{1.14}$$

Thus, a copula is the restriction to the unit square of a distribution function with uniform margins on $[0,1]$. Let $\mathcal{S}$ denote the set of functions $C : [0,1] \times [0,1] \to [0,1]$ satisfying Equations (1.12) and (1.13) such that C is increasing in each variables. As in Bassan and Spizzichino (2004), we shall call these functions extended semicopulae. Note that some properties of those h-copulae will be studied in Chapter 4.

Definition 1.3.9. *A function $C : [0,1] \times [0,1] \to [0,1]$ increasing in each variables, such that $C(0,v) = C(0,u) = 0$, $C(1,v) = v$ and $C(u,1) = u$, for all $0 \leq u, v \leq 1$ will be called a semicopula. The set of semicopulae will be denoted $\mathcal{S}$.*

Further, let $\mathcal{H}$ denote the set of continuous strictly increasing functions $[0,1] \to [0,1]$ such that $h(0) = 0$ and $h(1) = 1$. Those functions would be called distortion functions. Note that $h \in \mathcal{H}$ if and only if $h^{\leftarrow} \in \mathcal{H}$, so that $(\mathcal{H}, \circ)$ defines a group, where $\circ$ denotes the composition operation. The identity is the identical function on $[0,1]$. For all $h \in \mathcal{H}$ and $C \in \mathcal{C}$, define

$$\Psi_h(C)(u,v) = h^{\leftarrow}(C(h(u), h(v))), 0 \leq u, v \leq 1.$$

Such a function will be called a h-copula. $\mathcal{H}$-copulae will be functions $\Psi_h(C)$ for some distortion function h and some copula C. Further, notice that for $h, h' \in \mathcal{H}$,

$$\Psi_{h \circ h'}(C)(u,v) = (\Psi_h \circ \Psi_{h'})(C)(u,v), 0 \leq u, v \leq 1.$$

A copula will be said to be symmetric or exchangeable if

$$C(u,v) = C(v,u), 0 \leq u, v \leq 1.$$

Exchangeable $\mathcal{H}$-copulae are called (simply) semicopulae.

Definition 1.3.10. *If $h \in \mathcal{H}$ is a convex distortion function, and C is a copula, then*

$$\Psi_h(C)(u,v) = h(C(h^{\leftarrow}(u), h^{\leftarrow}(v))) \tag{1.15}$$

is a copula. It will be called distorted copula.

Example 1.3.11. *A particular case is when h is a power function, and when the power is the invert of an integer, $h(x) = x^{1/n}$, i.e.*

$$\Psi_h(C)(u,v) = C^n(u^{1/n}, v^{1/n}), 0 \leq u, v \leq 1 \text{ and } n \in \mathbb{N}.$$

Note that this copula is the survival copula of the componentwise maxima: if $(X_1, Y_1), ..., (X_n, Y_n)$ is an i.i.d. sample, where the (X_i, Y_i)'s have copula C, then the survival copula of $(X_{n:n}, Y_{n:n})$ is $\Psi_h(C)$ (see Chapter 6 of this thesis for more details on this functional equation).

Example 1.3.12. *Let ϕ denote a convex decreasing function on $(0,1]$ such that $\phi(1) = 0$, and define $C(u,v) = \phi^{\leftarrow}(\phi(u) + \phi(v))$. This function is a copula, called Archimedean copula, and function ϕ is a generator of that copula (see Section 1.5 for a detailed presentation). The class of Archimedean copulae is stable by distortion. Let C denote an Archimedean copula with generator ϕ, and $h \in \mathcal{H}$ a convex distortion function, then $\Psi_h(C)$ is also an Archimedean copula, with generator $\phi \circ h^{\leftarrow}$.*

Genest and Rivest (2001) called such a transformation the *multivariate probability integral transformation*. Wang, Nelsen and Valdez (2005) called this copula a *distorted copula*.

Definition 1.3.13. *A semicopula C which satisfies the Lipschitz condition*

$$|C(u_1, v_1) - C(u_2, v_2)| \leq |u_1 - u_2| + |v_1 - v_2|,$$

for all $0 \leq u_1, u_2, v_1, v_2 \in [0,1]$ will be called a quasicopula. The set of quasicopulae will be denoted $\mathcal{Q}$.

Definition 1.3.14. *An associative semicopula C, i.e. $C(C(u,v),w) = C(u, C(v,w))$, for all $0 \leq u, v, w \leq 1$) will be called a t-norm.*

Note (see Durante and Sempi (2004)) that $\mathcal{S}$ strictly includes the family of the so-called quasicopulae. For example, $\Psi_h\left(C^{\perp}\right)$ is a quasicopula if and only if h is a distortion function such that $-\log h$ is convex, in dimension 2. Furthermore $\mathcal{S}$ strictly includes t-norms.

For the description of aging, consider the following family of semicopulae,

$$\mathcal{A} = \left\{\Psi_h\left(C^{\perp}\right), h \in \mathcal{H}\right\}.$$

Thus, the elements of $\mathcal{A}$ can be written

$$C(u,v) = h^{\leftarrow}(h(u)\, h(v)) = \phi^{\leftarrow}(\phi(u) + \phi(v)),$$

where $h \in \mathcal{H}$ is called the multiplicative generator, and $\psi = -\log h$ the additive generator. Those elements of $\mathcal{A}$ are Archimedean t-norm (see Schweizer and Sklar (1983)), and will be called Archimedean semicopulae. As mentioned earlier, an Archimedean semicopula is a copula if and only if $-\log h$ is convex.

Further, as shown in Bassan and Spizzichino (2004), family $\mathcal{S}$ is closed under operation Ψ_h, in the sense that for all $C \in \mathcal{S}$ and all $h \in \mathcal{H}$, $\Psi_h(C) \in \mathcal{S}$. But this property does not hold for the family of copulae $\mathcal{C}$.

Definition 1.3.15. *A capacity on a measurable space $(\Omega, \mathcal{A})$ is a set function $C : \mathcal{A} \to [0,1]$ such that $C(\emptyset) = 0$, $C(\Omega) = 1$, and $A \subset B$ implies $C(A) \leq C(B)$. Further, a capacity is said to be convex if for all A, B, $C(A) + C(B) \leq C(A \cup B) + C(A \cap B)$.*

Note that a capacity is a weaker notion than a probability, and therefore can then be seen as a *non-additive probability measure* (see Denneberg (1994)). Recall further that if f denotes a distortion function, strictly convex, then for any probability measure $\mathbb{P}$, $f \circ \mathbb{P}$ is a convex capacity. Such a result can be extended in higher dimension: if T denotes a t-norm, then, for any probability measures $\mathbb{P}$ and $\mathbb{Q}$, $T(\mathbb{P}(\cdot), \mathbb{Q}(\cdot))$ is capacity.

Example 1.3.16. ***Interest of quasi-copulae*** *- As shown in Alsina, Nelsen and Schweizer (1993), if χ is a binary operation, derivable from a function ψ, and induced pointwise (i.e. $\chi(F_X, F_Y)(t) = \chi(F_X(t), F_Y(t))$ for all t), then either*

- $\psi(X,Y) = \max\{X,Y\}$ *and then χ is a quasi-copula,*
- $\psi(X,Y) = \min\{X,Y\}$ *and then χ is the dual of a quasi-copula (i.e. $u + v - \chi(u,v)$ is a quasi-copula),*
- $\psi(X,Y) = X$ *and $\chi(u,v) = u$, or $\psi(X,Y) = Y$ and $\chi(u,v) = v$.*

Note further (see Nelsen, Quesada Molina, Schweizer and Sempi (1996)) that all those results also hold in higher dimension.

Example 1.3.17. ***Interest of semi-copulae*** - *As mentioned in Bassan and Spizzichino (2004), the bivariate aging is the result of the interplay between dependence (copulae) and univariate aging. Hence, when describing dependence induced by joint survival function, three functions can be considered, in the particular case of exchangeable positive random variables. Set $G = F_X = F_Y$ for convenience, and let F denote the joint distribution function.*

1. *The survival copula,*

$$C^*(u, v) = \overline{F}\left(\overline{G}^{\leftarrow}(u), \overline{G}^{\leftarrow}(v)\right), 0 \leq u, v \leq 1.$$

2. *The multivariate aging function (see Bassan and Spizzichino (2000)),*

$$B(u, v) = \exp\left(-\overline{G}^{\leftarrow}\left(\overline{F}(-\log u, -\log v)\right)\right), 0 \leq u, v \leq 1.$$

   B is a semicopula. Such a function is related to function $h(x, y) = \overline{G}^{\leftarrow}\left(\overline{F}(x, y)\right)$ introduced in Barlow and Spizzichino (1993).

3. *The Archimedean semicopula with additive generator $\overline{G}^{\leftarrow}$,*

$$A(u, v) = \overline{G}\left(\overline{G}^{\leftarrow}(u) + \overline{G}^{\leftarrow}(v)\right) = \Gamma^{\leftarrow}(\Gamma(u)\Gamma(v)), 0 \leq u, v \leq 1,$$

   where $\Gamma(x) = \exp\left(-\overline{G}^{\leftarrow}(x)\right)$ is a function in $\mathcal{H}$, and thus $A = \Psi_\Gamma\left(C^\perp\right)$.

Based on those functions, several interpretations of aging dependencies can be considered (see Bassan and Spizzichino (2004)). For instance, the independence case. If $C^* = C^\perp$, lifetimes are independent. If $A = C^\perp$, then the one-dimensional marginal distributions G are exponential. And finally, if $B = C^\perp$ then $\overline{F}(x,y) = \overline{G}(x+y)$ and therefore, $\overline{F}$ Schur-constant (see Marshall and Olkin (1979)). From Barlow and Mendel (1992), this is equivalent to the requirement that for any $h > 0$, and all $x, y \geq 0$, the following inequality holds,

$$\mathbb{P}(X > x + h | X > x, Y > y) = \mathbb{P}(Y > y + h | X > x, Y > y).$$

Conditionally on the same history of survivals ($\{X > x, Y > y\}$), the marginals distribution of residual lifetimes are identical. Hence, residual lifetimes are exchangeable (see Spizzichino (2001)).

1.4 Fréchet-Hoeffding bounds

Hoeffding (1940, 1942) gave an explicit formulation of the statement "*there is a functional dependence between random variables X and Y*": there exits a function $g : \mathbb{R}^2 \to \mathbb{R}$ such that $g(X, Y) = 0$ almost surely. Equivalently, if $\mathcal{G} = \{(x, y), g(x, y) = 0\}$, this statement can be written as $\mathbb{P}((X, Y) \in \mathcal{G}) = 1$. This formulation is a more general concept than those studied in this section (and this thesis),

- $\mathcal{G}$ is a strictly monotone increasing curve, and there exists a strictly increasing function ϕ such that $Y = \phi(X)$,

- $\mathcal{G}$ is a strictly monotone decreasing curve, and there exists a strictly decreasing function ϕ such that $Y = \phi(X)$.

Hoeffding (1942) proved that those two cases yield bounds on Fréchet classes. In higher dimension, the following results holds, i.e. the family of copulae is bounded: for all copula C,

$$\begin{aligned} C^-(u_1, ..., u_d) &= \max\{0, u_1 + ... + u_d - (d-1)\} \\ &\leq C(u_1, ..., u_d) \\ &\leq C^+(u_1, ..., u_d) = \min\{u_1, ..., u_d\}, \end{aligned}$$

for all $\boldsymbol{u} = (u_1, ..., u_d) \in [0,1]^d$. C^- and C^+ the so-called *Fréchet-Hoeffding* lower and upper bounds.

Further, note that C^+ is (always) a copula, and C^- is a copula only if $d = 2$. In such a case, the supports of those copulae are the diagonal, the main one for the upper bound, the other one for the lower. Graphs of those two copulae (when $n = 2$) can be visualized on Figure 1.3.

For the lower Fréchet bound, the following assertions hold,

- If $d = 2$, C^- is the distribution function of $\boldsymbol{U} = (U, 1-U)$ where U is uniformly distributed on $[0,1]$.

- (X, Y) has copula C^- if and only if there is a non-decreasing function ϕ and a non-increasing function ψ such that $(X, Y) \overset{\mathcal{L}}{=} (\phi(Z), \psi(Z))$ for some random variable Z.

- (X, Y) has copula C^- if and only if there is a strictly decreasing function ϕ such that $Y = \phi(X)$.

For the upper bound,

- C^+ is the distribution function of $\boldsymbol{U} = (U, ..., U)$ where U is uniformly distributed on $[0,1]$.

- $\boldsymbol{X} = (X_1, ..., X_n)$ has copula C^+ if and only if there are non-decreasing functions ϕ_i's such that $(X_1, ..., X_n) \overset{\mathcal{L}}{=} (\phi_1(Z), ..., \phi_d(Z))$ for some random variable Z.

- (X, Y) has copula C^+ if and only if there is a strictly increasing function ϕ such that $Y = \phi(X)$

The particular case of the upper Fréchet-Hoeffding bound, leading to *comonotonic* random variables, will be developed in Section 4.1.6. The lower case in dimension 2 will be called either *anti-comonotonic* or *counter-comonotonic*.

Example 1.4.1. *Those bounds on copulae provide also bounds for several quantities. For instance, when $d = 2$, if $g : \mathbb{R}^2 \to \mathbb{R}$ is supermodular (or 2-increasing when $d = 2$), i.e.*

$$g(x_1, y_1) + g(x_2, y_2) \geq g(x_1, y_2) + g(x_2, y_1),$$

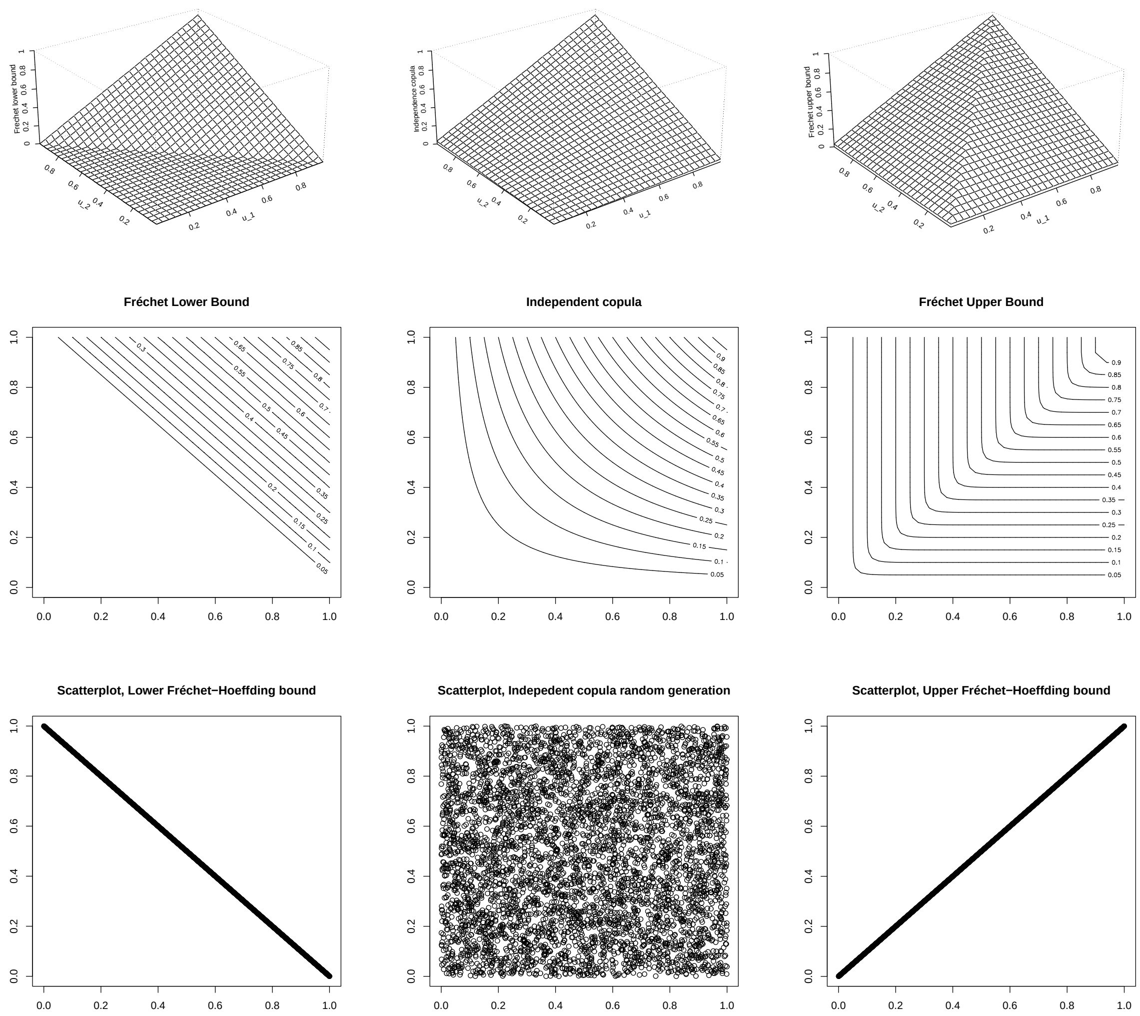


Figure 1.3: Lower bound, independent copula and upper bound, $n = 2$, with the surface of the distribution functions, and the associated level curves, including below scatterplot of random generations.

for all $x_1 \leq x_2$ and $y_1 \leq y_2$, then for all random vector (X, Y) with marginal distribution functions F_X and F_Y,

$$\mathbb{E}(g(F_X^{\leftarrow}(U), F_Y^{\leftarrow}(1-U))) \leq \mathbb{E}(g(X,Y)) \leq \mathbb{E}(g(F_X^{\leftarrow}(U), F_Y^{\leftarrow}(U))),$$

where U is uniformly distributed on $[0,1]$ (see Tchen (1980)). Stop-loss premium for the sum of two risks (e.g. $\mathbb{E}(X+Y-t)_+$) can be bounded by comonotonic and anti-comonotonic versions of X and Y since $(x,y) \mapsto (x+y-t)_+$ is supermodular. Moreover, most of the multiple-life insurance premiums can be written as the expected value of some supermodular function of time-until-death random variables for an husband and his wife.

Nelsen, Quesada Molina, Rodríguez Lallena and Úbeda Flores (2004) gave other bounding results when restricting the set of copulae. Let $\mathcal{S}$ denote a nonempty set of bivariate functions with a common domain, then $\overline{S}$ and $\underline{S}$ denote, respectively, the pointwise supremum and infimum of $\mathcal{S}$ i.e., for each (u, v)

$$\overline{S}(u,v) = \sup\{S(u,v), S \in \mathcal{S}\} \text{ and } \underline{S}(u,v) = \inf\{S(u,v), S \in \mathcal{S}\}.$$

As proved in Nelsen, Quesada Molina, Rodríguez Lallena and Úbeda Flores (2004), if $\mathcal{S}$ is a set of copulae, then $\overline{S}$ and $\underline{S}$ are both quasi-copulae. This can be extended in any dimension $d \geq 2$. Hence, if C^- is usually not a copula, it is always a quasi-copula.

1.5 Archimedean copulae

An important class of copulae are the so-called *Archimedean* copulae (from Ling (1965) and Genest and MacKay (1986a))

Definition 1.5.1. *Let ψ denote a convex decreasing function $(0,1] \rightarrow [0,\infty]$ such that $\psi(1) = 0$. Define the inverse (or quasi-inverse if $\psi(0) < \infty$) as*

$$\psi^{\leftarrow}(t) = \begin{cases} \psi^{-1}(t) & \text{for } 0 \leq t \leq \psi(0) \\ 0 & \text{for } \psi(0) < t < \infty. \end{cases}$$

If $d > 2$, assume more generally that $\psi^{\leftarrow}$ is d-completely monotonic (recall that ϕ is d-completely monotonic if it continuous and has derivatives which alternate in sign, i.e. for all $k = 0, 1, ..., d$, $(-1)^k d^k\phi(t)/dt^k \geq 0$ for all t). Function

$$C(u_1, ..., u_n) = \psi^{\leftarrow}(\psi(u_1) + ... + \psi(u_d)), u_1, ..., u_n \in [0,1],$$

is a copula, called an Archimedean copula, with generator ψ.

Let Ψ_d denote the set of Archimedean generators in dimension d (properties and characterizations of Archimedean in dimension d will be studied in Chapter 5). Note that ψ and $c \cdot \psi$ (where c a positive constant) yield the same copula, and conversely, two Archimedean copulae are equal if their generators are equal up to a multiplicative constant. If $\psi(t) \rightarrow \infty$ when $t \rightarrow 0$, the generator will be said to be strict.

Example 1.5.2. *The independent copula $C^{\perp}$ is an Archimedean copula, with generator $\psi(t) = -\log t$. The upper Fréchet-Hoeffding copula is not Archimedean (but can be obtained as the limit of some Archimedean copulae).*

Example 1.5.3. *From one Archimedean copula with generator ψ, it is easy to generate several other Archimedean copula. For instance, if $h : [0,1] \rightarrow [0,1]$ a concave distortion function, then $\psi \circ h$ is also an Archimedean copula (see also Section 1.3.2). Furthermore, Ψ_d is a stable family by linear combination: for all $\alpha, \beta \geq 0$, if ψ is an Archimedean generator, so is $\alpha\psi + \beta\psi$ (but note that a linear combination of Archimedean copulae is not Archimedean anymore). And finally, as mentioned in Genest, Ghoudi and Rivest (1995), Ψ_d is a stable family by scaling: if $0 < \kappa < 1$, then $\psi_\kappa(\cdot) = \psi(\kappa\cdot) - \psi(\kappa)$ also generates an Archimedean copula. Chapter 2 of this thesis will give an interpretation of this generator, in terms of conditioning: if (U, V) is a random vector with an Archimedean copula C generated by ϕ, the copula of the truncated vector (U, V) given $U \leq u$ and $V \leq v$, where $(u, v) \in [0,1]^2$ satisfies $C(u, v) = \kappa$, is the Archimedean copula generated by ψ_κ. Table 1.1 gives some generators of Archimedean copulae.*

ψ		$C(u,v)$	Reference
$\log(1/t)$		uv	Independence
$(-\log t)^\theta$	$\theta \geq 1$	$\exp\left(-[[-\log(u)]^\theta + [-\log(v)]^\theta]^{1/\theta}\right)$	Gumbel (1960), Section 1.3.8.
$\left(t^{-\theta} - 1\right)$	$\theta \geq 0$	$\left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$	Kimeldorf and Sampson (1975) Clayton (1978), Section 1.3.6
$-\log \dfrac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$\theta \in \mathbb{R}$	$\dfrac{-1}{\theta} \log\left(1 - \dfrac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	Frank (1979), Chapter 7

Table 1.1: Some parametric families of Archimedean copulae.

Remark 1.5.4. *One of the earliest result on Archimedean copulae is Ling's theorem, stating that, in dimension 2, Archimedean copulae are the only copulae satisfying $C(u,u) < u$ for all $u \in (0,1)$, and the associativity functional equation*

$$C(C(u,v),w) = C(u,C(v,w)) \text{ for all } u,v,w \in [0,1].$$

Note that this relationship allows to define some serial iterates (see Schweizer and Sklar (1983)), in the sense that

$$C_2(u_1,u_2) = \psi^{\leftarrow}(\psi(u_1)+\psi(u_2)) \text{ and } C_k(u_1,...,u_{k-1},u_k) = C_2(C_{k-1}(u_1,...,u_{k-1}),u_k),$$

for all $k \geq 3$.

Under some assumptions, the following proposition stated that the limit of Archimedean copulae is still Archimedean. More precisely, Genest and MacKay (1986a) proved that if C_n is a sequence of absolutely continuous Archimedean copulae, with twice differentiable generators ψ_n, the limit of C_n when $n \to \infty$ is Archimedean if and only if there exists $\psi \in \Psi$ such that, for all $s,t \in [0,1]$,

$$\lim_{n\to\infty} \frac{\psi_n(s)}{\psi_n'(t)} = \frac{\psi(s)}{\psi'(t)}. \tag{1.16}$$

In the case where the limit in Equation (1.16) is null, the limiting copula is the upper Fréchet-Hoeffding copula. As we shall see in Chapter 3, weaker necessary and conditions on the ψ_n's can be obtained to characterize the limiting copula.

Further, it is also possible to compare Archimedean for the pointwise order. Let C_1 and C_2 be two Archimedean copulae, with respective generator ψ_1 and ψ_2. Then $C_1 \preceq C_2$ for the pointwise order if and only if $\phi_1 \circ \phi_2^{\leftarrow}$ is subadditive, i.e. $\psi_1 \circ \psi_2^{\leftarrow}(x+y) \leq \psi_1 \circ \psi_2^{\leftarrow}(x) + \psi_1 \circ \psi_2^{\leftarrow}(y)$ for all $x,y \in [0,1]$. Note that the following conditions are sufficient,

- if $\psi_1 \circ \psi_2^{\leftarrow}$ is concave, $C_1 \preceq C_2$,
- if ψ_1/ψ_2 is increasing, $C_1 \preceq C_2$,
- if ψ_1 and ψ_2 both differentiable on $(0,1)$, and if ψ_1'/ψ_2' is increasing, $C_1 \preceq C_2$.

Note further that if Ling's theorem (see Ling (1965)) was historically the first way to introduce Archimedean copulae, other characterizations can also be used:

- the **frailty approach** and the use of the Laplace transform of some latent factor. The underlying idea is that random variables $X_1,...,X_d$ are conditionally independent, given a latent factor Θ, e.g. $X_i|\Theta \sim \mathcal{E}(\lambda_i\Theta)$, $i = 1,...,d$. The expression of the Archimedean generator is then related with the Laplace transform of the latent factor (see e.g. Clayton (1978), Oakes (1989), Zheng and Klein (1995), or Bandeen-Roche and Liang (1996)). We will detail this approach in Section 1.6.

- the **survival function approach**: assume that for a random vector $\boldsymbol{X} = (X_1, ..., X_d)$ there is there is a convex (univariate) survival function $\overline{F}$ such that $\overline{F}(0) = 1$ and

$$\mathbb{P}(X_1 > x_1, ..., X_n > x_d) = \overline{F}(x_1 + ... + x_d).$$

  Then, the joint survival copula of $\boldsymbol{X} = (X_1, ..., X_d)$ is given by

$$\boldsymbol{u} = (u_1, ..., u_d) \mapsto \overline{F}(\overline{F}^{\leftarrow}(u_1) + ... + \overline{F}^{\leftarrow}(u_d)),$$

  which defines an Archimedean copula with generator $\psi = \overline{F}^{\leftarrow}$,

- **Abel criterion**, see Genest and MacKay (1986a), when $d = 2$: C is Archimedean if and only if there exists a mapping $f : (0, 1) \to (0, \infty)$ such that

$$\frac{\partial C(u, v)/\partial u}{\partial C(u, v)/\partial v} = \frac{f(u)}{f(v)},$$

  for all $u, v \in [0, 1]$. In such a case, $\phi(t) = \int_t^1 f(s)ds.$

Example 1.5.5. *Archimedean copulae are interesting when modeling exchangeable binary variables, since for every infinite exchangeable sequence $X_1, ..., X_n, ...$ of Bernoulli variables, there is a generator such that the associated n-dimensional Archimedean copula is a copula of $(X_1, ..., X_n)$ (see Müller and Scarsini (2004)).*

If ψ is a generator, set $F_\psi(x) = 1 - \psi^{\leftarrow}(x)$, for all $x \geq 0$. As proved in Genest and Rivest (1993), $F_\psi(\cdot)$ is a cumulative distribution function, of a unimodal distribution on $\mathbb{R}^+$, with mode 0. E.g. for independence, $\psi(t) = -\log t$, and therefore, $F_\psi(\cdot)$ is the distribution function of the standard exponential distribution.

Analogously, if ψ is a generator, set $Q_\psi(x) = 1 - \log[\psi(x)]$, for all $x \in (0, 1)$. As proved in Vandenhende and Lambert (2004), $Q_\psi(\cdot)$ is quantile function (the inverse of some cumulative distribution function). With strict generators, it is the quantile function of an unbounded distribution. E.g. is the associated distribution is logistic ($\mathbb{P}(Z \leq z) = [1 + \exp(-z/\theta)]^{-1}$), the associated copula is Gumbel copula (see Section 1.3.8).

Exchangeability (also called *interchangeability* or *indistinguishability*) is a common way to introduce dependence when modeling a large number of risks. It simply means that a permutation of the components in a portfolio should not affect the risk of the portfolio. It can be the case for motor insurance, where all policies are identical, or in credit risk for some first-to-default contract. This notion is closely related to the notion of *portfolio homogeneity.*

A first notion of *symmetry* can be the following: $\boldsymbol{X}$ is said to be symmetric if

$$H\boldsymbol{X} \stackrel{\mathcal{L}}{=} \boldsymbol{X} \text{ for all } H \in \mathcal{P}(d),$$

where $\mathcal{P}(d)$ denotes the set of $d \times d$ permutation matrices.

Other notions of symmetry can be considered, for instance, if the distribution of $\boldsymbol{X}$ is invariant by rotations, or by symmetry. Due to this property, in dimension 2, it can be seen that such distribution have necessarily circular isodensity curves. More formally, $\boldsymbol{X}$ is said to have a *spherical* distribution if

$$H\boldsymbol{X} \stackrel{\mathcal{L}}{=} \boldsymbol{X} \text{ for all } H \in \mathcal{O}(d),$$

where $\mathcal{O}(d)$ denotes the set of $d \times d$ orthogonal matrices (i.e. $H^t H = \mathbb{I}$). Note that the $\mathcal{N}(\boldsymbol{0}, \mathbb{I})$ distribution is spherical. For spherical distributions, the following statements are equivalent (see Fang, Kotz and Ng (1990) for instance),

1. $H\boldsymbol{X} \overset{\mathcal{L}}{=} \boldsymbol{X}$ for all $H \in \mathcal{O}(d)$;

2. the characteristic function of $\mathbf{X}$ is of the form $\boldsymbol{x} \mapsto \phi\left(\boldsymbol{x}^t\boldsymbol{x}\right)$ for some $\phi \in \Phi_d$, where

$$\Phi_d = \left\{\phi : \boldsymbol{x} \longmapsto \phi\left(\boldsymbol{x}^t\boldsymbol{x}\right) \text{ is a } d\text{-dimensional characteristic function}\right\};$$

3. $\boldsymbol{X}$ has the representation $\boldsymbol{X} \overset{\mathcal{L}}{=} R\boldsymbol{U}$ for some positive random variable R, independent of U uniformly distributed on the unit sphere in $\mathbb{R}^d$, for the Euclidean norm (R denotes the norm of X, while $\boldsymbol{U}$ is the angle);

4. for any $\boldsymbol{a} \in \mathbb{R}^d$, $\boldsymbol{a}^t\boldsymbol{X} \overset{\mathcal{L}}{=} \|a\| X_i$ for all $i = 1, ..., d$,

Following the construction of the $\mathcal{N}(\mu, \Sigma)$ distribution from the standard normal distribution $\mathcal{N}(0, \mathbb{I})$ some extensions can be considered. Consider $\boldsymbol{X} \overset{\mathcal{L}}{=} \mu + A\boldsymbol{Y}$ where $\boldsymbol{Y}$ is spherically distributed, where $A^tA = \Sigma$. $\boldsymbol{X}$ will then be said to be spherically distributed with parameters μ and $\Sigma = A^tA$. As earlier, some characterizations of elliptical distributions can be used: $\boldsymbol{X}$ has the representation $\boldsymbol{X} \overset{\mathcal{L}}{=} \mu + RA^t\boldsymbol{U}$ for some positive random variable R, such that R^2 has a $\chi^2(n)$ distribution, independent of $\boldsymbol{U}$ uniformly distributed on the unit sphere in $\mathbb{R}^d$, and A satisfies $A^tA = \Sigma$.

The t-distribution, with parameters m, μ and Σ, is obtained by considering $\mu + A^t\sqrt{m}\boldsymbol{Z}/S$, where $\boldsymbol{Z} \sim \mathcal{N}(0, \mathbb{I})$ and $S \sim \chi^2(m)$ are independent. The density of the t-distribution with parameters m, μ and Σ, in $\mathbb{R}^d$ is

$$\boldsymbol{x} \mapsto \frac{\Gamma((n+m)/2)}{(\pi m)^{n/2}\,\Gamma(m/2)} |\Sigma|^{-1/2} \left(1 + \frac{1}{m}(\boldsymbol{x} - \mu)^t \Sigma^{-1} (\boldsymbol{x} - \mu)\right)^{-(n+m)/2}.$$

When $m = 1$, the distribution is called the multivariate Cauchy distribution.

Elliptical symmetric distributions have been quite popular since they could define a flexible class. Nevertheless, as pointed out clearly in Cook and Johnson (1981), they might fail in modeling data.

Exchangeability is the mathematical notion for interchangeability: it emphasizes the homogeneous aspects of exchangeability. Bäuerle and Müller (1987) called this notion *indistinguishable risks*. One of the most important application is decomposing heterogeneous populations into homogeneous subclasses. Further, heterogeneity in insurance portfolio is usually modeled through mixture models. As seen in this section, de Finetti's theorem allows to express exchangeable risks using mixture models.

Definition 1.5.6. *A finite sequence $\{X_1, ..., X_d\}$ of random variables is exchangeable, or d-exchangeable, if*

$$(X_1, ..., X_d) \overset{\mathcal{L}}{=} \left(X_{\sigma(1)}, ..., X_{\sigma(d)}\right), \tag{1.17}$$

for any permutation σ of $\{1, ..., d\}$. More generally, an infinite sequence $\{X_1, X_2...\}$ of random variables is infinitely exchangeable (or simply exchangeable) if

$$(X_1, X_2, ...) \overset{\mathcal{L}}{=} \left(X_{\sigma(1)}, X_{\sigma(2)}, ...\right), \tag{1.18}$$

for any finite permutation σ of $\mathbb{N}^*$ (that is $Card\{i, \sigma(i) \neq i\} < \infty$).

Definition 1.5.7. *A d-exchangeable sequence $\{X_1, ..., X_d\}$ is called m-extendible (for some $m > d$), if $(X_1, ..., X_d) \overset{\mathcal{L}}{=} (Z_1, ..., Z_d)$, where $\{Z_1, ..., Z_m\}$ is some m-exchangeable sequence.*

Example 1.5.8. *Let $X_1, ..., X_d$ be random variables such that*

$$Var(X_i) = \sigma^2 \text{ and } cov(X_i, X_j) = \rho\sigma^2,$$

for $i = 1, ...d$, and $j \neq i$. Then, the following inequality holds

$$0 \leq Var\left(\sum_{i=1}^{d} X_i\right) = \sum_{i=1}^{d} Var(X_i) + \sum_{i \neq j} cov(X_i, X_j) = d\sigma^2 + d(d-1)\rho\sigma^2,$$

and therefore,

$$\rho \geq -\frac{1}{d-1}.$$

Thus, infinite exchangeability implies positive correlation (exchangeability is a strong notion of positive dependence).

More formally, on some probabilistic space $(\Omega, \mathcal{A}, \mathbb{P})$ (see e.g. Pollard (2002)), a probability measure $\mathbb{P}$ on $\mathcal{A}^{\mathbb{N}}$ of the product space $\mathbb{R}^{\mathbb{N}}$ (the state of all sequences $X_1, ..., X_n, ...$ of real-valued random variables) is said to be exchangeable if it is invariant under any finite permutation σ, or equivalently, the random vector $(X_1, ..., X_n)$ has the same distribution as $\left(X_{\sigma(1)}, ..., X_{\sigma(n)}\right)$ for every n-permutation σ, and any n. The distribution of all sets in $\mathcal{A}^{\mathbb{N}}$ whose indicator functions are n-symmetric forms a sub-σ-field of $\mathcal{A}^{\mathbb{N}}$, denoted $\mathcal{F}_n$. Note that the $(\mathcal{F}_n, n \in \mathbb{N})$ is a decreasing filtration on $\mathbb{R}^{\mathbb{N}}$.

The so-called de Finetti's theorem (see de Finetti (1937)) asserts that all exchangeable distribution can be built up from mixtures of product measures, e.g. based on conditional independence: for any sets A_i in $\mathcal{A}$,

$$\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n | \mathcal{F}_\infty) = \mathbb{P}(X_1 \in A_1 | \mathcal{F}_\infty) \times ... \times \mathbb{P}(X_n \in A_n | \mathcal{F}_\infty).$$

Using the formulation of Aldous (1985), "*an infinite exchangeable sequence is a mixture of i.i.d. sequences*". In the context of Bernoulli variables, the conditioning σ-algebra $\mathcal{F}_\infty$ can be represented through a random variable. Hence, an infinite sequence $X_1, X_2, ...$ of Bernoulli random variables is exchangeable if and only there is a random variable Θ, taking values in $[0, 1]$ such that, given $\Theta = \theta$ the X_i's are independent, and $X_i \sim \mathcal{B}(\theta)$ (see Schervish (1995) or Chow and Teicher (1997)).

Example 1.5.9. *This result can be easily interpreted in credit risk, where variables of interest are dichotomous (default or non-default). Let $X_1, X_2, ...$ be an infinite exchangeable sequence of Bernoulli variables, and let $S_n = X_1 + ... + X_n$ the number of defaults within n companies (for a given period of time). Then, the distribution of S_n is a mixture of binomial distributions, i.e. there is a distribution function H on $[0, 1]$ such that*

$$\mathbb{P}(S_n = k) = \int_0^1 \binom{n}{k} \omega^k (1-\omega)^{n-k} dH(\theta).$$

In a more general context, the σ-algebra $\mathcal{F}_\infty$ cannot be generated by a random variable, and thus, an infinite sequence $X_1, X_2, ...$ of random variables is exchangeable if and only if there exists a random probability measure, such that, given the random probability measure, the X_i's are independent.

Nevertheless, if there exists a random variable Θ such that, given Θ variables X_i's are independent, the infinite sequence $\{X_1, X_2, ..., X_n, ...\}$ is exchangeable. But the converse is not necessarily true (see Schervish (1995)).

1.6 Clayton's dependence structure

For the purpose of the presentation, we shall focus here on the bivariate case, extending afterwards the approach in higher dimension.

1.6.1 Approach through the odds ratio

Following Clayton (1978) or Oakes (1989), define the following associated function, based on the copula function

$$T(x,y) = \frac{C(x,y)\,C_{12}(x,y)}{C_1(x,y)\,C_2(x,y)}, \text{ for all } x,y \in [0,1], \tag{1.19}$$

where $C_1(x,y) = \partial C(x,y)/\partial x = \partial_x C(x,y)$, $C_2(x,y) = \partial C(x,y)/\partial y = \partial_y C(x,y)$ and $C_{12}(x,y) = \partial^2 C(x,y)/\partial x \partial y = \partial_{xy} C(x,y)$. Clayton (1978) defined the above measure while seeking an index to express the influence of parental history of a given disease upon the incidence in the offspring. Consider the pairs of related individuals, say sons and fathers: denote by X the age at death of the father and Y the age at death of son. It is assumed that the association arises because the two individuals share some common influence (common environmental or genetic influence). Thus, the parameter $T(x,y)$ measures the degree of association between X and Y, independence being implied by $T(x,y) = 1$ (and the converse is also true), while positive independence by $T(x,y) > 1$ (see Clayton (1978)). More precisely, $T(x,y) > 1$ if and only if (X,Y) is left corner set increasing (LCSI, i.e. the distribution function is TP2, see Section 1.5). Further, if the joint distribution tends to the associated upper Fréchet-Hoeffding bound, T approaches infinity.

Remark 1.6.1. *Drouet-Mari and Kotz (2001) pointed out that function T defined as in Equation* (1.19) *can be seen as a ratio of conditional hazard functions: consider two survival times X and Y, and define the hazard of (X,Y),*

$$h(x,y) = -\frac{\partial_{xy}\mathbb{P}(X > x, Y > y)}{\mathbb{P}(X > x, Y > y)},$$

the hazard of Y given X surviving beyond x,

$$h_Y(x,y) = -\frac{\partial_{y}\mathbb{P}(X > x, Y > y)}{\mathbb{P}(X > x, Y > y)},$$

and the hazard of Y given X failed at x,

$$h_{Y|X}(x,y) = -\frac{\partial_{xy}\mathbb{P}(X > x, Y > y)}{\partial_x\mathbb{P}(X > x, Y > y)},$$

then

$$T(x,y) = \frac{h_{Y|X}(x,y)}{h_Y(x,y)} = \frac{h_{X|Y}(x,y)}{h_X(x,y)}.$$

Equivalently, as in Clayton (1978) or Oakes (1989), this ratio might be written

$$T(x,y) = \frac{h(x,y)}{h_X(x,y)h_Y(x,y)}.$$

This function was defined in Clayton (1978) as an odds-ratio, which is symmetric, marginal-free, such that T, h_X and h_Y determine completely the joint distribution.

1.6.2 The so-called *Pareto* copulae

Assume that X and Y are exponentially distributed, and independent, with parameter Θ so that $\mathbb{P}(X > x|\Theta = \theta) = \exp(-\theta y)$ and $\mathbb{P}(Y > y|\Theta = \theta) = \exp(-\theta y)$. Assume that Θ is gamma-distributed with parameters β and γ, then the non-conditional joint distribution is

$$\mathbb{P}(X > x, Y > y) = \int_0^\infty \exp(-\theta[x+y]) \frac{\theta^{\gamma-1}\exp(-\theta/\beta)}{\beta^\gamma \Gamma(\gamma)} d\theta = (1 + \beta x + \beta y)^{-\gamma}. \tag{1.20}$$

Observe that both variables are Pareto distributed, i.e. $\mathbb{P}(X > x) = (1-\beta x)^{-\gamma}$ and $\mathbb{P}(Y > y) = (1-\beta y)^{-\gamma}$. The survival copula of (X, Y) is then

$$C^*(u, v) = (u^{-1/\gamma} + v^{-1/\gamma} - 1)^{-\gamma},$$

with $\gamma > 0$, called Clayton copula.

1.6.3 The frailty approach

Frailty models have been introduced at the end on the 70's, and have been popularized by Oakes (1989). The idea is to introduce dependence between survival times using an unobserved random variable Θ, called the "*frailty*". Assume that X and Y are two lifetimes, independent, conditionally to some exogenous factor Θ. Assume further that conditional marginal survival distribution satisfy

$$\overline{F}_{X|\Theta}(x|\theta) = \mathbb{P}(X > x|\Theta = \theta) = \overline{G}_X(x)^\theta,$$

for some distribution function G_X, for all θ, and similarly for Y given $\Theta = \theta$. Equivalently, the so-called *frailty variable* Θ acts multiplicatively on the marginal hazard functions. Thus

$$\mathbb{P}(X > x, Y > y) = \mathbb{E}(\mathbb{P}(X > x, Y > y|\Theta)) = \mathbb{E}(\mathbb{P}(X > x|\Theta) \cdot \mathbb{P}(Y > y|\Theta)),$$

i.e. using conditional expression of survival distributions,

$$\mathbb{P}(X > x, Y > y) = \mathbb{E}(\exp[-\Theta(-\log \mathbb{P}(X > x))] \cdot \exp[-\Theta(-\log \mathbb{P}(Y > y))]),$$

and finally, if ψ denotes the Laplace transform of Θ, i.e. $\phi(t) = \mathbb{E}(\exp(-t\Theta))$,

$$\mathbb{P}(X > x, Y > y) = \phi(-\log \mathbb{P}(X > x) - \log \mathbb{P}(Y > y)).$$

Since marginal distributions can be written $\mathbb{P}(X > x) = \phi(-\log \overline{G}_X(x))$ and similarly for Y, the survival copula of (X, Y) is then

$$C^*(u, v) = \phi(\phi^{\leftarrow}(u) + \phi^{\leftarrow}(v)),$$

which is the Archimedean copula with generator $\psi = \phi^{\leftarrow}$. This representation of frailty models has been introduced in Marshall and Olkin (1988). In the particular case where Θ is Gamma distributed (with Laplace transform $\psi(t) = (1+t)^{1/\alpha}$), C^* is Clayton copula.

1.6.4 Properties of Clayton copulae

Definition 1.6.2. *Given $\theta \geq 0$, Clayton copula with parameter θ is defined on* $[0, 1] \times [0, 1]$ *as*

$$C_\theta(u, v) = (u^{-1/\theta} + v^{-1/\theta} - 1)^{-\theta}.$$

Notice that when $\theta \to 0$ and $\theta \to \infty$, C is respectively the independent copula, and the upper-Fréchet-Hoeffding bound. Further, if $0 \leq \theta_1 \leq \theta_2$, observe that $C_{\theta_1}(u, v) \leq C_{\theta_2}(u, v)$ for all $u, v \in [0, 1]$. This copula can also be used for simulations since partial derivative can be obtained easily,

$$C_1(u, v) = \frac{\partial C(u, v)}{\partial u} = (1 + u^\theta(v^{-\theta} - 1))^{-1-1/\theta},$$

which can be inverted easily (see Section 1.9).

1.6.5 Clayton copulae in dimension $d \geq 2$

Since Clayton copulae are Archimedean, they can easily be extended in higher dimension.

Definition 1.6.3. *Given $\theta \geq 0$, Clayton copula with parameter θ is defined on $[0,1]^d$ as*

$$C(u_1, ..., u_d) = (u_1^{-1/\theta} + ... + u_d^{-1/\theta} - (d-1))^{-\theta}. \tag{1.21}$$

This copula can be obtained as follows (the construction of this distribution gives a nice way to simulate such a copula): let

$$U_i = \left(1 + \frac{Y_i}{Z}\right)^{-\theta}, i = 1, ..., d,$$

where the Y_i's are independent exponential $\mathcal{E}(1)$ variables, independent of Z having Gamma distribution $\mathcal{G}(\theta, 1)$. Then the joint distribution of $(1 - U_1, ..., 1 - U_d)$ is given by Equation (1.21).

1.7 Marshall and Olkin's dependence structure

1.7.1 The common shock model

The class of copulae called Marshall and Olkin copulae has been derived from Marshall and Olkin's distribution (see Marshall and Olkin (1967), or Muliere and Scarsini (1987)).

Let X and Y denote two lifetimes of two components (x and y). Assume that the shocks follow three independent Poisson processes: consider the time of occurrence of the shocks, Z_x (with parameter λ_x) which affects component x, Z_y (with parameter λ_y) which affects component y, and Z_{xy} (with parameter λ_{xy}) which affects both components. The times of occurrence of these shocks are assumed to be independent, and exponentially distributed. Hence, since a shock is fatal to one or both components, the survival function of the pair $(X, Y) = (\min\{Z_x, Z_{xy}\}, \min\{Z_y, Z_{xy}\})$ is then

$$\overline{F}(x, y) = \mathbb{P}(X > x, Y > y) = \mathbb{P}(Z_x > x) \cdot \mathbb{P}(Z_y > y) \cdot \mathbb{P}(Z_{xy} > \min\{x, y\}).$$

Since those distributions are exponential, rewrite

$$\mathbb{P}(X > x, Y > y) = \exp\left(-\lambda_x x - \lambda_y y - \lambda_{xy} \max\{x, y\}\right), x, y > 0.$$

Note that marginally, X and Y are exponentially distributed, respectively with parameters $\lambda_x + \lambda_{xy}$ and $\lambda_y + \lambda_{xy}$. Furthermore, note that $\overline{F}$ satisfies the *weak* lack of memory property, i.e.

$$\overline{F}(x + t, y + t) = \overline{F}(x, y)\overline{F}(t, t), \text{ for all } x, t, y \geq 0.$$

Remark 1.7.1. *The strong lack of memory property (see Marshall and Olkin (1967)) is*

$$\overline{F}(x + s, y + t) = \overline{F}(x, y)\overline{F}(s, t), \textit{ for all } x, y, s, t \geq 0.$$

Marshall and Olkin (1967) (see also Aczél (1966)) proved that only vectors with independent margins can satisfy such a property. Notice that other functions satisfying the weak lack of property memory can be found in Ghurye and Marshall (1984). An extension of the univariate property of lack of memory to dimension 2 is

$$\overline{F}((1 + \alpha)x, (1 + \alpha)y) = \overline{F}(x, y)\overline{F}(\alpha x, \alpha y), \alpha > 0.$$

This functional equation was considered by Pickands (see Pickands (1976)). As shown in Marshall and Olkin (1991), those distributions are somehow related to functional equations obtained when focusing on multivariate extremes (see Chapter 6 of this thesis). For instance, observe that (X,Y) has a distribution with exponential scaled minima if and only if

$$\left[\overline{F}\left(\frac{tx}{k},\frac{ty}{k}\right)\right]^k=\overline{F}(tx,ty),$$

for all $x,y,t>0$ and $k\in\mathbb{N}^*$.

Set

$$\alpha=\frac{\lambda_{xy}}{\lambda_x+\lambda_{xy}}\text{ and }\beta=\frac{\lambda_{xy}}{\lambda_y+\lambda_{xy}},$$

then, the survival copula of (X,Y) is given by

$$C^*(u,v)=uv\min\{u^{-\alpha},v^{-\beta}\}=\min\{u^{1-\alpha}v,uv^{1-\beta}\}.$$

1.7.2 Copulae induced by the model

Definition 1.7.2. *Given α and β in $(0,1)$, the associated Marshall and Olkin copula is defined as*

$$C(u,v)=\min\{u^{1-\alpha}v,uv^{1-\beta}\}.$$

This copula is also sometimes called the Cuadras-Augé copula since the case $\alpha=\beta$ first appeared in Cuadras and Augé (1981).

Remark 1.7.3. *Observe that these copulae are obtained from mixture distributions, with an absolutely continuous and a singular component, the mass of the singular component being concentrated on the curve $\mathcal{C}=\{(x,y)\in[0,1]\times[0,1],x^\alpha=y^\beta\}$. Hence, if (U,V) has copula C*

$$\mathbb{P}(U=V)=\mathbb{P}(\textit{the first shock affect both components})=\frac{\lambda_{xy}}{\lambda_x+\lambda_y+\lambda_{xy}}>0.$$

1.7.3 Application of Marshall and Olkin's framework

This model is claimed to be widely used in reliability (see Harris (1978)), and in *competing risks* context (see David and Moeschberger (1978)). In actuarial science, this model appeared in Frees, Carriere and Valdez (1996) and Bowers et al. (1997). As pointed out, *common shock* models have the primary advantages to be easy to interpret, and computationally convenient. In multilife insurance contracts (e.g. an husband and his wife), Marshall and Olkin's model can be used to take into account a common *shock* that is common to both lives. Consider T_x and T_y, the time-until-death random variables, of a man of age x and of his wife of age y respectively. Denote, using standard actuarial notations ${}_kp_x$ and ${}_kp_y$ the marginal survival probabilities ${}_kp_x=1-\mathbb{P}(T_x\leq k)$ and ${}_kp_y=1-\mathbb{P}(T_y\leq k)$. Analogously, let ${}_kp_{\overline{xy}}$ denote the conditional probability that at least one life survives an additional k years, i.e.

$${}_kp_{\overline{xy}}=1-\mathbb{P}(T_x\leq k,T_y\leq k).$$

In Marshall and Olkin's model, if the shock variable is exponentially distributed, with parameter λ_{xy}, then straightforward calculations (see Frees, Carriere and Valdez (1996)) show that

$${}_kp_{\overline{xy}}={}_kp_x+{}_kp_y-\exp(-\lambda_{xy}k){}_kp_x\cdot{}_kp_y.$$

Recall that last-survivor annuities are defined as

$$a_{\overline{xy}} = \sum_{k=1}^{\infty} v^k \mathbb{P}(T_x > k \text{ or } T_y > k) = \sum_{k=1}^{\infty} v^k \ {}_k p_{\overline{xy}}.$$

Defining ${}_k p_x^* = \exp(\lambda_{xy} k)_k p_x$ and ${}_k p_y^* = \exp(\lambda_{xy} k)_k p_y$, the annuity can be derived as

$$a_{\overline{xy}} = \sum_{k=1}^{\infty} e^{-(\delta+\lambda_{xy})\cdot k} \left({}_k p_x^* +_k p_y^* -_k p_x^* \cdot_k p_y^*\right), \text{ with } \delta = \log(v)$$

which is the annuity calculated assuming independence between ${}_k p_x^*$ and ${}_k p_y^*$. The annuity is calculated at force of interest $\delta + \lambda_x y$. From this expression, notice that the last-survivor annuity is a decreasing function of λ_{xy}: the greater the dependency, the smaller the annuity.

1.7.4 On the generalization of Marshall and Olkin's approach

Recall that Marshall and Olkin's distribution are characterized by the exponential marginal distribution and the weak lack of memory property,

$$\overline{F}(x+t, y+t) = \overline{F}(x, y) \cdot \overline{F}(t, t), \text{ for all } x, y, t > 0, \tag{1.22}$$

Muliere and Scarsini (1987) considered the following extension,

$$\overline{F}(x \star t, y \star t) = \overline{F}(x, y) \cdot \overline{F}(t, t), \text{ for all } x, y, t > 0, \tag{1.23}$$

for some binary commutative and associative operation $\star$ (i.e. $(x \star y) \star z = x \star (y \star z)$). As shown in Aczél (1966) and Schweizer and Sklar (1983), the continuous operators satisfying the association property, and the reducible condition $\{x \star y = x \star z$ or $y \star x = z \star x$ implies $y = z\}$ can be written

$$x \star y = \psi^{\leftarrow}(\psi(x) + \psi(y)),$$

where ψ is a continuous strictly monotone function. Assume further that $\star$ admits an identity element e (i.e. $x \star e = x$). Then (see Muliere and Scarsini (1987)) the continuous solution of the univariate version of the lack of memory property $\overline{F}(x \star t) = \overline{F}(x)\overline{F}(t)$ is

$$\overline{F}(x) = \exp(-\lambda\psi(x)), \lambda > 0 \text{ and } \exp(1) = \psi^{\leftarrow}(0) < t.$$

1.8 Gumbel's dependence structure

Consider here the so-called Gumbel dependence structure, as introduced in Gumbel (1960, 1961). It might also be called the Gumbel-Hougaard family (see Hougaard (1968)) or the *bivariate logistic extreme value distribution.*

Remark 1.8.1. *The "logistic " term was explained as follows: since this copula is Archimedean, and its generator is the exponential of the quantile function of some logistic distribution. But actually, there exists also in the literature a bivariate logistic Gumbel distribution defined on $\mathbb{R}^2$ by the following distribution function*

$$F(x, y) = \frac{1}{1 + \exp(-x) + \exp(-y)}, \ \textit{for all } x, y \in \mathbb{R}.$$

The associated copula is $C(u, v) = \dfrac{uv}{u + v - uv}$.

1.8.1 The bivariate logistic distribution

Consider the bivariate logistic distribution function given by

$$F(x,y) = \exp\left(-(x^{-\theta}+y^{-\theta})^{1/\theta}\right), x,y>0,$$

where $\theta \geq 1$. This function can be written in a more convenient form,

$$F(x,y) = \exp\left(-\left([-\log(e^{-1/x})]^{\theta}+[-\log(e^{-1/y})]^{\theta}\right)^{1/\theta}\right), x,y>0,$$

The margins of a bivariate logistic distribution are standard Fréchet (with marginal distribution $F(x)=e^{-1/x}\mathbf{1}(x>0)$. The associated copula can be derived from this expression, leading to

$$C(u,v) = \exp\left(-\left([-\log u]^{\theta}+[-\log v]^{\theta}\right)^{1/\theta}\right), u,v\in[0,1],$$

where $\theta \geq 1$.

1.8.2 Multivariate failure distributions

Again, as in Section 1.6.3, consider survival times X and Y independent conditionally to some random factor Θ, having a positive stable distribution, with Laplace transform $\psi(t)=\exp(-t^{\theta})$, $\theta \geq 1$. Assume that $X|\Theta$ and $Y|\Theta$ are independent and exponentially distributed with parameter Θ. The joint distribution is then the bivariate logistic distribution. Therefore, Gumbel copula is Archimedean with generator $\psi(t)=(-\log t)^{\theta}$, $\theta>0$. Hence, such a copula can be extended in dimension higher than 2.

1.8.3 Properties of Gumbel copulae

Definition 1.8.2. *Given $\theta \geq 1$, Gumbel copula with parameter θ is defined on $[0,1]\times[0,1]$ as*

$$C(u,v) = \exp\left(-\left([-\log u]^{\theta}+[-\log v]^{\theta}\right)^{1/\theta}\right), u,v\in[0,1].$$

Observe that Gumbel copulae satisfy the following homogeneous property:

$$C^{z}(u,v)=C(u^{z},v^{z}),\text{ for all } u,v\in[0,1],\text{ and for all } z>0,$$

(see Chapter 6 for the implication in terms of extreme value of the property, and Joe (1997)). In dimension $d \geq 2$, note that those copulae are given by

$$C(u_1,...,u_d) = \exp\left(-\left([-\log u_1]^{\theta}+...+[-\log u_d]^{\theta}\right)^{1/\theta}\right), u_1,...,u_d\in[0,1].$$

1.9 On partial derivatives of copulae

As we shall see in this paragraph, there is a strong link between partial derivatives of the copula, and conditional distributions related to the copula. Let (X,Y) denote a random pair with

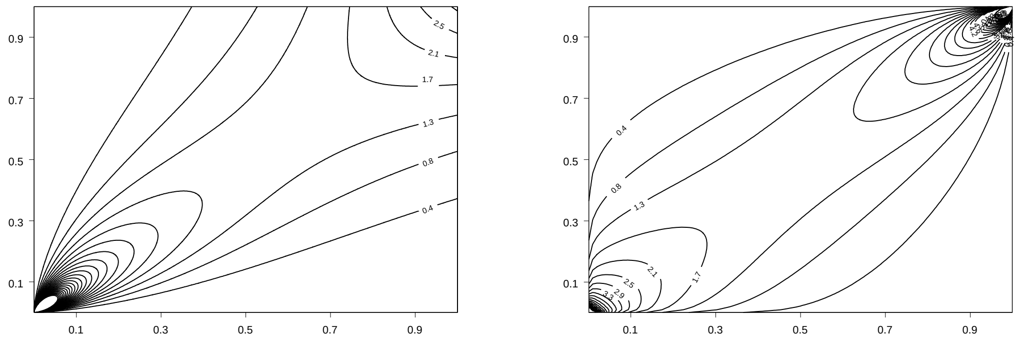


Figure 1.4: Clayton and Gumbel copulae density, level curves (uniform margins).

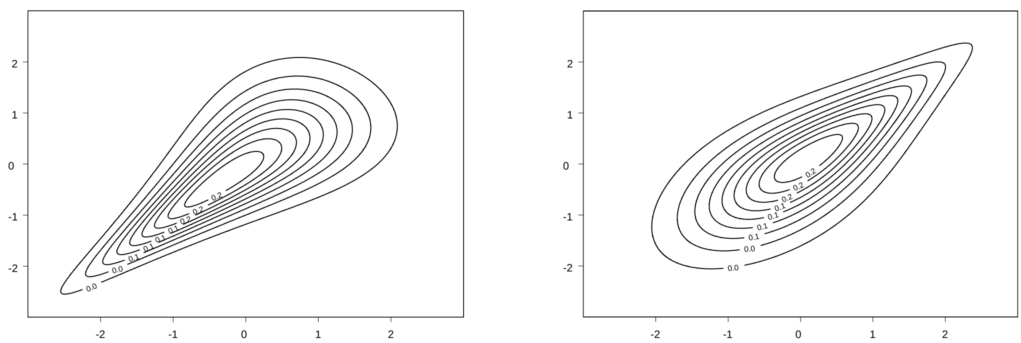


Figure 1.5: Clayton and Gumbel copulae density, level curves ($\mathcal{N}(0,1)$ margins).

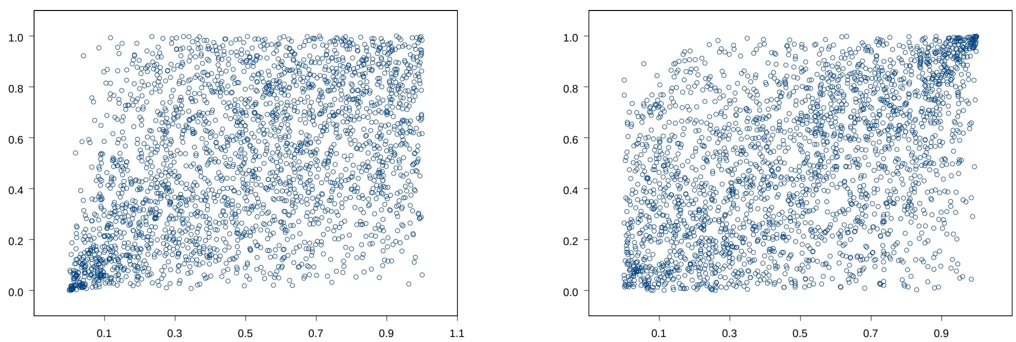


Figure 1.6: Some simulations of the Clayton, and Gumbel copulae.

continuous marginal distributions F_X and F_Y, joint distribution F_{XY} and copula C. Assume that $y \mapsto \mathbb{P}(X \leq x|Y = y)$ is continuous from the right, then

$$\begin{aligned}
\mathbb{P}(X \leq x|Y = y) &= \lim_{h \to 0} \mathbb{P}(X \leq x|y \leq Y < y + h) \\
&= \lim_{h \to 0} \frac{F_{XY}(x, y + h) - F_{XY}(x, y)}{F_Y(y + h) - F_Y(y)} \\
&= \lim_{h \to 0} \frac{C(F_X(x), F_Y(y + h)) - C(F_X(x), F_Y(y))}{F_Y(y + h) - F_Y(y)} \\
&= \lim_{h \to 0} \frac{C(F_X(x), F_Y(y) + \varphi(y, h)) - C(F_X(x), F_Y(y))}{\varphi(y, h)}
\end{aligned}$$

where $\varphi(y, h) = F_Y(y + h) - F_Y(y) \to 0$ for all y, as $h \to 0$, since F_Y is continuous. Hence,

$$\mathbb{P}(X \leq x|Y = y) = \frac{\partial C}{\partial v}(F_X(x), F_Y(y)). \tag{1.24}$$

From this relationship, partial derivatives of copulae also appear when copulae are used for temporal dependence. More precisely, consider some first order Markov process (X_t), in discrete time (the approach can be extended in continuous time as we will see briefly at the end of this paragraph), i.e. the conditional distribution of the response at any time t on the past history only depends on the last state of the process, i.e.

$$\mathbb{P}(X_t \leq x|X_{t-1} = x_1, X_{t-2} = x_2, ..., X_{t-k} = x_k) = \mathbb{P}(X_t \leq x|X_{t-1} = x_1).$$

Let $P(x, s; y, t)$ be a version of $\mathbb{P}(X_t \leq y|X_s \leq x)$ where $s < t$, satisfying $P(x, s; \cdot, t)$ is a distribution function on $\mathbb{R}$ (for real value Markov processes), and $P(\cdot, s; y, t)$ is-measurable. As a consequence of this property, the conditional probabilities satisfy the Chapman-Kolmogorov equations, which relate the state of the process at time t with an earlier time s through an intermediate time u,

$$P(x, s; y, t) = \int_{\mathbb{R}} P(z, u; y, t) P(x, s; dz, u), \tag{1.25}$$

$s < u < t$, $x, y \in \mathbb{R}$. Assume that (X_t) is a stationary process, and let C denote the copula of (X_t, X_{t-1}), for all $t \in \mathbb{Z}$. As shown by Darsow, Nguyen and Olsen (1992), the copula of (X_t, X_{t-2}) does not depend on t, and furthermore, using Equation (1.24), and Chapman-Kolmogorov equation, the copula is

$$(u, v) \mapsto \int_0^1 \frac{\partial C(u, t)}{\partial t} \frac{\partial C(t, v)}{\partial t} dt.$$

More generally, Darsow, Nguyen and Olsen (1992) introduced to product of copulae as follows: let C_1 and C_2 denote two copulae, then

$$C(u, v) = C_1 \star C_2(u, v) = \int_0^1 \frac{\partial C_1(u, t)}{\partial t} \frac{\partial C_2(t, v)}{\partial t} dt$$

is a copula. Note that this operation can be seen as the continuous analogous of the multiplication operator for transition matrices. Hence, some algebraic properties are preserved, e.g. $C^\perp \star C = C \star C^\perp = C^\perp$, while $C^+ \star C = C \star C^+ = C$. Consequently, $C^\perp$ is a null element, and C^+ is the identity. Moreover, $\star$ is associative, but not commutative. Using this operator, the following important result holds. Let (X_t) is a continuous stochastic process, and $C_{s,t}$ denotes the copula of (X_s, X_t). The following are equivalent,

- the conditional distributions satisfy the Chapman-Kolmogorov Equation (1.25) for all $s < u < t$, $x, y \in \mathbb{R}$

- for all $s < u < t$, $C_{s,t}(x,y) = C_{s,u} \star C_{u,t}(x,y)$, for all $s < u < t$, $x, y \in [0,1]$.

Therefore, (X_t) is a Markov process if and only if for all n and all $0 \leq t_1 < t_2 < ... < t_n$,

$$C_{t_1,t_2,...,t_n} = C_{t_1,t_2} \star C_{t_2,t_3} \star ... \star C_{t_{n-1},t_n}$$

Example 1.9.1. *An particular case is the one where copula C is idempotent i.e. $C = C \star C$. For instance the upper Fréchet-Hoeffding copula $C^+(u,v) = \min(u,v)$ and the independent copula $C^\perp(u,v) = uv$ are both idempotent. Indempotent families can also be defined, i.e. the product of copula within a family of copulae remains in the same family, e.g. Morgenstein family. C_θ is a Morgenstein copula, with parameter $\theta \in [-1,1]$, if $C_\theta(u,v) = uv + \theta uv(1-u)(1-v)$. Then*

$$\frac{\partial C_\theta(u,v)}{\partial v} = u + \theta u(1-u)(1-2v)$$

and the expression of copula $C_\theta \star C_\kappa$ is given by

$$\begin{aligned} C_\theta \star C_\kappa(u,v) &= \int_0^1 (u + \theta u[1-u][1-2t])(v + \kappa v[1-v][1-2t])\,dt \\ &= uv + \frac{\theta\kappa}{3} uv[1-u][1-v] = C_{\theta\kappa/3}(u,v). \end{aligned}$$

Example 1.9.2. *Figure 1.7 shows the case of Markov processes with $\mathcal{N}(0,1)$ margins, where C is a Gaussian copula on the left, and a Clayton copula on the right. Note that clusters of lower extremal values can be obtained much more frequently with Clayton copula.*

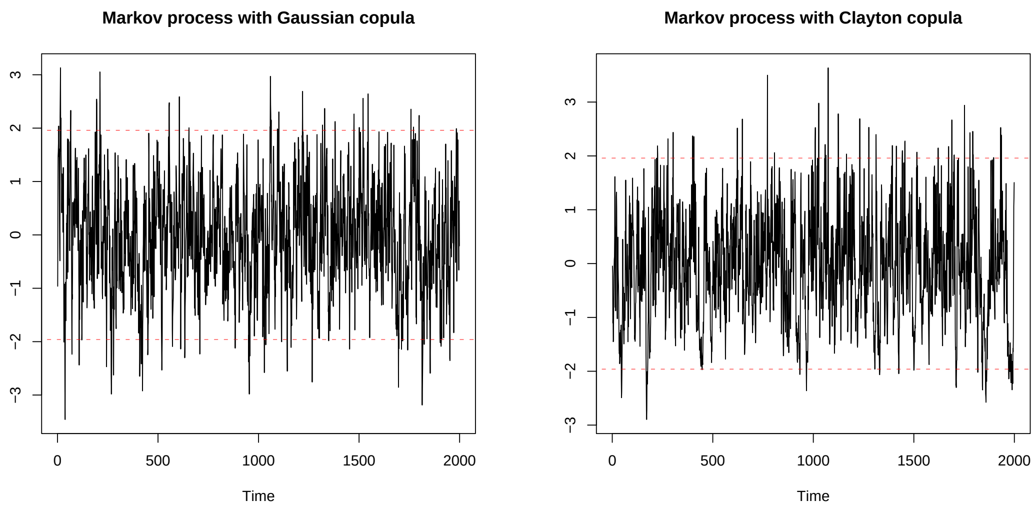


Figure 1.7: Some simulations of stationary Markov processes, with $\mathcal{N}(0,1)$ margins, where the copula of (X_t, X_{t-1}) is a Gaussian copula on the left, and a Clayton copula on the right.

Note that partial derivative are also interesting for simulation purpose. $\boldsymbol{U} = (U_1, .., U_n)$ with copula C could be simulated using the following algorithm,

- simulate U_1 uniformly on $[0,1]$,

$$u_1 \leftarrow \texttt{Random}_1,$$

- simulate U_2 from the conditional distribution $\partial_2 C(\cdot|u_1)$,

$$u_2 \leftarrow [\partial_2 C(\cdot|u_1)]^{-1}(\texttt{Random}_2),$$

- simulate U_k from the conditional distribution $\partial_k C(\cdot|u_1, ..., u_{k-1})$,

$$u_k \leftarrow [\partial_2 C(\cdot|u_1, ..., u_{k-1})]^{-1}(\texttt{Random}_k),$$

...etc, where the $\texttt{Random}_i$'s are independent calls of a `Random` function.

Chapter 2

Dynamical copulae models for credit risk

2.1 Introduction

The reasons for studying and modelling dependencies in finance and insurance are of different type. One motivation is that independence assumptions, which are typical of many stochastic models, are often due more to convenience rather than to the nature of the problem at hand. Furthermore, there are situations where neglecting dependence effects may incur into a (dramatic) risk underestimation (see e.g. Bäuerle and Müller (1987) and Daul, De Giorgi, Lindskog and McNeil (2003)). Besides this, widely used scalar dependence or risk measures such as linear correlation, tail dependence coefficients and Value-at-Risk generally do not provide a satisfactory description of the underlying dependence structure and have severe limitations when used for measuring (portfolio) risk outside the Gaussian world (see e.g. Embrechts, Hoeing and Juri (2002) and Juri and Wüthrich (2004) for counterexamples).

Taking care of dependencies becomes therefore important in order to extend standard models towards a more efficient risk management. However, relaxing the independence assumption yields much less tractable models. It is therefore not surprising that only recently, i.e. in the last ten years, the mathematical literature on the risk management of dependent risks showed significant developments. The main message sent by much of this research is the following (see e.g. Frees and Valdez (1998), Joe (1997), or Nelsen (1999) among others). It is (intuitively) clear that the probabilistic mechanism governing the interactions between random variables is completely described by their joint distribution. On the other hand, in most applied situations, the joint distribution may be unknown or difficult to estimate such that only the marginals are known (estimated or fixed a priori). To tackle this problem a flexible and powerful approach consists in trying to model the joint distribution by means of copulae. The latter, which are often called "*dependence structures*", can be viewed as marginal free versions of joint distribution functions capturing scale invariant dependence properties of the several random variables.

The reverse side of the medal of the copula approach is that it is usually difficult to chose or find the appropriate copula for the problem at hand. Often, the only possibility is to start with some guess such as a parametric family of copulae and then try to fit the parameters (as made e.g. in Daul, De Giorgi, Lindskog and McNeil (2003)). As a consequence, the models obtained may suffer a certain degree of arbitrariness. As shown by Juri and Wüthrich (2002, 2003), some remedy to this weakness of the copula approach is provided by dependence models for (bivariate) conditional joint extremes, where limiting results along the lines of the Pickands-Balkema-De Haan Theorem are obtained. Such copula-convergence theorems reflect a distributional approach to the modelling of dependencies in the tails and provide natural descriptions of multivariate extremal events. Moreover, they differ from classical bivariate extreme value results since the limits obtained are not bivariate extreme value distributions. A further advantage of this kind of results is that they may also allow to better face the problem of the lack of data which is typical for rare events. In fact, there are situations where the knowledge of the limiting dependence structure reduces the issue of modelling tail events to the estimation of one parameter solely (Juri and Wüthrich (2003)).

2.1.1 Outline of the chapter

This chapter is structured as follows. In Section 2.2.1 we briefly recall the copula concept and all its properties that we will need throughout the rest of the chapter. The idea of dependence structures for tail events is then formalized in Section 2.2.2, where the concept of *tail dependence copula* (LTDC) is introduced; the latter provides a natural description of conditional bivariate joint extremes. Sections 2.3 and 2.4 contain the main results, which extend part of the work of Juri and Wüthrich (2002, 2003). In particular, Theorem 2.3.4 identifies, under suitable regularity conditions, possible LTDC-limits, i.e. limit laws for bivariate joint extremes. Motivated by classical results such as the Central Limit Theorem and the Fisher-Tippett Theorem, we show in Section 2.4 that LTDC-limits are characterized by invariance properties (Theorems 2.4.6, 2.4.10 and Corollary 2.4.11). Section 2.5 will focus on the case of exchangeable portfolios. In Section 2.5, we show how the results of the preceding sections can be applied to the credit risk area, where, for intensity-based default models, dependence structures characterizing the behavior under stress scenarios of widely traded credit derivatives such as Credit Default Swap baskets or First-to-Default contract types are obtained.

2.2 Dependence structures for tail events

2.2.1 Preliminaries

As mentioned above, one of the main concepts used to describe scale invariant dependence properties of multivariate distributions is the copula one. In this work, we focus on bivariate continuous random vectors only and most of the following material can be found in Nelsen (1999) or Joe (1997).

Recall that a two-dimensional copula is a two-dimensional distribution function restricted to $[0,1]^2$ with standard uniform marginals. Hence, copulae can be equivalently defined as functions $C : [0,1]^2 \to [0,1]$ satisfying for $0 \le x \le 1$ and $(x_1, y_1), (x_2, y_2) \in [0,1]^2$ with $x_1 \le x_2$, $y_1 \le y_2$ the conditions

$$C(x,1) = C(1,x) = x, \quad C(x,0) = C(0,x) = 0, \tag{2.1}$$

$$C(x_2,y_2) - C(x_2,y_1) - C(x_1,y_2) + C(x_1,y_1) \ge 0, \tag{2.2}$$

where (2.1) translates into the uniformity of the marginals and that inequality (2.2), which is the 2-*increasing property*, can be interpreted as $\mathbb{P}(x_1 \le X \le x_2, y_1 \le Y \le y_2)$ for (X,Y) having distribution function C.

One of the most important and useful results about copulae is Sklar's Theorem stated below in its bivariate form (see Theorem 1.3.1). Let C be a copula and F_1, F_2 be univariate distribution functions. Then, for $(t_1, t_2) \in \mathbb{R}^2$,

$$F(t_1, t_2) = C(F_1(t_1), F_2(t_2)) \tag{2.3}$$

defines a distribution function with marginals F_1, F_2. Conversely, for a two-dimensional distribution function F with marginals F_1, F_2 there is a copula C satisfying (2.3). This copula is not necessarily unique, but it is if F_1 and F_2 are continuous, in which case for any $(x, y) \in [0,1]^2$,

$$C(x, y) = F(F_1^{\leftarrow}(x), F_2^{\leftarrow}(y)), \tag{2.4}$$

where $F_1^{\leftarrow}, F_2^{\leftarrow}$ denote the generalized left continuous inverses of F_1 and F_2.

2.2.2 Tail dependence copulae

A natural way to construct dependence structures (copulae) for bivariate (lower) tail events, is to consider first two-dimensional continuous conditional distribution functions, where the condition is that both variables fall below small thresholds. The second step is to get then the relative copula using the second part of Sklar's Theorem (Equation (2.4)).

Remark 2.2.1. *In the sequel, we will assume that the considered copula C is such that $x \mapsto C(x, y)$ and $y \mapsto C(x, y)$ are strictly increasing for all $x, y \in (0, 1]$. We denote by $\mathcal{C}$ the set of such copulae.*

Let (U, V) be a random vector with distribution function $C \in \mathcal{C}$. For any $(u, v) \in (0,1]^2$, the conditional distribution of (U, V) given $U \le u, V \le v$, denoted by $F(C, u, v)$, is given, for $0 \le x \le u$ and $0 \le y \le v$, by

$$F(C, u, v)(x, y) = \mathbb{P}(U \le x, V \le y | U \le u, V \le v) = \frac{C(x, y)}{C(u, v)}. \tag{2.5}$$

The marginal distribution functions of $F(C, u, v)$ in (2.5) are given for $0 \le x \le u$ and $0 \le y \le v$ respectively by

$$F_U(C, u, v)(x) = \frac{C(x, v)}{C(u, v)} \quad \text{and} \quad F_V(C, u, v)(y) = \frac{C(u, y)}{C(u, v)}. \tag{2.6}$$

Since, $F_U(C, u, v), F_V(C, u, v)$ are continuous, the unique copula relative to $F(C, u, v)$ is obtained from (2.4) and equals

$$F(C, u, v)(F_U(C, u, v)^{\leftarrow}(x), F_V(C, u, v)^{\leftarrow}(y)) = \frac{C(F_U(C, u, v)^{\leftarrow}(x), F_V(C, u, v)^{\leftarrow}(y))}{C(u, v)}. \tag{2.7}$$

Definition 2.2.2. *For $C \in \mathcal{C}$, we call the copula defined by (2.7) the lower tail dependence copula relative to C, LTDC for short, and we denote it by $\Phi(C, u, v)$.*

Note that the assumption that $C \in \mathcal{C}$ implies that $\{(u, v) \in [0,1]^2 : C(u, v) > 0\} = (0,1]^2$, i.e. it ensures that the LTDC $\Phi(C, u, v)$ is well defined for all $u, v \in (0, 1]$. Furthermore, $\lim_{u,v \to 0} \Phi(C, u, v)$ describes naturally the dependence structure underlying conditional bivariate random samples in the lower-tails.

Furthermore, starting with uniform marginals, i.e. with a copula C, is not a restriction since the dependence structure that would be obtained with different marginals is again of the type $\Phi(C, u, v)$. In fact, let X_1, X_2 have joint distribution function G, strictly increasing continuous marginals G_1, G_2 and copula C. Analogously to the above, consider for appropriate (i.e. such that the following expressions are well defined) $z_1, z_2 \in \mathbb{R}$ the conditional distribution function

$$G^{z_1,z_2}(x_1, x_2) = \mathbb{P}(X_1 \le x_1, X_2 \le x_2 | X_1 \le z_1, X_2 \le z_2). \tag{2.8}$$

Further, let $G_1^{z_1,z_2}(x_1) = G^{z_1,z_2}(x_1, z_2)$ and $G_2^{z_1,z_2}(x_2) = G^{z_1,z_2}(z_1, x_2)$, respectively. Because of Sklar's Theorem, we have that the copula relative to G^{z_1,z_2} is given by

$$\Phi(G, z_1, z_2)(u_1, u_2) = G^{z_1,z_2}((G_1^{z_1,z_2})^{\leftarrow}(u_1), (G_2^{z_1,z_2})^{\leftarrow}(u_2)). \tag{2.9}$$

Proposition 2.2.3. *In the above setting holds* $\Phi(C, G_1(z_1), G_2(z_2)) = \Phi(G, z_1, z_2)$.

Proof. For $w_i = G_i(z_i)$, $i = 1, 2$, we have by definition that

$$\Phi(C, G_1(z_1), G_2(z_2))(u_1, u_2) = \frac{C(F_{U_1}(C, w_1, w_2)^{\leftarrow}(u_1), F_{U_2}(C, w_1, w_2)^{\leftarrow}(u_2))}{C(G_1(z_1), G_2(z_2))}. \tag{2.10}$$

Further,

$$\begin{aligned} F_{U_1}(C, w_1, w_2)(v_1) &= \frac{C(v_1, w_2)}{C(w_1, w_2)} = \frac{C(v_1, G_2(z_2))}{C(G_1(z_1), G_2(z_2))} = \frac{G(G_1^{\leftarrow}(v_1), z_2)}{G(z_1, z_2)} \\ &= G_1^{z_1,z_2}(G_1^{\leftarrow}(v_1)), \end{aligned} \tag{2.11}$$

whence $F_{U_1}(C, w_1, w_2)^{\leftarrow}(u_1) = G_1((G_1^{z_1,z_2})^{\leftarrow}(u_1))$. Similarly, we have that $F_{U_2}(C, w_1, w_2)^{\leftarrow}(u_2) = G_2((G_2^{z_1,z_2})^{\leftarrow}(u_2))$. Thus,

$$\begin{aligned} \Phi(C, G_1(z_1), G_2(z_2))(u_1, u_2) &= \frac{C(G_1((G_1^{z_1,z_2})^{\leftarrow}(u_1)), G_2((G_2^{z_1,z_2})^{\leftarrow}(u_2)))}{G(z_1, z_2)} \\ = \frac{G((G_1^{z_1,z_2})^{\leftarrow}(u_1), (G_2^{z_1,z_2})^{\leftarrow}(u_2))}{G(z_1, z_2)} &= G^{z_1,z_2}((G_1^{z_1,z_2})^{\leftarrow}(u_1), (G_2^{z_1,z_2})^{\leftarrow}(u_2)) \\ = \Phi(G, z_1, z_2)(u_1, u_2). \end{aligned} \tag{2.12}$$

□

Remark 2.2.4. *Sometimes it may be more natural to look at dependencies in the upper-tails rather than in the lower-tails as e.g. in any situation where one is interested in the joint behavior of random variables conditional on high thresholds. To such an extent, one could consider in* (2.5) *the expression* $\mathbb{P}(U > x, V > y | U > u, V > v)$ *instead of* $\mathbb{P}(U \le x, V \le y | U \le u, V \le v)$ *yielding, through the analogous to* (2.7), *a dependence structure for upper-tail events. Such dependence structures can be also obtained replacing* C *in Definition 2.2.2 by the relative survival copula* $C^*(x, y) = x + y - 1 + C(1 - x, 1 - y)$, $x, y \in [0, 1]^2$. *Indeed, it is easily seen that for* (X, Y) *with distribution function* F, *marginals* F_1, F_2 *and copula* C, *the copula of* $(-X, -Y)$ *is precisely* C^* *and that for* $(x, y) \in \mathbb{R}^2$

$$\mathbb{P}(X > x, Y > y) = C^*(1 - F_1(x), 1 - F_2(y)). \tag{2.13}$$

2.2.3 An excursion in higher dimension

Most of the results in this Chapter will be proved in dimension 2, since they avoid too heavy notations, but also because limiting theorems can be obtained in that dimension. Nevertheless, the lower tail dependence copula can be defined in dimension $d \ge 2$.

Let $\boldsymbol{U} = (U_1, ..., U_d)$ be a random vector with uniform margins, so that its distribution function is a copula C. Assume that C is strictly increasing in all components. For any $\boldsymbol{u} = (u_1, ..., u_d) \in (0, 1]^d$, consider the random vector $(U_1, ..., U_d)$ given $\{U_1 \le u_1, ..., U_d \le u_d\}$, also denoted $\boldsymbol{U}$ given $\boldsymbol{U} \le \boldsymbol{u}$. If $F_{i|\boldsymbol{u}}(\cdot)$ denotes the marginal distribution function of U_i given $\{U_1 \le u_1, ..., U_i \le u_i, ..., U_d \le u_d\}$, note that

$$F_{i|\boldsymbol{u}}(x_i) = \frac{C(u_1, ..., u_{i-1}, x_i, u_{i+1}, ..., u_d)}{C(u_1, ..., u_{i-1}, u_i, u_{i+1}, ..., u_d)}, \tag{2.14}$$

and therefore the copula of the conditional vector $\boldsymbol{U}$ given $\boldsymbol{U} \le \boldsymbol{u}$ is

$$\Phi(C, u_1, ..., u_d)(x_1, ..., x_d) = \frac{C(F_{1|\boldsymbol{u}}^{\leftarrow}(x_1), ..., F_{d|\boldsymbol{u}}^{\leftarrow}(x_d))}{C(u_1, ..., u_d)}. \tag{2.15}$$

Definition 2.2.5. *For C denote a copula with strictly increasing components, the copula defined by Equation 2.15 is called lower tail dependence copula, LTDC for short, and denoted by $\Phi(C, \boldsymbol{u})$, $\boldsymbol{u} \in (0,1]^d$.*

2.3 A limit theorem

The main result of this section is given by Theorem 2.3.4 below, where limits of the type $\lim_{t\to 0} \Phi(C, r(t), s(t))$ are considered. An explicit form for the limit is provided under the assumption that the functions $r(\cdot), s(\cdot)$ defining the direction under which the limit is taken satisfy suitable regularity conditions. Further, an example of a non-symmetric LTDC-limit, i.e. a limit obtained under a direction (r, s) with $r \neq s$, is given in Proposition 2.3.10 where we show that a dependence model in the lower-tails may be given by the Marshall and Olkin copula of Example 1.3.3. As we will see in Section 2.5, this copula turns out to be a natural model for some credit derivatives.

For our purposes, the concept of regular variation appears to be the appropriate one. A standard reference to the topic of regular variation is Bingham, Goldie and Teugels (1987) and results for the multivariate case can also be found in de Haan, Omey and Resnick (1984). See also Section 3.4.3. of this thesis.

Definition 2.3.1. *A measurable function $f : (0,\infty) \to (0,\infty)$ is called regularly varying at 0 with index $\rho \in \mathbb{R}$, if for any $x > 0$,*

$$\lim_{t\to 0} \frac{f(tx)}{f(t)} = x^{\rho}. \tag{2.16}$$

We write $f \in \mathcal{R}^0_\rho$. In the case where $\rho = 0$, the function is said to be slow varying at 0.

Definition 2.3.2. *A measurable function $f : (0,\infty)^2 \to (0,\infty)$ is called regularly varying at 0 with auxiliary functions $r, s : (0,\infty) \to (0,\infty)$ if $\lim_{t\to 0} r(t) = \lim_{t\to 0} s(t) = 0$ and there is a positive measurable function $\phi : (0,\infty)^2 \to (0,\infty)$ such that*

$$\lim_{t\to 0} \frac{f(r(t)x, s(t)y)}{f(r(t), s(t))} = \phi(x,y) \quad \textit{for all } x, y > 0. \tag{2.17}$$

We write $f \in \mathcal{R}(r, s)$ and we call ϕ the limiting function under the direction (r, s).

Remark 2.3.3. *Definition 2.3.2 can be easily modified to include functions, such as copulae, having a domain different from $(0,\infty)^2$. This ensures in particular that the left hand side of (2.18) below is well-defined.*

Theorem 2.3.4. *Let $C \in \mathcal{C} \cap \mathcal{R}(r, s)$ with limiting function ϕ and assume that r, s are strictly increasing continuous functions such that $r \in \mathcal{R}^0_\alpha$ and $s \in \mathcal{R}^0_\beta$ for some $\alpha, \beta > 0$. Then, for any $(x, y) \in [0,1]^2$,*

$$\lim_{t\to 0} \Phi(C, r(t), s(t))(x,y) = \phi(\phi_X^{\leftarrow}(x), \phi_Y^{\leftarrow}(y)), \tag{2.18}$$

where $\phi_X(x) = \phi(x,1)$ and $\phi_Y(y) = \phi(1,y)$. Moreover, there is a constant $\theta > 0$ such that $\phi(x,y) = x^{\theta/\alpha} h(yx^{-\beta/\alpha})$ for $x > 0$, where

$$h(t) = \begin{cases} \phi_Y(t) & \textit{for } t \in [0,1] \\ t^{\theta/\beta}\phi_X(t^{-\alpha/\beta}) & \textit{for } t \in (1,\infty) \end{cases}. \tag{2.19}$$

Proof. The proof of this theorem is based on the following lemma.

Lemma 2.3.5. *Suppose that (X_d, Y_d) have continuous strictly increasing marginals and are such that $\lim_{n\to\infty}(X_d, Y_d) = (X, Y)$ in distribution for some (X, Y). Then,*

$$\lim_{n\to\infty} \|C_d - C\|_\infty = 0, \tag{2.20}$$

where C_d and C denote the copulae of (X_d, Y_d) and (X, Y), respectively.

Proof. Denote by $F_{X_d}, F_{Y_d}, F_X, F_Y, F_d$ and F the distribution functions of X_d, Y_d, X, Y, (X_d, Y_d) and (X, Y), respectively. Then, for $u, v \in [0, 1]$,

$$\begin{aligned} |C_d(u,v) - C(u,v)| &= |F_d(F^{\leftarrow}_{X_d}(u), F^{\leftarrow}_{Y_d}(v)) - F(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v))| \\ &\le |F_n(F^{\leftarrow}_{X_d}(u), F^{\leftarrow}_{Y_d}(v)) - F_d(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v))| \\ &\quad + |F_n(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v)) - F(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v))|. \end{aligned} \tag{2.21}$$

Because F_n is continuous and since F_{X_n} and F_{Y_n} are strictly increasing, $F^{\leftarrow}_{X_d}(u) \to F^{\leftarrow}_X(u)$ and $F^{\leftarrow}_{Y_d}(v) \to F^{\leftarrow}_Y(v)$ as $n \to \infty$ for any $u, v \in [0, 1]$. So, for any $\varepsilon > 0$ there is some positive integer N_1 such that for any $n \ge N_1$

$$|F_d(F^{\leftarrow}_{X_d}(u), F^{\leftarrow}_{Y_d}(v)) - F_d(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v))| \le \varepsilon/2. \tag{2.22}$$

Similarly, because $\lim_{n\to\infty} F_d(x, y) = F(x, y)$, there is N_2 such that for any $n \ge N_2$

$$|F_d(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v)) - F(F^{\leftarrow}_X(u), F^{\leftarrow}_Y(v))| \le \varepsilon/2. \tag{2.23}$$

Thus, for any $u, v \in [0, 1]$ and any $n \ge N = \max\{N_1, N_2\}$, we have that $|C_d(u, v) - C(u, v)| \le \varepsilon$, i.e. $\lim_{n\to\infty} C_d = C$ pointwise. Because $[0, 1]^2$ is compact and both C_d and C are continuous, this convergence is also uniform. This finishes the proof of Lemma 2.3.5. □

Let now (U, V) have distribution function C. Note that

$$\frac{C(r(t)x, s(t))}{C(r(t), s(t))} = \mathbb{P}(U \le r(t)x | U \le r(t), V \le s(t)), \tag{2.24}$$

$$\frac{C(r(t), s(t)y)}{C(r(t), s(t))} = \mathbb{P}(V \le s(t)y | U \le r(t), V \le s(t)), \tag{2.25}$$

$$\frac{C(r(t)x, s(t)y)}{C(r(t), s(t))} = \mathbb{P}(U \le r(t)x, V \le s(t)y | U \le r(t), V \le s(t)), \tag{2.26}$$

i.e. the distributions in (2.24)–(2.26) are respectively the conditional distributions of $U/r(t)$, $V/s(t)$ and $(U/r(t), V/s(t))$ given $U \le r(t), V \le s(t)$. Since copulae are invariant under strictly increasing transformations of the underlying variables, it follows that we can take the conditional distributions in (2.24)–(2.26) instead of $F_U(C, r(t), s(t))$, $F_V(C, r(t), s(t))$ and $F(C, r(t), s(t))$ in order to construct $\Phi(C, r(t), s(t))$. Further, since $C \in \mathcal{C}$ and because r, s are strictly increasing and continuous, we have that the distributions in (2.24)–(2.26) are continuous too and strictly increasing. By hypothesis $C \in \mathcal{R}(r, s)$, i.e.

$$\lim_{t\to 0} \frac{C(r(t)x, s(t)y)}{C(r(t), s(t))} = \phi(x, y) \quad \text{for all } x, y \in [0, 1], \tag{2.27}$$

so that the expressions in (2.26) converge to ϕ_X, ϕ_Y and ϕ as $t \to 0$ respectively. Thus, applying Lemma 2.3.5, we get

$$\lim_{t\to 0} \Phi(C, r(t), s(t))(x, y) = \phi(\phi^{\leftarrow}_X(x), \phi^{\leftarrow}_Y(y)), \tag{2.28}$$

whence (2.18) has been proved. Since $r \in \mathcal{R}^0_\alpha, s \in \mathcal{R}^0_\beta$, we have according to Theorem 2.1 in de Haan (1984) that there is $\theta > 0$ such that

$$\phi(t^\alpha x, t^\beta y) = t^\theta \phi(x,y), \text{ for all } t, x, y > 0. \tag{2.29}$$

Further, according to Aczél (1966) the most general solution to the functional equation (2.29) is given by

$$\phi(x,y) = \begin{cases} x^{\theta/\alpha} h(yx^{-\beta/\alpha}) & \text{if } x \neq 0 \\ cy^{\theta/\beta} & \text{if } x = 0 \text{ and } y \neq 0 \\ 0 & \text{if } x = y = 0 \end{cases}, \tag{2.30}$$

where c is a constant and h is function of one variable. Because $\phi(0,y) = 0$ and $\phi_Y(y) = \phi(1,y) = h(y)$, it follows that $c = 0$ and that the restriction of h on $[0,1]$ equals ϕ_Y, respectively. Further, we have for $x \in (0,1]$ that

$$\phi_X(x^{\alpha/\beta}) = \phi(x^{\alpha/\beta}, 1) = x^{\theta/\beta} h(1/x), \tag{2.31}$$

whence for $t = 1/x > 1$ we obtain $h(t) = h(1/x) = x^{-\theta/\beta}\phi_X(x^{\alpha/\beta}) = t^{\theta/\beta}\phi_X(t^{-\alpha/\beta})$, which shows (2.19) and finishes therefore the proof of Theorem 2.3.4. □

Remark 2.3.6. *Note that the limiting function ϕ in* (2.17) *is obtained from a pointwise convergence. Because the domain of a copula is the compact set $[0,1]^2$, it follows that the assumption $C \in \mathcal{C} \cap \mathcal{R}(r,s)$ implies that the convergence in* (2.18) *is also uniform, i.e. we have that* $\lim_{t\to 0} \|\Phi(C, r(t), s(t))(\cdot,\cdot) - \phi(\phi_X^{\leftarrow}(\cdot), \phi_Y^{\leftarrow}(\cdot))\|_\infty = 0$.

Remark 2.3.7. *Observe that the hypothesis that r, s are continuous functions is necessary, otherwise counterexamples such as "copulae with fractal support" as considered in Fredricks, Nelsen and Rodriguez-Lallena (2005) can be constructed. Let $T = (t_{ij})$ be a square matrix with non-negative entries whose sum equals to one determining the following subdivision of the unit square $[0,1]^2$ into rectangles: let c_i, $i = 0, \dots, n$ the sum of the entries of the first i columns of T with $c_0 = 0$ and let r_j, $j = 0, \dots, n$ be the sum of the entries in the first j rows of T with $r_0 = 0$. Then, the vectors $r = (r_0, \dots, r_d)$ and $c = (c_0, \dots, c_d)$ define partitions of* $[0,1]$*, whence* $[0,1]^2$ *is partitioned into the rectangles $R_{ij} = [c_{i-1}, c_i] \times [r_{i-1}, r_i]$. Further, for a given copula C and $(x,y) \in R_{ij}$, consider the new copula $T(C)$ defined by*

$$T(C)(x,y) = \sum_{u<i,v<j} t_{uv} + \frac{x - c_{i-1}}{c_i - c_{i-1}} \sum_{v<j} t_{iv} + \frac{y - r_{j-1}}{r_j - r_{j-1}} \sum_{u<i} t_{uj} + C\left(\frac{x - c_i}{c_i - c_{i-1}}, \frac{y - r_j}{r_j - r_{j-1}}\right) t_{ij}, \tag{2.32}$$

where empty sums are defined as zero. Fredricks, Nelsen and Rodriguez-Lallena (2005) show that for any copula C and any $T \neq 1$ there is a unique copula C_T that depends only on T such that $T(C_T) = C_T$. Moreover, they show that $C_T = \lim_{n\to\infty} T^n C$, where $T^n C = T(T^{n-1}C)$, $n \geq 1$, $T^1 C = T(C)$ and $T^0 C = C$. Consider now the case where the starting copula C is the independent copula, i.e. $C(x,y) = C^\perp(x,y) = xy$ and the transformation matrix T is given by

$$T = \begin{pmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.6 & 0 \\ 0.1 & 0 & 0.1 \end{pmatrix}, \tag{2.33}$$

whence $c = r = (0, 0.2, 0.8, 1)$. Then, we have for $t_k = 0.2^k$, $k \geq 1$ that

$$\Phi(C_T, t_k, t_k) = C_T = \lim_{n\to\infty} T^n C, \quad \textit{any } k \geq 1. \tag{2.34}$$

The fact that $\Phi(C_T, t_k, t_k) = C_T$ *can be explained with the help of Figure 2.1, where the support of* $T^n(C)$ *is plotted for* $n = 1, 2, 3, 4$ *and the colored regions are the ones where the measure relative to* $T^n(C)$ *concentrates its mass (indeed, we see from* (2.32) *that the support of* T^nC *is given by the rectangles corresponding to the non-zero elements of* T*). Observe that since* C *is the independent copula, the measure relative to* T^nC *spreads its mass uniformly on the colored squares. Taking for example the upper right picture in Figure 2.1, we see that restricting ourselves to* $[0, t_1]^2 = [0, 0.2]^2$ *we have exactly the same picture as in the upper left of Figure 2.1. This means that if* (U, V) *has copula* T^nC *for some* $n \geq 1$*, then* $(U, V)|U, V \leq t_1$ *has c.d.f.* $T^{n-1}C(xt_1, yt_1)$*,* $x, y \in [0, 1]$*. It follows that the copula of* $(U, V)|U, V \leq t_1$ *is exactly* $T^{n-1}C$*, i.e.* $\Phi(T^nC, t_1, t_1) = T^{n-1}C$*. Using the same arguments, we have in general that* $\Phi(T^nC, t_k, t_k) = T^{n-k}C$*. Finally, because* $\Phi(\cdot, t_k, t_k)$ *is continuous (see Lemma 2.4.13), it follows that* $\Phi(C_T, t_k, t_k) = \lim_{n\to\infty} \Phi(T^nC, t_k, t_k) = \lim_{n\to\infty} T^{n-k}C = C_T$*.*

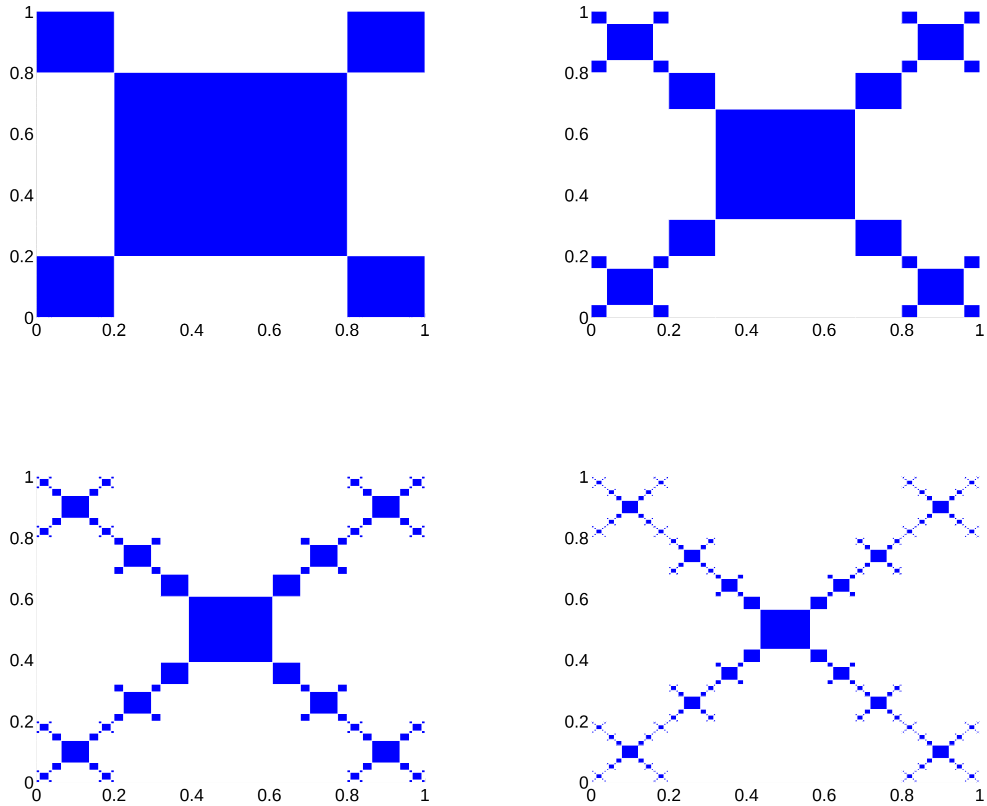


Figure 2.1: Support of $T^n(C)$ for $n = 1, 2, 3, 4$

Remark 2.3.8. *Letting* $\alpha = \beta = 1$*, we have that Theorem 2.3.4 generalizes Theorem 2.4 in Juri and Wüthrich (2004), the latter stating that*

$$\lim_{u\to 0} \Phi(C, u, u)(x, y) = G(g^{\leftarrow}(x), g^{\leftarrow}(y)), \tag{2.35}$$

where $g : [0, \infty) \to [0, \infty)$ *is the strictly increasing continuous function defined by* $g(x) = \lim_{u\to 0} C(xu, u)/C(u, u)$*,* $G(x, y) = y^\theta g(x/y)$ *for* $(x, y) \in (0, 1]^2$ *and* 0 *elsewhere and* θ *is a positive constant. In particular, Theorem 2.4 in Juri and Wüthrich (2003) applies to archimedean copulae having regularly varying generators in which case the LTDC-limit is the Clayton copula of Example 2.4.9 of the next section with parameter equal to minus the regular variation parameter (Theorem 3.4 in Juri and Wüthrich (2004), and Theorem 3.3 in Juri and Wüthrich (2003)).*

Following the previous remark, the analytical expression (2.35) for the limiting copula is due to the fact that homogeneous functions of order θ (in our case $G(x, y) = y^\theta g(x/y)$) have closed form expressions. Analogously, the closed form (2.19) comes from the fact that "*generalized homogeneous functions*" such as ϕ in Theorem 2.3.4 also have closed form representations (see the proof of Theorem 2.3.4 and Aczél (1966) for more details). Unfortunately, this is not the case in higher dimensions, so that, assuming that Theorem 2.3.4 could be extended case along the same lines to the multivariate, the limiting copula would not have a closed form expression.

Remark 2.3.9. *There are many papers in the literature concerning multivariate extremes. In particular, Bivariate Extreme Value (BEV) distributions are obtained as limit laws of suitably normalized componentwise maxima as it can be found e.g. in de Haan, Omey and Resnick (1984), Resnick (1987), and Joe (1997). It can be shown that the copula C of any BEV distribution satisfies the max-stability property*

$$C^t(u, v) = C(u^t, v^t) \quad \textit{for all } (u, v) \in [0, 1]^2 \textit{ and any } t > 0. \tag{2.36}$$

As mentioned in Juri and Wüthrich (2004), BEV copulae differ from LTDC-limits, the difference being similar to the one between the univariate Generalized Extreme Value (GEV) distributions and the Generalized Pareto Distribution (GPD). In fact, the GPD lives on the log-scaled compared to GEV distributions (Theorem 4.2 in Juri and Wüthrich (2004)). For instance, the Gumbel copula satisfies (2.36)*, but is not an LTDC-limit. For a more detailed discussion about relations with other results from the area of multivariate extremes we refer to Chapter* 6 *of this thesis*

We finish this section with an example of an LTDC-limit which is not of the form (2.35). We will see in Section 2.4 that Theorem 2.4.6 provides a whole family of other examples of this type.

Proposition 2.3.10. *Let $a, b : [0, 1] \to [0, 1]$ be two increasing functions with $a(0) = b(0) = 0$, $a(1) = b(1) = 1$ and such that $t \mapsto a(t)/t$, $t \mapsto b(t)/t$ are decreasing on $(0, 1]$. Then,*

$$C(x, y) = (a(x)y) \wedge (xb(y)) \tag{2.37}$$

defines a copula. Additionally, if $a \in \mathcal{R}^0_\alpha$, $b \in \mathcal{R}^0_\beta$, where $(\alpha, \beta) \in [0, 1]^2 \setminus \{(0, 0)\}$ and for directions r, s such that

$$\lim_{t \to 0} \frac{r(t)b(s(t))}{a(r(t))s(t)} = 1, \tag{2.38}$$

we have that

$$\lim_{t \to 0} \Phi(C, r(t), s(t))(x, y) = (x^\alpha y) \wedge (xy^\beta), \tag{2.39}$$

which is the Marshall and Olkin copula with parameters $1 - \alpha$ and $1 - \beta$.

Proof. In order to prove that (2.37) defines a copula, we have to show (2.1) and (2.2). For $x \in [0, 1]$, the conditions $C(x, 0) = C(0, x) = 0$ are satisfied because $a(0) = b(0) = 0$. Further, since $x \mapsto a(x)/x$ is decreasing with $a(1) = 1$, we have that $a(x) \geq x$ for any $x \in [0, 1]$. Thus, because $b(1) = 1$, we get $C(x, 1) = a(x) \wedge x = x$. Similarly, $C(1, x) = x$, $x \in [0, 1]$, which shows (2.1). Consider now $0 < x_1 \leq x_2 \leq 1$ and $0 < y_1 \leq y_2 \leq 1$. Then,

$$\begin{aligned} \Delta &= C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1) \\ &= x_2 y_2 \left(\frac{a(x_2)}{x_2} \wedge \frac{b(y_2)}{y_2} \right) - x_1 y_2 \left(\frac{a(x_1)}{x_1} \wedge \frac{b(y_2)}{y_2} \right) - x_2 y_1 \left(\frac{a(x_2)}{x_2} \wedge \frac{b(y_1)}{y_1} \right) \\ &\quad + x_1 y_1 \left(\frac{a(x_1)}{x_1} \wedge \frac{b(y_1)}{y_1} \right). \end{aligned} \tag{2.40}$$

Since $x \mapsto a(x)/x$ and $x \mapsto b(x)/x$ are decreasing, six different cases have to be considered:

1. Assume that $a(x_2)/x_2 \le a(x_1)/x_1 \le b(y_2)/y_2 \le b(y_1)/y_1$. Then, $\Delta = (y_2 - y_1)(a(x_2) - a(x_1)) \ge 0$ since a is increasing.

2. If $b(y_2)/y_2 \le b(y_1)/y_1 \le a(x_2)/x_2 \le a(x_1)/x_1$, then $\Delta = (x_2 - x_1)(b(y_2) - b(y_1))$, which is of course non-negative.

3. Suppose now that $a(x_2)/x_2 \le b(y_2)/y_2 \le b(y_1)/y_1 \le a(x_1)/x_1$. Then, $\Delta = x_1(b(y_1) - b(y_2)) + a(x_2)(y_2 - y_1)$ is non-negative if and only if
$$\frac{b(y_2) - b(y_1)}{y_2 - y_1} \le \frac{a(x_2)}{x_1}. \tag{2.41}$$
Since $x \mapsto b(x)/x$ is decreasing, the left hand side of (2.41) can be bounded as follows:
$$\frac{b(y_2) - b(y_1)}{y_2 - y_1} = \frac{b(y_2)}{y_2}\frac{y_2}{y_2 - y_1} - \frac{b(y_1)}{y_1}\frac{y_1}{y_2 - y_1} \le \frac{b(y_1)}{y_1}. \tag{2.42}$$
By hypothesis and since a is increasing, we have $b(y_1)/y_1 \le a(x_1)/x_1 \le a(x_2)/x_1$, whence (2.41).

4. The case $b(y_2)/y_2 \le a(x_2)/x_2 \le a(x_1)/x_1 \le b(y_1)/y_1$ yields $\Delta = (a(x_1) - a(x_2))y_1 + (x_2 - x_1)b(y_2)$, which can be shown to be non-negative using the same arguments as in (3).

5. If $a(x_2)/x_2 \le b(y_2)/y_2 \le a(x_1)/x_1 \le b(y_1)/y_1$, then $\Delta = (y_2 - y_1)a(x_2) - x_1 b(y_2) + y_1 a(x_1)$. By hypothesis,
$$x_1 b(y_2) \le a(x_1)y_2 = a(x_1)y_1 + a(x_1)(y_2 - y_1) \le a(x_1)y_1 + a(x_2)(y_2 - y_1), \tag{2.43}$$
where the last inequality follows because a is increasing. This shows that $\Delta \ge 0$.

6. The last case is given by $b(y_2)/y_2 \le a(x_2)/x_2 \le b(y_1)/y_1 \le a(x_1)/x_1$ and $\Delta = (x_2 - x_1)b(y_2) - y_1 a(x_2) + x_1 b(y_1)$. As in (5), it follows that $\Delta \ge 0$.

In order to prove (2.39), consider
$$\begin{aligned}
\frac{C(r(t)x, s(t)y)}{C(r(t), s(t))} &= \frac{[a(r(t)x)s(t)y] \wedge [r(t)xb(s(t)y)]}{[a(r(t))s(t)] \wedge [r(t)b(s(t))]} &(2.44)\\
&= \frac{\left[\frac{a(r(t)x)}{a(r(t))}y\right] \wedge \left[x\frac{b(s(t)y)}{b(s(t))}\frac{r(t)}{a(r(t))}\frac{b(s(t))}{s(t)}\right]}{1 \wedge \left[\frac{r(t)}{a(r(t))}\frac{b(s(t))}{s(t)}\right]}. &(2.45)
\end{aligned}$$

Since by hypothesis, $a \in \mathcal{R}^0_\alpha$, $b \in \mathcal{R}^0_\beta$ and $\lim_{t\to 0} r(t)b(s(t))/(a(r(t))s(t)) = 1$, it follows that
$$\lim_{t\to 0} \frac{C(r(t)x, s(t)y)}{C(r(t), s(t))} = (x^\alpha y) \wedge (xy^\beta) = \phi(x, y), \tag{2.46}$$
i.e. $C \in \mathcal{C} \cap \mathcal{R}(r, s)$ with limiting function ϕ. Since $0 < \alpha, \beta \le 1$, we have that $\phi_X(x) = \phi(x, 1) = x$ and $\phi_Y(y) = \phi(1, y) = y$, whence, because of Theorem 2.3.4, $\lim_{t\to 0} \Phi(C, r(t), s(t))(x, y) = \phi(\phi_X^{\leftarrow}(x), \phi_Y^{\leftarrow}(y)) = (x^\alpha y) \wedge (xy^\beta)$. □

Remark 2.3.11. *Condition* (2.38)*, is satisfied e.g. in the case where* $a(t) = t^\alpha$*,* $b(t) = t^\beta$*,* $r(t) = t^\gamma$ *and* $s(t) = t^\delta$ *with* $\beta\delta + \gamma = \alpha\gamma + \delta$*.*

2.4 Invariant copulae

There are many examples of (functional) limit theorems where the limit obtained is invariant under some kind of transformation. This is the case of the Central Limit Theorem, where stable laws (which coincide with the class of possible limit laws for sums of iid random variables) are invariant under the sum operator. A similar result holds for the GEV distribution, which is the limit of maxima of iid random variables as stated in the Fisher-Tippett Theorem (Embrechts, Klüppelberg and Mikosch (1997), Theorem 3.2.3).

In our context, we have that equation (2.7) can be seen as the result of a copula transformation mapping a copula $C \in \mathcal{C}$ to its LTDC $\Phi(C, u, v)$. Motivated by the above classical results, it seems therefore natural to look at copulae which are invariant under the LTDC-transformation (2.7).

Definition 2.4.1. *We say that $C \in \mathcal{C}$ is invariant on the unit square if $\Phi(C, u, v) = C$ for all $(u, v) \in (0, 1]^2$.*

Lemma 2.4.2. *Let (U, V) have distribution function $C \in \mathcal{C}$ and $(u, v) \in (0, 1]^2$. Then, $\Phi(C, u, v)$ satisfies for $(x, y) \in [0, u] \times [0, v]$ the identity*

$$\frac{C(x, y)}{C(u, v)} = \Phi(C, u, v) \left(\frac{C(x, v)}{C(u, v)}, \frac{C(u, y)}{C(u, v)} \right). \tag{2.47}$$

Proof. Because $C \in \mathcal{C}$, we have that $F_U(C, u, v)$ and $F_V(C, u, v)$ are strictly increasing. Because of Sklar's Theorem and using (2.5), (2.6), we get

$$\begin{aligned} \Phi(C, u, v) \left(\frac{C(x, v)}{C(u, v)}, \frac{C(u, y)}{C(u, v)} \right) &= \Phi(C, u, v)(F_U(C, u, v)(x), F_V(C, u, v)(y)) \\ &= F(C, u, v)(x, y) = \frac{C(x, y)}{C(u, v)}. \end{aligned} \tag{2.48}$$

This finishes the proof of Lemma 2.4.2. □

From Lemma 2.4.2, we have that C is invariant on the unit square if and only if for any $(u, v) \in (0, 1]^2$

$$\frac{C(x, y)}{C(u, v)} = C \left(\frac{C(x, v)}{C(u, v)}, \frac{C(u, y)}{C(u, v)} \right) \quad \text{for all } (x, y) \in [0, u] \times [0, v]. \tag{2.49}$$

A weaker type of invariance than the one of Definition 2.4.1, is given by copulae C such that $\Phi(C, u, v) = C$ holds only for a particular set of parameters $(u, v) \in (0, 1]^2$.

Definition 2.4.3. *A copula $C \in \mathcal{C}$ is said to be invariant on the diagonal if $\Phi(C, u, u) = C$ for all $u \in (0, 1]$. Similarly, $C \in \mathcal{C}$ is called "invariant under direction" $\mathcal{D} = \{(r(t), s(t)), t \in T\}$, $T \subset \mathbb{R}$ where $r, s : T \to (0, 1]$, whenever*

$$\Phi(C, r(t), s(t)) = C \quad \textit{for all } t \in T. \tag{2.50}$$

Invariant copulae on the diagonal have been considered by **?** (?, ?) and examples of such a copulae are given in Examples 2.4.4 and 2.4.9 below.

Example 2.4.4. *For $\alpha \in [0, 1]$ consider the Cuadras-Augé copula*

$$C_\alpha(x, y) = (x^{1-\alpha} y) \wedge (x y^{1-\alpha}).$$

The copula C_α can be seen as a particular case of a Marshall and Olkin copula of Example 1.3.3 with identical parameters and is a geometric mixture with weights α and $1-\alpha$ of the upper Fréchet bound $C^+(x,y) = x \wedge y$ and of the independent copula $C^\perp(x,y) = xy$. In fact,

$$C_\alpha(x,y) = C^+(x,y)^\alpha C^\perp(x,y)^{1-\alpha}.$$

For U,V with joint distribution function C_α, we have for $0 \le x,y \le u$ that

$$\begin{aligned} F_U(C_\alpha,u,u)(x) = F_V(C_\alpha,u,u)(x) &= \frac{C_\alpha(x,u)}{C_\alpha(u,u)} = \frac{x}{u}, \\ F(C_\alpha,u,u)(x,y) = \frac{C_\alpha(x,y)}{C_\alpha(u,u)} &= C_\alpha\left(\frac{x}{u},\frac{y}{u}\right). \end{aligned} \tag{2.51}$$

Thus, we immediately get from (2.7) that C_α is an invariant copula on the diagonal.

A particular family of curve-invariant copulae is the one of Definition 2.4.5 below. We will see in Corollary 2.4.11 that this family of copulae coincides with the LTDC-limits obtained in Theorem 2.3.4.

Definition 2.4.5. *Let α,β,θ be positive constants and P,Q be increasing continuous univariate distribution functions on $[0,1]$. We denote by $\mathcal{H}(\alpha,\beta,\theta)$ the set of two-dimensional distribution functions H on $[0,1]^2$ that can be expressed as*

$$H(x,y) = x^{\theta/\alpha} h(yx^{-\beta/\alpha}), \quad \textit{where} \quad h(t) = \begin{cases} Q(t) & \textit{if } t \in [0,1] \\ t^{\theta/\beta}P(t^{-\alpha/\beta}) & \textit{if } t \in (1,\infty) \end{cases}. \tag{2.52}$$

Theorem 2.4.6. *Let $\alpha,\beta,\theta > 0$ and $H \in \mathcal{H}(\alpha,\beta,\theta)$. Then,*

$$\Gamma(P,Q,\alpha,\beta,\theta)(u,v) = \begin{cases} Q^{\leftarrow}(v)^{\theta/\beta} P(P^{\leftarrow}(u)Q^{\leftarrow}(v)^{-\alpha/\beta}), & P^{\leftarrow}(u)^\beta \le Q^{\leftarrow}(v)^\alpha \\ P^{\leftarrow}(u)^{\theta/\alpha} Q(P^{\leftarrow}(u)^{-\beta/\alpha}Q^{\leftarrow}(v)), & P^{\leftarrow}(u)^\beta > Q^{\leftarrow}(v)^\alpha \end{cases} \tag{2.53}$$

defines an invariant copula on $\mathcal{D} = \{(P(t^\alpha), Q(t^\beta)), t \in (0,1]\}$.

Proof. We will first prove that $\Gamma(\alpha,\beta,\theta)$ defined by (2.53) is a copula and then show the invariance property. The function H defined by (2.52) can be rewritten as

$$\begin{aligned} H(x,y) &= \begin{cases} x^{\theta/\alpha}[yx^{-\beta/\alpha}]^{\theta/\beta} P([yx^{-\beta/\alpha}]^{-\alpha/\beta}) & \text{if } x^\beta < y^\alpha \\ x^{\theta/\alpha} Q(yx^{-\beta/\alpha}) & \text{if } x^\beta \ge y^\alpha \end{cases} \\ &= \begin{cases} y^{\theta/\beta} P(y^{-\alpha/\beta}x) & \text{if } x^\beta < y^\alpha \\ x^{\theta/\alpha} Q(yx^{-\beta/\alpha}) & \text{if } x^\beta \ge y^\alpha \end{cases}. \end{aligned} \tag{2.54}$$

By hypothesis, the marginals P,Q of H are strictly increasing continuous functions, whence it follows from Sklar's Theorem that the copula associated to H equals

$$H(P^{\leftarrow}(u), Q^{\leftarrow}(v)) = \begin{cases} Q^{\leftarrow}(v)^{\theta/\beta} P(P^{\leftarrow}(u)Q^{\leftarrow}(v)^{-\alpha/\beta}), & \text{if } P^{\leftarrow}(u)^\beta < Q^{\leftarrow}(v)^\alpha \\ P^{\leftarrow}(u)^{\theta/\alpha} Q(P^{\leftarrow}(u)^{-\beta/\alpha}Q^{\leftarrow}(v)), & \text{if } P^{\leftarrow}(u)^\beta \ge Q^{\leftarrow}(v)^\alpha \end{cases}, \tag{2.55}$$

which is precisely $\Gamma(P,Q,\alpha,\beta,\theta)$. We show now that $\Gamma(P,Q,\alpha,\beta,\theta)$ is invariant on the curve $\mathcal{D} = \{(P(t^\alpha), Q(t^\beta)), t \in (0,1]\}$. For notational convenience we denote $\Gamma(P,Q,\alpha,\beta,\theta)$ by C. In order to derive the LTDC associated to C, we first notice that from (2.55) it follows

$$\begin{aligned} C(P(t^\alpha), Q(t^\beta)) &= t^\theta, \\ C(x, Q(t^\beta)) &= \begin{cases} t^\theta P(P^{\leftarrow}(x)t^{-\alpha}) & \text{if } P^{\leftarrow}(x) < t^\alpha \\ P^{\leftarrow}(x)^{\theta/\alpha} Q(P^{\leftarrow}(x)^{-\beta/\alpha}t^\beta) & \text{if } P^{\leftarrow}(x) \ge t^\alpha \end{cases}, \\ C(P(t^\alpha), y) &= \begin{cases} Q^{\leftarrow}(y)^{\theta/\beta} P(t^\alpha Q^{\leftarrow}(y)^{-\alpha/\beta}) & \text{if } t^\beta < Q^{\leftarrow}(y) \\ t^\theta Q(P^{\leftarrow}(t^\alpha)^{-\beta/\alpha}Q^{\leftarrow}(y)) & \text{if } t^\beta \ge Q^{\leftarrow}(y) \end{cases}. \end{aligned} \tag{2.56}$$

Let now $(x, y) \in [0, P(t^\alpha)] \times [0, Q(t^\beta)]$. Because of (2.6) and (2.56), we have that the marginals of $F(C, P(t^\alpha), Q(t^\beta))$ are given respectively by

$$F_U(C, P(t^\alpha), Q(t^\beta))(x) = \frac{C(x, Q(t^\beta))}{C(P(t^\alpha), Q(t^\beta))} = \frac{t^\theta P(P^\leftarrow(x)t^{-\alpha})}{t^\theta} = P(P^\leftarrow(x)t^{-\alpha}), \tag{2.57}$$

$$F_V(C, P(t^\alpha), Q(t^\beta))(y) = \frac{C(P(t^\alpha), y)}{C(P(t^\alpha), Q(t^\beta))} = Q(t^{-\beta}Q^\leftarrow(y)). \tag{2.58}$$

Their inverses equal

$$\begin{aligned} &F_U(C, P(t^\alpha), Q(t^\beta))^\leftarrow(x) = P(P^\leftarrow(x)t^\alpha), \\ &F_V(C, P(t^\alpha), Q(t^\beta))^\leftarrow(y) = Q(t^\beta Q^\leftarrow(y)). \end{aligned} \tag{2.59}$$

Assume now that x, y are such that $P^\leftarrow(x)^\beta < Q^\leftarrow(y)^\alpha$. From (2.56) we obtain that

$$F(C, P(t^\alpha), Q(t^\beta))(x, y) = \frac{C(x, y)}{C(P(t^\alpha), Q(t^\beta))} = \frac{Q^\leftarrow(y)^{\theta/\beta} P(P^\leftarrow(x)Q^\leftarrow(y)^{-\alpha/\beta})}{t^\theta}. \tag{2.60}$$

Thus, for any $(x, y) \in (0, 1]^2$ such that $P^\leftarrow(F_U^\leftarrow(x))^\beta \leq Q^\leftarrow(F_V^\leftarrow(y))^\alpha$, i.e. $P^\leftarrow(x)^\beta \leq Q^\leftarrow(y)^\alpha$, we have that

$$\begin{aligned} &\Phi(C, P(t^\alpha), Q(t^\beta))(x, y) \\ &= F(C, P(t^\alpha), Q(t^\beta))(F_U(C, P(t^\alpha), Q(t^\beta))^\leftarrow(x), F_V(C, P(t^\alpha), Q(t^\beta))^\leftarrow(y)) \\ &= t^{-\theta}(t^\beta Q^\leftarrow(y))^{\theta/\beta} P(P^\leftarrow(x)t^\alpha(t^\beta Q^\leftarrow(y))^{-\alpha/\beta}) \\ &= Q^\leftarrow(y)^{\theta/\beta} P(P^\leftarrow(x)Q^\leftarrow(y)^{-\alpha/\beta}) = C(x, y). \end{aligned} \tag{2.61}$$

Similarly, if $(x, y) \in [0, P(t^\alpha)] \times [0, Q(t^\beta)]$ are such that $P^\leftarrow(x)^\beta \geq Q^\leftarrow(y)^\alpha$, then

$$F(C, P(t^\alpha), Q(t^\beta))(x, y) = \frac{P^\leftarrow(x)^{\theta/\alpha} Q(P^\leftarrow(x)^{-\beta/\alpha} Q^\leftarrow(y))}{t^\theta}. \tag{2.62}$$

Thus,

$$\begin{aligned} &\Phi(C, P(t^\alpha), Q(t^\beta))(x, y) \\ &= F(C, P(t^\alpha), Q(t^\beta))(F_U(C, P(t^\alpha), Q(t^\beta))^\leftarrow(x), F_V(C, P(t^\alpha), Q(t^\beta))^\leftarrow(y)) \\ &= t^{-\theta} P^\leftarrow(x)^{\theta/\alpha} Q(P^\leftarrow(y)^{-\beta/\alpha} Q^\leftarrow(x)) = C(x, y). \end{aligned} \tag{2.63}$$

Hence, for all $(x, y) \in [0, 1]^2$, $\Phi(C, P(t^\alpha), Q(t^\beta))(x, y) = C(x, y)$, i.e. C is invariant on $\mathcal{D} = \{(P(t^\alpha), Q(t^\beta)), t \in (0, 1]\}$. This finishes the proof of Theorem 2.4.6. □

Remark 2.4.7. *From Theorem 2.4.6, we immediately get* $\lim_{t\to 0} \Phi(\Gamma(P, Q, \alpha, \beta, \theta), P(t^\alpha), Q(t^\beta)) = \Gamma(P, Q, \alpha, \beta, \theta)$, *i.e.* $\Gamma(P, Q, \alpha, \beta, \theta)$ *is a LTDC-limit. Further, note that* $\Gamma(g, g, 1, 1, \theta)$ *is precisely the copula in* (2.35).

Example 2.4.8. *The copula* $\Gamma(\mathrm{Id}, \mathrm{Id}, \beta(\alpha + \beta - \alpha\beta)^\leftarrow, \alpha(\alpha + \beta - \alpha\beta)^\leftarrow, 1)$ *is the Marshall and Olkin copula which, because of Theorem 2.4.6, is invariant on*

$$\mathcal{D} = \{(t^{\beta/(\alpha+\beta-\alpha\beta)}, t^{\alpha/(\alpha+\beta-\alpha\beta)}), t \in (0, 1]\} = \{(t^\beta, t^\alpha), t \in (0, 1]\}. \tag{2.64}$$

Similarly, $\Gamma(\mathrm{Id}, \mathrm{Id}, \alpha, \beta, 1)$ *is also the Marshall and Olkin copula with parameters* $(\alpha + \beta - 1/\alpha$ *and* $(\alpha + \beta - 1)/\beta$.

Example 2.4.9. *For $P(x) = 2^{1/\theta}(1+x^{-\theta})^{-1/\theta}$ with $\theta = \alpha+\beta$, the copula $\Gamma(P,P,\alpha,\beta,\theta)$ is the Clayton copula with parameter θ, i.e. for $(x,y) \in [0,1]^2$*

$$\Gamma(P,P,\alpha,\beta,\theta)(x,y) = (x^{-\theta}+y^{-\theta}-1)^{-1/\theta}. \tag{2.65}$$

From Theorem 2.4.6, one has that this copula is invariant on $\mathcal{D} = \{(t^\alpha,t^\beta), t \in (0,1]\}$ for all α,β, i.e. $\Gamma(P,P,\alpha,\beta,\alpha+\beta)$ is invariant on $(0,1]^2$.

Theorem 2.4.10 below characterizes the possible LTDC-limits stating that they coincide with the set of invariant copulae on $(0,1]^2$. In particular, the family $\mathcal{H}(\alpha,\beta,\theta)$ characterizes LTDC-limits on curves $\mathcal{D} = \{(r(t),s(t)), t \in T\}$ provided that the starting copula C belongs to $\mathcal{C}\cap\mathcal{R}(r,s)$ and that r,s are strictly increasing continuous and regularly varying at 0 (Corollary 2.4.11).

Theorem 2.4.10. *If $C \in \mathcal{C}$ and C_0 are copulae such that $\lim_{u,v\to 0} \|\Phi(C,u,v) - C_0\|_\infty = 0$, then C_0 is invariant on the unit square.*

Proof. Let (u_n) and (v_n) be the two sequences defined recursively by the following relationship: Let α and β be two constants in $(0,1]$ with $(\alpha,\beta) \neq (1,1)$ so that, given u_n and v_n strictly positive, $C(u_{n+1},v_n)/C(u_n,v_n) = \alpha$ and $C(u_n,v_{n+1})/C(u_n,v_n) = \beta$ for all $n \geq 1$. Given u_n and v_n, u_{n+1} and v_{n+1}, we have from the continuity of C that are well defined (but not necessarily unique). Those sequences can be defined starting in $(1,1)$ so that $u_1 = \alpha$ and $v_1 = \beta$.
Because $\alpha,\beta \in (0,1]$, we have that $0 \leq u_{n+1} \leq u_n$ and $0 \leq v_{n+1} \leq v_n$. Let $u = \lim_{n\to\infty} u_d$ and $v = \lim_{n\to\infty} v_d$. If $u > 0$ and $v > 0$, then $C(u,v)/C(u,v) = \alpha = \beta$, i.e. $\alpha = \beta = 1$ contradicting the hypothesis $(\alpha,\beta) \neq (1,1)$ meaning that either $u = 0$ or $v = 0$.
Consider the copula $C_d = \Phi(C,u_d,v_d)$. Because of Lemma 2.4.12, it follows that

$$\Phi(C,u_{n+1},v_{n+1}) = \Phi(\Phi(C,u_d,v_d),u^*_{n+1},v^*_{n+1})$$

where u^*_{n+1} and v^*_{n+1} are given by $u^*_{n+1} = C(u_{n+1},v_d)/C(u_d,v_d)$ and $v^*_{n+1} = C(u_d,v_{n+1})/C(u_d,v_d)$ respectively. In other words, we have that $u^*_{n+1} = \alpha$ and $v^*_{n+1} = \beta$, whence

$$\Phi(C,u_{n+1},v_{n+1}) = \Phi(C_d,\alpha,\beta) = C_{n+1}.$$

Because $C_d = \Phi(C,u_n,v_n)$, then, as soon as either $u_n \to 0$ or $v_n \to 0$ when $n \to \infty$, C_n converges towards C_0 when $n \to \infty$. And so, because, given α and β, $\Phi(.,\alpha,\beta)$ is a continuous function, from Lemma 2.4.13, then necessarily, C_0 satisfies $\Phi(C_0,\alpha,\beta) = C_0$. This finishes the proof of Theorem 2.4.10. □

Corollary 2.4.11. *Assume that C satisfies the hypothesis of Theorem 2.3.4 and consider the copula $C_0 = \lim_{t\to 0}\Phi(C,r(t),s(t))$. Then, there is a constant $\theta > 0$ such that $C_0 = \Gamma(\phi_X,\phi_Y,\alpha,\beta,\theta)$ according to (2.53). As a consequence, C_0 is invariant on $\mathcal{D} = \{(\phi_X(t^\alpha),\phi_Y(t^\beta)), t \in (0,1]\}$.*

Proof. Since $C_0 = \lim_{t\to 0}\Phi(C,r(t),s(t))$, then it follows from Theorem 2.3.4 that $C_0(x,y) = \phi(\phi_X^{\leftarrow}(x),\phi_Y^{\leftarrow}(y))$, where for $x > 0$

$$\phi(x,y) = x^{\theta/\alpha}h(yx^{-\beta/\alpha}), \quad h(x) = \begin{cases} \phi_Y(x) & \text{if } x \in [0,1] \\ x^{\theta/\beta}\phi_X(x^{-\alpha/\beta}) & \text{if } x \in (1,\infty) \end{cases}. \tag{2.66}$$

In other words,

$$\begin{aligned} C_0(x,y) =& \phi(\phi_X^{\leftarrow}(x),\phi_Y^{\leftarrow}(y)) \\ =& \begin{cases} \phi_Y^{\leftarrow}(y)^{\theta/\beta}\phi_X(\phi_Y^{\leftarrow}(y)^{-\alpha/\beta}\phi_X^{\leftarrow}(x)) & \text{if } \phi_X^{\leftarrow}(x)^\beta < \phi_Y^{\leftarrow}(y)^\alpha \\ \phi_X^{\leftarrow}(x)^{\theta/\alpha}\phi_Y(\phi_Y^{\leftarrow}(y)\phi_X^{\leftarrow}(x)^{-\beta/\alpha}) & \text{if } \phi_X^{\leftarrow}(x)^\beta \geq \phi_Y^{\leftarrow}(y)^\alpha \end{cases}, \end{aligned} \tag{2.67}$$

i.e. $C_0 = \Gamma(\phi_X, \phi_Y, \alpha, \beta, \theta)$ according to (2.53). This finishes the proof of Corollary 2.4.11. □

The proof of Theorem 2.4.10 is based on the fact that $\Phi(C, u', v')$ can be seen as the LTDC obtained from another LTDC $\Phi(C, u, v)$, where $u \geq u'$ and $v \geq v$ (Lemma 2.4.12). The second ingredient in the proof is the continuity of $\Phi(\cdot, u, v)$ (Lemma 2.4.13). We state these preliminary results below and not only in the proof since we believe they are interesting in their own.

Lemma 2.4.12. *Let $C \in \mathcal{C}$. For $0 \leq u' \leq u \leq 1$ and $0 \leq v' \leq v \leq 1$ we have that*

1. *$\Phi(C, u', v') = \Phi(\Phi(C, u, v), u^*, v^*)$, where u^* and v^* are given by and $u^* = C(u', v)/C(u, v)$ and $v^* = C(u, v')/C(u, v)$ respectively,*
2. *$\Phi(\Phi(C, u, v), u', v') = \Phi(C, u^*, v^*)$ where u^* and v^* satisfy the relations $C(u^*, v) = u'C(u, v)$ and and $C(u, v^*) = v'C(u, v)$ respectively.*

Proof. (i) Let $C^* = \Phi(\Phi(C, u, v), u^*, v^*)$. Because of Lemma 2.4.2, we have for $0 \leq x \leq u^*$ and $0 \leq y \leq v^*$ that

$$\frac{\Phi(C, u, v)(x, y)}{\Phi(C, u, v)(u^*, v^*)} = C^* \left(\frac{\Phi(C, u, v)(x, v^*)}{\Phi(C, u, v)(u^*, v^*)}, \frac{\Phi(C, u, v)(u^*, y)}{\Phi(C, u, v)(u^*, v^*)} \right). \tag{2.68}$$

On the other hand, we have, again using Lemma 2.4.2, that $\Phi(C, u, v)(u^*, v^*)$ equals

$$\Phi(C, u, v)(u^*, v^*) = \Phi(C, u, v) \left(\frac{C(u', v)}{C(u, v)}, \frac{C(u, v')}{C(u, v)} \right) = \frac{C(u', v')}{C(u, v)}. \tag{2.69}$$

Further, $F_U(C, u, v)^{\leftarrow}(u^*) = u'$ and $F_V(C, u, v)^{\leftarrow}(v^*) = v'$ by definition of u^* and v^*. Because,

$$\Phi(C, u, v)(x, y) = \frac{C(F_U(C, u, v)^{\leftarrow}(x), F_V(C, u, v)^{\leftarrow}(y))}{C(u, v)}, \tag{2.70}$$

it follows multiplying (2.69) with (2.70) that

$$\frac{\Phi(C, u, v)(x, y)}{\Phi(C, u, v)(u^*, v^*)} = \frac{C(F_U(C, u, v)^{\leftarrow}(x), F_V(C, u, v)^{\leftarrow}(y))}{C(u', v')}. \tag{2.71}$$

Let $s = F_U(C, u, v)^{\leftarrow}(x)$ and $t = F_V(C, u, v)^{\leftarrow}(y)$, then, substituting into (2.68), we have

$$\frac{C(s, t)}{C(u', v')} = C^* \left(\frac{C(s, v')}{C(u', v')}, \frac{C(u', t)}{C(u', v')} \right) \tag{2.72}$$

for all x, y in $[0, u^*] \times [0, v^*]$. Because C is continuous, $F_U(C, u, v)$ and $F_V(C, u, v)$ are also continuous on $[0, u]$ and $[0, v]$ respectively. Hence, (2.72) holds for all s, t in $[0, u'] \times [0, v']$ because $F_U(C, u, v)^{\leftarrow}(u^*) = u'$ and $F_V(C, u, v)^{\leftarrow}(v^*) = v'$.
Finally, if $0 < u' \leq u \leq 1$ and $0 < v' \leq v \leq 0$, then $\Phi(C, u', v') = \Phi(\Phi(C, u, v), u^*, v^*)$, where u^* are v^* satisfy respectively $u^* = C(u', v)/C(u, v)$ and $v^* = C(u, v')/C(u, v)$.
(ii) Conversely, $C^* = \Phi(\Phi(C, u, v), u', v')$ satisfies, for $0 \leq x \leq u'$ and $0 \leq y \leq v'$

$$\frac{\Phi(C, u, v)(x, y)}{\Phi(C, u, v)(u', v')} = C^* \left(\frac{\Phi(C, u, v)(x, v')}{\Phi(C, u, v)(u', v')}, \frac{\Phi(C, u, v)(u', y)}{\Phi(C, u, v)(u', v')} \right). \tag{2.73}$$

Since

$$\frac{C(x, y)}{C(u, v)} = \Phi(C, u, v) \left(\frac{C(x, v)}{C(u, v)}, \frac{C(u, y)}{C(u, v)} \right), \tag{2.74}$$

we get that for all $x \le u'$ and $y \le v'$ that

$$\begin{aligned}&\frac{C(F_U(C,u,v)^{\leftarrow}(x), F_V(C,u,v)^{\leftarrow}(y))}{C(F_U(C,u,v)^{\leftarrow}(u'), F_V(C,u,v)^{\leftarrow}(v'))}\\ &= C^*\left(\frac{C(F_U(C,u,v)^{\leftarrow}(x), F_V(C,u,v)^{\leftarrow}(v'))}{C(F_U(C,u,v)^{\leftarrow}(u'), F_V(C,u,v)^{\leftarrow}(v'))}, \frac{C(F_U(C,u,v)^{\leftarrow}(u'), F_V(C,u,v)^{\leftarrow}(y))}{C(F_U(C,u,v)^{\leftarrow}(u'), F_V(C,u,v)^{\leftarrow}(v'))}\right)\end{aligned} \tag{2.75}$$

Let $u^* = F_U(C,u,v)^{\leftarrow}(u')$ and $v^* = F_V(C,u,v)^{\leftarrow}(v')$, i.e. u^* and v^* satisfy respectively $C(u^*,v) = u'C(u,v)$ and $C(u,v^*) = v'C(u,v)$. Then, for all $x \le u^*$ and $y \le v^*$

$$\frac{C(x,y)}{C(u^*,v^*)} = C^*\left(\frac{C(x,v^*)}{C(u^*,v^*)}, \frac{C(u^*,y)}{C(u^*,v^*)}\right), \tag{2.76}$$

i.e. $C^* = \Phi(C,u^*,v^*)$ from Sklar's Theorem since the functions $x \mapsto C(x,v^*)/C(u^*,v^*)$ and $y \mapsto C(u^*,y)/C(u^*,v^*)$ are continuous.
Finally, if $0 < u', u \le 1$ and $0 < v', v \le 0$, then $\Phi(\Phi(C,u,v),u',v') = \Phi(C,u^*,v^*)$ where u^* are v^* satisfy respectively $C(u^*,v) = u'C(u,v)$. Moreover, because $C(u^*,v) = u'C(u,v) \le C(u,v)$ and because $x \mapsto C(x,v)/C(u,v)$ is an increasing function, it follows that $u^* \le u$. Similarly, $v^* \le v$, which completes the proof of Lemma 2.4.12. □

Lemma 2.4.13. *For any $u,v \in (0,1]$, the map $\mathcal{C} \to \mathcal{C}$, $C \mapsto \Phi(C,u,v)$ is continuous with respect to the $\|\cdot\|_\infty$-norm.*

Proof. In order to show the continuity of $\Phi(\cdot,u,v)$, we have to bound differences of the form

$$|\Phi(C',u,v)(s,t) - \Phi(C,u,v)(s,t)|, \tag{2.77}$$

where $C, C' \in \mathcal{C}$ and $s,t \in [0,1]$. Since the functions $C(\cdot,v)/C(u,v)$ and $C(u,\cdot)/C(u,v)$ are continuous and take the values 0 and 1 at u, respectively v, we may assume without loss of generality that $s = C(x,v)/C(u,v)$ and $t = C(u,y)/C(u,v)$ for some $(x,y) \in [0,u] \times [0,v]$. Applying Lemma 2.4.2, it follows then

$$\Phi(C,u,v)(s,t) = \frac{C(x,y)}{C(u,v)}. \tag{2.78}$$

Let now $\Delta = C' - C$ and consider

$$\alpha_C(x,y) = \frac{C(x,y)}{C(u,v) + \Delta(u,v)} \quad \text{and} \quad \delta_\Delta(x,y) = \frac{\Delta(x,y)}{C(u,v) + \Delta(u,v)}. \tag{2.79}$$

We obtain that

$$\frac{C'(x,v)}{C'(u,v)} = \frac{C(x,y) + \Delta(x,y)}{C(u,v) + \Delta(u,v)} = \alpha_C(u,v)s + \delta_\Delta(x,v), \tag{2.80}$$

$$\frac{C'(u,y)}{C'(u,v)} = \alpha_C(u,v)t + \delta_\Delta(u,y). \tag{2.81}$$

Thus, using again Lemma 2.4.2, we get

$$\Phi(C',u,v)(\alpha_C(u,v)s + \delta_\Delta(x,v), \alpha_C(u,v)t + \delta_\Delta(u,y)) = \frac{C'(x,y)}{C'(u,v)}. \tag{2.82}$$

Now, the expression in (2.77) can be bounded as follows:

$$
\begin{aligned}
& |\Phi(C',u,v)(s,t)-\Phi(C,u,v)(s,t)| && (2.83)\\
\leq\ & |\Phi(C',u,v)(s,t)-\Phi(C',u,v)(\alpha_C(u,v)s+\delta_\Delta(x,v),\alpha_C(u,v)t+\delta_\Delta(u,y))| \\
\leq\ & +|\Phi(C',u,v)(\alpha_C(u,v)s+\delta_\Delta(x,v),\alpha_C(u,v)t+\delta_\Delta(u,y))-\Phi(C,u,v)(s,t)| \\
\leq\ & |\alpha_C(u,v)s+\delta_\Delta(x,v)-s|+|\alpha_C(u,v)t+\delta_\Delta(u,y)-t|+\left|\frac{C'(x,y)}{C'(u,v)}-\frac{C(x,y)}{C(u,v)}\right|,
\end{aligned}
$$

where the last inequality follows because any copula is Lipschitz-continuous with Lipschitz constant 1 and because of (2.79), (2.82). Further, from the definition of α_C, δ_Δ and because $x \leq u, s \leq 1$, we have that

$$
\begin{aligned}
|\alpha_C(u,v)s+\delta_\Delta(x,v)-s| &\leq \left|\frac{-\Delta(u,v)s}{C(u,v)+\Delta(u,v)}+\frac{\Delta(x,v)}{C(u,v)+\Delta(u,v)}\right| \leq \frac{2|\Delta(u,v)|}{C(u,v)+\Delta(u,v)} \\
&\leq \frac{2\|\Delta\|_\infty}{C(u,v)-\|\Delta\|_\infty}.
\end{aligned}
\tag{2.84}
$$

Similarly,

$$
|\alpha_C(u,v)t+\delta_\Delta(u,y)-t| \leq \frac{2\|\Delta\|_\infty}{C(u,v)-\|\Delta\|_\infty}. \tag{2.85}
$$

Further, since $x \leq u, y \leq v$, we have that

$$
\left|\frac{C'(x,y)}{C'(u,v)}-\frac{C(x,y)}{C(u,v)}\right| = \frac{|\Delta(x,y)C(u,v)-C(x,y)\Delta(u,v)|}{C'(u,v)C(u,v)} \leq \frac{2|C(u,v)\Delta(u,v)|}{C'(u,v)C(u,v)} \leq \frac{2\|\Delta\|_\infty}{C'(u,v)}. \tag{2.86}
$$

From (2.84), (2.84), (2.85) and (2.86), we get

$$
|\Phi(C',u,v)(s,t)-\Phi(C,u,v)(s,t)| \leq \frac{4\|\Delta\|_\infty}{C(u,v)-\|\Delta\|_\infty}+\frac{2\|\Delta\|_\infty}{C'(u,v)}, \tag{2.87}
$$

where the right hand side is independent from s, t and can be made arbitrarily small as $\|\Delta\|_\infty$ becomes small. This finishes the proof of proof of Lemma 2.4.13 □

Remark 2.4.14. *The parameters α, β of the LTDC-limit $\Gamma(P,Q,\alpha,\beta,\theta)$ can be interpreted as parameters describing the direction under which the limit is taken since, as stated in Theorem 2.4.6, $\Gamma(P,Q,\alpha,\beta,\theta)$ is invariant on $\mathcal{D}=\{(P(t^\alpha),Q(t^\beta)), t \in (0,1]\}$. However, such a distribution is not identifiable. In fact, α, β and θ are defined up to a positive multiplicative constant, thus $\Gamma(P,Q,\alpha,\beta,\theta)$ could be defined using two parameters solely. More precisely, for $\eta = \beta/\alpha$,*

$$
\Gamma(P,Q,\alpha,\beta,\theta)=\Gamma(P,Q,1,\eta,\theta)=\Gamma(P,Q,\eta,\theta). \tag{2.88}
$$

Moreover, for all $k > 0$, we have that

$$
\Gamma(P,Q,\eta,\theta)=\Gamma(P_k,Q_k,k\eta,k\theta), \tag{2.89}
$$

where $P_k(x)=P(x^k)$ and $Q_k(x)=Q(x^k)$, $x \in [0,1]$.

We finish this section with a Proposition stating that the only copula which is absolutely continuous and is also invariant on the unit square is the Clayton copula.

Proposition 2.4.15. *The only copula which is absolutely continuous and invariant on $[0,1]^2$ is the Clayton copula.*

Proof. Let C be an absolutely continuous and invariant copula on the unit square. Because of Lemma 2.4.2, we have for all $x, y, u, v \in]0, 1]$ that

$$\frac{C(xu, yv)}{C(u, v)} = C\left(\frac{C(xu, v)}{C(u, v)}, \frac{C(u, yv)}{C(u, v)}\right). \tag{2.90}$$

Since C is absolutely continuous, then derivating with respect to x and y yields

$$\frac{uvC_{12}(xu, yv)}{C(u, v)} = \frac{vC_2(u, yv)}{C(u, v)} \frac{uC_1(xu, v)}{C(u, v)} C_{12}\left(\frac{C(xu, v)}{C(u, v)}, \frac{C(u, yv)}{C(u, v)}\right), \tag{2.91}$$

where C_1, C_2 and C_{12} denote the partial derivatives of C with respect to the relative variables. The latter equation can be written as

$$\frac{C(u, v)C_{12}(xu, yv)}{C_2(u, yv)C_1(xu, v)} = C_{12}\left(\frac{C(xu, v)}{C(u, v)}, \frac{C(u, yv)}{C(u, v)}\right). \tag{2.92}$$

Inserting $x = y = 1$, we obtain that

$$\frac{C(u, v)C_{12}(u, v)}{C_2(u, v)C_1(u, v)} = C_{12}(1, 1) = \theta - 1. \tag{2.93}$$

The latter equation can be rewritten as

$$\frac{C_{12}(u, v)}{C_1(u, v)} = (\theta - 1)\frac{C_2(u, v)}{C(u, v)}. \tag{2.94}$$

Integrating with respect to v leads to

$$\log C_1(u, v) = (\theta - 1) \log C(u, v) + \kappa(u) \tag{2.95}$$

for some function κ of u. In order to determine the function κ, observe that

$$\log C_1(u, 1) = (\theta - 1) \log u + \kappa(u).$$

Substituting into equation (2.95) yields

$$\log \frac{C_1(u, v)}{C_1(u, 1)} = (\theta - 1) \log \frac{C(u, v)}{u}, \tag{2.96}$$

Taking the exponential on both sides produces the identity

$$\frac{C_1(u, v)}{C(u, v)^{\theta - 1}} = \frac{C_1(u, 1)}{u^{\theta - 1}}. \tag{2.97}$$

Integrating with respect to u, we obtain

$$\frac{C(u, v)^{-\theta}}{-\theta} = \frac{C(u, 1)^{-\theta}}{-\theta} + \lambda(v) = \frac{u^{-\theta}}{-\theta} + \lambda(v) \tag{2.98}$$

for some function λ of v. Because of symmetry, it follows that λ does not depend on v, i.e. that

$$\frac{C(u, v)^{-\theta}}{-\theta} = \frac{u^{-\theta}}{-\theta} + \frac{v^{-\theta}}{-\theta} + \text{constant}, \tag{2.99}$$

which can also be written as

$$C(u, v)^{-\theta} = u^{-\theta} + v^{-\theta} + \mathrm{c}, \text{ for all} 0 \leq u, v \leq 1, \tag{2.100}$$

where c is some constant. Finally, because C is a copula, it must be $C(1,1) = 1$, whence the constant in the equation above must be -1, i.e. C is the Clayton copula with parameter θ. Conversely, since the Clayton copula is absolutely continuous and also invariant on $[0,1]^2$, it follows that it is the only copula with this properties. $\square$

Note that this result can easily be extended in higher dimension. Hence, the relationship that should fulfill an invariant copula is that for all $x_1, ..., x_d, u_1, ..., u_d \in (0,1]$,

$$\frac{C(x_1u_1, ..., x_du_d)}{C(u_1, ..., u_d)} = C\left(\frac{C(x_1u_1, u_2..., u_d)}{C(u_1, ..., u_d)}, ..., \frac{C(u_1, ..., u_{d-1}, x_du_d)}{C(u_1, ..., u_d)}\right).$$

Theorem 2.4.16. *The only d-copula which is absolutely continuous and invariant on $[0,1]^d$ is Clayton copula.*

Proof. Since C is absolutely continuous, derivating with respect to x_i and x_j yields

$$\begin{aligned}\frac{u_iu_jC_{i,j}(x_1u_1, ..., x_du_d)}{C(u_1, ..., u_d)} &= \frac{u_iC_i(u_1, .., u_ix_i, ..., u_d)}{C(u_1, ..., u_d)}\frac{u_ijC_j(u_1, .., u_jx_j, ..., u_d)}{C(u_1, ..., u_d)} \\ &\times \; C_{i,j}\left(\frac{C(x_1u_1, u_2..., u_d)}{C(u_1, ..., u_d)}, ..., \frac{C(u_1, ..., u_{d-1}, x_du_d)}{C(u_1, ..., u_d)}\right).\end{aligned}$$

Note that the latter can be written

$$\frac{C(u_1, ..., u_d)C_{i,j}(x_1u_1, ..., x_du_d)}{C_i(u_1, .., u_ix_i, ..., u_d)C_j(u_1, .., u_jx_j, ..., u_d)} = C_{i,j}\left(\frac{C(x_1u_1, u_2..., u_d)}{C(u_1, ..., u_d)}, ..., \frac{C(u_1, ..., u_{d-1}, x_du_d)}{C(u_1, ..., u_d)}\right).$$

Inserting $x_1 = ... = x_d = 1$, we get

$$\frac{C(u_1, ..., u_d)C_{i,j}(u_1, ..., u_d)}{C_i(u_1, .., u_i, ..., u_d)C_j(u_1, .., u_j, ..., u_d)} = C_{i,j}\left(1, 1, ..., 1\right).$$

Set $\theta_{i,j} = C_{i,j}\left(1, 1, ..., 1\right) + 1$, so that the equation above can be written

$$\frac{C(u_1, ..., u_d)C_{i,j}(u_1, ..., u_d)}{C_i(u_1, .., u_i, ..., u_d)C_j(u_1, .., u_j, ..., u_d)} = \theta_{i,j} - 1.$$

The latter equation can be rewritten as

$$\frac{C_{i,j}(\boldsymbol{u})}{C_i(\boldsymbol{u})} = (\theta - 1)\frac{C_j(\boldsymbol{u})}{C(\boldsymbol{u})}, \text{ for all } \boldsymbol{u} \in (0,1]^d. \tag{2.101}$$

Integrating with respect to u_j leads to

$$\log C_j(\boldsymbol{u}) = (\theta_{i,j} - 1)\log C(\boldsymbol{u}) + \kappa(\boldsymbol{u}) \tag{2.102}$$

for some function κ of $\boldsymbol{u}$. In order to determine the function κ, observe that

$$\log C_i(1, ..., 1, u_i, 1, ..., 1) = (\theta_{i,j} - 1)\log u_i + \kappa(\boldsymbol{u}).$$

Substituting into equation (2.102) yields

$$\log\frac{C_i(u_1, ..., u_{i-1}, u_i, u_{i+1}, ..., u_d)}{C_i(1, ..., 1, u_i, 1, ..., 1)} = (\theta_{i,j} - 1)\log\frac{C(\boldsymbol{u})}{u_i}, \tag{2.103}$$

Taking the exponential on both sides produces the identity

$$\frac{C_i(\boldsymbol{u})}{C(\boldsymbol{u})^{\theta_{i,j}-1}} = \frac{C_i(1, ..., 1, u_i, 1, ..., 1)}{u_i^{\theta_{i,j}-1}}. \tag{2.104}$$

Integrating with respect to u_i, we obtain

$$\begin{aligned}\frac{C(u_1,...,u_i,...u_d)^{-\theta_{i,j}}}{-\theta_{i,j}} &= \frac{C(1,...,1,u_i,1,...,1)^{-\theta_{i,j}}}{-\theta_{i,j}} + \kappa(u_1,...,u_{i-1},u_{i+1},...,u_d) \\ &= \frac{u_i^{-\theta_{i,j}}}{-\theta_{i,j}} + \kappa(u_1,...,u_{i-1},u_{i+1},...,u_d)\end{aligned}$$

for some function κ of $(u_1,...,u_{i-1},u_{i+1},...,u_d)$. Since those term do not depend on j, using symmetry arguments, $\theta_{i,j} = \theta_i$. Using again some symmetry properties, it follows that necessarily the θ_i's have to be equal. Hence, $\theta_i = \theta$ for all $i = 1,2,...,n$, and

$$\frac{C(u_1,...,u_d)^{-\theta}}{-\theta} = \frac{u_1^{-\theta}}{-\theta} + ... + \frac{u_d^{-\theta}}{-\theta} + \text{constant}, \tag{2.105}$$

which can also be written as

$$C(u_1,...,u_d)^{-\theta} = u_1^{-\theta} + ... + u_d^{-\theta} + c, \text{ for all } u_1,...,u_d \in (0,1], \tag{2.106}$$

where c is some constant. Finally, because C is a copula, it must satisfy $C(1,1,...,1) = 1$, whence the constant in the equation above must be $-(d-1)$, i.e. C is the Clayton copula with parameter θ. Conversely, since the Clayton copula is absolutely continuous and also invariant on $[0,1]^d$, it follows that it is the only copula with this properties. □

Remark 2.4.17. *Note that the upper Fréchet-Hoeffding bound is not a an absolutely continuous copula, but it is an invariant copula.*

2.5 An application to credit risk

The main risk drivers of almost all credit derivatives such as e.g. *Credit Default Swap baskets* (CDS baskets) or *first-to-default* contract types are given by the relevant default times. Among the most popular (univariate) default time models we find intensity-based ones. As shown by Schönbucher and Schubert (2001) a copula approach allows to model naturally arbitrary dependence structures in such an intensity-based framework.

In this section we first review the setup of Schönbucher and Schubert (2001) and we then show how our LTDC-limits can be used as dependence structures for credit stress scenarios.

2.5.1 Intensity-based default models

For σ-algebras $\mathcal{A},\mathcal{B}$ with $\mathcal{A} \subset \mathcal{B}$ and for a set $B \in \mathcal{B}$, we will use in the sequel the notation $\mathcal{A} \wedge B = \{A \cap B, A \in \mathcal{A}\}$. Further, all filtrations are supposed to satisfy the usual conditions, i.e. they are assumed to be right continuous and such that the smallest σ-filed of the filtration is trivial. Finally, for a review of point process intensities we refer to **?**.

Schönbucher and Schubert (2001) propose the following intensity-based default model which we recall in the two-dimensional case. Let λ_i, $i = 1,2$ be non-negative càdlàg processes adapted to a filtration $(\mathcal{G}_t)_{t\geq 0}$ representing the general market information except explicit information on the occurrence of defaults. For U_1, U_2 standard uniformly distributed random variables, which are assumed to be independent from $\mathcal{G}_\infty = \cup_{t\geq 0}\mathcal{G}_t$, we define the default times as the random variables

$$\tau_i = \inf\{t > 0, \gamma_i(t) \leq U_i\}, \quad i = 1,2, \tag{2.107}$$

where $\gamma_i(t) = \exp(-\Lambda_i(t))$ is called *countdown processes* and $\Lambda_i(t) = \int_0^t \lambda_i(s)\,ds$. Note that, conditioned on $\mathcal{G}_\infty$, we have that

$$\mathbb{P}(\tau_1 \le t_1, \tau_2 \le t_2 | \mathcal{G}_\infty) = C^*(\gamma_1(t_1), \gamma_2(t_2)), \tag{2.108}$$

where C is the distribution function of (U_1, U_2). Thus, we see that defining default times as in (2.107) implies that, given general market information, the default dependence mechanism is completely described by C.

Remark 2.5.1. *The motivation behind* (2.107) *comes from the fact that, for a Cox process with intensity λ, the time τ of the first jump can be written*

$$\tau = \inf\left\{ t > 0 \middle| \int_0^t \lambda(s)\,ds \ge Z \right\}, \tag{2.109}$$

where Z is exponentially distributed with parameter 1 *(see Lando 1998).*

In general, the intensity of a point process depends on the information which is conditioned on. Denoting by N_i the default counting process of counterparty $i = 1, 2$ and by $\mathcal{F}_t^i$ the augmented filtration of $\sigma(N_i(s); 0 \le s \le t)$, we have that λ_i is the $\mathcal{F}_t^i$-intensity of N_i. However it is in the spirit of any multivariate model also consists in considering the information relative to the other counterparties such as the one given by C and $\mathcal{H}_t = \vee_{i=1,2}(\mathcal{F}_t^i \vee \mathcal{G}_t)$, $t \ge 0$. Indeed, we find in Schönbucher and Schubert (2001) that the $\mathcal{H}_t$-intensity h_i of N_i equals to

$$h_i(t) = \lambda_i(t) \cdot \gamma_i(t) \cdot \partial_i \log(C(\gamma_1(t), \gamma_2(t))). \tag{2.110}$$

Because of the term $\partial_i \log(C(\gamma_1(t), \gamma_2(t)))$, the intensity of a single counterparty is also affected by the dependence structure of the several counterparties. In the case where U_1, U_2 are independent, i.e. whenever $C = C^\perp$, we have that the right hand side of (2.110) reduces to $\lambda_i(t)$, i.e. to the $\mathcal{F}_t^i$-intensity of N_i. Further, under the additional information that the other obligor has already defaulted, i.e. $\{\tau_j = t_j\}$, $j \ne i$, $t_j > 0$, the default intensity of the survived counterparty takes the form

$$h_i^{-j}(t) = \lambda_i(t) \cdot \gamma_i(t) \cdot \frac{\partial_{ij} C(\gamma_1(t), \gamma_2(t))}{\partial_j C(\gamma_1(t), \gamma_2(t))}. \tag{2.111}$$

A special case of (2.110) and (2.111) is given by C equal to the Clayton copula with parameter θ of Example 2.4.9. In that case,

$$h_i(t) = \left(\frac{C(\gamma_1(t), \gamma_2(t))}{\gamma_i(t)} \right)^\theta \lambda_i(t) \quad \text{and} \quad h_i^{-j}(t) = (1+\theta) h_i(t). \tag{2.112}$$

As stated in Schönbucher and Schubert (2001), such a dependence structure reflects one of the main features of a model introduced by Davis and Lo ((2001), (1999b)), where knowledge of one obligor's default determines a jump in the spread of the other obligor by a factor $(1 + \theta)$.

2.5.2 Dependence structures for stress scenarios

Stress scenarios for default times arise in many different situations. For example, pension funds have to invest only in investment grade bonds because of regulatory reasons. Thus, a default (or downgrade) of a bond in the pension fund's portfolio determines the replacement of that bond, whence a possible (large) losses due to the bond's value decrease. Another example is given by first-to-default CDS baskets where in the case of an *early* default the protection seller receives the premium only for a short time but has to deliver the underlying very soon.

More generally, knowing or modelling the dependence structure of the several default times and in particular the joint behavior under averse market conditions, avoids risk underestimation allowing thus for a risk-adjusted pricing (for instance of credit derivatives). Such stress situations can be described by conditional distributions of the type

$$\mathbb{P}(\tau_1 \leq t_1, \tau_2 \leq t_2 | \mathcal{G}_\infty \wedge \{\tau_1 \leq T, \tau_2 \leq T\}), \tag{2.113}$$

as T tends to zero. Since the conditional distribution of τ_i given $\mathcal{H}_t^i$ equals $\gamma_i(t)$, it follows from Proposition 2.2.3 and Equation (2.108) that the copula relative to the conditional distribution in (2.113) is given by

$$\Phi(C^*, 1 - \gamma_1(T), 1 - \gamma_2(T)), \tag{2.114}$$

where C^* is the survival copula of C.

Example 2.5.2 (First-to-default)**.** *The conditional distribution of the first-to-default time $\tau = \tau_1 \wedge \tau_2$ conditioned on $\mathcal{G}_\infty \wedge \{\tau_1 \leq T, \tau_2 \leq T\}$ is given for $t \leq T$ by*

$$\begin{aligned} \mathbb{P}(\tau \leq t | \mathcal{G}_\infty \wedge \{\tau_1 \leq T, \tau_2 \leq T\}) &= 1 - \mathbb{P}(\tau_1 > t, \tau_2 > t | \mathcal{G}_\infty \wedge \{\tau_1 \leq T, \tau_2 \leq T\}) \\ &= 1 - C^*(1 - \gamma_1(t), 1 - \gamma_2(t)), \end{aligned} \tag{2.115}$$

where C^* is the survival copula of $\Phi(C^*, 1 - \gamma_1(T), 1 - \gamma_2(T))$.

Suppose now that λ_i is regularly varying at 0 with parameter $\delta_i \geq 0$ which, as it is easy to check, implies that $1 - \gamma_i \in \mathcal{R}^0_{1+\delta_i}$. Further, assume $C^* \in \mathcal{C} \cap \mathcal{R}(1 - \gamma_1, 1 - \gamma_2)$ with limiting function ϕ. Then, because of Corollary 2.4.11, there is a constant $\theta > 0$ such that

$$\lim_{T \to 0} \Phi(C^*, 1 - \gamma_1(T), 1 - \gamma_2(T)) = \Gamma(\phi_X, \phi_Y, 1 + \delta_1, 1 + \delta_2, \theta). \tag{2.116}$$

As a special case, we have for $\gamma_1 = \gamma_2 = \gamma$ and $\delta_1 = \delta_2 = 0$ that

$$\lim_{T \to 0} \Phi(C^*, 1 - \gamma(T), 1 - \gamma(T)) = \Gamma(g, g, 1, 1, \theta), \quad g = \phi_X, \tag{2.117}$$

which corresponds to the limiting copula (2.35) of Remark 2.3.8.

As we already mentioned at the end of Section 2.3, a special case of Theorem 2.3.4 is given by the situation where the starting copula is archimedean with a regularly varying generator. In this case, the LTDC-limit on the diagonal is the Clayton copula. Thus, the Davis-Lo-model can be seen as stress-scenario one.

Chapter 3

Lower tails for Archimedean copulae

3.1 Introduction and motivations

In risk management the choice of an appropriate dependence structure plays a crucial role (to estimate capital needed for hedging risks, or to price standard multiline products). For an overview of some of the recent developments and applications, in finance or actuarial science, we refer to Bäuerle and Müller (1987), Frees and Valdez (1998), Klugman and Parsa (1999) or the monograph of Denuit, Dhaene, Goovaerts and Kaaas (2005) and the references therein. Within the large set of copulae, it might be more convenient to restrict the study to the family of Archimedean copulae, introduced in Kimberling (1974) and intensively since; see for instance Genest and MacKay (1986a), Genest and Rivest (1993) or Müller and Scarsini (2004) among others. And because of the crucial importance of extremal events in insurance or finance, it became primordial to have a better understanding of the behavior of copulae in tails.

In Juri and Wüthrich (2003), tail dependence for bivariate Archimedean copulae is described using the concept of lower tail dependence copulae. The lower tail dependence copula of a copula C at level $0 < u < 1$ is defined as the copula of the conditional distribution of a random pair (U, V) with distribution function C when conditioned to be contained in the square $[0, u]^2$. If C is Archimedean, then the lower tail dependence copulae obtained from C must be Archimedean as well, and their generators admit simple expressions in terms of the generator of C.

The central topic in Juri and Wüthrich (2003) is the asymptotic behavior of the lower tail dependence copula of a strict Archimedean copula as the threshold u decreases to zero. The main result is that, under regularity conditions, the only possible limit of the lower tail dependence copula is the Clayton copula, the parameter of the latter being determined by the index of regular variation of the generator of the Archimedean copula at zero. The key rule of the Clayton copula was also mentioned in Charpentier (2004) or Bassan and Spizzichino (2004).

If section 3.2 will briefly recall some notations and results on Archimedean copulae, section 3.3 will extend some results of Genest and MacKay (1986b) on limiting behavior for a sequence of Archimedean copulae. Since Archimedean copulae are stable by truncature, those results will then be used to derive properties of lower tails for Archimedean copulae.

The topic of section 3.4 is the boundary case when the generator of the Archimedean copula is regularly varying at 0. In Theorem 3.5 in Juri and Wüthrich (2003), it is claimed that if C is a strict Archimedean copula whose generator is differentiable and regularly varying at 0, then the lower tail dependence copula C_u converges pointwise to the Clayton copula, including the two limiting case: tail independence if the generator is slowly varying, and tail comonotonicity if the generator is rapidly varying.

In section 3.4.1, we will study some necessary and sufficient conditions for Archimedean copulae to have some tail behavior. As we shall see, the case of tail independence is slightly different from the condition stated in Theorem 3.5 in Juri and Wüthrich (2003). In Section 3.4.2, we give an example of a strict Archimedean copula whose generator is continuously differentiable and slowly varying at the origin and such that the tail dependence copula C_u does not converge to the independence copula as u decreases to zero. The problem in Theorem 3.5 in Juri and Wüthrich (2003) seems to come from Lemma 3.4 in the same paper, which is shown by the same counterexample to be incorrect as well. Fortunately, the result can be fixed by imposing a stronger condition on the generator involving the de Haan's class Π, or equivalently, by assuming regular variation of the derivative of the generator, see Section 3.4.3.

And finally, in section 3.5, we will extend results obtained in section 3.4 to higher dimension. Again, tail dependence (section 3.5.1) and tail independence (section 3.5.2) be be considered separately.

3.2 Definitions and preliminaries

3.2.1 Bivariate Archimedean copulae

A function $C : [0,1]^2 \to [0,1]$ is called a *bivariate copula* if it is the restriction to $[0,1]^2$ of a bivariate distribution function whose marginals are given by the uniform distribution on the interval $[0,1]$. A function $\psi : [0,1] \to [0,\infty]$ is called a *strict generator* if it is decreasing, convex, $\psi(0) = \infty$ and $\psi(1) = 0$. The inverse function of a strict generator ψ is denoted by $\psi^{\leftarrow}$. A function $C : [0,1]^2 \to [0,1]$ is called a *strict Archimedean copula* if there exists a strict generator ψ such that

$$C(u,v) = \psi^{\leftarrow}\{\psi(u) + \psi(v)\}, (u,v) \in [0,1]^2.$$

Note that the generator is unique up to a multiplicative constant. A strict Archimedean copula is a copula. See the survey monograph by Nelsen (1999) and the references therein for more details.

3.2.2 Archimedean copulae in higher dimension

In higher dimension, as pointed out in section 1.5, some assumptions should added in order to define a proper copula. As pointed out in Nelsen, Quesada Molina, Rodríguez-Lallena and Úbeda-Flores (2002), in dimension $d > 2$, a generator which is simply decreasing and convex generates a quasi-copula, not necessarily a copula.

Hence, a function $\psi : [0,1] \to [0,\infty]$ is called a *generator* of order d if the following conditions hold:

- ψ is decreasing and $\psi(1) = 0$;
- the generalized inverse, $\psi^{\leftarrow} : [0,\infty] \to [0,1]$, of ψ, defined by

  $$\psi^{\leftarrow}(t) = \inf\{u \in [0,1] \mid \psi(u) \le t\} \text{ for all } t \in [0,\infty],$$

  is $d-2$ times continuously differentiable on $(0,\infty)$;
- the function $(-D)^{(d-2)}\psi^{\leftarrow}$ is convex.

The generator ψ is called *strict* if $\psi(0) = \infty$.

Under those assumption, a d-variate copula, C, is called *Archimedean* if there exists a generator, ψ, of order d such that

$$C(u_1, \dots, u_d) = \psi^{\leftarrow}\{\psi(u_1) + \cdots + \psi(u_d)\},$$

for all $(u_1, \dots, u_d) \in [0,1]^d$.

Characterization of generators in dimension $d \ge 2$ will be considered in Section 5.2.

3.2.3 Lower tail conditional copulae

Let C be a copula and let (U, V) be a random pair with joint distribution function C. Let $0 < u < 1$ be such that $C(u, u) > 0$. The *lower tail dependence copula relative to C at level u* is defined as the copula, C_u, of the joint distribution of (U, V) conditionally on the event $\{U \leq u, V \leq u\}$. Formally,

$$C_u(x, y) = \frac{C(x', y')}{C(u, u)}$$

where $0 \leq x' \leq u$ and $0 \leq y' \leq u$ are the solutions to the equations

$$\frac{C(x', u)}{C(u, u)} = x \text{ and } \frac{C(u, y')}{C(u, u)} = y;$$

see Definition 3.1 in Juri and Wüthrich (2003) or Definition 2.2 in Juri and Wüthrich (2004). *Upper tail dependence copulae* are defined in a similar way (see Definition 2.1 in Juri and Wüthrich (2004). Moreover, the definition can be extended by allowing the thresholds for the two margins to be different, that is, by conditioning on the event $\{U \leq u, V \leq v\}$, where $(u, v) \in (0, 1]^2$ are such that $C(u, v) > 0$, see Definition 2.5 in Charpentier (2004). In this note, we will be interested only in the diagonal.

If C is a strict bivariate Archimedean copula with generator ψ, then the lower tail dependence copula relative to C at level u is given by the strict Archimedean copula with generator ψ_u defined by

$$\psi_u(t) = \psi(tv) - \psi(v), 0 \leq t \leq 1, \tag{3.1}$$

where $v = v(u) = \psi^{\leftarrow}\{2\psi(u)\}$ (Proposition 3.2 in Juri and Wüthrich (2003)).

Remark 3.2.1. *Note that ψ_u is obtained using some translations of the original generator. On Figure 3.1,* (1) *is the original generator of the copula. Consider the restriction of the generator at* $(0, C(u, u)]$*, and an homothetic transformation so that its supports becomes* $[0, 1]$*, as in* (2)*. The equation of this curve is $t \mapsto \psi(t \cdot C(u, u))$ where $t \in (0, 1]$. Consider then a simple translation , so that this function becomes null in* 1*, i.e. $t \mapsto \psi(t \cdot C(u, u)) - \psi(C(u, u))$, as in* (3)*. Hence, the idea is to consider the restriction of ψ on the support* $(0, C(u, u)]$*, and to use homothetic transformations so that the support becomes* $(0, 1]]$*, an a translation to be null in* 1*, and so that ψ_u satisfies properties of an Archimedean generator.*

Since $v(u) \to 0$ as $u \to 0$, the asymptotic behavior of the lower tail dependence copula C_u as $u \to 0$ depends on the asymptotic behavior of ψ near the origin.

3.2.4 Regular variation and de Haan theory

A useful concept now is that of regular variation: A positive, measurable function f defined in a right-neighbourhood of zero is said to be *regularly varying at zero of index $\tau \in \mathbb{R}$* if

$$\lim_{u \to 0} \frac{f(ux)}{f(u)} = x^{\tau}, \qquad 0 < x < \infty,$$

with notation $f \in \mathcal{R}_\tau$ (or $f \in \mathcal{R}_\tau^0$ to specify that regular variation is considered at origin). If $\tau = 0$, then the limit is equal to one for all $0 < x < \infty$; in this case, f is said to be *slowly varying at zero.* A limiting case is obtained when $\tau = -\infty$: f is said to be *rapidly varying at zero of index $-\infty$*, notation $f \in \mathcal{R}_{-\infty}$, if

$$\lim_{u \to 0} \frac{f(ux)}{f(u)} = \begin{cases} 0 & \text{if } 1 < x < \infty, \\ 1 & \text{if } x = 1, \\ \infty & \text{if } 0 < x < 1. \end{cases}$$

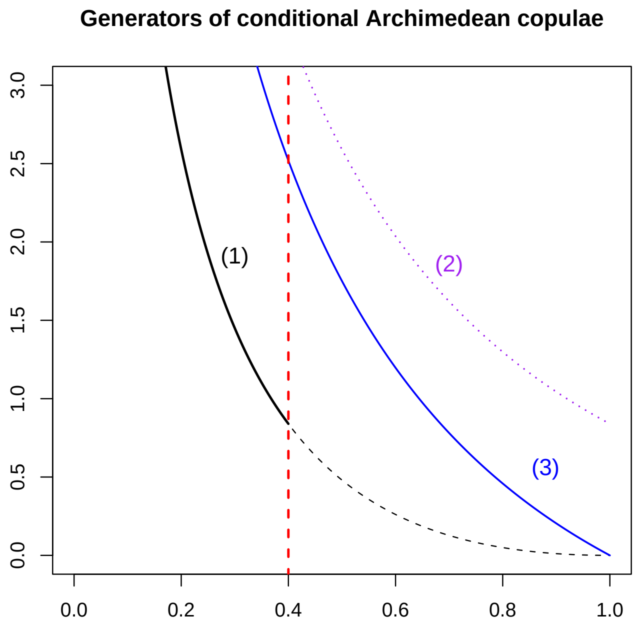


Figure 3.1: Geometric interpretation of the generator of the conditional copula.

Classically, regular variation is considered at infinity rather than at zero. However, it is typically straightforward to translate results from regular variation at infinity to regular variation at zero by considering the function $y \mapsto f(1/y)$ (see for instance Bingham, Goldie and Teugels (1987)).

Regular and slow variation

The definition of regular variation involves in principle an infinite set of limit relations. However, if a function is known to be convex, then regular variation of the function is equivalent to a single limit relation. Results of this type are known under the name "Monotone Density Theorem," see for instance section 1.7.3 in Bingham, Goldie and Teugels (1987). We will need the following two instances.

Lemma 3.2.2. *Let f be a positive, convex function of a real variable defined in a right-neighbourhood of zero. Let Df be a nondecreasing version of the Radon-Nikodym derivative of f. The function f is regularly varying at zero of index $\tau \in [-\infty, \infty]$ if and only if*

$$\lim_{s \to 0} \frac{sDf(s)}{f(s)} = \tau.$$

Proof. Let c be a positive number such that the domain of f includes the interval $(0, c]$. The function $\log f$ is absolutely continuous with Radon-Nikodym derivative $(Df)/f$. Denote $\tau(s) = sDf(s)/f(s)$. For $0 < s \leq c$, we have

$$f(s) = f(c) \exp\left(- \int_s^c \tau(t) \frac{\mathrm{d}t}{t} \right).$$

If additionally $0 < x < \infty$ with $x \neq 1$ and if s is such that also $sx \leq c$, then

$$\begin{aligned}\frac{f(sx)}{f(s)} &= \exp\left(\int_s^{sx} \tau(t)\frac{\mathrm{d}t}{t}\right) \\ &= \exp\left(\int_1^{x} \tau(st)\frac{\mathrm{d}t}{t}\right).\end{aligned}$$

The argument of the exponent converges to $\tau \log(x)$ as $s \to 0$. Hence indeed $f(sx)/f(s) \to x^\tau$ as $s \to 0$, as required.

Conversely, suppose that f is regularly varying at zero of index τ. By convexity, we have for all $0 < x < \infty$ and all sufficiently small s,

$$f(sx) - f(s) \geq s(x-1)Df(s).$$

If x is not equal to one, we can divide both sides of this inequality by $(x-1)$ and let s decrease to zero to get

$$\begin{aligned}\limsup_{s\to 0} \frac{sDf(s)}{f(s)} &\leq \frac{x^\tau - 1}{x-1}, \qquad \text{for all } 1 < x < \infty, \\ \liminf_{s\to 0} \frac{sDf(s)}{f(s)} &\geq \frac{x^\tau - 1}{x-1}, \qquad \text{for all } 0 < x < 1.\end{aligned}$$

Since $(x^\tau - 1)/(x-1) \to \tau$ as $x \to 1$ for all $\tau \in [-\infty, \infty]$, we conclude that $sDf(s)/f(s) \to \tau$ as $s \to 0$. □

Lemma 3.2.3. *Let f be a positive, convex function of a real variable defined in a neighbourhood of infinity. Let Df be a nondecreasing version of the Radon-Nikodym derivative of f. The function f is regularly varying at infinity of index $\tau \in [-\infty, \infty]$ if and only if*

$$\lim_{t\to\infty} \frac{tDf(t)}{f(t)} = \tau.$$

Proof. The proof of Lemma 5.1.4 is identical to the proof of Lemma 5.1.3. □

Lemma 3.2.4. *Let f be a positive, $k \geq 0$ times continuously differentiable function of a real variable defined in a neighbourhood of infinity. Assume that $(-D)^k f$ is convex and that $f(t) \to 0$ as $t \to \infty$. If f is regularly varying at infinity of index $-\tau \in [-\infty, 0]$, then for all integer $j = 1, \ldots, k+1$,*

$$\lim_{t\to\infty} \frac{t^j(-D)^j f(t)}{f(t)} = \tau(\tau+1)\cdots(\tau+j-1). \tag{3.2}$$

Proof. We proceed by induction on k. In case $k = 0$, the statement is trivially implied by Lemma 5.1.4.

So assume k is a positive integer. Note that by Lemma 5.1.1, the function $(-D)^j f$ is convex for every $j = 0, 1, \ldots, k$ and vanishes at infinity. Hence, by the induction hypothesis, (5.1) holds already for all $j = 1, \ldots, k$, so only the case $j = k+1$ remains to be shown.

First consider the case $0 < \tau < \infty$. Then we know that

$$(-D)^k f(t) \sim \tau(\tau+1)\cdots(\tau+k-1)t^{-k}f(t) \text{ as } t \to \infty.$$

In particular, the function $(-D)^k f$ is regularly varying at infinity of order $-\tau - k$. Apply Lemma 5.1.4 to get

$$(-D)^{k+1} f(t) \sim (\tau + k) t^{-1} (-D)^k f(t) \text{ as } t \to \infty.$$

Combine the two previous displays to see that (5.1) also holds for $j = k + 1$.

Next consider the case $\tau = 0$. Then we know that

$$(-D)^k f(t) = o\{t^{-k} f(t)\} \text{ as } t \to \infty.$$

Since $(-D)^k f$ is convex and since $(-D)^{k+1} f$ is nonnegative,

$$(-D)^k f(t/2) - (-D)^k f(t) \geq (t/2)(-D)^{k+1} f(t) \geq 0.$$

Combine the two previous displays with the fact that $f(t/2) \sim f(t)$ as $t \to \infty$ to see that $t^{k+1}(-D)^{k+1} f(t)/f(t) \to 0$ as $t \to \infty$.

Finally, consider the case $\tau = \infty$. For large enough t, we have by induction on k,

$$f(t) = \int_t^\infty \frac{(v-t)^k}{k!} (-D)^{k+1} f(v) \mathrm{d}v.$$

Let $1 < x < \infty$. For $v \geq tx^2$, we have $v - t \leq 2(v - tx)$ and thus

$$\begin{aligned} f(t) &\leq \int_t^{tx^2} \frac{(v-t)^k}{k!} (-D)^{k+1} f(v) \mathrm{d}v + \int_{tx^2}^\infty \frac{2^k (v-tx)^k}{k!} (-D)^{k+1} f(v) \mathrm{d}v \\ &\leq \frac{t^{k+1}(x^2-1)^{k+1}}{(k+1)!} (-D)^{k+1} f(t) + 2^k f(tx). \end{aligned}$$

Since $f(tx)/f(t) \to 0$ as $t \to \infty$, we find

$$\liminf_{t \to \infty} \frac{t^{k+1}(-D)^{k+1} f(t)}{f(t)} \geq \frac{(k+1)!}{(x^2-1)^{k+1}}.$$

Let $x \to 1$ to see that $t^{k+1}(-D)^{k+1} f(t)/f(t) \to \infty$ as $t \to \infty$. □

De Haan theory, and second order properties

The property that a function is slowly varying or rapidly varying is sometimes not informative enough. Versatile subclasses are the function classes Π and Γ due to L. de Haan and studied extensively in Bingham, Goldie and Teugels (1987), chapter 3. These classes turn up for instance in the study of the max-domain of attraction of the Gumbel distribution.

In the presence of convexity, the theory simplifies very much and can be reduced to ordinary regular variation. The following two lemmas describe the core theory of the classes Π and Γ restricted to convex functions. The proofs rely only on elementary properties of regularly varying functions, in particular the Uniform Convergence Theorem, see for instance Bingham, Goldie and Teugels (1987), Theorem 1.5.2.

Lemma 3.2.5. *Let f be a convex, decreasing, and positive function defined in a right-neighbourhood of zero. Let f' be a negative and nondecreasing version of the Radon-Nikodym derivative of f. The following statements are equivalent:*

(i) The function $-f'$ is regularly varying at zero of index -1.

(ii) If $x(\cdot)$ is a positive function defined in a right-neighbourhood of zero such that $sx(s) \to 0$ and $x(s) \to x \in [0, \infty]$ as $s \to 0$, then

$$\lim_{s \to 0} \frac{f(sx(s)) - f(s)}{sf'(s)} = \log(x).$$

(iii) There exists a positive function g defined in a neighbourhood of infinity such that

$$\lim_{s \to 0} \frac{f(sx) - f(s)}{g(s)} = -\log(x), \text{ for all } 0 < x < \infty.$$

In this case, $sf'(s)/f(s) \to 0$ as $s \to 0$ and f is slowly varying at zero.

Proof. (i) implies (ii). We have

$$f(sx(s)) - f(s) = \int_s^{sx(s)} f'(u)du = s \int_1^{x(s)} f'(su)du,$$

and thus

$$\frac{f(sx(s)) - f(s)}{sf'(s)} = \int_1^{x(s)} \frac{f'(su)}{f'(s)} du.$$

If $0 < x < \infty$, then by the Uniform Convergence Theorem, the right-hand side of the previous equation converges to $\int_1^x u^{-1}du = \log(x)$. If $x = 0$ or $x = \infty$, then by Fatou's Lemma, the right-hand side of the previous display converges to $-\infty$ or $+\infty$, respectively.

(ii) implies (iii). Trivial.

(iii) implies (i). Since $g(s) \sim f(se^{-1}) - f(s)$ as $s \to 0$, it is no loss of generality to assume that g is measurable. The function g is necessarily slowly varying at zero. To see why, pick $1 \neq x \in (0, \infty)$ and let λ be a limit point in $[0, \infty]$ of $g(sx)/g(s)$ as $s \to 0$. Since

$$\frac{f(sx^2) - f(s)}{g(s)} = \frac{f(sx^2) - f(sx)}{g(sx)} \cdot \frac{g(sx)}{g(s)} + \frac{f(sx) - f(s)}{g(s)}$$

we must have

$$-\log(x^2) = -\log(x)\lambda - \log(x)$$

whence $\lambda = 1$, confirming that g is slowly varying at zero.

Since f is convex, we have for all $0 < x < \infty$ and all sufficiently small, positive s,

$$f(sx) - f(s) \geq s(x - 1)f'(s).$$

For $1 < x < \infty$, this yields

$$\limsup_{s \to 0} \frac{sf'(s)}{g(s)} \leq \frac{-\log(x)}{x - 1},$$

while for $0 < x < 1$, we get

$$\liminf_{s \to 0} \frac{sf'(s)}{g(s)} \geq \frac{-\log(x)}{x - 1}.$$

Since $\log(x) \sim x - 1$ as $x \to 1$, we find $-sf'(s) \sim g(s)$ as $s \to 0$. Since g is slowly varying, it now follows that $-f'$ is regularly varying of index -1.

It remains to establish the final claim in the lemma. Let g be as in (iii); for instance $g(s) = -sf'(s)$. Let M be a positive constant. Since

$$\begin{aligned} 0 \leq \frac{f(s\exp(M))}{g(s)} &= \frac{f(s)}{g(s)} + \frac{f(s\exp(M)) - f(s)}{g(s)} \\ &= \frac{f(s)}{g(s)} - M + o(1) \text{ as } s \to 0, \end{aligned}$$

we must have $\liminf_{s\to 0} f(s)/g(s) \geq M$. Hence $g(s)/f(s) \to 0$ as $s \to 0$. But then, for every $0 < x < \infty$,

$$\frac{f(sx)}{f(s)} - 1 = \frac{g(s)}{f(s)} \cdot \frac{f(sx) - f(s)}{g(s)} \to 0, \text{ as } s \to 0$$

$\square$

Lemma 3.2.6. *Let f be a convex, decreasing, and positive function defined in a right-neighbourhood of zero. Assume $f(0) = \infty$ and let $f^{\leftarrow}$ be the inverse function of f, defined in a neighbourhood of infinity. Define $\phi(t) = -f^{\leftarrow}(t) f'\{f^{\leftarrow}(t)\}$ with f' a negative and nondecreasing version of the Radon-Nikodym derivative of f. The conditions (i)–(iii) in Lemma 3.2.5 are equivalent to each of the following ones:*

(iv) The function ϕ is self-neglecting, that is, $\phi(t) = o(t)$ as $t \to \infty$ and

$$\lim_{t\to\infty} \frac{\phi\{t + x\phi(t)\}}{\phi(t)} = 1$$

locally uniformly in $x \in \mathbb{R}$.

(v) We have $\phi(t) = o(t)$ as $t \to \infty$, and if $x(\cdot)$ is a real function defined in the neighbourhood of infinity such that $x(t) \to x \in [-\infty, \infty]$ and $t + x(t)\phi(t) \to \infty$ as $t \to \infty$, then

$$\lim_{t\to\infty} \frac{f^{\leftarrow}\{t + x(t)\phi(t)\}}{f^{\leftarrow}(t)} = \exp(-x).$$

(vi) There exists a positive function φ defined in a neighbourhood of infinity such that $\varphi(t) = o(t)$ as $t \to \infty$ and

$$\lim_{t\to\infty} \frac{f^{\leftarrow}\{t + x\varphi(t)\}}{f^{\leftarrow}(t)} = \exp(-x) \textit{ for all } x \in \mathbb{R}.$$

Proof. Note that $1/\phi$ in Lemma 3.2.6 is a version of the Radon-Nikodym derivative of $-\log f^{\leftarrow}$. Hence, there exists $0 < c < \infty$ such that

$$f^{\leftarrow}(t) = f^{\leftarrow}(c) \exp\left(-\int_c^t \frac{du}{\phi(u)}\right), \text{ for all } c \leq t < \infty.$$

(iv) implies (v). For sufficiently large t,

$$\log\left(\frac{f^{\leftarrow}\{t + x(t)\phi(t)\}}{f^{\leftarrow}(t)}\right) = -\int_t^{t+x(t)\phi(t)} \frac{du}{\phi(u)} = -\int_0^{x(t)} \frac{\phi(t)}{\phi\{t + v\phi(t)\}} dv$$

If $-\infty < x < \infty$, then the right-hand side of the previous display converges to $-x$. If $x = -\infty$ or $x = +\infty$, then by Fatou's lemma, the right-hand side of the previous display converges to $+\infty$ or $-\infty$, respectively.

(v) implies (vi). Trivial.

(vi) implies (iii). Define $g(s) = \varphi(f(s))$ for sufficiently small, positive s. Take $0 < y < \infty$ and put

$$h(y, s) = \frac{f(sy) - f(s)}{g(s)}.$$

We have to show that $\lim_{s\to 0} h(y, s) = -\log(y)$. Fix $0 < \varepsilon < y$. Since $f(s) \to \infty$ as $s \to 0$, we have

$$\lim_{s\to 0} \frac{f^{\leftarrow}\{f(s) - \log(y \pm \varepsilon) g(s)\}}{s} = y \pm \varepsilon.$$

Hence, there exists $s_\varepsilon > 0$ such that

$$\frac{f^{\leftarrow}\{f(s) - \log(y-\varepsilon)g(s)\}}{s} \leq y \leq \frac{f^{\leftarrow}\{f(s) - \log(y+\varepsilon)g(s)\}}{s},$$

for all $0 < s \leq s_\varepsilon$. However, we also have

$$y = \frac{f^{\leftarrow}\{f(s) + h(y,s)g(s)\}}{s}.$$

Since $f^{\leftarrow}$ is decreasing and g is positive, we find that

$$-\log(y-\varepsilon) \geq h(y,s) \geq -\log(y+\varepsilon) \text{ for all } 0 < s \leq s_\varepsilon.$$

Since ε can be taken arbitrarily small and since the logarithm is a continuous function, we find that $h(y,s) \to -\log(y)$ as $s \to 0$.

(i)–(ii) imply (iv). Observe that $\phi\{f(s)\} = -sf'(s)$. Hence $\phi\{f(s)\} = o\{f(s)\}$ as $s \to 0$, whence $\phi(t) = o(t)$ as $t \to \infty$.

Let $x(\cdot)$ be a function defined in a neighbourhood of infinity and such that $x(t) \to x \in \mathbb{R}$. Define

$$y(t) = \frac{f^{\leftarrow}\{t + x(t)\phi(t)\}}{f^{\leftarrow}(t)}.$$

Fix $\varepsilon > 0$. We have

$$\lim_{t\to\infty} \frac{f\{f^{\leftarrow}(t)\exp(-x \pm \varepsilon)\} - t}{\phi(t)} = x \mp \varepsilon.$$

On the other hand,

$$\frac{f\{f^{\leftarrow}(t)y(t)\} - t\}}{\phi(t)} = x(t) \to x \text{ as } t \to \infty.$$

Hence, there must exist $t_\varepsilon > 0$ such that

$$\begin{aligned} \frac{f\{f^{\leftarrow}(t)\exp(-x+\varepsilon)\} - t}{\phi(t)} &\leq \frac{f\{f^{\leftarrow}(t)y(t)\} - t}{\phi(t)} \\ &\leq \frac{f\{f^{\leftarrow}(t)\exp(-x-\varepsilon)\} - t}{\phi(t)} \end{aligned}$$

for all $t_\varepsilon \leq t < \infty$. Since f is decreasing and ϕ is positive, this implies

$$\exp(-x+\varepsilon) \geq y(t) \geq \exp(-x-\varepsilon), \text{ for all } t_\varepsilon \leq t < \infty.$$

Since ε can be taken arbitrarily small and since the exponential function is continuous, we find that $y(t) \to \exp(-x)$ as $t \to \infty$.

By (i), the function $s \mapsto \phi \circ f(s) = -sf'(s)$ is slowly varying at zero. But from

$$t + x(t)\phi(t) = f\{f^{\leftarrow}(t)y(t)\},$$

and the Uniform Convergence Theorem, it then follows that

$$\begin{aligned} \phi\{t + x(t)\phi(t)\} &= \phi \circ f\{f^{\leftarrow}(t)y(t)\} \\ &\sim \phi \circ f\{f^{\leftarrow}(t)\} = \phi(t), \text{ as } t \to \infty. \end{aligned}$$

$\square$

3.2.5 Clayton copula

The *Clayton copula* with parameter $\alpha \in [0, \infty)$ is the Archimedean copula with strict generator given by

$$\psi(x; \alpha) = \int_x^1 t^{-\alpha-1} dt = \begin{cases} \dfrac{x^{-\alpha} - 1}{\alpha} & \text{if } 0 < \alpha < \infty, \\ -\log(x) & \text{if } \alpha = 0 \end{cases}$$

for $0 < x \leq 1$; the corresponding copula is

$$C(x, y; \alpha) = \begin{cases} (x^{-\alpha} + y^{-\alpha} - 1)^{-1/\alpha} & \text{if } 0 < \alpha < \infty, \\ xy & \text{if } \alpha = 0 \end{cases}$$

for $(x, y) \in (0, 1]^2$. Note that $\lim_{\alpha \to 0} \psi(t; \alpha) = \psi(t; 0)$ and $\lim_{\alpha \to 0} C(x, y; \alpha) = C(x, y; 0)$. The comonotone copula, which is itself not an Archimedean copula, arises as the of the Clayton copula as $\alpha \to \infty$, that is,

$$C(x, y; \infty) = \lim_{\alpha \to \infty} C(x, y; \alpha) = \min(x, y)$$

for $(x, y) \in [0, 1]^2$.

The Clayton copula has the special property that at every level $0 < u < 1$, its lower tail dependence copula is again a Clayton copula and with the same parameter; see also Proposition 4.15 in Charpentier (2004). Moreover, the Clayton copula is the only copula which can arise as the limit of the lower tail dependence copula of an Archimedean copula whose generator is regulary varying at the origin of positive index (see Theorem 3.3 in Juri and Wüthrich (2003)).

3.3 Limiting copulae and limiting generators

Let C_n be a sequence of bivariate Archimedean copulae with generators ψ_n. In this note, we want to establish necessary and sufficient conditions for convergence of the sequence of copulae C_n to a limiting copula C in terms of asymptotic properties of the sequence of generators ψ_n. In particular, we seek to extend the results in Proposition 4.2 and 4.3 in Genest and MacKay (1986b) and Theorems 4.4.7 and 4.4.8 in Nelsen (1999) to generators which are possibly not everywhere differentiable, as well as give a number of alternative characterizations.

The characterizations are based in part upon an extension to general generators of Proposition 3.3 in Genest and MacKay (1986b), giving an expression for the joint distribution function of the pair of random variables $(X, C(X, Y))$, where (X, Y) is itself a random pair with distribution function given by the Archimedean copula C (section 3.3.1). The main results involve characterizations for the convergence of a sequence of Archimedean copulae to another Archimedean copula or to the comonotone copula. These two cases require a separate treatment, see sections 3.3.2 and 3.3.3, respectively.

Note that although the results are written down in dimension two for convenience (and to link them more easily with Genest and MacKay (1986b)), in section 3.3.4 we will see that in higher dimensions the results remain virtually unchanged.

From our results, one may get the impression that every limit copula of a sequence of Archimedean copulae is necessarily Archimedean or comonotone. This is not the case, however, as is shown by a counterexample in section 3.3.5.

If the generator is a natural way to identify the Archimedean copula, other functions can be considered as well. The *Kendall distribution function* K of a copula C is defined as the distribution function of the random variable $C(X, Y)$, where (X, Y) is a random pair with distribution function C, so

$$K(t) = \Pr[C(X, Y) \leq t], t \in [0, 1].$$

If the copula C is Archimedean with generator ψ, then $K(t) = t - \lambda(t)$ with $\lambda(t) = \psi(t)/\psi'(t)$ and ψ' is the right-hand derivative of ψ on $[0, 1)$ Proposition 1.1 in Genest and Rivest (1993). Conversely, from K or λ it is possible to reconstruct ψ up to a multiplicative constant via

$$\psi(u) = \psi(u_0) \exp\left(\int_{u_0}^{u} \frac{1}{\lambda(t)} \mathrm{d}t\right)$$

for $0 < u_0 < 1$ and $0 \leq u \leq 1$.

3.3.1 Auxiliary result

The following result is a useful device to deduce properties of the generator ψ of an Archimedean copula C from the copula itself. For twice continuously differentiable generators, the result can already be found in Proposition 3.3 in Genest and MacKay (1986b).

Proposition 3.3.1. *Let (X, Y) be a random pair with joint distribution function C, a bivariate Archimedean copula with generator ψ. Let ψ' be the right-hand derivative of ψ on $[0, 1)$. Put $Z = C(X, Y)$. For $(z, x) \in [0, 1]^2$,*

$$\mathbb{P}(X \leq x, Z \leq z) = \begin{cases} x & \text{if } x \leq z \leq 1, \\ z + \dfrac{\psi(x)}{\psi'(z)} - \dfrac{\psi(z)}{\psi'(z)} & \text{if } 0 < z < x \leq 1, \\ \dfrac{\psi(x) - \psi(0)}{\psi'(0)} & \text{if } z = 0 < x \text{ and } \psi'(0) > -\infty, \\ 0 & \text{if } z = 0 < x \text{ and } \psi'(0) = -\infty. \end{cases}$$

Proof. Since $Z = C(X, Y) \leq X$, we have $\mathbb{P}(X \leq x, Z \leq z) = \mathbb{P}(X \leq x) = x$ for $x \leq z \leq 1$. Hence we can restrict attention to $z < x$.

The case $z = 0 < x$ follows from the case $0 < z < x$ by the fact that $\psi'(0) = \lim_{z \to 0} \psi'(z)$ and the fact that $\lim_{z \to 0} \psi(z)/\psi'(z) = 0$ if $\psi'(z) = -\infty$, the latter property following from convexity.

Hence we can restrict attention to the case $0 < z < x$. Since both ψ' and the function $z \mapsto \mathbb{P}(X \leq x, Z \leq z)$ are right-continuous, it suffices to prove the stated equality for z such that ψ' is continuous in z.

We have

$$\begin{aligned} \mathbb{P}(X \leq x, Z \leq z) &= \mathbb{P}(X \leq z) + \mathbb{P}(z < X \leq x, Z \leq z) \\ &= z + \mathbb{P}(z < X \leq x, Z \leq z) \end{aligned}$$

We can focus on the last term on the right-hand side. Let n be a positive integer, and let

$$z = u_0 < u_1 < \cdots < u_n = x$$

be such that

$$\psi(u_i) = \left(1 - \frac{i}{n}\right) \psi(z) + \frac{i}{n} \psi(x), i = 0, 1, \ldots, n.$$

We have

$$\mathbb{P}(z < X \leq x, Z \leq z) = \sum_{i=1}^{n} \mathbb{P}(u_{i-1} < X \leq u_i, Z \leq z).$$

If $u_{i-1} < X \leq u_i$, then $C(u_{i-1}, Y) \leq Z \leq C(u_i, Y)$. Hence

$$\begin{aligned} &\sum_{i=1}^{n} \mathbb{P}(U_{i-1} < X \leq u_i, C(u_i, Y) \leq z) \\ &\leq \mathbb{P}(z < X \leq x, Z \leq z) \leq \sum_{i=1}^{n} \mathbb{P}(u_{i-1} < X \leq u_i, C(u_{i-1}, Y) \leq z). \end{aligned}$$

Further, for $z \le u \le 1$, since ψ and $\psi^{\leftarrow}$ are decreasing, $C(u,Y) \le z$ is equivalent to $Y \le \psi^{\leftarrow}\{\psi(z) - \psi(u)\}$. We find that

$$\begin{aligned}
&\mathbb{P}(z < X \le x,\, Z \le z) \\
&\quad \le \sum_{i=1}^{n} \mathbb{P}(u_{i-1} < X \le u_i,\, Y \le \psi^{\leftarrow}\{\psi(z) - \psi(u_{i-1})\}) \\
&\quad = \sum_{i=1}^{n} \left(C(u_i, \psi^{\leftarrow}\{\psi(z) - \psi(u_{i-1})\}) - C(u_{i-1}, \psi^{\leftarrow}\{\psi(z) - \psi(u_{i-1})\})\right) \\
&\quad = \sum_{i=1}^{n} \left(\psi^{\leftarrow}\{\psi(u_i) + \psi(z) - \psi(u_{i-1})\} - \psi^{\leftarrow}\{\psi(z)\}\right).
\end{aligned}$$

Our choice of the grid $\{u_i\}$ is such that

$$\psi(u_i) - \psi(u_{i-1}) = -\{\psi(z) - \psi(x)\}/n, i = 1, \dots, n.$$

Hence

$$\mathbb{P}(z < X \le x,\, Z \le z) \le n\big(\psi^{\leftarrow}[\psi(z) - \{\psi(z) - \psi(x)\}/n] - \psi^{\leftarrow}\{\psi(z)\}\big)$$

Since $\psi^{\leftarrow}$ is convex with nondecreasing derivative $1/(\psi' \circ \psi^{\leftarrow})$,

$$\psi^{\leftarrow}(a) - \psi(b) \le (a-b)\frac{1}{\psi'\{\psi^{\leftarrow}(a)\}}, 0 < a < b < \psi(0).$$

Combine the two previous displays to find

$$\mathbb{P}(z < X \le x,\, Z \le z) \le -\frac{\psi(z) - \psi(x)}{\psi'(\psi^{\leftarrow}[\psi(z) - \{\psi(z) - \psi(x)\}/n])}$$

Let n tend to infinity and use the fact that z is a continuity point of ψ' to find

$$\mathbb{P}(z < X \le x,\, Z \le z) \le -\frac{\psi(z) - \psi(x)}{\psi'(z)}.$$

The inequality in the other direction follows in a similar fashion. We give the steps here in full. By the same arguments as above,

$$\begin{aligned}
&\mathbb{P}(z < X \le x,\, Z \le z) \\
&\quad \ge \sum_{i=1}^{n} \mathbb{P}(u_{i-1} < X \le u_i,\, Y \le \psi^{\leftarrow}\{\psi(z) - \psi(u_i)\}) \\
&\quad = \sum_{i=1}^{n} \left(C(u_i, \psi^{\leftarrow}\{\psi(z) - \psi(u_i)\}) - C(u_{i-1}, \psi^{\leftarrow}\{\psi(z) - \psi(u_i)\})\right) \\
&\quad = \sum_{i=1}^{n} \left(\psi^{\leftarrow}\{\psi(z)\} - \psi^{\leftarrow}\{\psi(u_{i-1}) + \psi(z) - \psi(u_i)\}\right) \\
&\quad = n\big(\psi^{\leftarrow}\{\psi(z)\} - \psi^{\leftarrow}[\psi(z) + \{\psi(z) - \psi(x)\}/n]\big) \\
&\quad \ge -\frac{\psi(z) - \psi(x)}{\psi'(\psi^{\leftarrow}[\psi(z) + \{\psi(z) - \psi(x)\}/n])}.
\end{aligned}$$

Let n tend to infinity and use the fact that z is a continuity point of ψ' to arrive at

$$\mathbb{P}(z < X \le x,\, Z \le z) \ge -\frac{\psi(z) - \psi(x)}{\psi'(z)},$$

as required. □

3.3.2 Convergence to Archimedean copula

In this section, we investigate necessary and sufficient properties of a sequence of Archimedean copulae C_n with generators ψ_n to converge to an Archimedean copula C with generator ψ. For twice continuously differentiable generators, the equivalence of (i) and (ii) in Proposition 3.3.2 below was already established in Proposition 4.2 in Genest and MacKay (1986b). The claim that characterization (iii) is sufficient for copula convergence seems to be new.

Proposition 3.3.2. *The following five conditions are equivalent:*

(i) $\lim_{n\to\infty} C_n(x,y) = C(x,y)$ *for all* $(x,y) \in [0,1]^2$

(ii) $\lim_{n\to\infty} \psi_n(x)/\psi_n'(y) = \psi(x)/\psi'(y)$ *for every* $x \in (0,1]$ *and* $y \in (0,1)$ *such that* ψ' *is continuous in* y.

(iii) $\lim_{n\to\infty} \lambda_n(x) = \lambda(x)$ *for every* $x \in (0,1)$ *such that* λ *is continuous in* x.

(iv) There exist positive constants κ_n *such that* $\lim_{n\to\infty} \kappa_n\psi_n(x) = \psi(x)$ *for all* $x \in [0,1]$.

(v) $\lim_{n\to\infty} K_n(x) = K(x)$ *for every* $x \in (0,1)$ *such that* K *is continuous in* x.

Proof. (i) implies (ii). Let (X,Y) and (X_n,Y_n) be pairs of random variables with joint distribution functions C and C_n, respectively. Also, put $Z = C(X,Y)$ and $Z_n = C_n(X_n,Y_n)$. By (i), (X_n,Y_n) converges in distribution to (X,Y) as $n \to \infty$. Moreover, since C is a continuous distribution function, the convergence of C_n to C is necessarily uniform in $(x,y) \in [0,1]^2$. Hence (X_n,Z_n) converges in distribution to (X,Z) as $n \to \infty$. By Proposition 3.3.1, we have

$$\lim_{n\to\infty} \frac{\psi_n(x) - \psi_n(y)}{\psi_n'(y)} = \frac{\psi(x) - \psi(y)}{\psi'(y)}$$

for all $0 < y < x \le 1$ such that ψ' is continuous in y. Choose $x = 1$ to find

$$\lim_{n\to\infty} \frac{\psi_n(y)}{\psi_n'(y)} = \frac{\psi(y)}{\psi'(y)}.$$

Combine the two previous displays to get

$$\lim_{n\to\infty} \frac{\psi_n(x)}{\psi_n'(y)} = \frac{\psi(x)}{\psi'(y)}$$

for every $0 < y \le x \le 1$ such that $y < 1$ and ψ' is continuous in y. Let $0 < x_i < 1$ for $i = 1,2$ and apply the above display to (x_1,y) and (x_2,y) for some $0 < y < \min(x_1,x_2)$ in which ψ' is continuous to arrive at

$$\lim_{n\to\infty} \frac{\psi_n(x_1)}{\psi_n(x_2)} = \frac{\psi(x_1)}{\psi(x_2)}.$$

Combine the last two displays to arrive at (ii).

(ii) implies (iii). Trivial.

(iii) implies (iv). For $0 < x < y < 1$, we have

$$\log\psi_n(y) - \log\psi_n(x) = \int_x^y \frac{\psi_n'(z)}{\psi_n(z)}\mathrm{d}z.$$

Suppose that we can show that the limit of the integral of the right-hand side of the previous display is equal to the integral of the (almost everywhere) limit of the integrand. Then we have

$$\lim_{n\to\infty}\{\log\psi_n(y) - \log\psi_n(x)\} = \log\psi(y) - \log\psi(x).$$

This, in turn, obviously implies (iv).

In order to justify interchanging limit and integral in the previous paragraph, we will show that (iii) implies

$$\limsup_{n\to\infty} \sup_{z\in[x,y]} \left|\frac{\psi_n'(z)}{\psi_n(z)}\right| < \infty.$$

Let $0 < \varepsilon < x$ be such that

$$|\psi'(x-\varepsilon)| \le \psi(y)/(4\varepsilon).$$

By (iii), we have

$$\lim_{n\to\infty} \mathbf{1}\left(\frac{|\psi_n'(z)|}{\psi_n(z)} > 2\frac{|\psi'(z)|}{\psi(z)}\right) = 0$$

for almost every $z \in [x-\varepsilon, y]$. Since the above indicator variables are bounded and converge pointwise to zero, there exists a positive integer n_ε such that

$$\int_{x-\varepsilon}^{y} \mathbf{1}\left(\frac{|\psi_n'(z)|}{\psi_n(z)} > 2\frac{|\psi'(z)|}{\psi(z)}\right) \mathrm{d}z < \varepsilon$$

for all integer $n \ge n_\varepsilon$. Hence, for $z \in [x, y]$ and integer $n \ge n_\varepsilon$, there exist $z - \varepsilon < u < z$ such that

$$\frac{|\psi_n'(u)|}{\psi_n(u)} \le 2\frac{|\psi'(u)|}{\psi(u)} \le 2\frac{|\psi'(x-\varepsilon)|}{\psi(y)} \le \frac{1}{2\varepsilon}$$

But then, since ψ_n and $|\psi_n'|$ are both nonincreasing,

$$\frac{\psi_n(z)}{|\psi_n'(z)|} \ge \frac{\psi_n(u) - (z-u)|\psi_n'(u)|}{|\psi_n'(u)|} \ge 2\varepsilon - \varepsilon = \varepsilon,$$

as required.

(iv) implies (i). Let $\phi_n = \kappa_n \psi_n$. Then ϕ_n is a generator of C_n. Since each ϕ_n is monotone and since ψ is monotone and continuous, we have $\lim_{n\to\infty} \phi_n(x_n) = \psi(x)$ whenever $\lim_{n\to\infty} x_n = x$ in $[0, 1]$. Hence also $\lim_{n\to\infty} \phi_n^{\leftarrow}(t_n) = \psi^{\leftarrow}(t)$ whenever $\lim_{n\to\infty} t_n = t$ in $[0, \infty]$. Hence, for every $(x, y) \in [0, 1]^2$,

$$C_n(x, y) = \phi_n^{\leftarrow}\{\phi_n(x) + \phi_n(y)\} \to \psi^{\leftarrow}\{\psi(x) + \psi(y)\} = C(x, y),$$

as $n \to \infty$.

(v) implies (iii) and conversely. Trivial. □

3.3.3 Convergence to comonotone copula

The comonotone copula is itself not an Archimedean copula, so that Proposition 3.3.2 is not suitable for deciding whether a sequence of copulae converges to the comonotone copula. The following resulting, extending Theorem 4.4.8 in Nelsen (1999) to arbitrary generators, gives such a criterion.

Proposition 3.3.3. *The following four conditions are equivalent:*

(i) $\lim_{n\to\infty} C_n(x, y) = \min(x, y)$ *for all* $(x, y) \in [0, 1]^2$

(ii) $\lim_{n\to\infty} \lambda_n(x) = 0$ *for every* $x \in (0, 1)$.

(iii) $\lim_{n\to\infty} \psi_n(y)/\psi_n(x) = 0$ *for every* $0 \le x < y \le 1$.

(iv) $\lim_{n\to\infty} K_n(x) = x$ *for every* $x \in (0,1)$.

Proof. *(i) implies (ii).* Let (X_n, Y_n) be a pair of random variables with distribution function C_n. Since the limit of C_n is the comonotone copula, (X_n, Y_n) converges in distribution to (X, X), where X is a uniform random variable on $(0,1)$. But since the convergence in (i) is necessarily uniform, we find that $Z_n = C_n(X_n, Y_n)$ converges in distribution to $\min(X, X) = X$, whence $\lim_{n\to\infty} \Pr[Z_n \le z] = z$ for all $z \in [0,1]$. But by Proposition 3.3.1,

$$\Pr[Z_n \le z] = z + \frac{\psi_n(z)}{\psi_n'(z)}, \qquad 0 < z < 1.$$

Hence we arrive at (ii).

(ii) implies (iii). Let $0 < x < y < 1$ (the cases $x = 0$ or $y = 1$ follow by monotonicity of ψ_n). We have

$$\frac{\psi_n(x)}{\psi_n(y)} - 1 = \frac{\psi_n(x) - \psi_n(y)}{\psi_n(y)} \ge \frac{(y-x)|\psi_n'(y)|}{\psi_n(y)}.$$

By (ii), the right-hand side diverges to infinity as $n \to \infty$.

(iii) implies (i). Since each C_n is a symmetric copula, it suffices to consider $0 < x \le y < 1$. Take $0 < w < x$. By (ii), we have $\psi_n(w) \ge 2\psi_n(x) \ge \psi_n(x) + \psi_n(y)$ for all sufficiently large integer n, whence

$$w \le \psi_n^{\leftarrow}\{\psi_n(x) + \psi_n(y)\} = C_n(x,y) \le x.$$

Let first $n \to \infty$ and then $w \uparrow x$ to find that $\lim_{n\to\infty} C_n(x,y) = x$. $\square$

3.3.4 Extension to higher dimensions

Propositions 3.3.2 and 3.3.3 can be readily extended to the general multivariate case. Let d be an integer at least two. A d-variate copula C is the distribution function of a d-variate random vector $(X_1, \ldots, X_d)$,

$$C(x_1, \ldots, x_d) = \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d)$$

the components of which are uniformly distributed on the interval $[0,1]$, that is, $\mathbb{P}(X_j \le x) = x$ for $j = 1, \ldots, d$ and $x \in [0,1]$. A d-variate copula C is called Archimedean if there exists a generator ψ such that

$$C(x_1, \ldots, x_d) = \psi^{\leftarrow}\{\psi(x_1) + \cdots + \psi(x_d)\}$$

for all $(x_1, \ldots, x_d) \in [0,1]^d$. In general, extra conditions on the generator ψ are required to ensure that the expression in the above display defines a genuine copula. A sufficient condition is for instance that $\psi^{\leftarrow}$ is d-times differentiable and $(-D)^j \psi^{\leftarrow} \ge 0$ for every $j = 1, \ldots, d$; see for instance Theorems 1 and 2 in Kimberling (1974), Schweizer and Sklar (1983), Example 3 in Barlow and Proschan (1996), and Section 4.6 in Nelsen (1999).

Obviously, if the distribution function of the random vector $(X_1, \ldots, X_d)$ is given by the d-variate Archimedean copula C with generator ψ, then the distribution function of every bivariate subvector (X_i, X_j), with $i \ne j$, is given by the bivariate Archimedean copula with the same generator. This property can be used to upgrade Propositions 3.3.2 and 3.3.3 to the general multivariate case.

Let C_n be a sequence of d-variate Archimedean copulae with generators ψ_n. On the one hand, if C_n converges to another d-variate Archimedean copula C with generator ψ or to the d-variate comonotone copula, then the sequence of bivariate Archimedean copulae with generators ψ_n must converge to the bivariate Archimedean copula with generator ψ or to the bivariate comonotone copula, respectively. Hence, the stated conditions on the sequence of generators are certainly necessary for convergence of the sequence of copulae. On the other hand, they are also sufficient, as the proofs of the implications "(iv) implies (i)" in Proposition 3.3.2 and "(iii) implies (i)" in Proposition 3.3.3 carry over to the d-variate case with only notational changes.

3.3.5 Counterexample

From Propositions 3.3.2 and 3.3.3, one might get the impression that every limit copula of a sequence of Archimedean copulae is necessarily Archimedean or comonotone. This is not true, as is demonstrated by the following example.

For integer $n \geq 2$, define a generator ψ_n by

$$\psi_n(x) = \begin{cases} n - 2(n-1)x & \text{if } 0 \leq x \leq 1/2, \\ 2(1-x) & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

That is, ψ_n is piecewise linear with knots $\psi_n(0) = n$, $\psi_n(1/2) = 1$, and $\psi_n(1) = 0$. Denoting the right-hand derivative of ψ_n with ψ_n', we have

$$\lambda_n(x) = \frac{\psi_n(x)}{\psi_n'(x)} = \begin{cases} x - n/\{2(n-1)\} & \text{if } 0 \leq x < 1/2, \\ x - 1 & \text{if } 1/2 \leq x \leq 1, \end{cases}$$

and therefore

$$K_n(x) = x - \lambda_n(x) = \begin{cases} n/\{2(n-1)\} & \text{if } 0 \leq x < 1/2, \\ 1 & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

Let C_n be the Archimedean copula with generator ψ_n. By direct computation, one arrives at

$$\lim_{n\to\infty} C_n(x,y) = C(x,y) = \begin{cases} (x+y-1/2)_+ & \text{if } (x,y) \in [0,1/2]^2, \\ x & \text{if } 0 \leq x < 1/2 < y \leq 1, \\ y & \text{if } 0 \leq y < 1/2 < x \leq 1, \\ 1/2 + (x+y-3/2)_+ & \text{if } (x,y) \in [1/2,1]^2. \end{cases}$$

The copula C corresponds to the uniform distribution, with respect to one-dimensional Lebesgue measure, on the union of the two line segments $\{(x,y) \in [0,1]^2 \mid x+y = 1/2\}$ and $\{(x,y) \in [0,1]^2 \mid x+y = 3/2\}$. The copula C is not Archimedean, because the function

$$\lim_{n\to\infty} \frac{|\psi_n'(x)|}{\psi_n(x)} = \begin{cases} 1/(1/2 - x) & \text{if } 0 \leq x < 1/2, \\ 1/(1-x) & \text{if } 1/2 \leq x < 1 \end{cases}$$

is not integrable around $x = 1/2$. Note also that K_n converges towards K as n goes to infinity, where

$$K(x) = \begin{cases} 1/2 & \text{if } 0 \leq x < 1/2, \\ 1 & \text{if } 1/2 \leq x \leq 1. \end{cases}$$

Hence, $\lim_{x\uparrow 1/2} K(x) = 1/2$, and from Proposition 1.2 in Genest and Rivest (1993), the associated copula cannot be Archimedean.

3.4 Lower tail dependence copulae in dimension $d = 2$

3.4.1 Main result

Our main result, Theorem 3.4.1, can be seen as an extension of Theorems 3.3, 3.5 and 3.6 of Juri and Wüthrich (2003). The asymptotic behavior of lower tail dependence copulae for general symmetric bivariate copulae is studied in Juri and Wüthrich (2004), and for nonsymmetric bivariate copulae in Charpentier (2004).

Theorem 3.4.1. *Let C be a strict Archimedean copula with generator ψ. Let ψ' be the left-hand derivative of ψ at $(0,1]$. Let $0 \leq \alpha \leq \infty$. Consider the following four statements:*

(i) $\lim_{u \to 0} C_u(x,y) = C(x,y;\alpha)$ *for all* $(x,y) \in [0,1]^2$*;*

(ii) $-\psi' \in \mathcal{R}_{-\alpha-1}$.

(iii) $\psi \in \mathcal{R}_{-\alpha}$.

(iv) $\lim_{u \to 0} u\psi'(u)/\psi(u) = -\alpha$.

If $\alpha = 0$ *(tail independence),*

$$\textit{(i)} \iff \textit{(ii)} \implies \textit{(iii)} \iff \textit{(iv)},$$

and if $\alpha \in (0,\infty]$ *(tail dependence),*

$$\textit{(i)} \iff \textit{(ii)} \iff \textit{(iii)} \iff \textit{(iv)}.$$

As we will see in Theorem 3.4.2, if $\alpha = 0$, then *(iii)* does not imply *(i)* or *(ii)*, in contradiction to Lemma 3.4 and Theorem 3.5 of Juri and Wüthrich (2003).

Proof. The lower tail dependence copula of C at $0 < u < 1$ is the Archimedean copula with generator

$$\psi_u(x) = \psi(xv) - \psi(v), 0 \leq x \leq 1,$$

where $v = v(u) = \psi^{\leftarrow}\{2\psi(u)\}$. Note that $v(u)$ is continuous in u and decreases to 0 as u decreases to zero. The (left-hand) derivative of ψ_u at $x = 1$ is equal to

$$\psi_u'(1) = v\psi'(v) = -g(v).$$

- **The case** $\alpha \in (0,\infty)$ **(tail dependence)**

(i) implies (ii). By Proposition 2 in Charpentier and Segers (2006a) (which extends Theorem 4.4.7 in Nelsen (1999) and Proposition 4.2 in Genest and MacKay (1986b) to the case of generators which are not twice continuously differentiable, see Proposition 3.3.2 in this Chapter), for $0 < x \leq 1$,

$$\lim_{u \to 0} \frac{\psi_u(x)}{\psi_u'(1)} = \frac{\psi(x;\alpha)}{\psi'(1;\alpha)} = -\psi(x;\alpha).$$

Hence, for $0 < x \leq 1$,

$$\lim_{v \to 0} \frac{\psi(vx) - \psi(v)}{g(v)} = -\psi(x;\alpha).$$

For $0 < x < 1$, we get

$$\begin{aligned} \frac{g(vx)}{g(v)} &= \left(\frac{\psi(vx^2) - \psi(v)}{g(v)} - \frac{\psi(vx) - \psi(v)}{g(v)} \right) \Bigg/ \frac{\psi(vx^2) - \psi(vx)}{g(vx)} \\ &\to \{\psi(x^2;\alpha) - \psi(x;\alpha)\}/\psi(x;\alpha) = x^{-\alpha}, \text{ as } v \to 0. \end{aligned}$$

Hence, $g \in \mathcal{R}_{-\alpha}$, and thus $-\psi' \in \mathcal{R}_{-\alpha-1}$.

(ii) implies (i). If $-\psi' \in \mathcal{R}_{-\alpha-1}$, then $g \in \mathcal{R}_{-\alpha}$, whence

$$\begin{aligned} \frac{\psi(xv) - \psi(v)}{g(v)} &= \int_{xv}^{v} \frac{g(t)}{g(v)} \frac{dt}{t} \\ &= \int_{x}^{1} \frac{g(vt)}{g(v)} \frac{dt}{t} \\ &\to \int_{x}^{1} t^{-\alpha-1} dt = \psi(x;\alpha), \text{ as } v \to 0. \end{aligned}$$

Since $\psi'_u(y) = v\psi'(vy) = -y^{-1}g(vy)$ for $0 < y \leq 1$, and since $g \in \mathcal{R}_{-\alpha}$, we have

$$\lim_{u\to 0} \frac{\psi'_u(y)}{g(v)} = -y^{-\alpha-1} = \psi'(y;\alpha), 0 < y \leq 1,$$

and thus

$$\lim_{u\to 0} \frac{\psi_u(x)}{\psi'_u(y)} = \frac{\psi(x;\alpha)}{\psi'(y;\alpha)}.$$

for all $x, y \in (0,1]$. By Theorem 4.4.7 in Nelsen (1999), we find that (i) must hold.

(iii) implies (iv). This follows from the Monotone Density Theorem (Theorem 1.7.2 in Bingham, Goldie and Teugels (1987) applied to the function $x \mapsto \psi(1/x)$.

(iv) implies (iii). This follows from the Representation Theorem for regularly varying functions (see equation (1.5.2) in Bingham, Goldie and Teugels (1987)).

So far, we have established the equivalences *(i)* $\Longleftrightarrow$ *(ii)* and *(iii)* $\Longleftrightarrow$ *(iv)*.

(ii) implies (iii). This follows from Karamata's Theorem (see e.g. Proposition 1.5.8 in Bingham, Goldie and Teugels (1987)) applied to the function $x \mapsto \psi(1/x)$.

(iii) and (iv) imply (ii). This is immediate, since $-\psi'(x) \sim \alpha x^{-1}\psi(x)$ as $x \to 0$ and $\psi \in \mathcal{R}_{-\alpha}$.

• **The case $\alpha = 0$ (tail independence)**

The proofs of all the implications, except for the last one, also hold when $\alpha = 0$.

• **The case $\alpha = \infty$ (tail comonotonicity)**

According Proposition 3 in Charpentier and Segers (2006a) (which extends Theorem 4.4.8 in Nelsen (1999) and Proposition 4.3 in Genest and MacKay (1986b) to the case of generators which are not twice continuously differentiable), (1) is equivalent to

$$\lim_{u\to 0} \frac{\psi_u(x)}{\psi'_u(x)} = 0, 0 < x \leq 1.$$

Combine the above three displays to find that (1) is equivalent to

$$\lim_{v\to 0} \frac{\psi(vx) - \psi(v)}{v\psi'(vx)} = 0, 0 < x \leq 1. \tag{3.3}$$

We show first the circle of implications (i)⇒(ii)⇒(iv)⇒(i) and then the equivalence (iii)⇔(iv).

(i) implies (ii). Since ψ is decreasing and convex,

$$0 \leq (x-1)v\psi'(v) \leq \psi(vx) - \psi(v), 0 < x \leq 1; 0 < v \leq 1.$$

Since (i) is equivalent to (3.3), the above inequality implies

$$\lim_{v\to 0} \frac{\psi'(v)}{\psi'(vx)} = 0, 0 < x < 1.$$

Hence $\psi' \in \mathcal{R}_{-\infty}$.

(ii) implies (iv). Let $1 < x < \infty$. There exists $0 < u_0 \leq 1/x$ such that

$$\frac{\psi'(ux)}{\psi'(u)} \leq \frac{1}{2x}, 0 < u \leq u_0.$$

Let $0 < u \leq u_0$ and let $k = 0, 1, 2, \ldots$ be such that $ux^k < u_0 \leq ux^{k+1}$. Since ψ is decreasing and convex,

$$\begin{aligned}
\psi(u) &= \sum_{j=0}^{k}\{\psi(ux^j) - \psi(ux^{j+1})\} + \psi(ux^{k+1}) \\
&\leq \sum_{j=0}^{k} ux^j(1-x)\psi'(ux^j) + \psi(u_0) \\
&\leq u\psi'(u)(1-x)\sum_{j=0}^{k} x^j \frac{1}{(2x)^j} + \psi(u_0) \\
&\leq 2(1-x)u\psi'(u) + \psi(u_0).
\end{aligned}$$

Since ψ is strict, there exists $0 < u_1 < u_0$ such that $\psi(u) \geq 2\psi(u_0)$ for all $0 < u \leq u_1$. Hence, by the previous display,

$$\psi(u) \leq 4(1-x)u\psi'(u), 0 < u \leq u_1.$$

Let u decrease to zero to find

$$\limsup_{u\to 0} \frac{\psi(u)}{-u\psi'(u)} \leq 4(x-1).$$

Since x was an arbitrary element in $(1, \infty)$, we arrive at (iv).

(iv) implies (i). Since (i) is equivalent to (3.3), it is sufficient to show that (iv) implies (3.3). Let $0 < v \leq 1$ and $0 < x < 1$. We have

$$\left|\frac{\psi(vx) - \psi(v)}{v\psi'(vx)}\right| \leq \frac{\psi(vx)}{v|\psi'(vx)|} \leq \frac{\psi(vx)}{vx|\psi'(vx)|}.$$

By (iv), the right-hand side of this equation tends to zero as $v \to 0$, whence (3.3), as required.

(iii) implies (iv). Let $0 < u < 1$ and $1 < x < 1/u$. Since ψ is convex,

$$\psi(u) - \psi(ux) \leq (1-x)u\psi'(u).$$

By (iii), $\lim_{u\to 0}\psi(ux)/\psi(u) = 0$ for every $1 < x < \infty$. Divide both sides of the inequality in the previous display by $\psi(u)$ and let u decrease to zero to find

$$\liminf_{u\to 0} \frac{-u\psi'(u)}{\psi(u)} \geq \frac{1}{x-1}, 1 < x < \infty.$$

The right-hand side in the previous display becomes arbitrarily large as $x \to 1$, whence (iv).

(iv) implies (iii). Let $0 < x < 1$. Since ψ is convex, we have for $0 < u \leq 1$,

$$\psi(ux) - \psi(u) \geq (x-1)u\psi'(u),$$

whence

$$\frac{\psi(ux)}{\psi(u)} \geq (x-1)\frac{u\psi'(u)}{\psi(u)} + 1.$$

By (iv), the right-hand side side of this inequality tends to infinity as $u \to 0$, whence (iii), as required. $\square$

3.4.2 Counterexample

We claim in Theorem 3.4.1 that for general $\alpha \in [0, \infty]$, statements *(i)* and *(ii)* imply statements *(iii)* and *(4)*. If $\alpha > 0$, the converse is also true. However, if $\alpha = 0$, then the converse does not hold, as shown by the following counterexample, contradicting Theorem 3.5 in Juri and Wüthrich (2003)).

Theorem 3.4.2. *There exists a strict Archimedean copula C whose generator ψ is continuously differentiable and slowly varying at the origin, but such that the lower tail dependence copula of C at level u does not converge to the independence copula as $u \to 0$.*

Proof. Let $f : (0, 1] \to \mathbb{R}$ be the piece-wise linear function with knots

$$f(2^{-k}) = 2^k, k = 0, 1, 2, \dots.$$

That is, f is the linear interpolation of the function $(0, 1] \ni x \mapsto x^{-1}$ at the points $\{2^{-k} \mid k = 0, 1, 2, \dots\}$. Define the function $\psi : [0, 1] \to [0, \infty]$ by

$$\psi(s) = \int_s^1 f(x)dx, s \in [0, 1].$$

By construction, the function ψ is continuously differentiable with derivative $\psi' = -f$. Since f is decreasing, ψ' is increasing, whence ψ is convex. Hence, ψ is a strict generator.

As $s^{-1} \le f(s) \le 2s^{-1}$ for all $s \in (0, 1]$, we have $\psi(s) \ge \log(1/s)$ and thus

$$0 \le \frac{sf(s)}{\psi(s)} \le \frac{2}{\log(1/s)} \to 0, \text{ as } s \to 0.$$

Hence, as ψ is convex, for every $1 < x < \infty$,

$$0 \le 1 - \frac{\psi(sx)}{\psi(s)} \le \frac{s(1-x)f(s)}{\psi(s)} \to 0, \text{ as } s \to 0.$$

Therefore, ψ is slowly varying at the origin.

Let C_u be the tail dependence copula relative to C, the Archimedean copula with generator ψ, at level $0 < u < 1$. We will show that $C_{2^{-k}} = C$ for every positive integer k. Hence, C_u cannot converge to the independence copula as $u \to 0$.

By the definition of the function f,

$$\psi(2^{-k-1}) - \psi(2^{-k}) = \int_{2^{-k-1}}^{2^{-k}} f(x)dx = \frac{3}{4}$$

for all nonnegative integer k. Since also $\psi(1) = 0$, we get $\psi(2^{-k}) = \frac{3}{4}k$ and thus $\psi^{\leftarrow}\{2\psi(2^{-k})\} = 2^{-2k}$ for all nonnegative integer k. By (5.11), the tail dependence copula of C at level $u = 2^{-k}$ is therefore Archimedean with generator

$$\psi_{2^{-k}}(t) = \psi(2^{-2k}t) - \psi(2^{-2k}) = \int_{2^{-2k}t}^{2^{-2k}} f(x)dx = \int_t^1 2^{-2k} f(2^{-2k}x)dx$$

for $t \in [0, 1]$. The function $(0, 1] \ni x \mapsto f_k(x) = 2^{-2k} f(2^{-2k}x)$ is piece-wise linear with knots $f_k(2^{-j}) = 2^j$ for all nonnegative integer j. Hence, f_k must coincide with f. But then, $\psi_{2^{-k}}$ coincides with ψ, and thus $C_{2^{-k}}$ coincides with C for all nonnegative integer k, as required □

The copula C in the proof of the previous theorem has the fractal-like property that $C_u = C$ for all $u = 2^{-k}$ with nonnegative integer k. Still, since its generator ψ has a positive second derivative, the support of C is the whole unit square $[0, 1]^2$.

3.4.3 Discussion: Asymptotic independence

The problem with Theorem 3.5 in Juri and Wüthrich (2003) comes from the auxiliary Lemma 3.4 in the same paper. In this Lemma, it is claimed that if ψ is a strict generator, differentiable and slowly varying at the origin, then there exists a positive function g on $(0, 1)$ such that

$$\lim_{u\to 0} \frac{\psi(ux) - \psi(u)}{g(u)} = -\log(x) \tag{3.4}$$

for every $0 < x < \infty$. However, the generator ψ appearing in the proof of Theorem 3.4.2 satisfies $\psi(ux) - \psi(u) = \psi(x)$ for every $u = 2^{-k}$ with $k = 0, 1, 2, \ldots$, contradicting the claim.

The condition (3.4) states that the function ψ belongs to the de Haan class Π with auxiliary function g, notation $\psi \in \Pi_g$; see for instance Bingham, Goldie and Teugels (1987), chapter 3. [Here, we conveniently shift from asymptotics at infinity to asymptotics at zero by considering the function $y \mapsto \psi(1/y)$ for $y \geq 1$.] By the Monotone Density Theorem (Theorem 3.6.8 in Bingham, Goldie and Teugels (1987)), equation (3.4) is equivalent to

$$-\psi' \in \mathcal{R}_{-1} \tag{3.5}$$

and in this case, $g(s) \sim -s\psi'(s)$ as $s \to 0$. Moreover, by Karamata's theorem (Proposition 1.5.9a in Bingham, Goldie and Teugels (1987), (3.5) implies $\psi \in \mathcal{R}_0$. The converse is not true however, as demonstrated by our counterexample.

3.5 Lower tail dependence copulae in dimension $d \geq 2$

The joint lower tail of a d-variate Archimedean copula C is determined by the asymptotic behavior of its generator ψ near the origin. If ψ is not strict, that is, if $\psi(0)$ is finite, then there exists $0 < s < 1$ such that $d\psi(s) \geq \psi(0)$ and thus $C(s, \ldots, s) = \psi^{\leftarrow}\{d\psi(s)\} = 0$. Hence, the only interesting case occurs when ψ is strict, so $\psi(0) = \infty$, as we will assume henceforth in this section.

In the bivariate case, the coefficient of lower tail dependence, λ_L, is given by

$$\lambda_L = \lim_{s\to 0} \frac{C(s, s)}{s} = \lim_{s\to 0} \frac{\psi^{\leftarrow}\{2\psi(s)\}}{s} = \lim_{t\to\infty} \frac{\psi^{\leftarrow}(2t)}{\psi^{\leftarrow}(t)},$$

provided one and hence all of the limits exist. Hence, if ψ is regularly varying at zero of index $-\theta \in [-\infty, 0]$, then $\psi^{\leftarrow}$ is regularly varying at infinity of index $-1/\theta \in [-\infty, 0]$, and

$$\lambda_L = 2^{-1/\theta},$$

see also Theorem 3.9 in Juri and Wüthrich (2003). In particular, if $\theta > 0$, then $\lambda_L > 0$, so the joint lower tail of C exhibits asymptotic dependence. On the other hand, if $\theta = 0$, then $\lambda_L = 0$, so the joint lower tail of C exhibits asymptotic independence. These two cases persist in the general multivariate case and are treated in subsections 3.5.1 and 3.5.2, respectively.

3.5.1 Asymptotic dependence

Theorem 3.5.1. *Let C be a d-variate Archimedean copula with generator ψ. If ψ is regularly varying at zero of index $-\theta \in [-\infty, 0]$, then for all $\boldsymbol{x} \in (0, \infty)^d$,*

$$\lim_{s\to 0} s^{-1} C(sx_1, \ldots, sx_d) = \begin{cases} 0 & \text{if } \theta = 0, \\ (x_1^{-\theta} + \cdots + x_d^{-\theta})^{-1/\theta} & \text{if } 0 < \theta < \infty, \\ \min(x_1, \ldots, x_d) & \text{if } \theta = \infty. \end{cases}$$

Proof. First consider the case $\theta = 0$. Fix $0 < \varepsilon < 1$. Since ψ is slowly varying, $\psi(s\varepsilon) \leq 2\psi(s)$ and thus $s\varepsilon \geq \psi^{\leftarrow}\{2\psi(s)\}$ for all sufficiently small, positive s. Hence $\psi^{\leftarrow}\{2\psi(s)\} = o(s)$ as $s \to 0$. Hence, denoting $x = \max\{x_1, \ldots, x_d\}$, we have $C(sx_1, \ldots, sx_d) \leq \psi^{\leftarrow}\{d\psi(sx)\} \leq \psi^{\leftarrow}\{2\psi(sx)\} = o(s)$ as $s \to 0$.

Secondly, consider the case $0 < \theta < \infty$. We have

$$\begin{aligned} & s^{-1}C(sx_1, \ldots, sx_d) \\ &= \frac{1}{\psi^{\leftarrow}\{\psi(s)\}}\psi^{\leftarrow}\left\{\psi(s)\left(\frac{\psi(sx_1)}{\psi(s)} + \cdots + \frac{\psi(sx_d)}{\psi(s)}\right)\right\}. \end{aligned}$$

The function ψ is regularly varying at zero of index $-\theta$, and therefore, the function $\psi^{\leftarrow}$ is regularly varying at infinity of index $-1/\theta$. Moreover, $\psi(s) \to \infty$ as $s \to 0$. By the Uniform Convergence Theorem for regularly varying functions, see for instance Bingham, Goldie and Teugels (1987), Theorem 1.5.2, the right-hand side of the previous display therefore converges as $s \to 0$ to $(x_1^{-\theta} + \cdots + x_d^{-\theta})^{-1/\theta}$.

Finally, consider the case $\theta = \infty$. Denote $m = \min\{x_1, \ldots, x_d\}$. We have

$$s^{-1}\psi^{\leftarrow}\{d\psi(sm)\} \leq s^{-1}C(sx_1, \ldots, sx_d) \leq m.$$

Fix $0 < \lambda < 1$. Since ψ is regularly varying at zero with index $-\infty$, we have $\psi(\lambda sm) \geq d\psi(sm)$ and thus $\lambda m \leq s^{-1}\psi^{\leftarrow}\{d\psi(sm)\}$ for all sufficiently small, positive s. Let λ increase to one to see that $s^{-1}C(sx_1, \ldots, sx_d) \to m$ as $s \to 0$. □

Corollary 3.5.2. *Let C be a d-variate Archimedean copula with generator ψ. If ψ is regularly varying at zero of index $-\theta \in [-\infty, 0]$, then for all $\boldsymbol{x} \in (0, \infty)^d$,*

$$\begin{aligned} & \lim_{s \to 0} s^{-1}\mathbb{P}(U_1 \leq sx_i \text{ or } \ldots \text{ or } U_d \leq sx_d) \qquad (3.6) \\ &= \begin{cases} x_1 + \cdots + x_d & \text{if } \theta = 0, \\ \displaystyle\sum_{I \subset \{1,\ldots,d\}:|I| \geq 1} (-1)^{|I|-1}\left(\sum_{i \in I} x_i^{-\theta}\right)^{-1/\theta} & \text{if } 0 < \theta < \infty, \\ \max(x_1, \ldots, x_d) & \text{if } \theta = \infty, \end{cases} \end{aligned}$$

where $(U_1, \ldots, U_d)$ is a random vector with distribution function C.

Proof. By the inclusion-exclusion formula,

$$\mathbb{P}(\exists i = 1, \ldots, d : U_i \leq sx_i) = \sum_{I \subset \{1,\ldots,d\}:|I| \geq 1} (-1)^{|I|-1}\mathbb{P}(\forall i \in I : U_i \leq sx_i).$$

For every subset I of $\{1, \ldots, d\}$ of cardinality at least two, the distribution function of the vector $(U_i)_{i \in I}$ is given by the $|I|$-variate Archimedean copula with generator ψ. Now apply Theorem 3.5.1 to arrive at the stated limit relation. The expression for the limit in case $\theta = \infty$ follows from a well-known relation between minima and maxima. □

Remark 3.5.3. *By Corollary 3.5.2, an Archimedean copula whose generator is regularly varying at the origin is in the min-domain of attraction of the negative logistic dependence structure Joe (1990). That is, if $(U_{i1}, \ldots, U_{id})$ are independent random vectors with common joint distribution function given by C as in Corollary 3.5.2, then for all $\boldsymbol{x} \in (0, \infty)^d$,*

$$\lim_{n \to \infty} \mathbb{P}\left(\min_{i=1,\ldots,n} U_{i1} > x_1/n, \ldots, \min_{i=1,\ldots,n} U_{id} > x_d/n\right) = \exp\{-l(x_1, \ldots, x_d)\},$$

with $l(x_1,\dots,x_d)$ equal to the right-hand side of (3.6). *The case $\theta = 0$ corresponds to independence and the case $\theta = \infty$ to comonotonicity. In the bivariate case, this result was established for the broader class of Archimax copulae in Capéraà, Fougères and Genest (2000).*

Next, we study the limiting distribution of a random vector $(U_1,\dots,U_d)$ with distribution function C as in Theorem 3.5.1 conditionally on the event that some of its coordinates are small. More precisely, let J be a non-empty subset of $\{1,\dots,d\}$ and let $0 < y_j < \infty$ for all $j \in J$. Then we want to determine the limiting distribution of $(U_1,\dots,U_d)$ conditionally on $U_j \le sy_j$ for all $j \in J$ as $s \to 0$.

Of special interest is this limiting distribution's copula. In the bivariate case, if $J = \{1,2\}$ and $(y_1,y_2) = (1,1)$, then if the generator ψ is regularly varying of index $-\theta < 0$, then the limiting copula is necessarily the Clayton copula; see Juri and Wüthrich (2003), Theorem 3.3. This property turns out to extend to arbitrary dimensions and to arbitrary, non-empty subsets J of $\{1,\dots,d\}$. The proofs of the two following corollaries are straightforward applications of Theorem 3.5.1.

Corollary 3.5.4. *Let C be an Archimedean copula with generator ψ which is regularly varying at the origin of index $-\infty < -\theta < 0$, and let $(U_1,\dots,U_d)$ be a random vector with distribution function C. Let J be a non-empty subset of $\{1,\dots,d\}$ and let $0 < y_j < \infty$ for all $j \in J$. Then for every $(x_1,\dots,x_d) \in (0,\infty)^d$,*

$$\begin{aligned}
&\lim_{s\to 0} \mathbb{P}(\forall i = 1,\dots,d : U_i \le sx_i \mid \forall j \in J : U_j \le sy_j) \qquad (3.7)\\
&= \left(\frac{\sum_{j\in J^c} x_j^{-\theta} + \sum_{j\in J} \{\min(x_j,y_j)\}^{-\theta}}{\sum_{j\in J} y_j^{-\theta}} \right)^{-1/\theta}.
\end{aligned}$$

The copula of the distribution function on the right is the Clayton copula.

Corollary 3.5.5. *If in Corollary 3.5.4 we have instead $\theta = \infty$, then, denoting $y = \min\{y_j \mid j \in J\}$,*

$$\lim_{s\to 0} \mathbb{P}(\forall i = 1,\dots,d : U_i \le sx_i \mid \forall j \in J : U_j \le sy_j) = \min\left(\frac{x_1}{y},\dots,\frac{x_d}{y},1\right).$$

Inspired by Corollaries 3.5.4 and 3.5.5, one could conjecture that if ψ is a strict generator which is slowly varying at the origin, then the limit of the copula of the distribution of $(U_1,\dots,U_d)$ conditionally on $U_j \le s$ for all $j = 1,\dots,d$ is necessarily the independence copula. This is not true however, not even in dimension $d = 2$; see Example 3.5.6. In particular, Theorem 3.5 in Juri and Wüthrich (2003) is wrong.

Example 3.5.6. *Consider the function*

$$\psi(s) = \int_s^1 f(x)dx, \text{ for all } s \in [0,1],$$

where $f : (0,1] \to \mathbb{R}$ is the piece-wise linear function with knots

$$f(2^{-k}) = 2^k, \text{ for all } k = 0,1,2,\dots,$$

that is, f is the linear interpolation of the function $x \mapsto x^{-1}$ at the points 2^{-k} for $k = 0,1,2,\dots$. The function ψ is a strict, continuously differentiable generator of order two. Moreover, ψ is slowly varying at zero.

Let (U,V) be a random pair with distribution function C, the bivariate Archimedean copula with generator ψ. For $0 < s < 1$, let C_s be the copula of the distribution of (U,V) conditionally *on $U \leq s$ and $V \leq s$. Then C_s does not converge as $s \to 0$ to the independence copula. For instance, by some tedious calculations, for all $k = 0, 1, 2, \ldots$,*

$$C_{2^{-k}}(3/4, 3/4) = (3/2) - \sqrt{7/8} \approx 0.5645,$$

and this is different from $(3/4)^2 = 9/16 = 0.5625$.

Example 3.5.7. *Several parametric families of Archimedean copulae described in the literature satisfy the assumptions of Theorem 3.5.1. The generator of Clayton's copula is regularly varying at zero of index $-\theta = \alpha$. Among the parametric families listed in Table 4.1 in Nelsen (1999), generator number (19), defined as $\psi(t) = e^{\alpha/t} - e^{\alpha}$, with $\alpha \in (0, \infty)$, is regularly varying at zero of index $-\infty$. In the same Table, the generators with numbers (14) and (16), defined by $\psi(t) = (t^{-1/\alpha} - 1)^{\alpha}$ for all $\alpha \in [1, \infty)$ and $\psi(t) = (\alpha/t + 1)(1 - t)$ for all $\alpha \in [0, \infty)$ respectively, are both regularly varying at zero of index -1.*

Example 3.5.8. *An application of Theorem 3.5.1 can be obtained in the case where $\psi^{\leftarrow}$ is completely monotone. Then, as mentioned in section 1.5 the generator can be seen as the inverse of a Laplace transform of a random variable Θ called frailty. If the Laplace transform of Θ is asymptotically equivalent to the Laplace transform of a Gamma distribution, i.e $\phi(t) = t^{-1/\beta}\mathcal{L}(t)$ as $t \to \infty$, with $\beta > 0$ where $\mathcal{L}$ is a slowly varying function, then the limiting copula is Clayton copula with parameter β. Using the Tauberian theorem (see Feller (1971) or Bingham, Goldie and Teugels (1987)), the tail of Laplace transform of Θ at infinity is related to the shape of the distribution of Θ at origin. More precisely $\mathbb{P}(\Theta \leq \theta) \sim \theta^{1/\beta}\Gamma(1 + \theta^{-1})^{-1}$ as $\theta \to 0$.*

3.5.2 Asymptotic independence

If the generator, ψ, of an Archimedean copula, C, is slowly varying at the origin, then by the case $\theta = 0$ in Theorem 3.5.1, the lower tail of the copula exhibits asymptotic independence, that is,

$$\lim_{s \to 0} s^{-1} C(s, \ldots, s) = 0. \tag{3.8}$$

First, we explore in some depth the relation between asymptotic independence of the lower tail of C and the growth rate of ψ at zero.

Theorem 3.5.9. *Let C be a d-variate Archimedean copula with generator ψ. Consider the following three statements:*

(i) *ψ is slowly varying at zero;*

(ii) $\lim_{s \to 0} s^{-1} C(s, \ldots, s) = 0$;

(iii) $\lim_{s \to 0} \log\{\psi(s)\} / \log(s) = 0$.

Then (i) implies (ii), and (ii) implies (iii). None of the converse implications is true.

Proof. (i) implies (ii). See Theorem 3.5.1, case $\theta = 0$.

(ii) implies (iii). If $\psi(0) < \infty$, there is nothing to prove, so assume $\psi(0) = \infty$. Let $0 < \varepsilon < 1$. There exists $0 < s_\varepsilon < 1$ such that $s^{-1}\psi\{d\psi(s)\} \leq \varepsilon$ and thus $\psi(s\varepsilon)/\psi(s) \leq d$ for all $0 < s \leq s_\varepsilon$. Let $0 < x \leq s_\varepsilon$ and let k be the nonnegative integer for which $\varepsilon^{k+1}s_\varepsilon < x \leq \varepsilon^k s_\varepsilon$, that is, $k \leq \log(x/s_\varepsilon)/\log(\varepsilon) < k + 1$. Then $\psi(x) < \psi(\varepsilon^{k+1}s_\varepsilon) \leq d^{k+1}\psi(s_\varepsilon)$, whence

$$\begin{aligned} \log\psi(x) &< (k+1)\log(d) + \log\psi(s_\varepsilon) \\ &\leq \left(\frac{\log(x/s_\varepsilon)}{\log(\varepsilon)} + 1\right)\log(d) + \log\psi(s_\varepsilon). \end{aligned}$$

This implies $\liminf_{x\to 0} \log\{\psi(x)\}/\log(x) \geq \log(d)/\log(\varepsilon)$. Let ε decrease to zero to complete the proof of the statement.

(iii) does not imply (ii). Define $s_k = 2^{-3^k+1}$ for nonnegative integer k, so $s_0 = 1$ and $s_{k+1} = s_k^3/4$ for nonnegative integer k. Let ψ be the linear interpolation of the function $(0,1] \to [0,\infty) : s \mapsto -\log(s)$ in the points s_k; also, put $\psi(0) = \infty$. Then ψ is a strict generator of order $d = 2$. We claim that ψ satisfies (iii) but not (ii).

On the one hand, take $0 < x \leq 1$. There exists a nonnegative integer k such that $s_{k+1} < x \leq s_k$. Since $s_{k+1} = s_k^3/4 \geq x^3/4$, we get

$$\psi(x) < \psi(s_{k+1}) = -\log s_{k+1} \leq -\log(x^3/4) = \log(4) - 3\log(x).$$

Hence ψ satisfies (iii).

On the other hand, since $s_{k+1} < s_k/4 < s_k$ and ψ is linear on each interval $[s_{k+1}, s_k]$, we have

$$\frac{\psi(s_k/4) - \psi(s_k)}{s_k - s_k/4} = \frac{\psi(s_{k+1}) - \psi(s_k)}{s_k - s_{k+1}}.$$

Since $\psi(s_{k+1})/\psi(s_k) \to 3$ and $s_{k+1}/s_k \to 0$ as $k \to \infty$, we must have

$$\lim_{k\to\infty} \frac{\psi(s_k/4)}{\psi(s_k)} = \frac{5}{2}. \tag{3.9}$$

Hence, for all sufficiently large k, we have $\psi(s_k/4) > 2\psi(s_k)$, whence $s_k/4 < \psi^{\leftarrow}\{2\psi(s_k)\}$. Hence, ψ does not satisfy (ii).

(ii) does not imply (i). Choose $1 < a < \sqrt{2}$ and define $s_k = 2^{-a^k+1}$ for nonnegative integer k, so $s_0 = 1$ and $s_{k+1} = 2^{-a+1} s_k^a$ for nonnegative integer k. Let ψ be the linear interpolation of the function $(0,1] \to [0,\infty) : s \mapsto -\log(s)$ in the points s_k; also, put $\psi(0) = \infty$. Then ψ is a strict generator of order $d = 2$. We claim that ψ satisfies (ii) but not (i).

On the one hand, let $0 < \varepsilon < 1$. Since $s_{k+1}/s_k \to 0$ and $\psi(s_{k+1})/\psi(s_k) \to a$ as $k \to \infty$, there exists integer k_ε such that $s_{k+1}/s_k \leq \varepsilon$ as well as $\psi(s_{k+1})/\psi(s_k) < \sqrt{2}$ for all integer $k \geq k_\varepsilon$. For $0 < x \leq s_{k_\varepsilon}$, there exists an integer $k \geq k_\varepsilon$ such that $s_{k+1} < x \leq s_k$. Then $\psi(s_{k+2})/\psi(x) < 2$ and thus $\psi^{\leftarrow}\{2\psi(x)\} \leq s_{k+2} \leq \varepsilon x$. Let ε decrease to zero to see that ψ satisfies (ii).

On the other hand, by an argument similar to the one leading to (3.9), one can show that $\lim_{k\to\infty} \psi(s_k/2)/\psi(s_k) = (a-1)/2 + 1$. Hence, ψ does not satisfy (i). □

In case of asymptotic independence, the speed of convergence in (3.8) can be arbitrarily fast. As an extreme case, if ψ is not strict, then $C(s,\dots,s) = 0$ for all sufficiently small, positive s. In order to obtain more precise results, we need extra assumptions on the behavior of ψ in the neighbourhood of zero, or, equivalently, of $\psi^{\leftarrow}$ in the neighbourhood of infinity. First, we focus on the diagonal, that is, on the function

$$C(s,\dots,s) = \psi^{\leftarrow}\{d\psi(s)\},$$

more precisely on how fast this function converges to zero. A crude measure of this speed is the index of regular variation, $1/\eta$, of this function at zero. Note that $\psi^{\leftarrow}\{d\psi(s)\} \leq s$, so necessarily $1/\eta \geq 1$. In the bivariate case, this $\eta \in [0,1]$ corresponds to the "coefficient of tail dependence" introduced by Ledford and Tawn (1996, 1997).

If ψ is regularly varying at zero with index $-\infty \leq -\theta < 0$, then by Theorem 3.5.1 we have $C(s,\dots,s) \sim d^{-1/\theta} s$ as $s \to 0$, whence $\eta = 1$. On the other hand, if $\eta < 1$, then $C(s,\dots,s) = o(s)$ as $s \to 0$, whence the lower tail of C exhibits asymptotic independence.

If the function $C(s,\dots,s)$ is regularly varying with index $1/\eta$, then the value of η can be easily computed from

$$\lim_{s\to 0} \frac{\log s}{\log C(s,\dots,s)} = \eta. \tag{3.10}$$

This formula from the fact that $\log \ell(s) = o(\log s)$ whenever ℓ is a slowly varying function at zero. The converse is not true, however: equation (3.10) does not imply that $C(s,\dots,s)$ is regularly varying at zero with index $1/\eta$. A simple sufficient condition for the latter property is given in the following Theorem.

Theorem 3.5.10. *Let C a d-variate Archimedean copula with strict generator ψ. If*

$$\lim_{t\to\infty} \frac{D(\log\psi^{\leftarrow})(dt)}{D(\log\psi^{\leftarrow})(t)} = \frac{1}{d\eta},$$

then the function $s \mapsto C(s,\dots,s)$ is regularly varying at zero with index $1/\eta$.

Proof. Put $f(s) = C(s,\dots,s) = \psi^{\leftarrow}\{d\psi(s)\}$. We have

$$Df(s) = \frac{dD\psi(s)}{D\psi\{f(s)\}},$$

whence

$$\frac{sDf(s)}{f(s)} = d\frac{sD\psi(s)}{f(s)D\psi\{f(s)\}}.$$

On the other hand,

$$D(\log\psi^{\leftarrow})(t) = \frac{1}{\psi^{\leftarrow}(t)D\psi\{\psi^{\leftarrow}(t)\}}.$$

Hence

$$\frac{sDf(s)}{f(s)} = d\frac{D(\log\psi^{\leftarrow})\{d\psi(s)\}}{D(\log\psi^{\leftarrow})\{\psi(s)\}}.$$

The assumptions now imply that

$$\lim_{s\to 0}\frac{sDf(s)}{f(s)} = \frac{1}{\eta}.$$

This condition is sufficient for f to be regularly varying at zero with index $1/\eta$. □

Notice that the condition in Theorem 3.5.10 is fulfilled as soon as the function $-1/D(\log\psi^{\leftarrow})$ is regularly varying at infinity of index τ given by $d^{\tau} = d\eta$, that is, $\tau = 1 + \log(\eta)/\log(d) \leq 1$. However, in this case, a much stronger result is true.

Theorem 3.5.11. *Let C be a d-variate Archimedean copula with strict generator ψ. If the function $\phi = -1/D(\log\psi^{\leftarrow})$ is regularly varying of index $-\infty < \tau \leq 1$ and if $\phi(t) = o(t)$ as $t \to \infty$, then for every $\boldsymbol{x} \in (0,\infty)^d$,*

$$\lim_{s\to 0}\frac{C(sx_1,\dots,sx_d)}{C(s,\dots,s)} = \prod_{i=1}^{d} x_i^{d^{-\tau}}. \tag{3.11}$$

Proof. Since $\phi(t) = o(t)$ as $t \to \infty$ and since ϕ is regularly varying of finite index, we have

$$\frac{\phi\{t + x\phi(t)\}}{\phi(t)} = \frac{\phi[t\{1 + x\phi(t)/t\}]}{\phi(t)} \to 1, \text{ as } t \to \infty$$

locally uniformly in $x \in \mathbb{R}$. By Lemmas 3.2.5–3.2.6, we infer that

$$\lim_{s\to 0}\frac{\psi(sx) - \psi(s)}{\phi\{\psi(s)\}} = -\log(x), \tag{3.12}$$

$$\lim_{t\to\infty}\frac{\psi^{\leftarrow}\{t + y\phi(t)\}}{\psi^{\leftarrow}(t)} = \exp(-y). \tag{3.13}$$

Moreover, the above limit relations hold locally uniformly in $x \in (0,\infty)$ and $y \in \mathbb{R}$. In the terminology of Bingham, Goldie and Teugels (1987), chapter 3, the function ψ belongs to the de Haan class Π, while $\psi^{\leftarrow}$ belongs to the class Γ.

Now we have gathered all the necessary information about the asymptotic properties of ψ and $\psi^{\leftarrow}$. For $\boldsymbol{x} \in (0,\infty)^d$ and for $s \to 0$,

$$\begin{aligned} C(sx_1,\ldots,sx_d) &= \psi^{\leftarrow}\{\psi(sx_1)+\cdots+\psi(sx_d)\} \\ &= \psi^{\leftarrow}\left(d\psi(s)+\sum_{i=1}^{d}\frac{\psi(sx_i)-\psi(s)}{\phi\{\psi(s)\}}\frac{\phi\{\psi(s)\}}{\phi\{d\psi(s)\}}\phi\{d\psi(s)\}\right) \end{aligned}$$

In view of (3.12)–(3.13) and the fact that ϕ is regularly varying at infinity of index τ, the right-hand side of the previous expression is for $s \to 0$ asymptotically equivalent to

$$\psi^{\leftarrow}\{d\psi(s)\}\exp\left(-\sum_{i=1}^{d}(-\log x_i)d^{-\tau}\right) = C(s,\ldots,s)\prod_{i=1}^{d}x_i^{d^{-\tau}},$$

as required. □

Theorem 3.5.12. *Under the conditions of Theorem 3.5.11, for non-empty subset J of $\{1,\ldots,d\}$ and for every $\boldsymbol{x} \in (0,\infty)^d$,*

$$\begin{aligned} &\mathbb{P}(\forall j \in J : U_j \le sx_j; \forall j \in J^c : U_j \le \psi^{\leftarrow}\{x_j^{-1}\phi(\psi(s))\}) \qquad (3.14)\\ &\sim \quad \psi^{\leftarrow}\{|J|\psi(s)\}\prod_{j\in J}x_j^{|J|^{-\tau}}\prod_{j\in J^c}\exp\left(-|J|^{-\tau}x_j^{-1}\right) \quad \text{as } s \to 0. \end{aligned}$$

Proof. For $J = \{1,\ldots,d\}$, the statement reduces to Theorem 3.5.11. So assume that J^c is not empty. The probability on the left-hand side of (3.14) is equal to

$$\psi^{\leftarrow}\left(\sum_{j\in J}\psi(sx_j)+\sum_{j\in J^c}x_j^{-1}\phi(\psi(s))\right) = \psi^{\leftarrow}\{|J|\psi(s)+u(\boldsymbol{x};s)\phi(|J|\psi(s))\}$$

with

$$u(\boldsymbol{x};s) = \left(\sum_{j\in J}\frac{\psi(sx_j)-\psi(s)}{\phi(\psi(s))}+\sum_{j\in J^c}x_j^{-1}.\right)\frac{\phi(\psi(s))}{\phi(|J|\psi(s))}$$

We now proceed as in the proof of Theorem 3.5.11. By equation (3.12) and since $\psi(0) = \infty$ and ϕ is regularly varying at infinity of index τ,

$$u(\boldsymbol{x};s) \to \left(-\sum_{j\in J}\log(x_j)+\sum_{j\in J^c}x_j^{-1}\right)|J|^{-\tau} \text{ as } s \to 0.$$

The stated asymptotic relation now follows from equation (3.13). □

Corollary 3.5.13. *Under the conditions of Theorem 3.5.11, for every $\boldsymbol{x} \in (0,\infty)^d$ and every $(u_j)_{j\in J} \in (0,1]^{|J|}$,*

$$\begin{aligned} &\mathbb{P}(\forall j \in J : U_j \le su_jx_j; \forall j \in J^c : U_j \le \psi^{\leftarrow}\{x_j^{-1}\phi(\psi(s))\} \mid \forall j \in J : U_j \le sx_j) \\ &\to \prod_{j\in J}u_j^{|J|^{-\tau}}\prod_{j\in J^c}\exp\left(-|J|^{-\tau}x_j^{-1}\right), \quad \text{as } s \to 0. \end{aligned}$$

Remark 3.5.14. *In Theorem 3.5.11, if the index of regular variation, τ, of the function ϕ is smaller than one, then automatically $\phi(t) = o(t)$ as $t \to \infty$.*

Remark 3.5.15. *For fixed $0 < s < 1$, the map $x \mapsto \psi^{\leftarrow}\{x^{-1}\phi(\psi(s))\}$ is an increasing homeomorphism from $[0, \infty]$ to $[0, 1]$. For fixed $0 < x < \infty$, the asymptotics of $\psi^{\leftarrow}\{x^{-1}\phi(\psi(s))\}$ as $s \to 0$ depend on those of $\phi(t)$ as $t \to \infty$; in particular, if the latter converges to zero or infinity, then the former converges to one or zero, respectively. However, in all cases, since $\phi(\psi(s)) = o(\psi(s))$ as $s \to 0$ and since the function $\psi^{\leftarrow}$ is regularly varying at infinity of index $-\infty$,*

$$\limsup_{s\to 0} \frac{s}{\psi^{\leftarrow}\{x^{-1}\phi(\psi(s))\}} \leq \limsup_{s\to 0} \frac{\psi^{\leftarrow}\{\psi(s)\}}{\psi^{\leftarrow}\{\psi(s)/2\}} = 0.$$

Hence, in equation (3.14), the variables U_j for $j \in J^c$ are always of larger order of magnitude than the variables U_j for $j \in J$.

Remark 3.5.16. *The limiting distribution function in Corollary 3.5.13 is equal to the product of its marginal distribution functions. Hence, the corresponding limit copula is the independent one.*

Example 3.5.17. *Let $0 < \alpha \leq 1$. The generator $\psi(s) = (-\log s)^{1/\alpha}$, corresponding to the Gumbel copula,*

$$C(u_1, \ldots, u_d) = \exp[-\{(-\log u_1)^{1/\alpha} + \cdots + (-\log u_d)^{1/\alpha}\}^{\alpha}],$$

satisfies the assumptions of Theorem 3.5.11 with $\psi^{\leftarrow}(t) = \exp(-t^{\alpha})$, $\phi(t) = \alpha^{-1}t^{1-\alpha}$, and $\tau = 1 - \alpha$. For $0 < s < 1$ and positive integer k, we have

$$\psi^{\leftarrow}\{k\psi(s)\} = s^{k^{\alpha}},$$

whence, by (3.11), for all $\boldsymbol{x} \in (0, \infty)^d$,

$$C(sx_1, \ldots, sx_d) \sim s^{d^{\alpha}} \prod_{i=1}^{d} x_i^{d^{\alpha-1}}, \text{ as } s \to 0.$$

Moreover, for $0 < s < 1$ and $0 < x < \infty$,

$$\psi^{\leftarrow}\{x^{-1}\phi(\psi(s))\} = \exp\{-(-\log s)^{1-\alpha}(\alpha x)^{-\alpha}\},$$

whence, by (3.14), for all $\boldsymbol{x} \in (0, \infty)^d$ and every non-empty subset J of $\{1, \ldots, d\}$,

$$\begin{aligned} &\mathbb{P}(\forall j \in J : U_j \leq sx_j; \forall j \in J^c : U_j \leq \exp\{-(-\log s)^{1-\alpha}(\alpha x)^{-\alpha}\}) \\ \sim\ & s^{|J|^{\alpha}} \prod_{j\in J} x_j^{|J|^{\alpha-1}} \prod_{j\in J^c} \exp\left(-|J|^{\alpha-1}x_j^{-1}\right) \text{ as } s \to 0. \end{aligned}$$

These relations continue to hold for $\alpha > 1$ provided we modify ψ outside a neighbourhood of zero so that it becomes a proper generator.

Example 3.5.18. *Let ψ be a generator of order d such that there exists $1 < p < \infty$ such that $\psi(s) = \exp\{(-\log s)^{1/p}\}$ for all s in a neighbourhood of zero. Then for sufficiently large t, we have $\psi^{\leftarrow}(t) = \exp\{-(\log t)^p\}$ and $\phi(t) = p^{-1}t(\log t)^{1-p}$. Note that the function ϕ is regularly varying at infinity of index $\tau = 1$ and satisfies $\phi(t)/t \to 0$ as $t \to \infty$. For positive integer k and for positive s in a neighbourhood of zero,*

$$\psi^{\leftarrow}\{k\psi(s)\} = \exp[-\{(-\log s)^{1/p} + \log k\}^p].$$

This expression, for fixed integer $k \geq 2$ and considered as a function of s, is at the same time $o(s)$ as $s \to 0$ and regularly varying at zero of index one. By Theorem 3.5.11, denoting the d-variate Archimedean copula with generator ψ by C, we have for all $\boldsymbol{x} \in (0, \infty)^d$,

$$C(sx_1, \ldots, sx_d) \sim \exp[-\{(-\log s)^{1/p} + \log d\}^p] \prod_{i=1}^{d} x_i^{1/d}, \text{ as } s \to 0.$$

Moreover, for $0 < x < \infty$ and for positive s in a neighbourhood of zero,

$$\begin{aligned} &\psi^{\leftarrow}\{x^{-1}\phi(\psi(s))\} \\ &= \exp\left\{-\Big((-\log s)^{1/p} - (1 - 1/p)\log(-\log s) - \log(px)\Big)^p\right\}. \end{aligned}$$

This expression, for fixed positive x and considered as a function of s, is at the same time regularly varying zero of index one and is of larger order of magnitude than s as $s \to 0$, conform with Remark 3.5.15. By (3.14), it describes the order of magnitude of the variables U_j for $j \in J^c$ conditionally on U_j for $j \in J$ being of order $O(s)$.

Chapter 4

Dynamic dependence ordering for Archimedean copulae

In the context of insurance (and reinsurance) of large claims, Geluk and de Vries (2006) pointed out that "*in case of heavy tailed random variables, apart from the fact that the coefficient of correlation may not be defined, its main disadvantage is that it does not capture very well possible dependence in the tails*". In finance and yield curve modeling, Junker, Szimayer and Wagner (2006) observed that "*dependence in the center of the distribution may be treated separately from the dependence in the distribution tails*", and that symmetric as well as asymmetric tail dependence should be considered.

In this chapter, we will investigate how to compare the strength of the dependence globally and in upper tails. Section 4.1 will recall some basics on stochastic orderings. In section 4.1.1 stochastic ordering of random variables will be considered, and extended to random vectors in 4.1.2. But in order to compare the strength of the dependence, those concepts will be not be sufficient, and the notion of *dependence orderings* will be considered in section 4.1.3. The motivation in that section is to state that dependence orderings should be based on copulae, and only on copulae. From this concept, section 4.2 will focus on the case of Archimedean copulae, and to study properties of the Archimedean generator needed to assess whether random vector $\boldsymbol{X}$ given $\boldsymbol{X} \leq \boldsymbol{x}_1$ is more or less dependent than $\boldsymbol{X}$ given $\boldsymbol{X} \leq \boldsymbol{x}_2$, when $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$. Sections 4.2.1 and 4.2.3 will study properties of $\boldsymbol{X}$ given $\boldsymbol{X} \leq \boldsymbol{x}$, when the copula of $\boldsymbol{X}$ is Archimedean. And then, a characterization of Archimedean copulae which are more and more dependent (in tails) will be given in section 4.2.4.

4.1 Dependence ordering: comparing dependent risks

A natural way to introduce a partial order to compare random variables would be the relation $X \leq_{a.s.} Y$ which holds if and only if $X(\omega) \leq Y(\omega)$ for $\mathbb{P}$-almost all ω. However, this ordering does not only depend on the distributions. In this section, we will discuss some orderings on distribution functions, and abusively call them order relations on random variables, or random vectors.

Historically, stochastic orderings have been introduced in Hardy, Littlewood and Pólya (1934), since the *majorization* concept (used for comparing positive vectors in $\mathbb{R}^d$) can be related to ordering between probability measures (see also Marshall and Olkin (1979)). The dilatation concept extended the majorization one to more general probability measures (see e.g. Strassen (1965)).

Definition 4.1.1. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{R}^d$. Let $a_{i:d}$ denote the i-smallest component of*

$\boldsymbol{a}$, $i = 1, ..., d$, *i.e.* $a_{1:d} \leq a_{2:d} \leq ... \leq a_{d:d}$. *Then* $\boldsymbol{a}$ *is said to be majorized by* $\boldsymbol{b}$*, denoted* $\boldsymbol{a} \prec \boldsymbol{b}$*, if and only if*

$$\sum_{i=1}^{k} a_{i:d} \leq \sum_{i=1}^{k} b_{i:d} \text{ for all } k = 1, ..., d-1, \text{ and } \sum_{i=1}^{d} a_{i:d} = \sum_{i=1}^{d} b_{i:d}.$$

The standard stochastic ordering (stochastic dominance) was only introduced in Lehmann (1955), and most of the other have been studied intensively in the 90's, motivated by different areas of application (queuing theory, reliability, economics, actuarial sciences...).

4.1.1 Stochastic orderings for random variables

Stochastic orderings are partial order relations on the set of distribution functions (see Shaked and Shantikumar (1994) or Müller and Stoyan (2001)). Recall that a partial order relation $\preceq$ on $\mathcal{E}$ is a binary relation $\mathcal{E}$ such that the following conditions hold,

1. (reflexivity) $x \preceq x$ for all $x \in \mathcal{E}$,
2. (transitivity) $x \preceq y$ and $y \preceq z$ imply $x \preceq z$ for all $x, y, z \in \mathcal{E}$,
3. (antisymmetry) $x \preceq y$ and $y \preceq x$ imply $x = y$ for all $x, y \in \mathcal{E}$.

Notice that we might abuse of notions, writing $X \preceq Y$ for two random variables, instead of $\mathbb{P}_X \preceq \mathbb{P}_Y$ (more specifically, $\mathbb{P}_X = \mathbb{P}_Y$ does not imply $X = Y$). Let $\sim$ denote the equality in distribution, i.e. $X \sim Y$ if and only if $X \preceq Y$ and $Y \preceq X$.

Definition 4.1.2. *Consider two random variables X and Y, defined on the same space,*

- *X is said to be smaller than Y for the usual stochastic order ($X \leq_{ST} Y$) if $F_X(t) \geq F_Y(t)$, for all $t \in \mathbb{R}$, or equivalently $\overline{F}_X(t) \leq \overline{F}_Y(t)$, for all $t \in \mathbb{R}$. Another characterization can be the following: $X \leq_{ST} Y$ if and only if*

$$\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y)),$$

  for all real-valued increasing functions ϕ for which both expectations exist (see Müller and Stoyan (2001)).

- *X is said to be smaller than Y for the convex order ($X \leq_{CX} Y$) if*

$$\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y)),$$

  for all real-valued convex functions ϕ for which both expectations exists.

- *X is said to be smaller than Y for the increasing convex order ($X \leq_{ICX} Y$) if*

$$\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y)),$$

  for all real-valued increasing convex functions ϕ for which both expectations exists.

- *X is said to be smaller than Y for the Laplace transform order ($X \leq_{LT} Y$) if X and Y are positive, and*

$$\mathbb{E}(\exp(-uY)) \leq \mathbb{E}(\exp(-uX)), \text{ for all } u \in \mathbb{R}_+.$$

- *X is said to be smaller than Y for the hazard rate order ($X \leq_{HR} Y$) if for all $t \in \mathbb{R}$, $(X|X > t) \leq_{ST} (Y|Y > t)$. If X and Y are absolutely continuous, $X \leq_{HR} Y$ if $r_X(t) \geq r_Y(t)$, $t \geq 0$, where r_X denotes the hazard rate function of X, where $r_X(x) = f_X(x)/\overline{F}_X(x)$.*

- *X is said to be smaller than Y for the likelihood ratio order ($X \leq_{LR} Y$) if for all $a < b$, $(X|X \in [a, b]) \leq_{ST} (Y|Y \in [a, b])$. If X and Y are absolutely continuous, $X \leq_{LR} Y$ if f_X/f_Y is decreasing over the union of the supports of X and Y.*

- *X is said to be smaller than Y for the Lorenz order ($X \leq_L Y$) if $L_X(u) \geq L_Y(u)$ for all $u \in [0, 1]$, where*
$$L_X(u) = \left(\int_0^1 F_X^{\leftarrow}(\omega)d\omega\right)^{\leftarrow} \cdot \int_0^u F_X^{\leftarrow}(\omega)d\omega.$$

Mention here that some of the stochastic orders cannot be used to compare any random variables X and Y, e.g. the Laplace transform or the hazard rate have to be well defined.

Remark 4.1.3. *Some definitions here might differ from definition in the literature. E.g. for the Laplace transform order, we follow here Alzaid, Kim and Proschan (1991) and Shaked and Shantikumar (1994), but some authors say that $X \leq_{LT} Y$ if $\mathbb{E}(\exp(-tX)) \leq \mathbb{E}(\exp(-tY))$ (i.e. $X \leq_{LT} Y$ means that X is larger than Y). Defined as in Definition 4.1.2, note that $X \leq_{ST} Y$ implies that $X \leq_{LT} Y$ (definitions are consistent).*

Example 4.1.4. *If $X \leq_{ST} Y$ and $\mathbb{E}(X) = \mathbb{E}(Y)$ then $X \overset{\mathcal{L}}{=} Y$:*
$$\mathbb{E}(Y) - \mathbb{E}(X) = \int_{\mathbb{R}} [1 - F_Y(t)]\, dt - \int_{\mathbb{R}} [1 - F_X(t)]\, dt = \int_{\mathbb{R}} [F_X(t) - F_Y(t)]\, dt = 0,$$
but since $F_X(t) \geq F_Y(t)$ it does imply that, necessarily $F_X(t) = F_Y(t)$, i.e. $X \overset{\mathcal{L}}{=} Y$.

Example 4.1.5. *The convex order is said to be a variability order, since $X \leq_{CX} Y$ implies $\mathbb{E}(X) = \mathbb{E}(Y)$ and $Var(X) \leq Var(Y)$, if variances exist. Orders $\leq_{CX}$ and $\leq_{ICX}$ have several applications in insurance and economics ($\leq_{ST}$ and $\leq_{ICX}$ are respectively called first and second order stochastic dominance, and denoted $\leq_{VaR}$ and $\leq_{TVaR}$ in Denuit and Charpentier (2004), and denoted $\leq_{SL,=}$ and $\leq_{SL}$ in Denuit, Dhaene, Goovaerts and Kaaas (2005)). Moreover, the convex order is also called the stop-loss order in actuarial sciences, since $X \preceq_{CX} Y$ if and only if for all t, $\mathbb{E}(X - t)_+ \leq \mathbb{E}(Y - t)_+$. Further, note that $X \preceq_L Y$ if and only if $X/\mathbb{E}(X) \preceq_{CX} Y/\mathbb{E}(Y)$. In the standard Cramér-Lundberg model (see Denuit and Charpentier (2004)), with compound Poisson risk processes, with a given premium rate π, and a given frequency λ for the Poisson process (N_t), recall that the time of ruin is the stopping time defined as*
$$\tau_X(u) = \inf\left\{t, u + \pi t - \sum_{i=1}^{N_t} X_i < 0\right\},$$
where the X_i's are i.i.d. positive random variables, and the associated ruin probability is $\psi_X(u) = \mathbb{P}(\tau_X(u) < \infty)$. Then (see Kaas, van Heerwaarden and Goovaerts (1994)), if $X \leq_{ICX} Y$ then $\psi_X(u) \leq \psi_Y(u)$ for all $u \geq 0$.

Example 4.1.6. *The Laplace Transform order has been intensively used in actuarial science (see Heilmann (1986), Cai and Garrido (1998) or Denuit (2001)). In actuarial models, exponential mixtures ore often considered when modeling the cost of some claims, or the duration of some* *disability. Assume (see e.g. Hesselager, Wang and Willmot (1998)) that* $(X|\Theta = \theta) \sim \mathcal{E}(\theta)$*, so that the distribution of* X *can be written*

$$\overline{F}(x) = \mathbb{P}(X > x) = \int_0^\infty \exp(-\theta t)\, d\Pi(t)\,,\ x \geq 0,$$

for some distribution function Π*, then it appears as the Laplace transform of the mixing distribution* Π*. Therefore, it can be seen easily that if* $\Theta_1 \preceq_{LT} \Theta_2$*, then* $\overline{F}_1(x) \leq \overline{F}_2(x)$ *for all* $x \geq 0$*. In life insurance, assume that* X *and* Y *are two remaining lifetimes, then* $X \preceq_{LT} Y$ *means that the whole life premium related to* X *is always smaller than the whole life premium related to* Y *(whatever the interest rate), since*

$$\frac{1 - \mathbb{E}(\exp(-tX))}{t} = \int_0^\infty \exp(-tx)\, \mathbb{P}(X > x)\, dx.$$

Example 4.1.7. *Functions* ϕ *can be interpreted in terms of utility function in von Neuman and Morgenstein model. An agent with (increasing) utility* u *is said to be more risk averse than agent with utility function* v *if* u *is a concave transformation of* v *(the so-called Arrow-Pratt theorem).*

4.1.2 Stochastic orderings for random vectors

Multivariate extensions of some of the orderings defined above can be considered. Note that based on one univariate ordering, several multivariate ordering can be defined as "*natural extensions*" (e.g. for the $\leq_{ST}$ order, the first three orders in Definition 4.1.8 can be considered, see Marshall and Olkin (1979)). The "*natural*" order of $\mathbb{R}^d$ is such that $\boldsymbol{x} \leq \boldsymbol{y}$ means $x_i \leq y_i$ for all $i = 1, ..., d$. Hence, f is an increasing function $\mathbb{R}^d \to \mathbb{R}^k$ if $\boldsymbol{x} \leq \boldsymbol{y}$ implies $f(\boldsymbol{x}) \leq g(\boldsymbol{y})$.

Definition 4.1.8. *Consider two random vectors* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$*, defined on the same space,*

- $\boldsymbol{X}$ *is said to be smaller than* $\boldsymbol{Y}$ *for the usual stochastic order* ($\boldsymbol{X} \leq_{ST} \boldsymbol{Y}$) *if and only if*

$$\mathbb{E}(\phi(\boldsymbol{X})) \leq \mathbb{E}(\phi(\boldsymbol{Y}))\,,$$

  for all real-valued increasing functions $\phi : \mathbb{R}^d \to \mathbb{R}$ *for which both expectations exist.*

- $\boldsymbol{X}$ *is said to be smaller than* $\boldsymbol{Y}$ *for the upper orthant stochastic order* ($\boldsymbol{X} \leq_{UO} \boldsymbol{Y}$) *if and only if*

$$\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) \leq \overline{F}_{\boldsymbol{Y}}(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \mathbb{R}^d.$$

- $\boldsymbol{X}$ *is said to be smaller than* $\boldsymbol{Y}$ *for the lower orthant stochastic order* ($\boldsymbol{X} \leq_{LO} \boldsymbol{Y}$) *if and only if*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \geq F_{\boldsymbol{Y}}(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \mathbb{R}^d.$$

- $\boldsymbol{X}$ *is said to be smaller than* $\boldsymbol{Y}$ *for the convex order* ($\boldsymbol{X} \leq_{CX} \boldsymbol{Y}$) *if*

$$\mathbb{E}(\phi(\boldsymbol{X})) \leq \mathbb{E}(\phi(\boldsymbol{Y}))\,,$$

  for all real-valued convex functions $\phi : \mathbb{R}^d \to \mathbb{R}$ *for which both expectations exists.*

- $\boldsymbol{X}$ *is said to be smaller than* $\boldsymbol{Y}$ *for the increasing convex order* ($\boldsymbol{X} \leq_{ICX} \boldsymbol{Y}$) *if*

$$\mathbb{E}(\phi(\boldsymbol{X})) \leq \mathbb{E}(\phi(\boldsymbol{Y}))\,,$$

  for all real-valued increasing convex functions $\phi : \mathbb{R}^d \to \mathbb{R}$ *for which both expectations exists.*

- *$\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ for the Laplace transform order ($\boldsymbol{X} \leq_{LT} \boldsymbol{Y}$) if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are positive, and*

$$\mathbb{E}\left(\exp\left(-\boldsymbol{u}^t\boldsymbol{X}\right)\right) \leq \mathbb{E}\left(\exp\left(-\boldsymbol{u}^t\boldsymbol{Y}\right)\right), \text{ for all } \boldsymbol{u} \in \mathbb{R}_+^d.$$

- *$\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the likelihood order denoted $\boldsymbol{X} \leq_{LR} \boldsymbol{Y}$ if $\boldsymbol{X}$ and $\boldsymbol{Y}$ have densities, with*

$$f_{\boldsymbol{X}}(\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{x}) \geq f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y}),$$

for all $\boldsymbol{x} \leq \boldsymbol{y}$ (componentwise ordering in $\mathbb{R}^d$).

- *$\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the TP2 order (totally positive of order 2, or strong likelihood order) denoted $\boldsymbol{X} \leq_{TP2} \boldsymbol{Y}$ if $\boldsymbol{X}$ and $\boldsymbol{Y}$ have densities, with*

$$f_{\boldsymbol{X}}(\boldsymbol{x} \wedge \boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{x} \vee \boldsymbol{y}) \geq f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y}), \text{for all } \boldsymbol{x} \text{ and } \boldsymbol{y}.$$

- *$\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in the supermodular order denoted $\boldsymbol{X} \leq_{SM} \boldsymbol{Y}$ if*

$$\mathbb{E}(\phi(\boldsymbol{X})) \leq \mathbb{E}(\phi(\boldsymbol{Y})),$$

for all supermodular functions $\phi : \mathbb{R}^d \to \mathbb{R}$ for which both expectations exists, i.e.

$$\phi(\boldsymbol{x} \wedge \boldsymbol{y}) \phi(\boldsymbol{x} \vee \boldsymbol{y}) \geq \phi(\boldsymbol{x}) \phi(\boldsymbol{y}).$$

Note that for several multivariate orderings, it might be interesting to have some characterizations based on univariate stochastic order. For instance, $\boldsymbol{X} \leq_{ST} \boldsymbol{Y}$ if and only if

$$\Phi(\boldsymbol{X}) \leq_{ST} \Phi(\boldsymbol{Y}),$$

for all increasing function $\Phi : \mathbb{R}^n d \to \mathbb{R}$.

Example 4.1.9. *The supermodular order for random vectors is important in insurance, since $\boldsymbol{X} \leq_{SM} \boldsymbol{Y}$ implies that $X_1 + ... + X_d \leq_{ICX} Y_1 + ... + Y_d$, i.e. the stop-loss premium for some portfolios $\boldsymbol{X}$ and $\boldsymbol{Y}$ of size n should be ordered as $\boldsymbol{X}$ and $\boldsymbol{Y}$ for the supermodular order (see Müller (1997)).*

Example 4.1.10. *Hu, Khaledi and Shaked (2003) introduced the following stochastic order: $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ for the "weak multivariate hazard rate order", denoted $\boldsymbol{X} \leq_{WHR} \boldsymbol{Y}$ if and only if $\overline{F}_{\boldsymbol{Y}}(\boldsymbol{x})/\overline{F}_{\boldsymbol{X}}(\boldsymbol{x})$ is increasing in $\boldsymbol{x}$. Note that the name of this ordering comes from another characterization: $\boldsymbol{X} \leq_{WHR} \boldsymbol{Y}$ if and only if $X_i \leq_{HR} Y_i$ for all $i = 1, ..., n$. Equivalently, it can be written more explicitly*

$$(\boldsymbol{X}|\boldsymbol{X} > \boldsymbol{h}) \leq_{UO} (\boldsymbol{Y}|\boldsymbol{Y} > \boldsymbol{h}) \text{ for all } \boldsymbol{h}.$$

See also Example 4.1.13 for a similar relationship.

Note that some of those stochastic orders are stronger than others.

$$\boldsymbol{X} \leq_{TP2} \boldsymbol{Y} \Rightarrow \left\{ \begin{array}{l} \boldsymbol{X} \leq_{LR} \boldsymbol{Y} \\ \boldsymbol{X} \leq_{ST} \boldsymbol{Y} \Rightarrow \left\{ \begin{array}{l} \boldsymbol{X} \leq_{UO} \boldsymbol{Y} \\ \boldsymbol{X} \leq_{LO} \boldsymbol{Y} \end{array} \right. \end{array} \right.$$

and independently

$$\left\{ \begin{array}{l} \boldsymbol{X} \leq_{SM} \boldsymbol{Y} \Rightarrow \boldsymbol{X} \leq_{UO} \boldsymbol{Y} \\ \boldsymbol{X} \leq_{CX} \boldsymbol{Y} \Rightarrow \boldsymbol{X} \leq_{ICX} \boldsymbol{Y} \end{array} \right.$$

Note further that those stochastic orders imply some orderings on margins. For instance, if $\boldsymbol{X} \leq_{ST} \boldsymbol{Y}$, then $X_i \leq_{ST} Y_i$ for all $i \in \{1,...,n\}$ (closure under marginalization). And this property is also true for $\leq_{CX}$, $\leq_{ICX}$, or $\leq_{SM}$.

Hence, those orders might not be interesting to compare the strength of the dependence since the orderings are mainly influenced by marginal behaviors.

In dimension 2, the following conditions should hold to define a stochastic order, as in Kimeldorf and Sampson (1989),

Definition 4.1.11. *Let F_1 and F_2 denote two distribution functions, and $\mathcal{F}(F_1, F_2)$ the associated Fréchet class. A binary function $\preceq$ is said to be a stochastic order on $\mathcal{F}(F_1, F_2)$ if it fulfills the following properties,*

1. *(identical margins) $F \preceq G$ implies that $F, G \in \mathcal{F}(F_1, F_2)$,*
2. *(concordance) $F \preceq G$ implies $F(x,y) \leq G(x,y)$ for all x, y,*
3. *(transitivity) $F \preceq G$ and $G \preceq H$ implies $F \preceq H$,*
4. *(reflexivity) $F \preceq F$,*
5. *(antisymmetry) $F \preceq G$ and $G \preceq F$ implies that $F = G$,*
6. *(bounds) $F^- \preceq F \preceq F^+$ where F^- and F^+ are respectively the lower and the upper Fréchet-Hoeffding bounds of $\mathcal{F}(F_1, F_2)$,*
7. *(closure, weak convergence) $F_n \preceq G_n$ for all $n \in \mathbb{N}$, $F_n \to F$ and $G_n \to G$ imply $F \preceq G$,*
8. *(index order) $(X_1, X_2) \preceq (Y_1, Y_2)$ implies $(X_2, X_1) \preceq (Y_2, Y_1)$,*
9. *(increasing transform) $(X_1, X_2) \preceq (Y_1, Y_2)$ implies $(\phi(X_1), X_2) \preceq (\phi(Y_1), Y_2)$ for all strictly increasing functions $\phi : \mathbb{R} \to \mathbb{R}$,*
10. *(decreasing transform) $(X_1, X_2) \preceq (Y_1, Y_2)$ implies $(\phi(X_1), X_2) \succeq (\phi(Y_1), Y_2)$ for all strictly decreasing functions $\phi : \mathbb{R} \to \mathbb{R}$.*

Such a definition can be extend in higher dimension (see e.g. Joe (1997) or Müller and Stoyan (2001)), on some Fréchet-space: random vectors can be compared if and only if they have identical margins.

Example 4.1.12. *The so-called concordance order (see Tchen (1980) or Joe (1997)), defined as $(X_1, X_2) \leq_C (Y_1, Y_2)$ if and only if*

$$cov(\phi(X_1), \psi(X_2)) \leq cov(\phi(Y_1), \psi(Y_2)),$$

for all increasing functions ϕ and ψ. Note that this order is an integral order (for supermodular functions), and moreover, it is the only integral order satisfying all the properties of Definition 4.1.11 (see Müller and Stoyan (2001)).

Example 4.1.13. *Let $F_{\boldsymbol{X}}$ and $F_{\boldsymbol{Y}}$ denote two distributions in the same Fréchet class. Colangelo, Scarsini and Shaked (2004) introduced some stochastic orders for tails. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two random vectors with distribution function $F_{\boldsymbol{X}}$ and $F_{\boldsymbol{Y}}$ respectively, $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ for the "upper orthant increasing ratio order", denoted $\boldsymbol{X} \leq_{UOIR} \boldsymbol{Y}$ if and only if $\overline{F}_{\boldsymbol{Y}}(\boldsymbol{x})/\overline{F}_{\boldsymbol{X}}(\boldsymbol{x})$ is increasing in $\boldsymbol{x}$. Equivalently, it can be written more explicitly*

$$(\boldsymbol{X} - \boldsymbol{h} | \boldsymbol{X} > \boldsymbol{h}) \leq_{UO} (\boldsymbol{Y} - \boldsymbol{h} | \boldsymbol{Y} > \boldsymbol{h}) \text{ for all } \boldsymbol{h} \text{ such that } \overline{F}_{\boldsymbol{X}}(\boldsymbol{h}), \overline{F}_{\boldsymbol{Y}}(\boldsymbol{h}) > 0.$$

Note that this ordering satisfies most of the axioms of Definition 4.1.11, except the 6th one: the upper upper Fréchet bound does not always dominate every distribution, hence, we do not necessarily have

$$(\boldsymbol{X} - \boldsymbol{h} | \boldsymbol{X} > \boldsymbol{h}) \leq_{UO} (\boldsymbol{X}^+ - \boldsymbol{h} | \boldsymbol{X}^+ > \boldsymbol{h}) \text{ for all } \boldsymbol{h}$$

4.1.3 Dependence and copulae

As pointed out in Definition 4.1.11, stochastic orders for random vectors can be interpreted only in the case where random vectors are in the same Fréchet space. But in some situations, it could be interesting to say that, in some sense, $\boldsymbol{X}$ is more dependent, or more positively dependent, than $\boldsymbol{Y}$, without taking into account any marginal behavior. As we shall see in the next section, as well as concordance measures should not dependent on marginal behavior, dependence orderings should not dependent on marginal orderings.

Example 4.1.14. *In Chapter 2 of this thesis, the interest is to compare the strength of dependence between a random vector $\boldsymbol{X}$ of individual X_i's time before default, at time 0, and the same vector at time t given that no default occurred, i.e. $\boldsymbol{X}$ given $\{\boldsymbol{X}|t \cdot \mathbf{1}\} = \{X_1 > t, ..., X_d > t\}$, for some $t > 0$. Since those vectors obviously do not have the same margins, in order to compare the strength of dependence, some dependence order $\preceq$ should be properly define, to see whether $\boldsymbol{X} \preceq (\boldsymbol{X}|\boldsymbol{X} > t \cdot \mathbf{1})$.*

Example 4.1.15. *Analogously, in Chapter 3 of this thesis, the interest is to compare the strength of dependence between a random vector $\boldsymbol{X}$ of risk variables, $\boldsymbol{X}$ given that $\boldsymbol{X} \in \mathcal{A}$ where $\mathcal{A}$ is an "extremal" subset of $\mathbb{R}^d$. Again, margins can be sensibly different, and usual stochastic orderings can be be used directly to assess whether $\boldsymbol{X}$ is more dependent in its tails.*

Dependence order, or copula-based ordering

Definition 4.1.16. *Let $\mathcal{C}$ denote the set of copulae. A binary function $\preceq$ is said to be a positive dependence order if it fulfills the following properties,*

1. *(concordance) $C_{\boldsymbol{X}} \preceq C_{\boldsymbol{Y}}$ implies $C_{\boldsymbol{X}}(u,v) \leq C_{\boldsymbol{Y}}(u,v)$ for all $u,v \in [0,1]$,*
2. *(transitivity, reflexivity and antisymmetry) $\preceq$ is an order relation,*
3. *(bounds) $C^- \preceq C \preceq C^+$ where C^- and C^+ are respectively the lower and the upper Fréchet-Hoeffding bounds of $\mathcal{C}$,*
4. *(transposition) $C_{\boldsymbol{X}} \preceq C_{\boldsymbol{Y}}$ implies $C'_{\boldsymbol{X}} \preceq C'_{\boldsymbol{Y}}$ where $C'_{\cdot}(u,v) = C_{\cdot}(v,u)$ for all $u,v \in [0,1]$,*
5. *(closure, weak convergence) $C_{\boldsymbol{X},n} \preceq C_{\boldsymbol{Y},n}$ for all $n \in \mathbb{N}$, $C_{\boldsymbol{X},n} \xrightarrow{w} C_{\boldsymbol{X}}$ and $C_{\boldsymbol{Y},n} \to C_{\boldsymbol{Y}}$ imply $C_X \preceq C_{\boldsymbol{Y}}$.*

Based on this definition, let us prove now the the bijection between ordering on some Fréchet spaces $\mathcal{F}(F_1, F_2)$, and ordering on the set of copulae.

Proposition 4.1.17. *There is a correspondence between the ordering on the Fréchet space (for continuous marginal distributions), and the induced ordering on copulae functions. More precisely, let $\preceq$ denote a dependence relation on $\mathcal{C}$, and $\leq_*$ denote the relation defined as follows: $F \leq_* G$ if and only if $F, G \in \mathcal{F}(F_1, F_2)$ and $C_F \preceq C_G$ (where C_F and C_G are the copulae of F and G respectively). Then $\leq_*$ is a positive dependence order (in the sense of Definition 4.1.11).*

Proof. Let us prove that all the items are satisfied

1. is satisfied by construction, $F, G \in \mathcal{F}(F_1, F_2)$.
2. (concordance) $F \leq_* G$ implies $C_F \preceq C_G$ and so, $C_F(u,v) \leq C_G(u,v)$ for all $u,v \in [0,1]$, and so.$C_F(F_1(x), F_2(y)) \leq C_G(F_1(x), F_2(y))$ for all x, y, i.e. $F(x,y) \leq G(x,y)$.
3. (transitivity, reflexivity and antisymmetry) trivially obtained.

6. (bounds) since $C^{-}(u,v) \leq C(u,v) \leq C^{+}(u,v)$, for all $u,v \in [0,1]$, it follows that $C^{-}(F_1(x), F_2(y)) \leq C(F_1(x), F_2(y)) \leq C^{+}(F_1(x), F_2(y))$ for all x,y, i.e. $F^{-} \leq F \leq F^{+}$.

7. (closure, weak convergence) Consider $F_n \leq_* G_n$ for all $n \in \mathbb{N}$, $F_n \to F$ and $G_n \xrightarrow{w} G$, and copulae $C_{F,n} \preceq C_{G,n}$. Because $F_n = C_{F,n}(F_1, F_2) \to F = C_F(F_1, F_2)$ then $C_{F,n} \to C_F$ and analogously, $C_{G,n} \to C_G$. Since $\preceq$ is a dependence relation on $\mathcal{C}$, then $C_F \preceq C_G$, and so $F = C_F(F_1, F_2) \leq_* C_G(F_1, F_2) = G$.

8. (index order) Consider (X_1, X_2) and (Y_1, Y_2) with respective distribution functions F and G, and respective copulae C_F and C_G. Assume that $(X_1, X_2) \preceq (Y_1, Y_2)$, i.e. $F \leq_* G$. Denote by F' and G' the distribution functions of (X_2, X_1) and (Y_2, Y_1) respectively. Then $F' = C'_F(F_2, F_1)$ where $C'_F(u,v) = C_F(v,u)$ and analogously, $G' = C'_G(F_2, F_1)$. Since $\preceq$ is a dependence relation on $\mathcal{C}$, then $C'_F \preceq C'_G$, and so $F' \leq_* G'$.

9. (increasing transform) because (X_1, X_2) and $(\phi(X_1), X_2)$ have the same copula for all strictly increasing functions $\phi : \mathbb{R} \to \mathbb{R}$, item 9 holds,

10. (decreasing transform) If ϕ is a strictly decreasing function $\mathbb{R} \to \mathbb{R}$, the copula of $(\phi(X_1), X_2)$ is $C"_F(u,v) = v - C_F(1-u, v)$, where C_F is the copula of (X_1, X_2). And if $C_F \preceq C_G$, then $C"_F \succeq C"_G$. So finally, item 10 holds.

□

Conversely, it is possible to define some "*copula based stochastic orderings*", here in dimension 2 (but this can easily be extended in higher dimension),

Proposition 4.1.18. *Let $\leq_*$ denote some stochastic ordering defined on some Fréchet space (see Definition 4.1.11), and define $\preceq_*$ on the Fréchet space of copulae, i.e. $C_1 \preceq_* C_2$ if and only if for any (univariate) distribution functions F_X, F_Y, $C_1(F_X, F_Y) \leq_* C_2(F_X, F_Y)$. Then $\preceq_*$ is a stochastic ordering (as in Definition 4.1.11).*

Proof. Consider two distributions F_X and F_Y.

1. (concordance)trivial, by construction,

2. (transitivity, reflexivity and antisymmetry) easy to check,

5. (bounds) Since $\leq_*$ defines a stochastic ordering on $\mathcal{F}(F_X, F_Y)$, then, due to Fréchet-Hoeffding bounds on $\mathcal{F}(F_X, F_Y)$,

$$\max\{F_X + F_Y - 1, 0\} \leq_* C(F_X, F_Y) \leq_* \min\{F_X, F_Y\}$$

   Therefore, $C^{-} \preceq_* C \preceq_* C^{+}$.

6. (transposition) the transposition property of $\preceq_*$ is directly deduced from the *index property* of $\leq_*$.

7. (closure, weak convergence) Consider some sequences $C_{1,n}$ and $C_{2,n}$ such that $C_{1,n} \preceq C_{2,n}$ for all $n \in \mathbb{N}$, with $C_{1,n} \to C_1$ and $C_{2,n} \to C_2$. Given F_X and F_Y, set $F_{i,n} = C_{i,n}(F_X, F_Y)$, $i = 1, 2$. Recall that $C_{1,n} \preceq_* C_{2,n}$ if and only if $F_{1,n} \leq_* F_{2,n}$ for all n. And moreover, due to the uniform continuity of copulae, $C_{i,n}(F_X, F_Y) \to C_i(F_X, F_Y)$ as $n \to \infty$, i.e. $F_{i,n} \to F_i$. Finally, the *closer property* of $\leq_*$ allows to conclude.

□

To avoid confusion, $\preceq_*$ will also be denoted $\preceq_{d-*}$, for *dependence* order

Definition 4.1.19. *Let $\leq_*$ denote some stochastic ordering defined on some Fréchet space (see Definition 4.1.11). Then $\boldsymbol{X}$ is said to be dominated by $\boldsymbol{Y}$ for the dependence order induced by $\leq_*$ (or copula-based stochastic order induced by $\leq_*$), denoted $\boldsymbol{X} \preceq_{d-*} \boldsymbol{Y}$ if and only if $F_{\boldsymbol{X}}(\boldsymbol{X}) \leq_* F_{\boldsymbol{Y}}(\boldsymbol{Y})$, i.e. in dimension n, $(F_{X,1}(X_1), ..., F_{X,d}(X_d)) \leq_* (F_{Y,1}(Y_1), ..., F_{Y,d}(Y_d))$.*

Example 4.1.20. *A dependence ordering can be deduced from the PQD order. Let X and Y be two continuous random vectors in dimension $d = 2$, then $\boldsymbol{X} \preceq_{d-PQD} \boldsymbol{Y}$ if and only if $C_{\boldsymbol{X}} \leq C_{\boldsymbol{Y}}$, or equivalently $C^*_{\boldsymbol{X}} \leq C^*_{\boldsymbol{Y}}$.*

Example 4.1.21. *A dependence ordering can be deduced from the CI order. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two continuous random vectors, then $\boldsymbol{X} \preceq_{d-PQD} \boldsymbol{Y}$ if and only if $C_{\boldsymbol{X}} \leq C_{\boldsymbol{Y}}$, for the pointwise order.*

From this definition, the following result can be immediately deduced,

Corollary 4.1.22. *Given a stochastic order $\leq_*$, $\boldsymbol{X} = (X_1, X_2) \preceq_{d-*} \boldsymbol{Y} = (Y_1, Y_2)$ if and only if for any $\phi, \psi : \mathbb{R} \to \mathbb{R}$ strictly increasing,*

$$(\phi(X_1), \psi(X_2)) \preceq_{d-*} (\phi(Y_1), \psi(Y_2)).$$

Dependence order for Archimedean copulae

Recall that a copula C in dimension n is said to be Archimedean if it can be written

$$C(x_1, ..., x_d) = \phi^{\leftarrow}(\phi(x_1) + ... + \phi(x_d)),$$

where $\phi : [0, 1] \to [0, +\infty)$ is a decreasing convex function, such that $\phi(1) = 0$ and $\phi(x) \to \infty$ as $x \to 0$, and such that $\phi^{\leftarrow}$ is a d-completely monotone function (see sections 1.5 for a presentation, and 3.2 or 5.1 for alternative characterizations). Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote two continuous random vectors, with respective copulae C_X and C_Y respectively, assumed to be Archimedean, with generator ϕ_X and ϕ_Y. Then, from Genest and MacKay (1986b), if $\phi_X \circ \phi_Y^{\leftarrow}$ is concave, or if ϕ_X/ϕ_Y is increasing, then $\boldsymbol{X} \preceq_{d-ST} \boldsymbol{Y}$.

Example 4.1.23. *From the upper orthant increasing ratio order introduced in Example 4.1.13, the associated dependence order can be considered: $\boldsymbol{X} \leq_{d-UOIR} \boldsymbol{Y}$ if and only if $C^*_{\boldsymbol{Y}}(\boldsymbol{u})/C^*_{\boldsymbol{X}}(\boldsymbol{u})$ is increasing in $\boldsymbol{u}$. Colangelo, Scarsini and Shaked (2004) gave a nice characterization of the inequality $\boldsymbol{X}^{\perp} \leq_{d-UOIR} \boldsymbol{X}$ in the bivariate case. Define the RTI notion as follows: X is said to be right tail increasing in Y ($RTI(X|Y)$) if $y \mapsto \mathbb{P}(X > x|Y > y)$ is increasing, for all x (this is a concept of positive dependence, define in higher dimension in Definition 4.1.24). Then $(X^{\perp}, Y^{\perp}) \leq_{d-UOIR} (X, Y)$ if and only if $RTI(F_X(X)|F_Y(Y))$ and $RTI(F_Y(Y)|F_X(X))$. And in the case of Archimedean copulae with generator ϕ, $(X^{\perp}, Y^{\perp}) \leq_{d-UOIR} (X, Y)$ if and only if $\log \phi^{\leftarrow}$ is concave. As we will see in section 4.2.4 of this chapter, the log-concavity of the first derivative of the Archimedean copula has also an interpretation in terms of stochastic orderings, in upper tails.*

4.1.4 Positive dependence

Several concepts of "*positive*" and "*negative*" dependence can be define, most of them being derived from dependence orders, based on the comparison between $\boldsymbol{X}$ and $\boldsymbol{X}^{\perp}$ (see e.g. Lehmann (1966)), where $\boldsymbol{X}^{\perp}$ denotes an independent version of $\boldsymbol{X}$ (i.e. $X_i \stackrel{\mathcal{L}}{=} X_i^{\perp}$, $i = 1, ..., n$, and the components of $\boldsymbol{X}^{\perp}$ are independent).

Definition 4.1.24. *Let $\boldsymbol{X}$ denote some stochastic random vector, and $\boldsymbol{X}^{\perp}$ an independent version,*

- *$\boldsymbol{X}$ is said to be positively upper orthant dependent (PUOD) if $\boldsymbol{X}^{\perp} \preceq_{UO} \boldsymbol{X}$, i.e.*

$$F(\boldsymbol{x}) = F(x_1, ..., x_d) \geq F_1(x_1) \cdot ... \cdot F_d(x_d) = F^{\perp}(\boldsymbol{x}).$$

  and $\boldsymbol{X}$ is said to be positively lower orthant dependent (PLOD) if $\boldsymbol{X} \preceq_{LO} \boldsymbol{X}^{\perp}$

$$\overline{F}((x) = \overline{F}(x_1, ..., x_n) \geq \overline{F}_1(x_1) \cdot ... \cdot \overline{F}_d(x_d) = \overline{F}^{\perp}(\boldsymbol{x}).$$

  If $n = 2$ those two notions are equivalent, and $\boldsymbol{X}$ is then said to be PQD (positive quadrant dependent).

- *$\boldsymbol{X}$ is said to be supermodular dependent (SMD) if $\boldsymbol{X}^{\perp} \leq_{SM} \boldsymbol{X}$.*

- *$\boldsymbol{X}$ is said to be (strongly) positively associated if*

$$Cov(\phi(\boldsymbol{X}), \psi(\boldsymbol{X})) \geq 0,$$

  for all increasing bounded functions ϕ and $\psi : \mathbb{R}^d \to \mathbb{R}$.

- *$\boldsymbol{X}$ is said to be weakly positively associated if*

$$Cov\left(\phi(\boldsymbol{X}_*), \psi\left(\boldsymbol{X}'_*\right)\right) \geq 0,$$

  for all subset of components $\boldsymbol{X}_*$ ($\boldsymbol{X}'_*$ denoting the complementary) increasing bounded functions $\phi : \mathbb{R}^k \to \mathbb{R}$ and $\psi : \mathbb{R}^{n-k} \to \mathbb{R}$.

- *$\boldsymbol{X}$ is said to be MTP2 if $\boldsymbol{X}^{\perp} \preceq_{LR} \boldsymbol{X}$, or equivalently, if $\boldsymbol{X}$ has density f such that*

$$f(\boldsymbol{x} \wedge \boldsymbol{y}) f(\boldsymbol{x} \vee \boldsymbol{y}) \geq f(\boldsymbol{x}) f(\boldsymbol{y}),$$

  for all $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e. its density is supermodular.

- *$\boldsymbol{X}$ is said to be positive dependent through stochastic ordering (PDS) if $(X_i|X_j = x) \leq_{ST} (X_i|X_j = y)$ for all $i \neq j$, $x < y$.*

- *$\boldsymbol{X}$ is said to be conditional increasing in sequence (CIS) if*

$$(X_i|X_1 = x_1, ..., X_{i-1} = x_{i-1}) \leq_{ST} (X_i|X_1 = y_1, ..., X_{i-1} = y_{i-1})$$

  for all $i = 2, ..., d$, $x_k \leq y_k$.

- *$\boldsymbol{X}$ is said to be conditional increasing (CI) if*

$$\left(X_i|X_{j_1} = x_1, ..., X_{j_{i-1}} = x_{i-1}\right) \leq_{ST} \left(X_i|X_{j_1} = y_1, ..., X_{j_{i-1}} = y_{i-1}\right)$$

  for all $(j_1, ..., j_{i-1}) \subset \{1, ..., d\}$, $i = 2, ..., d$, $x_k \leq y_k$.

- $\boldsymbol{X}$ *is said to be right tail increasing (RTI, or right corner set increasing, RCSI) if* $\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}|\boldsymbol{X} > \boldsymbol{x}')$ *is increasing in* $\boldsymbol{x}'$ *for all* $\boldsymbol{x}$.
- $\boldsymbol{X}$ *is said to be left tail increasing (LTD) if* $\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}|\boldsymbol{X} \leq \boldsymbol{x}')$ *is decreasing in* $\boldsymbol{x}'$ *for all* $\boldsymbol{x}$.

As we shall see in Section 4.1.5 most of those ordering and induced positive dependence concepts appear when modeling and comparing risks.

Some of the relations between various concepts of positive dependence can be summarized as (see Müller and Stoyan (2001) or Christofides and Vaggelatou (2004)),

$$MTP2 \Longrightarrow \left\{ \begin{array}{l} CI \Longrightarrow CIS \Longrightarrow \text{Associated} \Longrightarrow SMD \Longrightarrow \left\{ \begin{array}{l} PUOD \\ PLOD \end{array} \right. \\ \left\{ \begin{array}{l} RTI \Longrightarrow \left\{ \begin{array}{l} PUOD \\ PLOD \end{array} \right. \\ LTD \Longrightarrow \left\{ \begin{array}{l} PUOD \\ PLOD \end{array} \right. \end{array} \right. \end{array} \right.$$

Remark 4.1.25. *Note that if* $\mathcal{F}(F_1, ..., F_d)$ *defines the Fréchet class associated with marginal distributions* $F_1, .., F_d$*, one can also define the positive-Fréchet class* $\mathcal{F}^+_{PLOD}(F_1, ..., F_d)$*, of n-dimensional distributions, with marginal distributions* $F_1, .., F_d$ *that are positively dependent in the PLOD sense.*

Observe that several dependence properties for random vectors $\boldsymbol{X}$ can be expressed as properties on the underlying copula C.

Example 4.1.26. $\boldsymbol{X} = (X_1, X_2)$ *is PQD if and only if* $C^\perp \leq C$.

In fact, several concepts of positive dependence can be simply transposed on copula functions since they do not depend on marginal behaviors.

4.1.5 Application to competing default risks

Standard competing risk models usually assume independent competing events and focus on the distribution of the date of occurrence of the first (or more generally, the k-th) event. This leads to the study of the distribution of the corresponding order statistics.

Consider a credit portfolio, with homogeneous credits (i.e. the same design: initial balance, interest rate, maturity...). The borrowers may default and the defaults can be characterized by the duration variables giving the time before default for each individual. The distribution of those order statistics should be known to price derivatives, and assuming the independence between the competing risks (e.g. individual defaults), we neglect the possibility of default correlation, and as a consequence, the derivatives are likely mispriced.

Assume that $X_1, ..., X_n$ have identical distribution, characterized by c.d.f. F_X, assumed to be absolutely continuous. Let $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ denote the associate order statistics. Set $U_i = F_X(X_i)$, $i = 1, ..., n$, then $X_{i:n} = F_X^{\leftarrow}(U_{i:n})$, so that dependence properties can be studied through standardized durations U_i. Moreover, consider the following aging concepts,

Definition 4.1.27. *Let* X *denote a random variable with distribution function* $F_X(\cdot)$*.* X *is said to be IFR (Increasing Failure Rate) if* $t \mapsto \overline{F}_X(t + x)/\overline{F}_X(t)$ *is decreasing for all* $x \geq 0$*, or equivalently* $-\log \overline{F}_X$ *is convex. Respectively, it is said to be DFR (Decreasing Failure Rate) if* $t \mapsto \overline{F}_X(t + x)/\overline{F}_X(t)$ *is increasing for all* $x \geq 0$*, or equivalently* $-\log \overline{F}_X$ *is concave. Further,* X *is said to be NBU (New Better than Used) if* $\overline{F}_X(y + x) \leq \overline{F}_X(x) \cdot \overline{F}_X(y)$ *for all* $x, y \geq 0$*, or equivalently* $-\log \overline{F}_X$ *is subadditive.*

If $n = 2$, one gets

$$\mathbb{P}(U_{1:2} \leq u) = 1 - \mathbb{P}(U_1 > u, U_2 > u) = 2u - C(u, u)$$

$$\mathbb{P}(U_{2:2} \leq u) = \mathbb{P}(U_1 \leq u, U_2 \leq u) = C(u, u)$$

Thus, both marginal distributions of order statistics depend on the value of the copula on the diagonal.

Example 4.1.28. *Let $X_1, ...X_n$ be a sequence of random variables. For all $0 < i < j < n$, then $X_{i:n} \leq_{ST} X_{j:n}$. Further, if variables X_i's are independent, then $X_{i:n} \leq_{HR} X_{j:n}$ (see Shaked and Yao (1991)). And moreover, if variables X_i's are independent and identically distributed, then $X_{i:n} \leq_{LR} X_{j:n}$.*

If variables X_i's are independent and identically distributed, with an absolutely continuous distribution, then $X_{i:n}$ and $X_{j:n}$ are TP2 for all $0 \leq i < j \leq n$. Further, $X_{j:n}$ is RTI in $X_{i:n}$ (see Boland, Hollander, Joag-Dev and Kochar (1994)).

Example 4.1.29. *Let $X_1, ...X_n$ and $Y_1, ..Y_n$ be two sequences of independent random variables, such that $X_i \leq_{ST} Y_i$ for all $i = 1, .., n$. Then $X_{i:n} \leq_{ST} Y_{i:n}$ for $i = 1, .., n$. Further, if the variables are exponentially distributed, with parameters $\lambda_{X,i}$ and $\lambda_{Y,i}$ respectively, so that*

$$(\lambda_{X,1}, ..., \lambda_{X,n}) \prec (\lambda_{Y,1}, ..., \lambda_{Y,n}),$$

($\prec$ being the majorization order), then

$$(X_{1:n}, ..., X_{n:n}) \leq_{ST} (Y_{1:n}, ..., Y_{n:n}),$$

and in particular $X_{i:n} \leq_{ST} Y_{i:n}$. Moreover (see Avérous (2002)), if the X_i's have a common distribution, as well as the Y_i's, then

$$(X_{i:n}, X_{j:n}) \leq_{ST} (Y_{i:n}, Y_{j:n}) \text{ if and only if } X \leq_{HR} Y.$$

Example 4.1.30. *Let N denote a counting variable, and define $X_{1:N} = \min\{X_1, X_2, ..., X_N\}$ where $X_1, X_2, ...$ is a sequence of independent and identically distributed random variables, with common c.d.f. F. Analogously, define $X_{N:N} = \max\{X_1, X_2, ..., X_N\}$, and let $F_{1:N}$ and $F_{N:N}$ denote the c.d.f. of the maxima and the minima. If F is DFR then $F_{1:N}$ is also DFR, and if F is IFR so is $F_{N:N}$ (see Shaked (1977)). Further, if F is NBU then $F_{N:N}$ is also NBU (see Bartoszewicz (2001)).*

4.1.6 The strongest dependence concept: comonotonicity

Recall from Section 1.4 that comonotonicity between two risks is obtained when their underlying copula is the upper-Fréchet-Hoeffding bound. In the following Proposition, some equivalent characterization are given for the comonotonicity of a random vector

Proposition 4.1.31. *A random vector $\boldsymbol{X} = (X_1, ..., X_d)$ is comonotonic if and only if one of the following equivalent conditions holds,*

1. *$\boldsymbol{X}$ has a comonotonic copula, i.e. for all $x_1, ..., x_d$,*

$$F(x_1, ..., x_d) = \min\{F_1(x_1), F_2(x_2), ..., F_d(x_d)\}, \tag{4.1}$$

2. *For U uniformly distributed on $[0, 1]$,*

$$\boldsymbol{X} \stackrel{\mathcal{L}}{=} (F_1^{\leftarrow}(U), F_2^{\leftarrow}(U), ..., F_d^{\leftarrow}(U)), \tag{4.2}$$

3. *There is a random variable Z and some increasing functions $f_1, ..., f_d$ such that*

$$\boldsymbol{X} \overset{\mathcal{L}}{=} \left(f_1\left(Z\right), f_2\left(Z\right), ..., f_d\left(Z\right)\right), \tag{4.3}$$

4. *All pairs (X_i, X_j) are comonotonic, $i, j = 1, ..., d$.*

Proof. See Schmeidler (1986), **?** or Müller and Stoyan (2001). □

Remark 4.1.32. *Note that comonotonicity is a fundamental concept in economics (see Chateauneuf (1991, 1999), where the definition is usually written as follows: two real-valued random variables X_1 and X_2 defined on the same probability space are comonotonic if*

$$[X_1(\omega) - X_1(\omega')] \cdot [X_2(\omega) - X_2(\omega')] \geq 0, \text{ for all } \omega, \omega'.$$

Example 4.1.33. *The concept of comonotonicity is an important concept in actuarial science. Consider for instance an insurance portfolio of individual risks $X_1, ..., X_n$, which are not assume to be independent. Financial hedging techniques might be used by the insurer to reduce the aggregate risk. In Alternative Risk Transfer techniques, compensation will be obtained as the pay-off of the financial will increase if the loss $X_1 + ... + X_n$ increases (portfolio based contract). But in the case of hurricane or earthquakes, the insurer could also by some (call) options, labeled on the CAT-index (index of catastrophe losses, of the Chicago Board of Trade). These options will be exercised when the CAT-index reaches a sufficiently high level, and in such a case, investors replace traditional reinsurer. Cat bounds can also be used, where the payment of the coupons and the principal could be conditioned on the occurrence of a catastrophe. In such a case, the notion of 'catastrophe' can also be based on some exogenous index (e.g. Richter level, or maximum wind speed). Those financial product may be interesting substitute to traditional reinsurance in the case where the financial compensation and the loss $X_1 + ... + X_n$ are comonotonic, or at least as comonotonic as possible (see (Dhaene et al. (2002)) for more details on such applications).*

Example 4.1.34. *More generally, recall that if X denotes the claim amount for the insured, and $I(X)$ denote the indemnity, which is the amount paid by the insurance company, $I(\cdot)$ is usually assumed to be increasing (see Denuit and Charpentier (2004)). Therefore, X and $I(X)$ are comonotonic. In the context of risk sharing schemes (e.g. coinsurance), one insurer pays $I(X)$ while $X - I(X)$ can be paid by an other insurer. It is also usually assumed that both variates are comonotonic, i.e. $I(x)$ and $x - I(x)$ are both increasing (usually written $0 \leq I'(\cdot) \leq 1$). Several examples can be considered*

- *stop-loss coverage, or deductible coverage, defined by $I(x) = (x-d)_+$ for some $n \geq 0$,*
- *quota-share coverage (e.g. coinsurance), defined by $I(x) = \alpha x$ where $\alpha \in (0,1)$,*
- *coverage with a maximal limit, defined by $I(x) = x \wedge d$ for some $n \geq 0$,*

  Those three example lead to comonotonic schemes, as well as the combination of those three, $I(x) = \left(\alpha (x-d)_+\right) \wedge u$. But notice that the deductible, is not comonotonic , when $I(x) = x\mathbf{1}(x \geq d)$.

Example 4.1.35. *The Value-at-Risk of a sum of comonotonic random variables, with distribution functions $F_1, ..., F_d$ is given by*

$$VaR\left(X_1^+ + ... + X_d^+, \alpha\right) = VaR\left(X_1, \alpha\right) + ... + VaR\left(X_d, \alpha\right), \; 0 < \alpha < 1, \tag{4.4}$$

(see Dhaene et al. (2001, 2002)). From this relationship, some stable properties can be derived: the comonotonic sum of Pareto variables $\mathcal{P}(\alpha, \beta_i)$ is also Pareto distributed, with parameters α and $\beta = \beta_1 + ... + \beta_d$; the comonotonic sum of exponential variables $\mathcal{E}(\lambda_i)$ is also exponentially distributed, with parameter $\lambda = \lambda_1 + ... + \lambda_d$ (and for Gaussian, Gumbel or Gamma, among others). Further, the associated stop-loss premium is

$$\mathbb{E}\left(\left(S^+ - d\right)_+\right) = \sum_{i=1}^{d} \mathbb{E}\left(\left(X_i - VaR\left(X_i, F_{S^+}(d)\right)\right)_+\right),$$

where F_{S^+} denotes the c.d.f. of $S^+ = X_1^+ + ... + X_d^+$ (see Dhaene et al. (2001) and Vyncke (2003)).

For more results and applications of the concepts of comonotonicity in finance and actuarial science, see Dhaene et al. (2001), Dhaene et al. (2002) or Vyncke (2003).

Remark 4.1.36. *The elements of $\mathcal{F}(F_1, ..., F_d)$ are bounded above by the upper Fréchet-Hoeffding distribution, but also below, by the so-called Fréchet-Hoeffding lower bound, defined as*

$$M(x_1, .., x_d) = \max\{F_1(x_1) + ... + F_d(x_d) - (d-1), 0\},$$

and moreover, for all $x_i \in \mathbb{R}$, $M(x_1, .., x_d) \leq F(x_1, .., x_d)$ for all $F \in \mathcal{F}(F_1, ..., F_d)$. Note that in the bivariate case (and only in the bivariate case), $M \in \mathcal{F}(F_1, F_2)$. More specifically, M is the distribution function of $(F_1^{\leftarrow}(U), F_2^{\leftarrow}(1-U))$ where $U \sim \mathcal{U}([0,1])$. And further, in the case of positive variables in $\mathcal{F}^+(F_1, ..., F_d)$, $\boldsymbol{X}$ has distribution function M if and only if $\boldsymbol{X}$ is mutually exclusive, i.e. at most one of the component can be different from 0, $\mathbb{P}(X_i > 0, X_j > 0) = 0$ for all $i \neq j$ (see also (Dhaene and Denuit (1999)))

4.2 Ordering with Archimedean dependence structure

In this section, we will focus on orderings and conditioning. More precisely, we shall consider conditioning by lower orthant, i.e. when $\boldsymbol{X} \leq F_{\boldsymbol{X}} \int(\boldsymbol{u})$, as in Chapter 2 and 3. As pointed out in the introduction, this many not be relevant for actuaries who should be interested in large claim dependencies. But using the notion of survival copula (see section 1.3.1), upper and lower orthant are closely related (see chapter 5 and 6 for a more detailed study of upper tails).

4.2.1 Conditioning with Archimedean copulae

As noticed in section 3.2.3, if C denotes a d-dimensional Archimedean copula, with generator ϕ, given $\boldsymbol{u} \in [0,1]^d$, with at least one non-null component, the conditional copula $C_{\boldsymbol{u}}$ is Archimedean, with generator

$$\phi_{\boldsymbol{u}}(t) = \phi(t \cdot C(\boldsymbol{u})) - \phi(C(\boldsymbol{u})), \text{ for all } t \in (0,1].$$

Example 4.2.1. *Gumbel copulae have generator $\phi(t) = [-\ln t]^\theta$ where $\theta \geq 1$. For any $\boldsymbol{u} \in (0,1]^d$, the corresponding conditional copula has generator*

$$\phi_{\boldsymbol{u}}(t) = \left[M^{1/\theta} - \ln t\right]^\theta - M \text{ where } M = [-\ln u_1]^\theta + ... + [-\ln u_d]^\theta. \tag{4.5}$$

Example 4.2.2. *Clayton copulae C have generator $\phi(t) = t^{-\theta} - 1$ where $\theta > 0$. Hence,*

$$\phi_{\boldsymbol{u}}(t) = [t \cdot C(\boldsymbol{u})]^{-\theta} - 1 - \phi(C\boldsymbol{u}) = t^{-\theta} \cdot C(\boldsymbol{u})^{-\theta} - 1 - [C\boldsymbol{u}^\theta - 1] = C(\boldsymbol{u})^{-\theta} \cdot [t^{-\theta} - 1], \tag{4.6}$$

hence $\phi_{\boldsymbol{u}}(t) = C(\boldsymbol{u})^{-\theta} \cdot \phi(t)$. Since the generator of an Archimedean copula is unique up to a multiplicative constant, $\phi_{\boldsymbol{u}}$ is also the generator of Clayton copula, with parameter θ (see also Section 2.4.).

Note that this invariance property can be obtained in the subclass of Archimedean copulae with a factor representation, obtained using the frailty approach (see section 1.5). Assume that variables X_i's are independent, conditionally on Θ, a positive random variable, such that $\mathbb{P}(X_i \leq x_i|\Theta) = G_i(x)^\Theta$ where G_i denotes a distribution function. The joint distribution function of $\boldsymbol{X}$ is given by

$$\begin{aligned}
F_{\boldsymbol{X}}(x_1, ..., x_d) &= \mathbb{E}(\mathbb{P}(X_1 \leq x_1, ..., X_d \leq X_d|\Theta)) \\
&= \mathbb{E}\left(\prod_{i=1}^d \mathbb{P}(X_i \leq x_i|\Theta)\right) = \mathbb{E}\left(\prod_{i=1}^d G_i(x_i)^\Theta\right) \\
&= \mathbb{E}\left(\prod_{i=1}^d \exp\left[-\Theta(-\log G_i(x_i))\right]\right) = \psi\left(-\sum_{i=1}^d \log G_i(x_i)\right),
\end{aligned}$$

where ψ is the Laplace transform of the distribution of Θ, i.e. $\psi(t) = \mathbb{E}(\exp(-t\Theta))$. Because the marginal distributions are given respectively by

$$F_i(x_i)\mathbb{P}(X_i \leq x_i) = \psi(-\log G_i(x_i)), \tag{4.7}$$

the copula of $\boldsymbol{X}$ is

$$C(\boldsymbol{u}) = F_{\boldsymbol{X}}\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right) = \psi\left(\psi^{\leftarrow}(u) + \ldots + \psi^{\leftarrow}(u_d)\right)$$

This copula is an Archimedean copula with generator $\phi = \psi^{\leftarrow}$.

Example 4.2.3. *Gumbel copulae could be obtained when factor Θ has its Laplace transform equal to $\psi(t) = \exp\left[-t^{1/\theta}\right]$. Furthermore, Clayton copulae are obtained when the heterogeneity factor Θ has a Laplace transform equal to $\psi(t) = [1-t]^{-1/\theta}$. The heterogeneity distribution is a Gamma distribution with degrees of freedom $1/\theta$ (see Section 1.3 of this thesis).*

Theorem 4.2.4. *Consider $\boldsymbol{X}$ with Archimedean copula, with a factor representation, and let ψ denote the Laplace transform of the heterogeneity factor Θ. Let $\boldsymbol{u} \in (0,1]^d$, then $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$ (in the sense that $\{X_1 \leq F_1^{\leftarrow}(u_1), ..., X_d \leq F_d^{\leftarrow}(u_d))$ is an Archimedean copula with a factor representation, where the factor has Laplace transform*

$$\psi_{\boldsymbol{u}}(t) = \frac{\psi(t + \psi^{\leftarrow}(C(\boldsymbol{u})))}{C(\boldsymbol{u})}, \tag{4.8}$$

given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$.

Proof. Note that $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$ will be said to have an Archimedean copula with a factor representation if all the components are independent, given a positive factor Θ', and if marginal distribution functions can be written $G_i'(x_i)^{\Theta'}$.

Consider a random vector $\boldsymbol{Y}$ such that $\boldsymbol{Y} \overset{\mathcal{L}}{=} \boldsymbol{X}|\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$. The joint distribution function of $\boldsymbol{Y}$, denoted F', is

$$\begin{aligned}
F'(\boldsymbol{x}) &= \mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}|\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})) \\
&= \frac{\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x})}{\mathbb{P}(\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u}))} \text{ on } (-\infty, F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})], \\
&= \frac{\psi(\psi^{\leftarrow}(F_1(x_1)) + ...\psi^{\leftarrow}(F_d(x_d)))}{C(\boldsymbol{u})} \\
&= \frac{\psi(-\log G_1(x_1) - .... - \log G_d(x_d))}{C(\boldsymbol{u})},
\end{aligned}$$

since $F_i(x_i) = \psi(-\log G_i(x_i))$. Hence, from this relationship one gets that the marginal distribution of $\boldsymbol{Y}$ is

$$\begin{aligned} F_i'(x_i) &= \lim_{x_j \to F_j^{\leftarrow}(u_j), j \neq i} F(\boldsymbol{x}) \\ &= \frac{\psi(-\log(G_i(x_i))) + \psi^{\leftarrow}(u_1) + ... + \psi^{\leftarrow}(u_{i-1}) + \psi^{\leftarrow}(u_{i+1}) + ... + \psi^{\leftarrow}(u_d)}{C(\boldsymbol{u})} \\ &= \frac{\psi\left([-\log(G_i(x_i))) - \psi^{\leftarrow}(u_i)] + \psi^{\leftarrow}(u_1) + ... + \psi^{\leftarrow}(u_d)\right)}{\psi\left(\psi^{\leftarrow}(u_1) + ... + \psi^{\leftarrow}(u_d)\right)}. \end{aligned}$$

Recall (see Feller (1966)) that if ψ is the Laplace transform of random variable Z, so that $\psi(t) = \mathbb{E}(\exp(-tZ))$, where Z has distribution function F_Z, then ϕ defined as $\phi(t) = \psi(t+c)/\psi(c)$ is the Laplace transform of some random variable Z' with cumulative distribution function $F_{Z'}(t) = \exp(-ct) F_Z(t)$.

Hence, the marginal distribution function of Y_i can be written

$$F_i'(x_i) = \psi_{\boldsymbol{u}}([-\log(G_i(x_i)) - \psi^{\leftarrow}(u_i)]),$$

where $\psi_{\boldsymbol{u}}$ is the Laplace transform defined as

$$\psi_{\boldsymbol{u}}(t) = \frac{\psi\left(t + \psi^{\leftarrow}(u_1) + ... + \psi^{\leftarrow}(u_d)\right)}{\psi\left(\psi^{\leftarrow}(u_1) + ... + \psi^{\leftarrow}(u_d)\right)} = \frac{\psi(t + \psi^{\leftarrow}(C(\boldsymbol{u})))}{C(\boldsymbol{u})}.$$

Set further $G_i'(x_i) = \exp(\log(G_i(x_i)) + \psi(u_i))$ on $(-\infty, F_i^{\leftarrow}(u_i)]$. One gets easily that G_i' is an increasing function, with $G_i'(x_i) \to 0$ as $x_i \to -\infty$ and $G_i'(F_i^{\leftarrow}(u_i)) = \exp(0) = 1$. Hence, G_i' is a cumulative distribution function. Similarly for all $i \in \{1, ..., d\}$.

As at now, we have that there exists a random variable Θ' with Laplace transform $\psi_{\boldsymbol{u}}$, such that $\mathbb{P}(Y_i \leq x_i | \Theta') = G_i(x_i)^{\Theta'}$ for all $i \in \{1, ..., d\}$. Let us prove that given Θ', the components of $\boldsymbol{Y}$ are independent.

On the one hand, we have obtained that the joint distribution function of $\boldsymbol{Y}$ is

$$F'(\boldsymbol{x}) = \frac{\psi(-\log G_1(x_1) - .... - \log G_d(x_d))}{C(\boldsymbol{u})}.$$

From the expression of ψ', note that this expression becomes

$$F'(\boldsymbol{x}) = \psi_{\boldsymbol{u}}(-\log G_1(x_1) - .... - \log G_d(x_d) - \psi^{\leftarrow}(u_1) - ... - \psi^{\leftarrow}(u_d)).$$

On the other hand,

$$\begin{aligned} & \mathbb{E}\left(\mathbb{P}\left(Y_1 \leq x_1 | \Theta'\right) \cdot ... \cdot \mathbb{P}\left(Y_d \leq x_d | \Theta'\right)\right) \\ = \ & \mathbb{E}\left(G_1'(x_1)^{\Theta'} \cdot ... \cdot G_d'(x_d)^{\Theta'}\right) \\ = \ & \mathbb{E}\left(\exp[-\Theta'(-\log G_1'(x_1))] \cdot ... \cdot \exp[-\Theta'(-\log G_d'(x_d))]\right) \\ = \ & \mathbb{E}\left(\exp[-\Theta'(-[\log G_1(x_1) + \psi^{\leftarrow}(u_1)])] \cdot ... \cdot \exp[-\Theta'(-[\log G_d(x_d) + \psi^{\leftarrow}(u_d)])]\right) \\ = \ & \psi_{\boldsymbol{u}}\left(-\log(G_1(x_1)) - \psi^{\leftarrow}(u_1) - ... - \log(G_d(x_d)) - \psi^{\leftarrow}(u_d)\right), \end{aligned}$$

and therefore, one gets that

$$\begin{aligned} \mathbb{E}\left(\mathbb{P}\left(Y_1 \leq x_1 | \Theta'\right) \cdot ... \cdot \mathbb{P}\left(Y_d \leq x_d | \Theta'\right)\right) &= F'(\boldsymbol{x}) \\ &= \mathbb{E}\left(\mathbb{P}\left(Y_1 \leq x_1, ..., Y_d \leq x_d | \Theta'\right)\right), \end{aligned}$$

i.e. given Θ', the components of $\boldsymbol{Y}$ are independent.

In order to conclude, let us just observe that Θ' is a positive random variable, since

$$\mathbb{P}(\Theta' < 0) = \lim_{t\to\infty} \psi_{\boldsymbol{u}}(t) = \lim_{t\to\infty} \psi(t) = 0,$$

since Θ is a positive variable. Finally, the conditional vector $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$ will be said to have an Archimedean copula with a factor representation . This finishes the proof of Theorem 4.2.4. □

4.2.2 Invariance by truncation

We have obtained in the previous section that Clayton copulae were the only invariant copula, in the sense that $\Phi(C, \boldsymbol{u}) = C$ for all $\boldsymbol{u} \in (0,1]^d$. As proved in earlier chapters, when the truncation is under a specific direction a wide class of copulae can be obtained. More formally if $\mathcal{D} = \{(r_1(z), ..., r_d(z), z \geq 0\}$, C is invariant under this direction if $\Phi(C, r_1(z), ..., r_d(z)) = C$ for all $z \geq 0$. Let $\mathcal{C}_{\mathcal{D}}$ denote the class of invariant copulae under direction $\mathcal{D}$. Given a direction $\mathcal{D}$, one can wonder if some Archimedean copulae (apart from Clayton's) belong to class $\mathcal{C}_{\mathcal{D}}$. As proved in the following theorem, only Clayton copula can be invariant, whatever the direction considered.

Theorem 4.2.5. *Consider a continuous direction* $\mathcal{D} = \{(r_1(z), ..., r_d(z)) \in [0,1]^d, z \geq 0\}$, *from* $(1, ..., 1)$ *to* $(0, ..., 0)$. *The only invariant Archimedean copula under direction* $\mathcal{D}$ *is Clayton copula.*

Proof. Let C be an Archimedean copula, with generator ϕ, invariant copula under direction $\mathcal{D}$. Let $\boldsymbol{U}$ denote a random vector with distribution function C, and $t \geq 0$. Let ϕ_z denote the generator of the Archimedean copula of $\boldsymbol{U}$ given $\{U_1 \leq r_1(z), ..., U_d \leq r_d(z)\}$. From Theorem 4.2.7, note that

$$\phi_z(t) = \phi(t \cdot C(r_1(z), ..., r_d(z))) - \phi(C(r_1(z), ..., r_d(z))).$$

C is an invariant copula if and only if for all $z \geq 0$, ϕ_z is proportional with ϕ (the Archimedean generator being defined up to a multiplicative constant). Since C is absolutely continuous $C(r_1(\cdot), ..., r_d(\cdot))$ covers the range $[0,1]$, and therefore, ϕ is the generator of an invariant Archimedean copula if and only if

$$\phi(t \cdot c) - \phi(c) \propto \phi(t) \text{ for all } c, t \in [0,1],$$

hence, if $\kappa(c)$ denotes the proportionally coefficient,

$$\phi(t \cdot c) - \phi(c) = \kappa(c) \cdot \phi(t) \text{ for all } c, t \in [0,1]. \tag{4.9}$$

Set $x = -\log t$ and $y = -\log c$, and $f(t) = \phi(\exp(-t))$ where $t \geq 0$, and $h(t) = K(\exp(-t))$. Solving equation (4.9) is equivalent with solving

$$f(x+y) = h(y) f(x) + f(y) \text{ where } x, y \geq 0, \tag{4.10}$$

for some positive functional h, where $f(0) = 0$. Interchanging the variables in (4.10), we write

$$f(x+y) = h(x) f(y) + f(x) \tag{4.11}$$

which, together with (4.10), leads to

$$h(y) f(x) + f(y) = h(x) \cdot f(y) + f(x) \tag{4.12}$$

that is,

$$f(x)[h(y) - 1] = f(y) \cdot [h(x) - 1] \tag{4.13}$$

- If $h(t) = 1$, then (4.10) reduces to Cauchy's basic equation. Then $f(t) = \alpha t$, and $\phi(t) = -\alpha \log t = -\log t^{\alpha}$ on $]0,1]$. This case yields the independent case,

$$\begin{aligned} C(x,y) &= \phi^{\leftarrow}(\phi(x) + \phi(y)) \\ &= \exp\left(-\left[-\alpha \log x - \alpha \log y\right]/\alpha\right) = \exp\left(\log(xy)\right) \\ &= xy = C^{\perp}(x,y). \end{aligned}$$

- If, however, there exists a t_0 such that $h(y_0) \neq 1$, then it follows from (4.13) that

$$f(x) = \frac{f(y_0)}{h(y_0) - 1}\left[h(x) - 1\right] = \gamma\left[h(x) - 1\right], \tag{4.14}$$

where γ is a constant. If $\gamma = 0$, then f is constant (null) and then, ϕ is null on $[0,1]$. So we assume that $\gamma \neq 0$, and substitute (4.14) into (4.10),

$$\gamma\left[h(x+y) - 1\right] = \gamma\left[h(x) - 1\right] + \gamma\left[h(y) - 1\right] \tag{4.15}$$

so that we can obtain the following equation

$$h(x+y) = h(x)\,h(y) \text{ where } x, y \geq 0, \tag{4.16}$$

The most general solutions of Cauchy-type functional equation (4.16) are, according to Theorem (2.1.2.1) of Aczél (1966),

$$h(t) = 0 \text{ and } h(t) = exp(\alpha t) \tag{4.17}$$

So finally, from (4.14) and because $f(t) = \phi(\exp(-t))$, functions ϕ have to satisfy

$$\phi(t) = \gamma\left[\exp(\alpha \log t) - 1\right] = \gamma\left[t^{\alpha} - 1\right] \tag{4.18}$$

which is the general form of Clayton's copula generator.

□

4.2.3 $\mathcal{H}$-copulae and the factor model

In Section 1.3.2, we have introduced the class of $\mathcal{H}$-copulae, defined as

$$\Phi_h(C)(u,v) = h^{\leftarrow}(C(h(u),h(v))), 0 \leq u, v \leq 1,$$

where C is a copula, and $h \in \mathcal{H}$ is a convex distortion function. As noticed earlier, copulae $\Phi_h(C^{\perp})$ are Archimedean copulae. An idea can be to focus on the factor interpretation of Archimedean copulae, and to extend it in the non-independent case.

Assume that there exists a positive random variable Θ, such that, conditionally on Θ, random vector $\boldsymbol{X} = (X_1, ..., X_d)$ has copula C, which does not depend on Θ. Assume moreover that C is in extreme value copula, (see Joe (1997)). The following results holds,

Lemma 4.2.6. *Let Θ be a random variable with Laplace transform ψ, and consider a random vector $\boldsymbol{X} = (X_1, ..., X_d)$ such that $\boldsymbol{X}$ given Θ has copula C, an extreme value copula (in the sense that $C\left(x_1^h, ..., x_d^h\right) = C^h(x_1, ..., x_d)$ for all $h \geq 0$). Assume that, for all $i = 1, ..., d$, $\mathbb{P}(X_i \leq x_i|\Theta) = G_i(x_i)^{\Theta}$ where the G_i's are distribution functions. Then $\boldsymbol{X}$ has copula*

$$C_{\boldsymbol{X}}(x_1, ..., x_d) = \psi\left(-\log\left(C\left(\exp\left[-\psi^{\leftarrow}(x_1)\right], ..., \exp\left[-\psi^{\leftarrow}(x_d)\right]\right)\right)\right), \tag{4.19}$$

which is copula $\Psi_h(C)$ with $h(\cdot) = \exp\left[-\psi^{\leftarrow}(\cdot)\right]$.

Proof. Let $\boldsymbol{X}$ be a random vector such that $\boldsymbol{X}$ given Θ has copula C and $\mathbb{P}(X_i \leq x_i|\Theta) = G_i(x_i)^{\Theta}$, $i = 1, ..., d$. Then, the (unconditional) joint distribution function of $\boldsymbol{X}$ is given by

$$\begin{aligned} F(\boldsymbol{x}) &= \mathbb{E}(\mathbb{P}(X_1 \leq x_1, ..., X_d \leq x_d|\Theta)) = \mathbb{E}(C(\mathbb{P}(X_1 \leq x_i|\Theta), ..., \mathbb{P}(X_d \leq x_d|\Theta))) \\ &= \mathbb{E}\left(C\left(G_1(x_1)^{\Theta}, ..., G_d(x_d)^{\Theta}\right)\right) = \mathbb{E}\left(C^{\Theta}(G_1(x_1), ..., G_d(x_d))\right) \\ &= \psi(-\log C(G_1(x_1), ..., G_d(x_d))), \end{aligned}$$

where ψ is the Laplace transform of the distribution of Θ, i.e. $\psi(t) = \mathbb{E}(\exp(-t\Theta))$. Because C is an extreme value copula,

$$C\left(G_1(x_1)^{\Theta}, ..., G_d(x_d)^{\Theta}\right) = C^{\Theta}(G_1(x_1), ..., G_d(x_d)).$$

One gets finally that the unconditional marginal distribution functions are $F_i(x_i) = \psi(-\log G_i(x_i))$, and therefore

$$C_{\boldsymbol{X}}(x_1, ..., x_d) = \psi(-\log(C(\exp[-\psi^{\leftarrow}(x)], \exp[-\psi^{\leftarrow}(y)]))) \tag{4.20}$$

This finishes the proof of Lemma 4.2.6. □

We will see with the Theorem below that, in the case where the copula of $\boldsymbol{X}$ is an $\mathcal{H}$-copula, that the stability of exchangeable Archimedean copulae with a factor representation can be extended to $\mathcal{H}$-copula, with additional assumptions.

Theorem 4.2.7. *Let $\boldsymbol{X}$ be a random vector with an $\mathcal{H}$-copula, with a convex distortion generator, with a factor representation, let ψ denote the Laplace transform of the heterogeneity factor Θ, C denote the underlying copula, and G_i's the marginal parameters.*

(1) *Let $\boldsymbol{u} \in (0,1]^d$, then, the copula of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$ is*

$$C_{\boldsymbol{X},\boldsymbol{u}}(\boldsymbol{x}) = \psi_{\boldsymbol{u}}(-\log(C_{\boldsymbol{u}}(\exp[-\psi_{\boldsymbol{u}}^{\leftarrow}(x_1)], ..., \exp[-\psi_{\boldsymbol{u}}^{\leftarrow}(x_d)]))) = \Phi_{h_{\boldsymbol{u}}}(C_{\boldsymbol{u}})(\boldsymbol{x}), \tag{4.21}$$

where $h_{\boldsymbol{u}}(\cdot) = \exp[-\psi_{\boldsymbol{u}}^{\leftarrow}(\cdot)]$, and where

- *$\psi_{\boldsymbol{u}}$ is the Laplace transform defined as $\psi_{\boldsymbol{u}}(t) = \psi(t+\alpha)/\psi(\alpha)$ where $\alpha = -\log(C(\boldsymbol{u}^*))$, $\boldsymbol{u}_i^* = \exp[-\psi^{\leftarrow}(u_i)]$ for all $i = 1, ..., d$. Hence, $\psi_{\boldsymbol{u}}$ is the Laplace transform of Θ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$,*

- *$\mathbb{P}(X_i \leq x_i|\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u}), \Theta) = G_i'(x_i)^{\Theta}$ for all $i = 1, ..., d$, where*

$$G_i'(x_i) = \frac{C(u_1^*, u_2^*, ..., G_i(x_i), ..., u_d^*)}{C(u_1^*, u_2^*, ..., u_i^*, ..., u_d^*)}, \tag{4.22}$$

- *and $C_{\boldsymbol{u}}$ is the following copula*

$$C_{\boldsymbol{u}}(\boldsymbol{x}) = \frac{C(G_1(G_1'^{\leftarrow}(x_1)), ..., G_d(G_d'^{\leftarrow}(x_d)))}{C(G_1(F_1^{\leftarrow}(u_1)), ..., G_d(F_d^{\leftarrow}(u_d)))}. \tag{4.23}$$

(2) *Furthermore, the copula of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$ is an $\mathcal{H}$-copula with a factor representation if and only if $C_{\boldsymbol{u}}$ is an extreme value copula.*

Proof. (1) Let $C_{\boldsymbol{X}}$ be the copula of $\boldsymbol{X}$, that is

$$C_{\boldsymbol{X}}(u_1,...,u_d) = \psi\left(-\log\left(C\left(\exp\left[-\psi^{\leftarrow}(u_1)\right],...,\exp\left[-\psi^{\leftarrow}(u_d)\right]\right)\right)\right).$$

(i) The marginal distribution of X_i given $\boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u})$, and given $\Theta = \theta$ is

$$\begin{aligned}
&\mathbb{P}\left(X_i \leq x_i | \boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u}), \Theta = \theta\right) \\
= &\frac{\mathbb{P}\left(X_1 \leq F_1^{\leftarrow}(u_1),...,X_{i-1} \leq F_1^{\leftarrow}(u_{i-1}), X_i \leq x_i, X_{i+1} \leq F_1^{\leftarrow}(u_{i+1}),...,X_d \leq F_d^{\leftarrow}(u_d)|\Theta=\theta\right)}{\mathbb{P}\left(X_1 \leq F_1^{\leftarrow}(u_1),...,X_{i-1} \leq F_1^{\leftarrow}(u_{i-1}), X_i \leq F_i^{\leftarrow}(u_i), X_{i+1} \leq F_1^{\leftarrow}(u_{i+1}),...,X_d \leq F_d^{\leftarrow}(u_d)|\Theta=\theta\right)} \\
= &\frac{C\left(\mathbb{P}\left(X_1 \leq F_1^{\leftarrow}(u_1)|\Theta=\theta\right),...,\mathbb{P}\left(X_i \leq x_i|\Theta=\theta\right),...,\mathbb{P}\left(X_d \leq F_d^{\leftarrow}(u_d)|\Theta=\theta\right)\right)}{C\left(\mathbb{P}\left(X_1 \leq F_1^{\leftarrow}(u_1)|\Theta=\theta\right),...,\mathbb{P}\left(X_i \leq F_i^{\leftarrow}(u_i)|\Theta=\theta\right),...,\mathbb{P}\left(X_d \leq F_d^{\leftarrow}(u_d)|\Theta=\theta\right)\right)}
\end{aligned}$$

since C is the copula of $\boldsymbol{X}$ given Θ, i.e.

$$\mathbb{P}\left(X_1 \leq x_1,...,X_d \leq x_d|\Theta=\theta\right) = C\left(\mathbb{P}\left(X_1 \leq x_1|\Theta=\theta\right),...,\mathbb{P}\left(X_d \leq x_d\right)|\Theta=\theta\right).$$

Hence,

$$\mathbb{P}\left(X_i \leq x_i | \boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u}), \Theta = \theta\right) = \frac{C(G_1(F_1^{\leftarrow}(u_1))^{\theta},...,G_i(x_i)^{\theta},...,G_d(F_d^{\leftarrow}(u_d))^{\theta})}{C(G_1(F_1^{\leftarrow}(u_1))^{\theta},...,G_i(F_i^{\leftarrow}(u_i))^{\theta},...,G_d(F_d^{\leftarrow}(u_d))^{\theta})}$$

because C is an extreme value copula. Since $F_j(x_j) = \psi\left(-\log G_j(x_j)\right)$, set $u_j^* = G_j\left(F_j^{\leftarrow}(u_j)\right) = \exp\left[-\psi^{\leftarrow}(u_j)\right]$ for all $j = 1,...,d$. The marginal distribution satisfies,

$$\mathbb{P}\left(X_i \leq x_i | \boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u}), \Theta = \theta\right) = \left(\frac{C(u_1^*,...,u_{i-1}^*, G_i(x_i), u_{i+1}^*,...,u_d^*)}{C(u_1^*,...,u_{i-1}^*, u_i^*, u_{i+1}^*,...,u_d^*)}\right)^{\theta} \tag{4.24}$$

One can get easily that

$$G_i^*(x_i) = \frac{C(u_1^*,...,u_{i-1}^*, G_i(x_i), u_{i+1}^*,...,u_d^*)}{C(u_1^*,...,u_{i-1}^*, u_i^*, u_{i+1}^*,...,u_d^*)}$$

is (univariate) distribution function, since C and G_i are both increasing, and moreover $G_i^*\left(F_i^{\leftarrow}(u_i)\right) = u_i^*$.

(ii) The joint distribution function of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u})$ is

$$\begin{aligned}
\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x} | \boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u})\right) &= \frac{\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x}\right)}{\mathbb{P}\left(\boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u})\right)} = \frac{\mathbb{E}\left(\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x}|\Theta\right)\right)}{C(\boldsymbol{u})} \\
&= \frac{\mathbb{E}\left(C\left(G_1(x_1)^{\Theta},...,G_d(x_d)^{\Theta}\right)\right)}{C(\boldsymbol{u})} \\
&= \frac{\mathbb{E}\left(C\left(G_1(x_1),...,G_d(x_d)\right)\right)^{\Theta}}{C(\boldsymbol{u})}
\end{aligned}$$

From the expression of copula $C_{\boldsymbol{X}}$,

$$C_{\boldsymbol{X}}(\boldsymbol{u}) = \psi\left(-\log\left(C\left(\exp\left[-\psi^{\leftarrow}(u_1)\right],...,\exp\left[-\psi^{\leftarrow}(u_d)\right]\right)\right)\right) = \psi\left(-\log\left(C\left(u_1^*,...,u_d^*\right)\right)\right),$$

one gets,

$$\begin{aligned}
\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x} | \boldsymbol{X} \leq F^{\leftarrow}_{\boldsymbol{X}}(\boldsymbol{u})\right) &= \frac{\psi(-\log C(G_1(x_1),...,G_d(x_d)))}{\psi(-\log C(u_1^*,...,u_d^*))} \\
&= \frac{\psi[-\log C(u_1^*,...,u_d^*) - \alpha] + \alpha}{\psi(\alpha)}
\end{aligned}$$

where $\alpha = -\log\left(C\left(u_1^*, ..., u_d^*\right)\right)$. Set $\psi_{\boldsymbol{u}}(t) = \psi(t+\alpha)/\psi(\alpha)$. From this expression, $\psi_{\boldsymbol{u}}$ is also a Laplace transform. Furthermore, the expression above could be written

$$\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x} | \boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})\right) = \psi_{\boldsymbol{u}}\left(-\log \frac{C\left(G_1(x_1), ..., G_d(x_d)\right)}{C\left(u_1^*, ..., u_d^*\right)}\right).$$

We can then write the conditional marginal distribution function as

$$\begin{aligned}\mathbb{P}\left(X_i \leq x_i | \boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})\right) &= \psi_{\boldsymbol{u}}\left(-\log \frac{C\left(u_1^*, ..., G_i(x_i), ..., u_d^*\right)}{C\left(u_1^*, ..., u_d^*\right)}\right) \\ &= \psi_{\boldsymbol{u}}(-\log G_i^*(x_i)),\end{aligned}$$

i.e.,

$$\mathbb{P}\left(X_i \leq x_i | \boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})\right) = \mathbb{E}\left(G_i^*(x_i)^{\Theta}\right), \tag{4.25}$$

where Θ has Laplace transform $\psi_{\boldsymbol{u}}$.

(iii) Let $C_{\boldsymbol{u}}$ be the functional defined on $[0,1]^d$ by

$$C_{\boldsymbol{u}}(x_1, ..., x_d) = \frac{C\left(G_1\left(G_1^{*\leftarrow}(x_1)\right), ..., G_d\left(G_d^{*\leftarrow}(x_d)\right)\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right), ..., G_d\left(F_d^{\leftarrow}(u_d)\right)\right)}. \tag{4.26}$$

Because C is d-increasing (C is a copula) and the G_i's are increasing, $C_{\boldsymbol{u}}$ is d-increasing. Furthermore,

$$C_{\boldsymbol{u}}(x_1, ..., x_{i-1}, 0, x_{i+1}, ..., x_d) = \frac{C\left(G_1\left(G_1^{*\leftarrow}(x_0)\right), ..., G_i\left(G_i^{*\leftarrow}(0)\right)\right), ..., G_d\left(G_d^{*\leftarrow}(x_d)\right)\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right), ..., G_d\left(F_d^{\leftarrow}(u_d)\right)\right)} = 0,$$

and

$$\begin{aligned}C_{\boldsymbol{u}}(1, ..., 1, x_i, 1, ..., 1) &= \frac{C\left(G_1\left(G_1^{*\leftarrow}(1)\right), ..., G_i\left(G_i^{*\leftarrow}(x_i)\right), ..., G_d\left(G_d^{*\leftarrow}(1)\right)\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right), ..., G_d\left(F_d^{\leftarrow}(u_d)\right)\right)} \\ &= \frac{C\left(u_1^*, ..., u_{i-1}^*, G_i\left(G_i^{*\leftarrow}(x_i)\right), u_{i+1}^*, ..., u_d^*\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right), ..., G_d\left(F_d^{\leftarrow}(u_d)\right)\right)}\end{aligned}$$

so, finally, $C_{\boldsymbol{u}}(1, ..., 1, G_i^*(x_i), 1, ..., 1) = G_i^*(x_i)$, that is, since G_i^* is bijective on $[0,1]$, for all z_i in $[0,1]$, $C_{\boldsymbol{u}}(1, ..., 1, z_i, 1, ..., 1) = z_i$. So, finally, $C_{\boldsymbol{u}}$ is a copula.

(iv) Using the results obtained above, one gets that the copula of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$ is $C_{\boldsymbol{X},\boldsymbol{u}}$ defined as

$$C_{\boldsymbol{X},\boldsymbol{u}}(x_1, ..., x_d) = \psi_{\boldsymbol{u}}\left(-\log\left(C_{\boldsymbol{u}}\left(\exp\left[-\psi_{\boldsymbol{u}}^{-1}(x_1)\right], \exp\left[-\psi_{\boldsymbol{u}}^{\leftarrow}(x_d)\right]\right)\right)\right) = \Psi_{h_{\boldsymbol{u}}}(C_{\boldsymbol{u}})(x_1, ..., x_d).$$

which is the analogous of the result of Proposition (4.2.6).

(2) Assume that $\boldsymbol{X} = (X_1, ..., X_d)$ has an $\mathcal{H}$-copula. Using the notions of the beginning of the prof, let $C_{\boldsymbol{u}}$ denote the copula of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$) and given Θ. Then, for all $\theta \geq 0$

$$\begin{aligned}C_{\boldsymbol{u}}(\boldsymbol{x})^{\theta} &= \frac{C\left(G_1\left(G_1^{*\leftarrow}(x_1)\right), ..., G_d\left(G_d^{*\leftarrow}(x_d)\right)\right)^{\theta}}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right), ..., G_d\left(F_d^{\leftarrow}(u_d)\right)\right)^{\theta}} \\ &= \frac{C\left(G_1\left(G_1^{*\leftarrow}(x_1)\right)^{\theta}, ..., G_d\left(G_d^{*\leftarrow}(x_d)\right)^{\theta}\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right)^{\theta}, ..., G_d\left(F_d^{\leftarrow}(u_d)\right)^{\theta}\right)} \\ &= \frac{C\left(\mathbb{P}\left(X_1 \leq G_1^{*\leftarrow}(x_1) | \Theta = \theta\right), ..., \mathbb{P}\left(X_d \leq G_d^{*\leftarrow}(x_d) | \Theta = \theta\right)\right)}{C\left(\mathbb{P}\left(X_1 \leq F_1^{\leftarrow}(u_1) | \Theta = \theta\right), ..., \mathbb{P}\left(X_d \leq F_d^{\leftarrow}(u_d) | \Theta = \theta\right)\right)} \\ &= \frac{\mathbb{P}\left(X_1 \leq G_1^{*\leftarrow}(x_1), ..., X_d \leq G_d^{*\leftarrow}(x_d) | \Theta = \theta\right)}{C\left(u_1^*, ..., u_d^*\right)}.\end{aligned} \tag{4.27}$$

Note that the numerator could be written

$$\begin{aligned}
& \mathbb{P}\left(\boldsymbol{X} \leq G^{*\leftarrow}(\boldsymbol{x}) \,|\Theta=\theta\right) \\
= \;& \mathbb{P}\left(\boldsymbol{X} \leq G^{*\leftarrow}(\boldsymbol{x}) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right) \cdot \mathbb{P}\left(\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}) \,|\Theta=\theta\right) \\
= \;& \mathbb{P}\left(\boldsymbol{X} \leq G^{*\leftarrow}(\boldsymbol{x}) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right) \cdot C(\boldsymbol{u}^{*}),
\end{aligned}$$

and therefore

$$C_{\boldsymbol{u}}(\boldsymbol{x})^{\theta} = \mathbb{P}\left(\boldsymbol{X} \leq G^{*\leftarrow}(\boldsymbol{x}) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right). \tag{4.28}$$

From this expression, using the fact that $C_{\boldsymbol{u}}$ is the copula of $\boldsymbol{X} \leq G^{*-1}(\boldsymbol{x})$ and $\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u})$ and $\Theta=\theta$, we get

$$\begin{aligned}
& \mathbb{P}\left(\boldsymbol{X} \leq G^{*\leftarrow}(\boldsymbol{x}) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right) \\
= \;& C_{\boldsymbol{u}}\left(\mathbb{P}\left(X_1 \leq G_1^{*\leftarrow}(x_1) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right), ..., \mathbb{P}\left(X_d \leq G_d^{*\leftarrow}(x_d) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right)\right) \\
= \;& C_{\boldsymbol{u}}\left(\mathbb{P}\left(X_1 \leq G_1^{*-1}(x_1) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u})\right)^{\theta}, ..., \mathbb{P}\left(X_d \leq G_d^{*\leftarrow}(x_d) \,|\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u})\right)^{\theta}\right) \\
= \;& C_{\boldsymbol{u}}(x_1^{\theta}, ..., x_d^{\theta}).
\end{aligned}$$

Hence, for all $\theta \geq 0$, $C_{\boldsymbol{u}}(\boldsymbol{x})^{\theta} = C_{\boldsymbol{u}}(\boldsymbol{x}^{\theta})$ and therefore, $C_{\boldsymbol{u}}$ is an extreme value copula.

Conversely, assume that $C_{\boldsymbol{u}}$ is an extreme value copula. The conditional joint distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u})$, and $\Theta=\theta$ is

$$\begin{aligned}
& \mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x} | \boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right) && (4.29)\\
= \;& \frac{\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x} | \Theta\theta\right)}{\mathbb{P}\left(\boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta\right)} \\
= \;& \frac{C\left(\mathbb{P}\left(X_1 \leq x_1 | \Theta=\theta\right), ..., \mathbb{P}\left(X_d \leq x_d | \Theta=\theta\right)\right)}{C\left(\mathbb{P}\left(X_1 \leq F_1^{\leftarrow}(u_1) \,|\Theta=\theta\right), ..., \mathbb{P}\left(X_d \leq F_d^{\leftarrow}(u_d) \,|\Theta=\theta\right)\right)} \\
= \;& \frac{C\left(G_1(x_1)^{\theta}, ..., G_d(x_d)^{\theta}\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right)^{\theta}, ..., G_d\left(F_d^{\leftarrow}(u_d)\right)^{\theta}\right)} \\
= \;& \left[\frac{C\left(G_1(x_1), ..., G_d(x_d)\right)}{C\left(G_1\left(F_1^{\leftarrow}(u_1)\right), ..., G_d\left(F_d^{\leftarrow}(u_d)\right)\right)}\right]^{\theta} \\
= \;& C_{\boldsymbol{u}}\left(G_1^{*}(x_1), ..., G_d^{*}(x_d)\right)^{\theta} = C^{*}\left(G_1^{*}(x_1)^{\theta}, ..., G_d^{*}(x_d)^{\theta}\right) && (4.30)\\
= \;& C_{\boldsymbol{u}}\left(\mathbb{P}(X_1 \leq x_1 | \boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta), ..., \mathbb{P}(X_d \leq x_d | \boldsymbol{X} \leq F^{\leftarrow}(\boldsymbol{u}), \Theta=\theta)\right), && (4.31)
\end{aligned}$$

because $C_{\boldsymbol{u}}$ is an extreme value copula. So finally, $C_{\boldsymbol{u}}$ is the copula of $\boldsymbol{X}$ given $\boldsymbol{X} \leq F_{\boldsymbol{X}}^{\leftarrow}(\boldsymbol{u})$) and given Θ. This finishes the proof of Theorem 4.2.7. □

As well as Archimedean copulae are stable (in the sense that conditional copulae are still Archimedean copula), those copulae define a stable family of copulae, by conditioning.

4.2.4 Comparing tails for Archimedean copulae

From Theorem 4.2.7, one can notice that the generator of the conditional copula is the same on a given level curve of the copula C : if $C(u_1, v_1) = C(u_2, v_2)$, that is $\phi(u_1)+\phi(v_1) = \phi(u_2)+\phi(v_2)$, then $\Phi(C, u_1, v_1) = \Phi(C, u_2, v_2)$. Moreover, ordering $\Phi(C, t_1, t_1)$ and $\Phi(C, t_2, t_2)$ where $t_1 \leq t_2$ is equivalent with ordering $\Phi(C, u_1, v_1)$ and $\Phi(C, u_2, v_2)$ where

$$\phi(t_1) + \phi(t_1) = \phi(u_1) + \phi(v_1) \leq \phi(u_2) + \phi(v_2) = \phi(t_2) + \phi(t_2).$$

For expository purpose (and the convenient interpretation), we will focus only on results on the diagonal.

When studying the evolution of the conditional copula on the diagonal, one can expect a dependence structure which is all the more positively dependent as t decreases, or similarly, all the less dependent. In the first case, if $0 < t_2 \leq t_1 \leq 1$, $\Phi(C, t_1) \preceq \Phi(C, t_2)$, in the sense that $\Phi(C, t_1)(x, y) \leq \Phi(C, t_2)(x, y)$ for all x, y in $[0, 1] \times [0, 1]$, which could be seen as the PQD-ordering, as induced by the PQD property (see Lehmann (1966)).

Theorem 4.2.8. *Let t_1 and t_2 such that $0 < t_2 \leq t_1 \leq 1$, and let C be an Archimedean copula with generator ϕ. Let*

$$f_{12}(x) = \phi\left(\frac{C_1}{C_2}\phi^{\leftarrow}(x + \phi(C_2))\right) - \phi(C_1) \text{ and } f_{21}(x) = \phi\left(\frac{C_2}{C_1}\phi^{\leftarrow}(x + \phi(C_1))\right) - \phi(C_2), \tag{4.32}$$

where $C_1 = C(t_1, t_1)$ and $C_2 = C(t_2, t_2)$. Then

- *$\Phi(C, t_2, t_2)(x, y) \leq \Phi(C, t_1, t_1)(x, y)$ for all x, y in $[0, 1] \times [0, 1]$ if and only if $f_{21}(x)$is sudadditive,*
- *$\Phi(C, t_2, t_2)(x, y) \geq \Phi(C, t_1, t_1)(x, y)$ for all x, y in $[0, 1] \times [0, 1]$ if and only if $f_{12}(x)$is sudadditive.*

Proof. As shown in Nelsen (1999), if C_1 and C_2 are two Archimedean copulae with generator ϕ_1 and ϕ_2, then $C_2 \preceq C_1$, in the sense that $C_2(x, y) \leq C_1(x, y)$ for all x, y in $[0, 1] \times [0, 1]$, if and only if $\phi_2 \circ \phi_1^{\leftarrow}$ is subadditive, that is

$$\phi_2 \circ \phi_1^{\leftarrow}(x + y) \leq \phi_2 \circ \phi_1^{\leftarrow}(x) + \phi_2 \circ \phi_1^{\leftarrow}(y) \text{ for all } x, y \geq 0 \tag{4.33}$$

In the case of conditional copulae, $\phi_2(x) = \phi(C_2 x) - \phi(C_2)$ and $\phi_1(x) = \phi(C_1 x) - \phi(C_1)$, and so, $\Phi(C, t_2) = C_2 \preceq C_1 = \Phi(C, t_1)$ if and only if $f_{21}(x)$is sudadditive, where

$$f_{21}(x) = \phi\left(\frac{C_2}{C_1}\phi^{\leftarrow}(x + \phi(C_1))\right) - \phi(C_2). \tag{4.34}$$

One gets analogous results for f_{12}.

This finishes the proof of Theorem 4.2.8.

□

Example 4.2.9. *The case of Clayton copulae could be seen as a limiting case, in the sense that $\phi(t) = t^{-\theta} - 1$ and so, f_{12} is linear, i.e.*

$$f_{12}(x) = ax + b \text{ where } a = C_1^{\theta}/C_2^{\theta}. \tag{4.35}$$

We obtain here the particular case mentioned in Lemma 5.5.8. in Schweizer and Sklar (1983).

In the case were ϕ is twice differentiable, a sufficient condition for uniform ordering of conditional copula is the following.

Lemma 4.2.10. *If ϕ is twice differentiable, let $\psi(x) = \log -D\phi(t)$,*

(i) If ψ is concave on $]0, 1]$, then $\Phi(C, t_2, t_2)(x, y) \leq \Phi(C, t_1, t_1)(x, y)$ for all x, y in $[0, 1] \times [0, 1]$, for all $0 < t_2 \leq t_1 \leq 1$.

(ii) Similarly, if $\psi(x)$ is convex on $]0, 1]$, then $\Phi(C, t_2, t_2)(x, y) \geq \Phi(C, t_1, t_1)(x, y)$ for all x, y in $[0, 1] \times [0, 1]$, for all $0 < t_2 \leq t_1 \leq 1$.

Proof. (i) Let $0 \leq t_2 \leq t_1 \leq 1$, and $\beta = C(t_2, t_2)$, $\gamma = C(t_1, t_1)$ and $\alpha = \gamma/\beta, \alpha \leq 1$. Let $f(x) = \phi(\alpha\phi^{\leftarrow}(x + \phi(\beta))) - \phi(\gamma)$, then

$$\frac{d}{dx} f(x) = \frac{\alpha}{\phi'(\phi^{\leftarrow}(x + \phi(\beta)))} \phi'(\alpha\phi^{\leftarrow}(x + \phi(\beta)))$$

$$\frac{d^2}{dx^2} f(x) = \alpha \frac{\alpha\phi''(\alpha\phi^{\leftarrow}(x + \phi(\beta))).\phi'(\phi^{\leftarrow}(x + \phi(\beta))) - \phi'(\alpha\phi^{\leftarrow}(x + \phi(\beta))).\phi''(\phi^{\leftarrow}(x + \phi(\beta)))}{\phi'(\phi^{\leftarrow}(x + \phi(\beta)))^3}$$

Because ϕ is a generator of an Archimedean copula, ϕ is positive, and $\phi^{'}$ is negative. So, finally, $d^2 f_{12}(x)/dx^2$ is negative if and only if

$$\alpha\phi''(\alpha\phi^{\leftarrow}(x + \phi(\beta))).\phi'(\phi^{\leftarrow}(x + \phi(\beta))) - \phi'(\alpha\phi^{\leftarrow}(x + \phi(\beta))).\phi''(\phi^{\leftarrow}(x + \phi(\beta))) \geq 0 \tag{4.36}$$

for all x, that is $\alpha\phi''(\alpha y).\phi'(y) - \phi'(\alpha y).\phi''(y) \geq 0$ for all y,or, dividing by $\phi'(y).\phi'(\alpha y)$,

$$\frac{\alpha\phi''(\alpha y)}{\phi'(\alpha y)} - \frac{\phi''(y)}{\phi'(y)} \geq 0 \text{ or } \frac{-\alpha\phi''(\alpha y)}{-\phi'(\alpha y)} \geq \frac{-\phi''(y)}{-\phi'(y)} \text{ for all } y,\ \alpha \leq 1 \tag{4.37}$$

Because $\alpha\phi''(\alpha y) = (\phi'(\alpha y))'$ and $\phi''(y) = (\phi'(y))'$, let $g(t) = D\log -D\phi(t) = D\psi(t)$, then $D^2 f_{12}(x)$ is negative if and only if $g(\alpha y) \geq g(y)$ for all y and $\alpha \leq 1$, that is g is decreasing, or ψ is concave. In this case, f is concave, and, furthermore, $f(0) = 0$. From Lemma 4.4.3 in Nelsen (1999) one gets that f is subadditive.

(ii) Same proof holds : $D^2 f_{21}(x)$ is negative if and only if $g(\alpha y) \geq g(y)$ for all y and $\alpha \geq 1$, that is g is increasing, or ψ is convex.

This finishes the proof of Lemma 4.2.10.

□

Example 4.2.11. *Let C be a Ali-Mikhail-Haq copula (Ali, Mikhail and Haq* (1978)*), with generator $\phi(x) = \log(1 - \theta(1 - x)) - \log x$. Then*

$$\phi'(x) = \frac{\theta}{1 - \theta(1 - x)} - \frac{1}{x} \text{ and } \psi(x) = \log\left(\frac{1}{x} - \frac{\theta}{1 - \theta(1 - x)}\right) \tag{4.38}$$

One gets that

$$\psi''(x) = \frac{-2(1 - \theta)}{\phi'(x)^2} \left[\frac{3\theta^2 x^2 + 3\theta(1 - \theta)x + (1 - \theta)^2}{x^3(1 - \theta(1 - x))^3}\right] \tag{4.39}$$

which is positive. So finally, ψ is a concave function on $[0, 1]$, and so $\Phi(C, t_2)(x, y) \leq \Phi(C, t_1)(x, y)$ for all x, y in $[0, 1] \times [0, 1]$, for all $0 < t_2 \leq t_1 \leq 1$: (X, Y) given $X \leq F_X(t)$ and $Y \leq F_Y(t)$ is less and less positively dependent, as t decreases towards 0.

Example 4.2.12. *Let C be the copula given by (4.2.19) in Nelsen (1999), that is with generator $\phi(x) = \exp(\theta/x) - \exp(\theta)$. Then, for all t_1 and t_2 such that $0 < t_2 \leq t_1 \leq 1$, and let $C_i = \theta/\log[2\exp(\theta/t_i) - \exp(\theta)]$ where $i = 1, 2$. One gets*

$$f_{12}(x) = \exp\left(\frac{\log[2\exp(\theta/t_1) - \exp(\theta)]}{\log[2\exp(\theta/t_2) - \exp(\theta)]} \log(x + 2\exp(\theta/t_2) - \exp(\theta))\right) - 2\exp(\theta/t_1) + \exp(\theta) \tag{4.40}$$

After derivating two times with respect to x, one gets $D^2 f_{12}(x) \geq 0$ and $f_{12}(x)$ is concave. Hence, because $f_{12}(0) = 0$ and $f_{12}(x)$ is convex, then $f_{12}(x)$ is subadditive. For all t_1 and t_2 such that $0 < t_2 \leq t_1 \leq 1$, $f_{12}(x)$ is subadditive : (X, Y) given $X \leq F_X(t)$ and $Y \leq F_Y(t)$ is more and more positively dependent, as t decreases towards 0.

One can notice that this case is an application of Lemma 4.2.10 *:*

$$\psi(x) = \log -\phi'(t) = \frac{\theta}{x} + \log\theta - 2\log x \tag{4.41}$$

is a convex function on $[0,1]$*, and so* $\Phi(C,t_2)(x,y) \geq \Phi(C,t_1)(x,y)$ *for all* x,y *in* $[0,1]\times[0,1]$*, for all* $0 < t_2 \leq t_1 \leq 1$*.*

Example 4.2.13. *Let* C *be a copula in the Gumbel-Barnett family (Gumbel* (1960*a*)*), that is* $\phi(x) = \log(1-\theta\log x)$*. Then*

$$\phi'(x) = \frac{-\theta}{x(1-\theta\log x)} \text{ and } \psi(x) = \log\theta - \log x - \log(1-\theta\log x), \tag{4.42}$$

which is a convex function on $[0,1]$*, and so* $\Phi(C,t_2)(x,y) \geq \Phi(C,t_1)(x,y)$ *for all* x,y *in* $[0,1]\times[0,1]$*, for all* $0 < t_2 \leq t_1 \leq 1$*. In that case* (X,Y) *given* $X \leq F_X(t)$ *and* $Y \leq F_Y(t)$ *is more and more positively dependent as* t *decreases towards* 0 *should be understood as* (X,Y) *given* $X \leq F_X(t)$ *and* $Y \leq F_Y(t)$ *is less and less negatively dependent as* t *decreases towards* 0*. This is a direct implication of the fact that the conditional copula of a Gumbel-Barnett copula remains in this family, with a smaller parameter.*

Example 4.2.14. *Let* C *be a Frank copula, with generator*

$$\phi(x) = -\log\left[(\exp(-\theta x)-1)/(\exp(-\theta)-1)\right],$$

then

$$\phi'(t) = \frac{\theta\exp(-\theta x)}{\exp(-\theta x)-1} \text{ and } \psi(t) = \log\theta - \theta x - \log(1-\exp(-\theta x)), \tag{4.43}$$

which satisfies $\psi''(x) = -\theta^2\exp(-\theta x)/\left[\exp(-\theta x)-1\right]^2 \leq 0$ *:* ψ *is concave, and so* $\Phi(C,t_2)(x,y) \leq \Phi(C,t_1)(x,y)$ *for all* x,y *in* $[0,1]\times[0,1]$*, for all* $0 \leq t_2 \leq t_1 \leq 1$*.*

Example 4.2.15. *Let* C *be a Gumbel copula, with generator* $\phi(x) = (-\log x)^\theta$*,* $\theta \geq 1$*, then*

$$\phi'(x) = -\theta(-\log x)^{\theta-1}/x, \text{ and } \psi(x) = \log\theta - \log x + (\theta-1)\log(\log[-x]) \tag{4.44}$$

This function being twice differentiable, one gets

$$\psi''(x) = \frac{(\log x)^2 - [\theta-1]\log x - [\theta-1]}{x^2[\log x]^2} = \frac{h(\log x)}{x^2[\log x]^2}, \tag{4.45}$$

where $h(y) = y^2 - [\theta-1]y - [\theta-1]$ *: this polynomial has two (real) roots, and one is negative. So finally,* $\psi''(x) \leq 0$ *on* $]0,x_0]$ *and* $\psi''(x) \geq 0$ *on* $[x_0,1]$ *for some* x_0 *:* ψ *is neither concave nor convex.*

4.2.5 Aging and dependence concepts for Archimedean copulae

As mentioned in Section 1.3 of this thesis, if ϕ is a generator, set $F_\phi(x) = 1 - \phi^{\leftarrow}(x)$, for all $x \geq 0$. Further, F_ϕ is cumulative distribution function, of a unimodal distribution on $\mathbb{R}^+$, with mode 0. Note that for independence, $\phi(t) = -\log x$, and therefore, $F_\phi(\cdot)$ is the distribution function of the standard exponential distribution.

Proposition 4.2.16. *Let* C *denote an bivariate Archimedean copula with generator* ϕ*. Then* C *is PQD (i.e.* $C^\perp \preceq C$ *for the pointwise order) if and only if* F_ϕ *is NBU (new better than used).*

Proof. Recall that given C_1 and C_2 two Archimedean copulae, with respective generator ϕ_1 and ϕ_2, $C_1 \preceq C_2$ for the pointwise order if and only if $\phi_1 \circ \phi_2^{\leftarrow}$ is subadditive, i.e. $\phi_1 \circ \phi_2^{\leftarrow}(x+y) \leq \phi_1 \circ \phi_2^{\leftarrow}(x) + \phi_1 \circ \phi_2^{\leftarrow}(y)$ for all $x, y \in [0,1]$ (see e.g. Nelsen (1999)). Hence, if $\phi_1(\cdot) = -\log(\cdot)$ and $\phi_2(\cdot) = \phi(\cdot)$, this yields $-\log(\phi^{\leftarrow})$ is subadditive. Hence, since $F_\phi(x) = 1 - \phi^{\leftarrow}(x)$, it means that $x \mapsto -\log(1 - F_\phi(x)$ is subadditive, which is a characterization of NBU distributions (see Barlow and Proschan (1975)). □

Note that those positive dependence concept and aging properties can be transposed in terms of orderings:

Corollary 4.2.17. *Let C_1 and C_2 denote two bivariate Archimedean copula with respective generators ϕ_1 and ϕ_2. Then $C_1 \preceq_{PQD} C_2$ if and only if $F_{\phi_1} \preceq_{NBU} F_{\phi_2}$.*

Proposition 4.2.18. *Let C denote an bivariate Archimedean copula with generator ϕ. Then C is LTD if and only if F_ϕ is IFR (increasing failure rate).*

Proof. From Definition 4.1.24, recall that given $\boldsymbol{X}$ is said to be left tail increasing (LTD) if $\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x} | \boldsymbol{X} \leq \boldsymbol{x}')$ is decreasing in $\boldsymbol{x}'$ for all $\boldsymbol{x}$. In terms of copulae, it means that

$$\frac{C(x,y)}{x} \leq \frac{C(x',y)}{x'} \text{ for all } y \in [0,1], 0 < x \leq x' \leq 1,$$

or equivalently

$$\frac{\phi^{\leftarrow}(\phi(x) + \phi(y))}{x} \leq \frac{\phi^{\leftarrow}(\phi(x') + \phi(y))}{x'} \text{ for all } y \in [0,1], 0 < x \leq x' \leq 1. \tag{4.46}$$

Since ϕ is strictly decreasing and continuous, it is bijective, and therefore, Equation (4.46) can be written equivalently, setting $u = \phi(x)$, $u' = \phi(x')$ and $v = \phi(y)$,

$$\frac{\phi^{\leftarrow}(u+v)}{\phi^{\leftarrow}(u)} \geq \frac{\phi^{\leftarrow}(u'+v)}{\phi^{\leftarrow}(u')} \text{ for all } v \in [0,\infty), 0 < u \leq u' < \infty, \tag{4.47}$$

because $\phi^{\leftarrow}$ is a decreasing function. Hence, from $F_\phi(x) = 1 - \phi^{\leftarrow}(x)$, note that Equation (4.47) becomes

$$\frac{1 - F_\phi(u+v)}{1 - F_\phi(u)} \geq \frac{1 - F_\phi(u'+v)}{1 - F_\phi(u')} \text{ for all } v \in [0,\infty), 0 < u \leq u' < \infty,$$

or, taking the logarithm,

$$\log[1 - F_\phi(u+v)] - \log[1 - F_\phi(u)] \geq \log[1 - F_\phi(u'+v)] - \log[1 - F_\phi(u')],$$

for all $v \in [0,\infty)$ and $0 < u \leq u' < \infty$, which can be written also

$$\frac{-\log[1 - F_\phi(u+v)] + \log[1 - F_\phi(u)]}{v} \geq \frac{-\log[1 - F_\phi(u'+v)] + \log[1 - F_\phi(u')]}{v},$$

i.e. the increasing rate of function $-\log[1 - F_\phi(\cdot)$ is an increasing function, hence, $-log[1 - F_\phi]$ is convex which is a characterization of IFR distributions (see Barlow and Proschan (1975)). □

Again those positive dependence concept and aging properties can be transposed in terms of orderings:

Corollary 4.2.19. *Let C_1 and C_2 denote two bivariate Archimedean copula with respective generators ϕ_1 and ϕ_2. Then $C_1 \preceq_{LTD} C_2$ if and only if $F_{\phi_1} \preceq_{IFR} F_{\phi_2}$.*

Chapter 5

Upper tails for Archimedean copulae

In chapter 3, we characterized the lower tail behavior of Archimedean copulae. We focused on lower tails for convenience, but studying upper tails should also be interesting (e.g. in terms of extreme values, as we will see in the next chapter, without the Archimedean assumption).

In this chapter, the aim is to obtain analogous results to the third chapter, about tails of Archimedean copulae, but in the upper corner **1**, instead of **0**. In order to derive properties for upper tails, we need alternative characterizations of generators of Archimedean copulae in dimension $d \geq 2$. Section 5.2.

If the study of lower tails was possible using only regular variation (of first and second order) to characterize tail dependence or tail independence, the study of upper tails will be more complicated. More precisely, in Section 5.3, if tail dependence we be considered as in chapter 3 (section 5.3.1), in order to characterize tail independence (section 5.3.2), we will have to separate dependence in independence, and independence in independence (as called in Draisma, Drees, Ferreira and de Haan (2004)). This chapter we will concluded by a short section on possible extensions of tail study, in off-diagonal corners (section 5.4).

5.1 Some additional results on regular variation for Archimedean generators

5.1.1 Multiply monotone functions

Lemma 5.1.1. *Let f be $k \geq 0$ times continuously differentiable function of a real variable defined in a neighbourhood of infinity and such that $f(t) \to 0$ as $t \to \infty$. If $(-D)^k f$ is convex, then for all $j = 0, 1, \ldots, k$, the function $(-D)^j f$ is nonnegative, nonincreasing, convex, and vanishes at infinity. Moreover, there exists a version $D^{k+1} f$ of the Radon-Nikodym derivative of $D^k f$ such that $(-D)^{k+1} f$ is nonnegative, nonincreasing, and vanishes at infinity.*

Proof. We proceed by induction on k. First assume $k = 0$. The assumption is then simply that f is convex and vanishes at infinity. Hence it must be nonnegative and nonincreasing. Moreover, $f(t) = \int_t^\infty (-D) f(s) ds$ for all large enough t, where Df is the right-hand derivative of f. Clearly, $(-D)f$ must be nonnegative and nonincreasing and must vanish at infinity.

Next assume that k is an integer larger than one. Since $(-D)^k f$ is convex, it must converge at infinity to a limit in $[-\infty, \infty]$. This limit must be zero, because otherwise $|f(t)| \to \infty$ as $t \to \infty$. Hence, the function $(-D)^k f$ satisfies the assumptions of the lemma, so that by the induction hypothesis, $(-D)^{k+1} f$ is nonnegative, nonincreasing, and vanishes at infinity. Moreover, since the derivative of $(-D)^{k-1} f$ is equal to $-(-D)^k f$, which is a nondecreasing function, the function $(-D)^{k-1} f$ must be convex. Apply the induction hypothesis once more to complete the proof. □

Remark 5.1.2. *Up to minor changes in terminology, a function f which satisfies the conditions of Lemma 5.1.1 is called multiply monotone or k-times monotone Gneiting (1997) or Williamson (1956). Lemma 5.1.1 is a simplified version of Proposition 4.4 in Gneiting (1997).*

5.1.2 Regular variation and second order properties

The definition of regular variation involves in principle an infinite set of limit relations. However, if a function is known to be convex, then regular variation of the function is equivalent to a single limit relation. Results of this type are known under the name "Monotone Density Theorem," see for instance section 1.7.3 in Bingham, Goldie and Teugels (1987). We will need the following two instances.

Lemma 5.1.3. *Let f be a positive, convex function of a real variable defined in a right-neighbourhood of zero. Let Df be a nondecreasing version of the Radon-Nikodym derivative of f. The function f is regularly varying at zero of index $\tau \in [-\infty, \infty]$ if and only if*

$$\lim_{s \to 0} \frac{sDf(s)}{f(s)} = \tau.$$

Proof. Let c be a positive number such that the domain of f includes the interval $(0, c]$. The function $\log f$ is absolutely continuous with Radon-Nikodym derivative $(Df)/f$. Denote $\tau(s) = sDf(s)/f(s)$. For $0 < s \leq c$, we have

$$f(s) = f(c) \exp\left(- \int_s^c \tau(t) \frac{dt}{t} \right).$$

If additionally $0 < x < \infty$ with $x \neq 1$ and if s is such that also $sx \leq c$, then

$$\begin{aligned} \frac{f(sx)}{f(s)} &= \exp\left(\int_s^{sx} \tau(t) \frac{dt}{t} \right) \\ &= \exp\left(\int_1^{x} \tau(st) \frac{dt}{t} \right). \end{aligned}$$

The argument of the exponent converges to $\tau \log(x)$ as $s \to 0$. Hence indeed $f(sx)/f(s) \to x^\tau$ as $s \to 0$, as required.

Conversely, suppose that f is regularly varying at zero of index τ. By convexity, we have for all $0 < x < \infty$ and all sufficiently small s,

$$f(sx) - f(s) \geq s(x-1)Df(s).$$

If x is not equal to one, we can divide both sides of this inequality by $(x-1)$ and let s decrease to zero to get

$$\begin{aligned} \limsup_{s \to 0} \frac{sDf(s)}{f(s)} &\leq \frac{x^\tau - 1}{x - 1}, & &\text{for all } 1 < x < \infty, \\ \liminf_{s \to 0} \frac{sDf(s)}{f(s)} &\geq \frac{x^\tau - 1}{x - 1}, & &\text{for all } 0 < x < 1. \end{aligned}$$

Since $(x^\tau - 1)/(x-1) \to \tau$ as $x \to 1$ for all $\tau \in [-\infty, \infty]$, we conclude that $sDf(s)/f(s) \to \tau$ as $s \to 0$. □

Lemma 5.1.4. *Let f be a positive, convex function of a real variable defined in a neighbourhood of infinity. Let Df be a nondecreasing version of the Radon-Nikodym derivative of f. The function f is regularly varying at infinity of index $\tau \in [-\infty, \infty]$ if and only if*

$$\lim_{t\to\infty} \frac{tDf(t)}{f(t)} = \tau.$$

Proof. The proof Lemma 5.1.4 is identical to the proof of Lemma 5.1.3. □

Lemma 5.1.5. *Let f be a positive, $k \geq 0$ times continuously differentiable function of a real variable defined in a neighbourhood of infinity. Assume that $(-D)^k f$ is convex and that $f(t) \to 0$ as $t \to \infty$. If f is regularly varying at infinity of index $-\tau \in [-\infty, 0]$, then for all integer $j = 1, \ldots, k+1$,*

$$\lim_{t\to\infty} \frac{t^j (-D)^j f(t)}{f(t)} = \tau(\tau+1)\cdots(\tau+j-1). \tag{5.1}$$

Proof. We proceed by induction on k. In case $k = 0$, the statement is trivially implied by Lemma 5.1.4.

So assume k is a positive integer. Note that by Lemma 5.1.1, the function $(-D)^j f$ is convex for every $j = 0, 1, \ldots, k$ and vanishes at infinity. Hence, by the induction hypothesis, (5.1) holds already for all $j = 1, \ldots, k$, so only the case $j = k+1$ remains to be shown.

First consider the case $0 < \tau < \infty$. Then we know that

$$(-D)^k f(t) \sim \tau(\tau+1)\cdots(\tau+k-1)t^{-k} f(t) \text{ as } t \to \infty.$$

In particular, the function $(-D)^k f$ is regularly varying at infinity of order $-\tau - k$. Apply Lemma 5.1.4 to get

$$(-D)^{k+1} f(t) \sim (\tau + k)t^{-1}(-D)^k f(t) \text{ as } t \to \infty.$$

Combine the two previous displays to see that (5.1) also holds for $j = k+1$.

Next consider the case $\tau = 0$. Then we know that

$$(-D)^k f(t) = o\{t^{-k} f(t)\} \text{ as } t \to \infty.$$

Since $(-D)^k f$ is convex and since $(-D)^{k+1} f$ is nonnegative,

$$(-D)^k f(t/2) - (-D)^k f(t) \geq (t/2)(-D)^{k+1} f(t) \geq 0.$$

Combine the two previous displays with the fact that $f(t/2) \sim f(t)$ as $t \to \infty$ to see that $t^{k+1}(-D)^{k+1} f(t)/f(t) \to 0$ as $t \to \infty$.

Finally, consider the case $\tau = \infty$. For large enough t, we have by induction on k,

$$f(t) = \int_t^\infty \frac{(v-t)^k}{k!}(-D)^{k+1} f(v) dv.$$

Let $1 < x < \infty$. For $v \geq tx^2$, we have $v - t \leq 2(v - tx)$ and thus

$$\begin{aligned} f(t) &\leq \int_t^{tx^2} \frac{(v-t)^k}{k!}(-D)^{k+1} f(v) dv + \int_{tx^2}^\infty \frac{2^k (v-tx)^k}{k!}(-D)^{k+1} f(v) dv \\ &\leq \frac{t^{k+1}(x^2-1)^{k+1}}{(k+1)!}(-D)^{k+1} f(t) + 2^k f(tx). \end{aligned}$$

Since $f(tx)/f(t) \to 0$ as $t \to \infty$, we find

$$\liminf_{t\to\infty} \frac{t^{k+1}(-D)^{k+1} f(t)}{f(t)} \geq \frac{(k+1)!}{(x^2-1)^{k+1}}.$$

Let $x \to 1$ to see that $t^{k+1}(-D)^{k+1} f(t)/f(t) \to \infty$ as $t \to \infty$. □

5.1.3 Miscellaneous results

Lemma 5.1.6. *Let k be positive integer, I be an open real interval, and $f : I \mapsto \mathbb{R}$ be a $(k-1)$ times continuously differentiable function. If $D^{k-1}f$ is absolutely continuous with Radon-Nikodym derivative $D^k f$, then for every $x \in I$ and $(x_1, \ldots, x_k) \in [0,\infty)^k$ for which $x+x_1+\cdots+x_k$ is in I,*

$$\sum_{K\subset\{1,\ldots,k\}} (-1)^{|K|} f\left(x + \sum_{i\in K} x_i\right) = \int_0^{x_1} \cdots \int_0^{x_k} (-D)^k f(x + t_1 + \cdots + t_k) dt_1 \cdots dt_k.$$

Proof. We prove the stated formula by induction on k.

Let k be equal to one. The assumption is simply that f is absolutely continuous with Radon-Nikodym derivative f', and the formula reduces to

$$f(x) - f(x + x_1) = -\int_0^{x_1} f'(x + t_1) dt_1,$$

which is just the definition of absolute continuity.

Let k be larger than one. We have

$$\begin{aligned}
&\sum_{K\subset\{1,\ldots,k\}} (-1)^{|K|} f\left(x + \sum_{i\in K} x_i\right) \\
&= \left(\sum_{\substack{K\subset\{1,\ldots,k\}\\ k\notin K}} + \sum_{\substack{K\subset\{1,\ldots,k\}\\ k\in K}}\right) (-1)^{|K|} f\left(x + \sum_{i\in K} x_i\right) \\
&= \sum_{K\subset\{1,\ldots,k-1\}} (-1)^{|K|} f\left(x + \sum_{i\in K} x_i\right) \\
&\quad + \sum_{K\subset\{1,\ldots,k-1\}} (-1)^{|K|+1} f\left(x + \sum_{i\in K} x_i + x_k\right) \\
&= \sum_{K\subset\{1,\ldots,k-1\}} (-1)^{|K|} \left\{ f\left(x + \sum_{i\in K} x_i\right) - f\left(x + \sum_{i\in K} x_i + x_k\right) \right\}.
\end{aligned}$$

Fix x_k and apply the induction hypothesis to the function $y \mapsto g(y) = f(y) - f(y + x_k)$ to arrive at

$$\sum_{K\subset\{1,\ldots,k\}} (-1)^{|K|} f\left(x + \sum_{i\in K} x_i\right) = \int_0^{x_1} \cdots \int_0^{x_{k-1}} (-D)^{k-1} g(x + t_1 + \cdots + t_{k-1}) dt_1 \cdots dt_{k-1}.$$

Since $D^{k-1}f$ is absolutely continuous with Radon-Nikodym derivative $D^k f$, the integrand in the previous display is equal to

$$\begin{aligned}
&(-D)^{k-1} g(x + t_1 + \cdots + t_{k-1}) \\
&= (-D)^{k-1} f(x + t_1 + \cdots + t_{k-1}) - (-D)^{k-1} f(x + t_1 + \cdots + t_{k-1} + x_k) \\
&= \int_0^{x_k} (-D)^k f(x + t_1 + \cdots + t_k) dt_k.
\end{aligned}$$

Substitute this expression for the integrand into the $(k-1)$-tuple integral above to arrive at the desired formula. $\square$

Lemma 5.1.7. *For every positive integer d and for every $(x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$\begin{aligned}\sum_{J \subset \{1,\ldots,d\}:|J|\geq 1} (-1)^{|J|-1} \max\{x_j \mid j \in J\} &= \min(x_1, \ldots, x_d), \\ \sum_{J \subset \{1,\ldots,d\}:|J|\geq 1} (-1)^{|J|-1} \min\{x_j \mid j \in J\} &= \max(x_1, \ldots, x_d),\end{aligned}$$

Proof. Note that the second formula follows from applying the first one to the vector $(-x_1, \ldots, -x_d)$. So it suffices to show the first formula. We proceed by induction on d.

If $d = 1$, then both sides of the stated equation are equal to x_1.

Let $d \geq 2$ and assume the hypothesis holds for dimension $d-1$. Without loss of generality, assume that $\max(x_1, \ldots, x_d) = x_d$. Then

$$\begin{aligned}&\sum_{J \subset \{1,\ldots,d\}:|J|\geq 1} (-1)^{|J|-1} \max\{x_j \mid j \in J\} \\ &= \sum_{J \subset \{1,\ldots,d-1\}:|J|\geq 1} (-1)^{|J|-1} \max\{x_j \mid j \in J\} + \sum_{J \subset \{1,\ldots,d\}:d \in J} (-1)^{|J|-1} x_d.\end{aligned}$$

By the induction hypothesis, the first term on the right-hand side of the previous display is equal to $\min(x_1, \ldots, x_{d-1}) = \min(x_1, \ldots, x_d)$. The second term on the right-hand side of the previous display is equal to x_d times

$$\sum_{J \subset \{1,\ldots,d-1\}} (-1)^{|J|} = \sum_{k=0}^{d-1} (-1)^k \binom{d-1}{k} = 0.$$

$\square$

5.2 Characterizations of Archimedean copulae in dimension $d \geq 2$

Archimedean copulae have been introduced in the first chapter, in section 1.5. But as we will see in this section, alternative characterizations can be considered.

Definition 5.2.1. *Let d be an integer, at least two. A function $\psi : [0,1] \to [0,\infty]$ is called a* generator *of order d if the following conditions hold:*

(i) ψ is decreasing and $\psi(1) = 0$;

(ii) the generalized inverse, $\psi^{\leftarrow} : [0,\infty] \to [0,1]$, of ψ, defined by

$$\psi^{\leftarrow}(t) = \inf\{u \in [0,1] \mid \psi(u) \leq t\} \text{ for all } t \in [0,\infty],$$

is $d-2$ times continuously differentiable on $(0,\infty)$;

(iii) the function $(-D)^{(d-2)}\psi^{\leftarrow}$ is convex. The generator ψ is called strict *if $\psi(0) = \infty$.*

The definition is stronger than it looks at first sight.

Lemma 5.2.2. *If ψ is a generator of order d, then for each $k = 2, \dots, d$:*

(i) ψ is a generator of order k;

(ii) the function $(-D)^{k-2}\psi^{\leftarrow}$ is convex, nonnegative, nonincreasing and converges at infinity to zero.

Proof. Note that if $0 \le s \le t \le \infty$, then $\{u \in [0,1] \mid \psi(u) \le s\}$ is a subset of $\{u \in [0,1] \mid \psi(u) \le t\}$, whence $\psi^{\leftarrow}(t) \le \psi^{\leftarrow}(s)$; so $\psi^{\leftarrow}$ is nonincreasing. Moreover, for arbitrary $\delta \in (0,1]$, if $t \in [\psi(\delta), \infty)$, then $\psi^{\leftarrow}(t) \le \delta$; hence $\psi^{\leftarrow}(t) \to 0$ as $t \to \infty$.

Assume first $d = 2$. By assumption, $\psi^{\leftarrow}$ is convex, so there is nothing to prove.

Second, assume $d \ge 3$. By Taylor's formula, for all $0 < s < t < \infty$ there exists $u \in [s,t]$ such that

$$\psi^{\leftarrow}(t) = \sum_{j=0}^{d-3} \frac{(t-s)^j}{j!} D^j \psi^{\leftarrow}(s) + \frac{(t-s)^{d-2}}{(d-2)!} D^{d-2}\psi^{\leftarrow}(u).$$

Since $(-D)^{d-2}\psi^{\leftarrow}$ is convex, it converges at infinity to some element a of $[-\infty, \infty]$. If a is different from zero, then the function $D^{d-2}\psi^{\leftarrow}$ is ultimately of constant sign and there exist $0 < s < \infty$ and $0 < \varepsilon < \infty$ such that $|D^{d-2}\psi^{\leftarrow}(u)| \ge \varepsilon$ for all $u \in [s, \infty)$. In that case, the expression on the right-hand side of the previous display must diverge to infinity as $t \to \infty$, which is a contradiction to $\psi^{\leftarrow}(t) \to 0$ as $t \to \infty$.

Hence $(-D)^{d-2}\psi^{\leftarrow}(t) \to 0$ as $t \to \infty$. Since $(-D)^{d-2}\psi^{\leftarrow}$ is also convex, this forces $(-D)^{d-2}\psi^{\leftarrow}$ to be nonnegative and nonincreasing.

Finally, since the function $(-D)^{d-3}\psi^{\leftarrow}$ is continuously differentiable and since its derivative, $-(-D)^{d-2}\psi^{\leftarrow}$, is nondecreasing, $(-D)^{d-3}\psi^{\leftarrow}$ must be convex. Hence $\psi^{\leftarrow}$ is also a generator of order $d-1$. By induction, $\psi^{\leftarrow}$ must be a generator of order $k = 2, \dots, d$. $\square$

Remark 5.2.3. *An easy sufficient condition for a decreasing function $\psi : [0,1] \to [0,\infty]$ with $\psi(1) = 0$ to be a generator of order d is that its generalized inverse $\psi^{\leftarrow}$ is d times differentiable and $(-D)^d\psi^{\leftarrow}$ is nonnegative.*

Remark 5.2.4. *If ψ is a generator of order d, then $D^{d-2}\psi^{\leftarrow}$ is convex and nonincreasing if d is even and $D^{d-2}\psi^{\leftarrow}$ is concave and nondecreasing if d is odd. In all cases, $D^{d-2}\psi^{\leftarrow}$ is absolutely continuous, and there exists a version of its Radon-Nikodym derivative, $D^{d-1}\psi^{\leftarrow}$, such that the function $(-D)^{d-1}\psi^{\leftarrow} = (-1)^{d-1}D^{d-1}\psi^{\leftarrow}$ is nonincreasing and nonnegative.*

Remark 5.2.5. *If ψ is a generator, then $\psi^{\leftarrow}(t) \to 1$ as $t \to 0$. For, if $0 < \delta < 1$ and $0 < t < \psi(1-\delta)$, then $\{u \in [0,1] \mid \psi(u) \le t\} \subset (1-\delta, 1]$, whence $\psi^{\leftarrow}(t) \ge 1 - \delta$.*

Remark 5.2.6. *If a generator ψ is strict, then $\psi^{\leftarrow}$ is just the ordinary inverse function of ψ. If ψ is not strict, that is, if $\psi(0)$ is finite, then $\psi^{\leftarrow}(t) = 0$ for all $t \ge \psi(0)$.*

Definition 5.2.7. *A d-variate copula, C, is called* Archimedean *if there exists a generator, ψ, of order d such that*

$$C(u_1, \dots, u_d) = \psi^{\leftarrow}\{\psi(u_1) + \cdots + \psi(u_d)\}, \tag{5.2}$$

for all $(u_1, \dots, u_d) \in [0,1]^d$. An Archimedean copula is called strict *if its generator is strict.*

If ψ is a generator of order d, then the right-hand side of (5.2) defines a genuine copula. The C-volume of hyperrectangles can be expressed directly in terms of the derivatives of $\psi^{\leftarrow}$.

Theorem 5.2.8. *If ψ is a generator of order d, then the right-hand side of equation (5.2) defines a genuine copula. For $\boldsymbol{u}$ and $\boldsymbol{v}$ in $[0,1]^d$ such that $u_i < v_i$ for all $i = 1, \ldots, d$ and such that the set $J = \{j = 1, \ldots, d \mid u_j > 0\}$ is not empty, the C-volume of the hyperrectangle $(\boldsymbol{u}, \boldsymbol{v}] = \prod_1^d (u_i, v_i]$ is given by the following formulas: if $J \neq \{1, \ldots, d\}$, then $C((\boldsymbol{u}, \boldsymbol{v}])$ is equal to*

$$\int_{\prod_{j \in J}[\psi(v_j), \psi(u_j)]} (-D)^{|J|} \psi^{\leftarrow} \left(\sum_{j \in J} y_j + \sum_{j \in J^c} \psi(v_j) \right) d(y_j)_{j \in J}, \tag{5.3}$$

and if $J = \{1, \ldots, d\}$, then $C((\boldsymbol{u}, \boldsymbol{v}])$ is equal to

$$\int_{\psi(v_1)}^{\psi(u_1)} \cdots \int_{\psi(v_{d-1})}^{\psi(u_{d-1})} h(y_1 + \cdots + y_{d-1}) dy_1 \cdots dy_{d-1} \tag{5.4}$$

with

$$h(y) = (-D)^{d-1} \psi^{\leftarrow} \{y + \psi(v_d)\} - (-D)^{d-1} \psi^{\leftarrow} \{y + \psi(u_d)\}.$$

If $\psi^{\leftarrow}$ is d times continuously differentiable, then (5.3) also holds for $J = \{1, \ldots, d\}$.

Proof. Let C be given by (5.2). It is immediately clear that C is grounded, that is, $C(u_1, \ldots, u_d) = 0$ as soon as $\min(u_1, \ldots, u_d) = 0$, and that the marginals of C are uniform, that is, if there exist $i = 1, \ldots, d$ such that $u_j = 1$ if $i \neq j$, then $C(u_1, \ldots, u_d) = u_i$.

It remains to show that the C-volume of any hyperrectangle in $[0,1]^d$ is non-negative. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be points in $[0,1]^d$ such that $u_i < v_i$ for every $i = 1, \ldots, d$. The C-volume of the hyperrectangle $(\boldsymbol{u}, \boldsymbol{v}] = \prod_1^d (u_i, v_i]$ is defined by

$$C((\boldsymbol{u}, \boldsymbol{v}]) = \sum_{I \subset \{1, \ldots, d\}} (-1)^{|I|} C(\boldsymbol{w}_I)$$

where the vector $\boldsymbol{w}_I$ is defined by $(\boldsymbol{w}_I)_i = u_i$ if $i \in I$ and $(\boldsymbol{w}_I)_i = v_i$ if $i \in \{1, \ldots, d\} \setminus I$.

Let $J = \{j \in \{1, \ldots, d\} \mid u_j > 0\}$. Since C is grounded, $C(\boldsymbol{w}_I)$ is equal to zero if I is not a subset of J. Hence the formula in the previous display reduces to

$$C((\boldsymbol{u}, \boldsymbol{v}]) = \sum_{I \subset J} (-1)^{|I|} C(\boldsymbol{w}_I).$$

If J is empty, then $\boldsymbol{u}$ is equal to the origin, and $C((\boldsymbol{u}, \boldsymbol{v}]) = C(\boldsymbol{v}) \geq 0$, as required.

Assume J is not empty. Then

$$C((\boldsymbol{u}, \boldsymbol{v}]) = \sum_{I \subset J} (-1)^{|I|} \psi^{\leftarrow} \left(\sum_{i \in I} \psi(u_i) + \sum_{i \in I^c} \psi(v_i) \right).$$

Denote $\Delta_j = \psi(u_j) - \psi(v_j)$ for $j \in J$; note that $0 < \Delta_j < \infty$. We have

$$C((\boldsymbol{u}, \boldsymbol{v}]) = \sum_{I \subset J} (-1)^{|I|} \psi^{\leftarrow} \left(\sum_{i=1}^d \psi(v_i) + \sum_{i \in I} \Delta_i \right).$$

First, if $J \neq \{1, \ldots, d\}$, then we can apply Lemma 5.1.6 to the right-hand side of the previous display, finding that $C((\boldsymbol{u}, \boldsymbol{v}])$ is equal to

$$\int_{\prod_{j \in J}[0, \Delta_j]} (-D)^{|J|} \psi^{\leftarrow} \left(\sum_{i=1}^d \psi(v_i) + \sum_{j \in J} t_j \right) d(t_j)_{j \in J}.$$

By Lemma 5.2.2 and Remark 5.2.4, this expression is nonnegative, as required. Substitute $y_j = t_j + \psi(v_j)$ to arrive at equation (5.3).

Second, suppose that $J = \{1, \ldots, d\}$. Then $C((\boldsymbol{u}, \boldsymbol{v}])$ is equal to

$$\begin{aligned}
&\left(\sum_{\substack{I \subset \{1,\ldots,d\} \\ d \notin I}} + \sum_{\substack{I \subset \{1,\ldots,d\} \\ d \in I}} \right) (-1)^{|I|} \psi^{\leftarrow} \left(\sum_{i=1}^{d} \psi(v_i) + \sum_{i \in I} \Delta_i \right) \\
&= \sum_{I \subset \{1,\ldots,d-1\}} (-1)^{|I|} \psi^{\leftarrow} \left(\sum_{i=1}^{d} \psi(v_i) + \sum_{i \in I} \Delta_i \right) \\
&\quad + \sum_{I \subset \{1,\ldots,d-1\}} (-1)^{|I|+1} \psi^{\leftarrow} \left(\sum_{i=1}^{d} \psi(v_i) + \sum_{i \in I} \Delta_i + \Delta_d \right) \\
&= \sum_{I \subset \{1,\ldots,d-1\}} (-1)^{|I|} g \left(\sum_{i=1}^{d} \psi(v_i) + \sum_{i \in I} \Delta_i \right),
\end{aligned}$$

where, for fixed Δ_d,

$$g(y) = \psi^{\leftarrow}(y) - \psi^{\leftarrow}(y + \Delta_d).$$

Apply Lemma 5.1.6 to the function g and substitute $y_j = t_j + \psi(v_j)$ to arrive at equation (5.4). By Remark 5.2.4, this expression is nonnegative. $\square$

5.3 Joint upper tail

In this section, we study the upper tail of a general multivariate Archimedean copula. In particular, we are interested in the asymptotic behavior of the joint survival function of such a copula in the neighbourhood of the upper vertex of the d-dimensional hypercube, and also in the conditional distribution of the corresponding random vector given that some but not necessarily all of its components are close to one.

It turns out that the crucial ingredient here is the behavior of the generator ψ in the neighbourhood of one, or equivalently of the function $s \mapsto \psi(1-s)$ in the neighbourhood of zero. Since $\psi(1) = 0$ and since ψ is convex, we have always $\psi(1-s) = -sD\psi(1) + o(s)$ as $s \to 0$. Hence, if the function $\psi(1 - \cdot)$ is regularly varying at zero of some index θ, then necessarily $1 \leq \theta \leq \infty$. Moreover, if $(-D)\psi(1) > 0$, then necessarily $\theta = 1$, although the converse is not true, that is, it may happen that $\theta = 1$ and $(-D)\psi(1) = 0$.

There are two major cases. On the one hand, if $\theta > 1$, then the copula is in the max-domain of attraction of the Gumbel or logistic max-stable dependence structure (subsection 5.3.1). On the other hand, if $\theta = 1$, then all bivariate coefficients of upper tail dependence are equal to zero (subsection 5.3.2). The theory then branches further into two cases according to whether $(-D)\psi(1)$ is positive or zero.

5.3.1 Asymptotic dependence

Theorem 5.3.1. *Let $\boldsymbol{U}$ be a d-variate random vector with distribution function C, an Archimedean copula with generator ψ. If the function $s \mapsto \psi(1 - s)$ is regularly varying at* *zero with index* $\theta \in [1, \infty]$, *then for all* $\boldsymbol{x} \in (0, \infty)^d$,

$$\lim_{s \to 0} s^{-1} \mathbb{P}(\exists i = 1, \ldots, d : U_i \geq 1 - s x_i) = \mathcal{L}(\boldsymbol{x}), \tag{5.5}$$
$$\lim_{s \to 0} s^{-1} \mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - s x_i) = r(\boldsymbol{x}). \tag{5.6}$$

with

$$\mathcal{L}(\boldsymbol{x}) = \begin{cases} (x_1^\theta + \cdots + x_d^\theta)^{1/\theta} & \text{if } 1 \leq \theta < \infty, \\ \max(x_1, \ldots, x_d) & \text{if } \theta = \infty, \end{cases}$$

and

$$r(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \theta = 1, \\ \displaystyle\sum_{I \subset \{1,\ldots,d\} : |I| \geq 1} (-1)^{|I|-1} \left(\sum_{i \in I} x_i^\theta \right)^{1/\theta} & \text{if } 1 \leq \theta < \infty, \\ \min(x_1, \ldots, x_d) & \text{if } \theta = \infty. \end{cases}$$

Proof. Equation (5.6) follows straightforwardly from equation (5.5) by the inverse inclusion-exclusion formula,

$$\begin{aligned} &\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - s x_i) \\ &= \sum_{I \subset \{1,\ldots,d\} : |I| \geq 1} (-1)^{|I|-1} \mathbb{P}(\exists i \in I : U_i \geq 1 - s x_i), \end{aligned}$$

and by the fact that for every subset I of $\{1, \ldots, d\}$ of cardinality at least two, the distribution function of the vector $(U_i)_{i \in I}$ is given by the $|I|$-variate Archimedean copula with generator ψ. For the case $\theta = \infty$, see also Lemma 5.1.7. So it remains to show (5.5).

First, consider the case $1 \leq \theta < \infty$. We have

$$\begin{aligned} & s^{-1} \mathbb{P}(\exists i = 1, \ldots, d : 1 - U_i \leq s x_i) \\ &= s^{-1} [1 - \psi^{\leftarrow} \{\psi(1 - s x_1) + \cdots + \psi(1 - s x_d)\}] \\ &= \frac{1}{1 - \psi^{\leftarrow}(\psi(1 - s))} \\ &\quad \times \left[1 - \psi^{\leftarrow} \left\{ \psi(1 - s) \left(\frac{\psi(1 - s x_1)}{\psi(1 - s)} + \cdots + \frac{\psi(1 - s x_d)}{\psi(1 - s)} \right) \right\} \right]. \end{aligned}$$

The function $x \mapsto 1/\psi(1 - 1/x)$ is regularly varying at infinity with index θ. Therefore, its inverse function, the function $t \mapsto 1/\{1 - \psi^{\leftarrow}(1/t)\}$ is regularly varying at infinity with index $1/\theta$ (Bingham, Goldie and Teugels (1987), Theorem 1.5.12), and thus the function $1 - \psi^{\leftarrow}$ is regularly varying at zero with index $1/\theta$. By the Uniform Convergence Theorem (Bingham, Goldie and Teugels (1987), Theorem 1.5.2), the right-hand side of the previous display converges to the stated expression for $\mathcal{L}(\boldsymbol{x})$.

Second, consider the case $\theta = \infty$. Pick $1 < \lambda < \infty$. Since $\psi(1 - \cdot)$ is regularly varying at zero of index ∞, we have $\psi(1 - \lambda t)/\psi(1 - t) \to \infty$ as $t \to 0$ and thus

$$\begin{aligned} \psi\{1 - s \max(x_1, \ldots, x_d)\} &\leq \psi(1 - s x_1) + \cdots + \psi(1 - s x_d) \\ &\leq d \psi\{1 - s \max(x_1, \ldots, x_d)\} \\ &\leq \psi\{1 - \lambda s \max(x_1, \ldots, x_d)\} \end{aligned}$$

for all s in a right-neighbourhood of zero. Apply the function $1 - \psi^{\leftarrow}$ to the various parts of this inequality, multiply by s^{-1} and let s decrease to zero to find

$$\begin{aligned}
\max(x_1, \ldots, x_d) &\leq \liminf_{s \to 0} s^{-1} \mathbb{P}(\exists i = 1, \ldots, d : U_i \geq 1 - sx_i) \\
&\leq \limsup_{s \to 0} s^{-1} \mathbb{P}(\exists i = 1, \ldots, d : U_i \geq 1 - sx_i) \\
&\leq \lambda \max(x_1, \ldots, x_d).
\end{aligned}$$

Let λ decrease to one to obtain the stated result. □

Corollary 5.3.2. *Under the conditions of Theorem 5.3.1, if $1 < \theta \leq \infty$, then for every non-empty subset J of $\{1, \ldots, d\}$, every $(x_1, \ldots, x_d) \in (0, \infty)^d$ and every $(y_j)_{j \in J} \in (0, \infty)^{|J|}$,*

$$\begin{aligned}
&\mathbb{P}(\forall j = 1, \ldots, d : U_j \geq 1 - sx_j \mid \forall j \in J : U_j \geq 1 - sy_j) \\
&\to \frac{r_d(z_1, \ldots, z_d)}{r_{|J|}\left((y_j)_{j \in J}\right)} \text{ as } s \to 0,
\end{aligned} \tag{5.7}$$

where $z_j = min(x_j, y_j)$ for $j \in J$ and $z_j = x_j$ for $j \in J^c$, and

$$r_k(u_1, \ldots, u_k) = \begin{cases} \displaystyle\sum_{I \subset \{1, \ldots, k\} : |I| \geq 1} (-1)^{|I|-1} \left(\sum_{i \in I} u_i^{\theta} \right)^{1/\theta} & \text{if } 1 < \theta < \infty, \\ \min(u_1, \ldots, u_d) & \text{if } \theta = \infty, \end{cases}$$

for all positive integer k and all $(u_1, \ldots, u_k) \in (0, \infty)^k$.

Proof. The corollary follows from Theorem 5.3.1 in the same way as Corollary 3.5.2 follows from Theorem 3.5.1. □

Remark 5.3.3. *The max-stable dependence structure corresponding to the limit in Theorem 5.3.1 is a special case of the so-called Gumbel or logistic dependence structure Joe (1990). The case $\theta = 1$ corresponds to independence and the case $\theta = \infty$ to comonotonicity.*

Note that the case $d = 2$ and $1 \leq \theta < \infty$ of Theorem 5.3.1 was already established in Capéraà, Fougères and Genest (2000), Proposition 4.1, but then for the more general class of Archimax copulae.

Example 5.3.4. *Again, several Archimedean copulae satisfy assumptions of Theorem 5.3.1. Gumbel's copula, with generator $\psi(t) = (-\log t)^{\alpha}$, with $\alpha \in [1, \infty)$ is such that $\psi(1-t)$ regularly varying at origin with index $\theta = \alpha$. Among Archimedean copulae described in Nelsen (1999), generator denoted (18) defined as $\psi(t) = e^{\alpha/(t-1)}$, with $\alpha \in (2, \infty)$, is regularly varying with infinite tail index, and therefore $\theta = \infty$. Note that generators of Joe copula (6) and Genest and Ghoudi copula (15), defined respectively by $\psi(t) = -\log(1 - (1-t)^{\alpha})$ for all $\alpha \in [1, \infty)$ and $\psi(t(1 - t^{1/\alpha})^{\alpha}$ for all $a \in [1, \infty)$ both satisfy assumptions of Theorem 5.3.1, with tail index $\theta = \alpha$. Most of the other Archimedean copulae have less weight in upper tails, and $\theta = 0$, e.g. Frank copula (5), with generator $\phi(t) = -\log((e^{-\alpha t} - 1)/(e^{-\alpha} - 1))$ with $\alpha \in \mathbb{R}/\{0\}$, or Clayton copula.*

Note that assumptions in Theorem 5.3.1 can be related to *standard* results on upper tails for Archimedean copulae (see Joe (1997)): $\lambda_U = 2 - D(\psi^{\leftarrow} \circ 2\psi)(1)$ and therefore upper tail dependence implies $D\psi(1) = 0$, while $-D\psi(1) > 0$ implies upper tail independence.

Example 5.3.5. *As in Example 3.5.8, an application of Theorem 5.3.1 can be obtained in a frailty model. If Θ denotes the frailty, and if $\mathbb{E}(\Theta) = -D\psi(0)$ is infinite, then the associated Archimedean copula has upper tail dependence. For instance, if Θ is Pareto distributed, $\mathbb{P}(\Theta > \theta) = \theta^{-\beta}\mathcal{L}(\theta)$ with $\beta \in [0,1]$ and where $\mathcal{L}$ is a slowly varying function, then, using the Tauberian theorem (see Feller (1971) or Bingham, Goldie and Teugels (1987)), the Laplace transform of Θ at origin satisfies $\phi(t) \sim 1 - t^{\beta}\mathcal{L}(1/t)$ as $t \to 0$, and therefore, this copula has upper tail dependence. Note that more precisely, $\lambda_U = 2 - 2^{\beta}$.*

5.3.2 Asymptotic independence

If $\theta = 1$ in Theorem 5.3.1, then the only information in (5.6) is that for all $\boldsymbol{x} \in (0,\infty)^d$,

$$\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - sx_i) = o(s) \text{ as } s \to 0. \tag{5.8}$$

The convergence to zero of the probability on the left-hand side of this display can be arbitrarily fast. For instance, if $d = 2$ and $\psi(s) = 1 - s$, the corresponding Archimedean copula being the countermonotonic one, then this probability even vanishes for all s in a right-neighbourhood of zero.

Hence, if $\theta = 1$, then Theorem 5.3.1 is not very informative. The present subsection attempts to give more precise results. It turns out there are two qualitatively different subcases, depending on whether $\lim_{s\to 0}\psi(1-s)/s = (-D)\psi(1)$ is positive (first paragraph below) or zero (second paragraph below).

Independence in independence

Theorem 5.3.6. *Let ψ be a generator of order d such that $\psi^{\leftarrow}$ is d times continuously differentiable and let $\boldsymbol{U}$ be a d-variate random vector with joint distribution function given by the Archimedean copula with generator ψ. If $(-D)^d\psi^{\leftarrow}(0) < \infty$ then $(-D)\psi(1) > 0$ and for all $\boldsymbol{x} \in (0,\infty)^d$*

$$\begin{aligned}&\mathbb{P}(\forall j = 1, \ldots, d : U_j \geq 1 - sx_j) \\ &= s^d(-D)^d\psi^{\leftarrow}(0)\{(-D)\psi(1)\}^d\prod_{j=1}^{d} x_j + o(s^d) \text{ as } s \to 0.\end{aligned} \tag{5.9}$$

Proof. By Theorem 5.2.8, the probability on the left-hand side of (5.9) is equal to

$$\int_{\prod_1^d[0,\psi(1-sx_j)]}(-D)^d\psi\left(\sum_{j=1}^{d} y_j\right)d(y_j)_{j=1}^d.$$

Since $\psi(1 - sx_j) = -sx_jD\psi(1) + o(s)$ as $s \to 0$ and since the integrand converges uniformly to $(-D)^d\psi^{\leftarrow}(0)$, we find (5.9). By induction on d, if $(-D)^d\psi^{\leftarrow}(0)$ is finite then necessarily $(-D)^k\psi^{\leftarrow}(0)$ is finite for all $k = 0, \ldots, d$. Since $D\psi^{\leftarrow}(0) = 1/D\psi(1)$, $D\psi(1)$ must be negative. □

Remark 5.3.7. *In Theorem 5.3.6, it can happen that $(-D)^d\psi^{\leftarrow}(0) = 0$, in which case the right-hand side of (5.9) simplifies to $o(s^d)$.*

Theorem 5.3.8. *Let ψ be a generator of order d and let $\boldsymbol{U}$ be a d-variate random vector with joint distribution function given by the Archimedean copula with generator ψ. Let J be a subset of*

$\{1,\ldots,d\}$ *such that both* J *and* J^c *are non-empty. If* $0 < (-D)^{|J|}\psi^{\leftarrow}(0) < \infty$, *then* $(-D)\psi(1) > 0$, *and for all* $(x_j)_{j\in J} \in (0,\infty)^{|J|}$ *and* $\boldsymbol{v} \in (0,1]^d$,

$$\begin{aligned}&\mathbb{P}(\forall j \in J : U_j \geq 1 - s v_j x_j; \forall j \in J^c : U_j \leq v_j \mid \forall j \in J : U_j \geq 1 - s x_j) \\ &\quad \to \frac{(-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J^c}\psi(v_j)\right)}{(-D)^{|J|}\psi^{\leftarrow}(0)} \prod_{j\in J} v_j \text{ as } s \to 0.\end{aligned} \tag{5.10}$$

Proof. For $(y_j)_{j\in J} \in (0,\infty)^{|J|}$ and $(v_j)_{j\in J^c} \in (0,1]^{|J|^c}$, we have by Theorem 5.2.8, for all sufficiently small, positive s,

$$\begin{aligned}&\mathbb{P}(\forall j \in J : U_j \geq 1 - s y_j; \forall j \in J^c : U_j \leq v_j) \\ &= \int_{\prod_J [0,\psi(1-sy_j)]} (-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J^c}\psi(v_j) + \sum_{j\in J} t_j\right) d(t_j)_{j\in J}.\end{aligned}$$

From the assumption that $(-D)^{|J|}\psi^{\leftarrow}(0)$ is finite, it follows by induction that $(-D)\psi^{\leftarrow}(0)$ is finite and hence that $(-D)\psi(1)$ is non-zero, whence positive. In particular, $\psi(1-t) \sim t(-D)\psi(1)$ as $t \to 0$. If $|J| \leq d-2$, then $(-D)^{|J|}\psi^{\leftarrow}$ is convex and thus continuous; if $|J| = d-1$, then $(-D)^{|J|}\psi^{\leftarrow}$ is by definition equal to minus the right-hand derivative of $(-D)^{d-2}\psi^{\leftarrow}$. In all cases, $(-D)^{|J|}\psi^{\leftarrow}$ is continuous from the right, and thus, as $s \to 0$,

$$\begin{aligned}&\mathbb{P}(\forall j \in J : U_j \geq 1 - s y_j; \forall j \in J^c : U_j \leq v_j) \\ &= s^{|J|}\{(-D)\psi(1)\}^d (-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J^c}\psi(v_j)\right)\prod_{j\in J} y_j + o\left(s^{|J|}\right).\end{aligned}$$

Now write the conditional probability on the left-hand side of (5.10) as a ratio of two probabilities and on each of those apply the asymptotic equivalence in the previous display to arrive at the stated formulas. □

In Theorem 5.3.8, the asymptotic distribution function of the vector $(U_j)_{j\in J^c}$ conditionally on $U_j \geq 1 - s x_j$ for all $j \in J$ is given by

$$F\left((v_j)_{j\in J^c}\right) = \frac{(-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J^c}\psi(v_j)\right)}{(-D)^{|J|}\psi^{\leftarrow}(0)}.$$

with $0 < v_j \leq 1$ for all $j \in J^c$. If $|J| = d-1$, then J^c is a singleton, and F is the left-continuous version of a univariate distribution function with support included in $(0,1]$; remember $(-D)^{d-1}\psi^{\leftarrow}$ is continuous from the right. On the other hand, if $|J| \leq d-2$, then $|J^c| \geq 2$, and F is a $|J^c|$-variate distribution function with support included in $(0,1]^{|J^c|}$. Its marginal distribution functions are continuous and identical, while its copula is Archimedean with generator given by the function

$$\psi_{|J|} = \left(\frac{(-D)^{|J|}\psi^{\leftarrow}(\,\cdot\,)}{(-D)^{|J|}\psi^{\leftarrow}(0)}\right)^{\leftarrow}. \tag{5.11}$$

Dependence in independence

If the function $s \mapsto \psi(1-s)$ is regularly varying at zero of index one but at the same time $D\psi(1) = 0$, that is, $\psi(1-s) = o(s)$ as $s \to 0$, then Theorem 5.3.1 only implies (5.8) while Theorems 5.3.6 and 5.3.8 are not applicable.

Theorem 5.3.9. *Let $\boldsymbol{U}$ be a d-variate random vector with distribution function C, an Archimedean copula with generator ψ. If the function $s \mapsto f(s) = \psi(1-s)$ is regularly varying at zero of index one and if $f(s) = o(s)$ as $s \to 0$, then the function $s \mapsto \ell(s) = s^{-1}f(s)$ is increasing and slowly varying at zero, and for all $\boldsymbol{x} \in (0,1]^d$,*

$$\begin{aligned} &\mathbb{P}(U_1 > 1 - sx_1; \forall j = 2, \ldots, d : U_j \leq 1 - \ell^{\leftarrow}(x_j^{-1}\ell(s)) \mid U_1 > 1 - s) \\ &\to \quad x_1 \min(x_2, \ldots, x_d), \quad \text{as } s \to 0. \end{aligned} \tag{5.12}$$

Proof. Because the function f is positive, convex, and vanishes at zero, the function ℓ is positive and nondecreasing. Moreover, if there would exist $0 < s < t$ such that $\ell(s) = \ell(t)$, then f would be linear on the interval $[0, t]$, contradicting the assumption that $f(s) = o(s)$ as $s \to 0$; hence ℓ is increasing. Since f is regularly varying at zero of index one, ℓ must be slowly varying at zero.

Write $z(s, x) = \ell^{\leftarrow}(x^{-1}\ell(s))$. Since $\ell(s)/\ell(z(s, x)) \to x$ as $s \to 0$ for fixed $0 < x < \infty$, we have

$$\lim_{s\to 0} z(s, x) \quad = \quad 0 \text{ for all } 0 < x < \infty, \tag{5.13}$$

$$\lim_{s\to 0} \frac{s}{z(s, x)} \quad = \quad 0 \text{ for all } 0 < x < 1. \tag{5.14}$$

It is sufficient to prove (5.12) in case all x_j are smaller than one. The probability on the left-hand side of (5.12) can be rewritten as

$$\begin{aligned} &\mathbb{P}(U_1 > 1 - sx_1; \forall j = 2, \ldots, d : U_j \leq 1 - z(s, x) \mid U_1 > 1 - s) \\ &= \quad s^{-1}\big(\mathbb{P}(\forall j = 2, \ldots, d : U_j \leq 1 - z(s, x_j)) \\ &\qquad - \mathbb{P}(U_1 \leq 1 - sx_1; \forall j = 2, \ldots, d : U_j \leq 1 - z(s, x_j))\big) \\ &= \quad s^{-1}\left\{\psi\left(\textstyle\sum_2^d \psi(1 - z(s, x_j))\right) - \psi\left(\psi(1 - sx_1) + \textstyle\sum_2^d \psi(1 - z(s, x_j))\right)\right\} \\ &= \quad s^{-1}\left\{f^{\leftarrow}\left(f(sx_1) + \textstyle\sum_2^d f(z(s, x_j))\right) - f^{\leftarrow}\left(\textstyle\sum_2^d f(z(s, x_j))\right)\right\} \\ &= \quad s^{-1}\int_{\sum_2^d f(z(s,x_j))}^{f(sx_1)+\sum_2^d f(z(s,x_j))} \frac{1}{f'(f^{\leftarrow}(t))}dt. \end{aligned}$$

By (5.13), the upper integration limit tends to zero, and by (5.14), the lower integration limit is asymptotically equivalent to the upper one. Moreover, because the function f is convex, vanishes at zero and is regularly varying at zero of index one,

$$f(s) = f(s) - f(0) \leq sf'(s) \leq f(2s) - f(s) \sim f(s) \text{ as } s \to 0,$$

whence $f'(s) \sim s^{-1}f(s)$ as $s \to 0$ and thus $1/f'(f^{\leftarrow}(t)) \sim t^{-1}f^{\leftarrow}(t)$ as $t \to 0$. By the uniform convergence theorem for regularly varying functions,

$$\begin{aligned} &\mathbb{P}[(U_1 > 1 - sx_1; \forall j = 2, \ldots, d : U_j \leq 1 - z(s, x) \mid U_1 > 1 - s) \\ &\sim \quad s^{-1}f(sx_1)\frac{f^{\leftarrow}\left(\sum_2^d f(z(s, x_j))\right)}{\sum_2^d f(z(s, x_j))} \\ &\sim \quad x_1 \frac{\ell(s)}{\ell \circ f^{\leftarrow}\left(\sum_2^d f(z(s, x_j))\right)}, \text{ as } s \to 0. \end{aligned}$$

Denote $m = \min(x_2, \ldots, x_d)$. Since

$$f(z(s, m)) \leq \textstyle\sum_2^d f(z(s, x_j)) \leq (d - 1)f(z(s, m))$$

and since the function $\ell \circ f^{\leftarrow}$ is slowly varying at zero, we find

$$\ell \circ f^{\leftarrow}\left(\textstyle\sum_2^d f(z(s,x_j))\right) \sim \ell \circ f^{\leftarrow}(f(z(s,m)))) = m^{-1}\ell(s) \text{ as } s \to 0.$$

Combine the two previously displayed asymptotic equivalencies to arrive at (5.12). □

Remark 5.3.10. *As shown in the proof of Theorem 5.3.9, for fixed $0 < x < 1$, the function $s \mapsto \ell^{\leftarrow}(x^{-1}\ell(s))$ converges to zero but at a slower rate than s, that is, $s/\ell^{\leftarrow}(x^{-1}\ell(s)) \to 0$ as $s \to 0$. Thus, conditionally on $U_1 > 1-s$, every U_j with $j \geq 2$ converges in law to one but at a slower rate than s, that is, for every $0 < \varepsilon < 1$, $1 < \lambda < \infty$, and $j = 2, \ldots, d$, we have*

$$\mathbb{P}(1-\varepsilon < U_j < 1 - s\lambda \mid U_1 > 1-s) \to 1 \textit{ as } s \to 0.$$

Remark 5.3.11. *In Theorem 5.3.9, conditionally on the event $U_1 > 1-s$, the remaining variables $U_2, \ldots, U_d$ are asymptotically independent from U_1 but completely dependent on each other.*

Remark 5.3.12. *Since the law of the random vector $\boldsymbol{U}$ is exchangeable, Theorem 5.3.9 obviously generalizes to the case where the conditioning event is $U_j > 1-s$ for some $j = 1, \ldots, d$.*

Next, we study the joint survival function of the vector $\boldsymbol{U}$ in Theorem 5.3.9. A precise asymptotic result on the probability that all U_j are close to the upper end-point is possible under a certain refinement of the condition that the function $\psi(1-\cdot)$ is regularly varying at zero of index one.

Theorem 5.3.13. *Let $\boldsymbol{U}$ be a d-variate random vector with distribution function given by an Archimedean copula with generator ψ. Define $f(s) = \psi(1-s)$. If $s^{-1}f(s) \to 0$ as $s \to 0$ and if the function $s \mapsto \mathcal{L}(s) = s(d/ds)\{s^{-1}f(s)\}$ is positive and slowly varying at zero, then the function $g(s) = sf'(s)/f(s) - 1$ is positive and slowly varying at zero as well, $g(s) \to 0$ as $s \to 0$, and for all $\boldsymbol{x} \in (0,\infty)^d$,*

$$\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - sx_i) \sim sg(s)r(x_1, \ldots, x_d), \textit{ as } s \to 0,$$

with

$$\begin{aligned} r(x_1, \ldots, x_d) &= \sum_{I \subset \{1,\ldots,d\}:|I|\geq 1} (-1)^{|I|}\left(\textstyle\sum_I x_i\right) \log\left(\textstyle\sum_I x_i\right) \\ &= (d-2)! \int_0^{x_1} \cdots \int_0^{x_d} \left(\sum_{i=1}^d t_i\right)^{-(d-1)} dt_1 \cdots dt_d. \end{aligned}$$

Proof. Denote $\mathcal{L}(s) = s(d/ds)\{s^{-1}f(s)\}$. Since $s^{-1}f(s) \to 0$ as $s \to 0$, we have

$$f(s) = s\int_0^s \mathcal{L}(t)\frac{dt}{t}, \text{ for all } 0 \leq s < 1.$$

Note that the function g can be written as

$$g(s) = \frac{sf'(s)}{f(s)} - 1 = \frac{s\mathcal{L}(s)}{f(s)}.$$

Hence, the function g is positive and slowly varying at zero. Moreover, by Fatou's lemma,

$$g(s) = \frac{s\ell(s)}{f(s)} = 1 \Bigg/ \int_0^1 \frac{\mathcal{L}(st)}{\mathcal{L}(s)}\frac{dt}{t} \to 0 \text{ as } s \to 0. \tag{5.15}$$

Also, for every $0 < x < \infty$ and every sufficiently small, positive s,

$$\begin{aligned} f(sx) &= sx \int_0^{sx} \mathcal{L}(t) \frac{dt}{t} \\ &= xf(s) + sx \int_s^{sx} \mathcal{L}(t) \frac{dt}{t} \\ &= xf(s) + xs\mathcal{L}(s) \int_1^x \frac{\mathcal{L}(st)}{\mathcal{L}(s)} \frac{dt}{t}. \end{aligned} \tag{5.16}$$

For sufficiently small, positive s, define $y(\boldsymbol{x}, s)$ by

$$f(sx_1) + \cdots + f(sx_d) = f\{s(x_1 + \cdots + x_d) + sg(s)y(\boldsymbol{x}, s)\}.$$

Since $f(0) = 0$ and f is increasing and convex, $y(\boldsymbol{x}, s)$ is well defined and nonpositive. Now by (5.16), we have on the one hand

$$f(sx_1) + \cdots + f(sx_d) = f(s) \sum_{i=1}^d x_i + s\ell(s) \sum_{i=1}^d x_i \int_1^{x_i} \frac{\mathcal{L}(st)}{\mathcal{L}(s)} \frac{dt}{t},$$

and on the other hand

$$\begin{aligned} & f\{s(x_1 + \cdots + x_d) + sg(s)y(\boldsymbol{x}, s)\} \\ = \quad & f\{sa(\boldsymbol{x}, s)\} = f(s)a(\boldsymbol{x}, s) + s\mathcal{L}(s)a(\boldsymbol{x}, s) \int_1^{a(\boldsymbol{x},s)} \frac{\mathcal{L}(st)}{\mathcal{L}(s)} \frac{dt}{t} \end{aligned}$$

where

$$a(\boldsymbol{x}, s) = \sum_{i=1}^d x_i + g(s)y(\boldsymbol{x}, s).$$

From these equations it follows that

$$\sum_{i=1}^d x_i \int_1^{x_i} \frac{\mathcal{L}(st)}{\mathcal{L}(s)} \frac{dt}{t} = y(\boldsymbol{x}, s) + a(\boldsymbol{x}, s) \int_1^{a(\boldsymbol{x},s)} \frac{\mathcal{L}(st)}{\mathcal{L}(s)} \frac{dt}{t}.$$

The left-hand side of this equation converges to $\sum_1^d x_i \log(x_i)$ by the Uniform Convergence Theorem (Theorem 1.2.1 in Bingham, Goldie and Teugels (1987)). Since $0 < a(\boldsymbol{x}, s) \leq \sum_1^d x_i$, the second term on the right-hand side of the previous equation remains bounded from above as $s \to 0$. Therefore, $y(\boldsymbol{x}, s)$ must remaind bounded from below as $s \to 0$. Since we already knew that $y(\boldsymbol{x}, s)$ is nonpositive, we get $y(\boldsymbol{x}, s) = O(1)$ as $s \to 0$. But since $g(s) \to 0$ as $s \to 0$, , it then follows that

$$a(\boldsymbol{x}, s) \to \sum_{i=1}^d x_i \text{ as } s \to 0.$$

Combine the two previous displays to conclude that, denoting $k(x) = x \log(x)$,

$$y(\boldsymbol{x}, s) \to y(\boldsymbol{x}) = \sum_{i=1}^d k(x_i) - k\left(\sum_{i=1}^d x_i\right), \text{ as } s \to 0.$$

Next, observe that

$$\begin{aligned} & \mathbb{P}(\exists i = 1, \ldots, d : U_i \geq 1 - sx_i) \\ = \quad & 1 - \psi^{\leftarrow}\{\psi(1 - sx_1) + \cdots + \psi(1 - sx_d)\} \\ = \quad & f^{\leftarrow}\{f(sx_1) + \cdots + f(sx_d)\} \\ = \quad & s(x_1 + \cdots + x_d) + sg(s)y(\boldsymbol{x}, s) \\ = \quad & s(x_1 + \cdots + x_d) + sg(s)y(\boldsymbol{x}) + o\{sg(s)\} \text{ as } s \to 0. \end{aligned}$$

By the inverse inclusion-exclusion formula,

$$\begin{aligned}
&\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - sx_i) \\
&= \sum_{I \subset \{1,\ldots,d\}: |I| \geq 1} (-1)^{|I|-1} \mathbb{P}(\exists i \in I : U_i \geq 1 - sx_i) \\
&= \sum_{I \subset \{1,\ldots,d\}: |I| \geq 1} (-1)^{|I|-1} \{ s \textstyle\sum_I x_i + sg(s) \sum_I k(x_i) - sg(s) k(\sum_I x_i) \} \\
&\quad + o\{sg(s)\}, \text{ as } s \to 0.
\end{aligned}$$

Now for every vector $(y_1, \ldots, y_d) \in (0, \infty)^d$,

$$\sum_{I \subset \{1,\ldots,d\}: |I| \geq 1} (-1)^{|I|-1} \textstyle\sum_I y_i = 0,$$

by an elementary combinatorial argument. Combine these two displays to arrive at

$$\begin{aligned}
&\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - sx_i) \\
&= sg(s) \sum_{I \subset \{1,\ldots,d\}: |I| \geq 1} (-1)^{|I|} k(\textstyle\sum_I x_i) + o\{sg(s)\} \text{ as } s \to 0.
\end{aligned}$$

This yields the first expression for $r(\boldsymbol{x})$. The second expression for $r(\boldsymbol{x})$ follows from Lemma 5.1.6 applied to the function k; note that $(-D)k(x) = -\log(x) - 1$ and $(-D)^d k(x) = (d-2)! x^{-(d-1)}$ for all integer $d \geq 2$. $\square$

Remark 5.3.14. *Under the conditions of Theorem 5.3.13, it follows from* (5.16) *that for all* $0 < x < \infty$,

$$f(sx)/f(s) = x + g(s) x \log(x) + o\{g(s)\} \text{ as } s \to 0.$$

Since also $g(s) \to 0$ *as* $s \to 0$, *we see that* f *is second-order regularly varying at zero of index one and auxiliary function* g.

Remark 5.3.15. *A simple sufficient condition for the function* L *in Theorem 5.3.13 to be positive and slowly varying is that the function* f *is twice continuously differentiable and that* f'' *is positive and regularly varying at zero of index* -1.

To see that this condition is sufficient, argue as follows. Note that $(d/ds)\{s\mathcal{L}(s)\} = sf''(s)$. *Since* $0 \leq s\mathcal{L}(s) \leq sf'(s) \to 0$ *as* $s \to 0$, *we get* $s\mathcal{L}(s) = \int_0^s t f''(t) dt$. *In particular,* L *is positive. From the fact that the function* $t \mapsto tf''(t)$ *is slowly varying at zero, it then follows that the function* $s \mapsto s\mathcal{L}(s)$ *must be regularly varying at zero of index one. Hence* L *is indeed slowly varying at zero.*

Remark 5.3.16. *Under the assumptions of Theorem 5.3.13, if* J *is a subset of* $\{1, \ldots, d\}$ *of cardinality at least two, then for all* $\boldsymbol{x} \in (0, \infty)^d$ *and* $(y_j)_{j \in J} \in (0, \infty)^{|J|}$,

$$\mathbb{P}(\forall i = 1, \ldots, d : U_i \geq 1 - sx_i \mid \forall j \in J : U_j \geq 1 - sy_j) = \frac{r(z_1, \ldots, z_d)}{r((y_j)_{j \in J})}$$

as $s \to 0$, *where* $z_j = \min(x_j, y_j)$ *for* $j \in J$ *and* $z_j = x_j$ *for* $j \in J^c$, *and with the function* r *as in Theorem 5.3.13.*

Recall from Theorem 5.3.9 that if J *is a singleton, then the asymptotic conditional distribution of* $\boldsymbol{U}$ *is qualitatively different from the one obtained here.*

An interesting special case is when $d = 2$ and $J = \{1, 2\}$, in which case the conclusion of Theorem 5.3.13 specializes to

$$\begin{aligned}&\mathbb{P}(U \geq 1 - sx, V \geq 1 - sy)\\ &\quad \sim \quad csg(s)\{(x + y)\log(x + y) - x\log(x) - y\log(y)\} \text{ as } s \to 0\end{aligned}$$

for all $(x, y) \in (0, \infty)^2$. In particular, if additionally $(u, v) \in (0, 1]^2$, then

$$\begin{aligned}&\lim_{s\to 0} \mathbb{P}(U \geq 1 - sux, V \geq 1 - svy \mid U \geq 1 - sx, V \geq 1 - sy)\\ &\quad = \quad \frac{(ux + vy)\log(ux + vy) - ux\log(ux) - vy\log(vy)}{(x + y)\log(x + y) - x\log(x) - y\log(y)}.\end{aligned}$$

For fixed (x, y), the expression on the right-hand side of this display is a bivariate distribution function in (u, v) with support included in $(0, 1]^2$.

Remark 5.3.17. *The case $d = 2$ of Theorem 5.3.13 provides examples of distributions for which the coefficient of upper tail dependence is equal to zero and at the same time Ledford and Tawn's index of tail dependence, η, is equal to one, see Ledford and Tawn (1997).*

The case of general d in Theorem 5.3.13 provides examples of distributions exhibiting hidden regular variation with a non-trivial hidden angular measure, see Resnick (2002b), or Maulik and Resnick (2003).

Example 5.3.18. *If ψ is a generator such that there exists $0 < \alpha < \infty$ such that $f(s) = \psi(1 - s) = s(-\log s)^{-\alpha}$ for all positive s in a neighbourhood of zero, then the conditions of Theorem 5.3.9 are satisfied with $\ell(s) = \{\log(1/s)\}^{-\alpha}$, $\ell^{\leftarrow}(t) = \exp(-t^{-1/\alpha})$ and thus*

$$\ell^{\leftarrow}(x^{-1}\ell(s)) = s^{x^\alpha}$$

for all $0 < x < \infty$ and all sufficiently small, positive s. In accordance to Remark 5.3.10, we have $\ell^{\leftarrow}(x^{-1}\ell(s)) \to 0$ and $s/\ell^{\leftarrow}(x^{-1}\ell(s)) \to 0$ as $s \to 0$ for every $0 < x < 1$.

Moreover, the conditions of Theorem 5.3.13 are satisfied with $\mathcal{L}(s) = \alpha(-\log s)^{-\alpha-1}$ and $g(s) = \alpha(-\log s)^{-1}$. In particular, if (U, V) is a random pair with distribution function given by the bivariate Archimedean copula with generator ψ, then for all $(x, y) \in (0, \infty)^2$,

$$\begin{aligned}&\mathbb{P}(U \geq 1 - sx, V \geq 1 - sy)\\ &\quad \sim \quad \alpha s(-\log s)^{-1}\{(x + y)\log(x + y) - x\log(x) - y\log(y)\}, \text{ as } s \to 0.\end{aligned}$$

5.4 Tail behavior at off-diagonal corners

5.4.1 Asymptotic dependence

For a generator ψ of order two or higher, we denote its left derivative on $(0, 1]$ by ψ'. Since ψ is convex and decreasing, the function ψ' is negative on $(0, 1)$, nondecreasing, and continuous from the left; we denote its limit at zero, which exists in $[-\infty, 0)$, by $\psi'(0)$.

Theorem 5.4.1. *Let ψ be a generator of order two, and let (U, V) be a random pair with joint distribution function given by the bivariate Archimedean copula with generator ψ. For all $(x, y) \in (0, \infty)^2$,*

$$\mathbb{P}(U \leq sx, V \geq 1 - sy) = \min\left(x, \frac{\psi'(1)}{\psi'(0)}y\right)s + o(s), \text{ as } s \to 0.$$

Proof. We have

$$\begin{aligned}\mathbb{P}(U \le sx, V \ge 1 - sy) &= \mathbb{P}(U \le sx] - \mathbb{P}[U \le sx, V < 1 - sy) \\ &= sx - \psi^{\leftarrow}\{\psi(sx) + \psi(1 - sy)\} \\ &= sx - \psi^{\leftarrow}\{\psi(sx) - sy\psi'(1) + o(s)\} \text{ as } s \to 0\end{aligned}$$

First, assume that $\psi'(1)/\psi'(0) = 0$, that is, $\psi'(0) = -\infty$ or $\psi'(1) = 0$. Fix $0 < \lambda < 1$. Since ψ is convex, $\psi(sx\lambda) \ge \psi(sx) + sx(\lambda - 1)\psi'(sx)$, and thus

$$\mathbb{P}(U \le sx, V \ge 1 - sy) \le sx - \psi^{\leftarrow}\{\psi(sx\lambda) + sx(1 - \lambda)\psi'(sx) - sy\psi'(1) + o(s)\}$$

as $s \to 0$. The argument in curly brackets in the above display is for all s in a neighbourhood of zero bounded from above by $\psi(sx\lambda)$, and thus

$$\liminf_{s \to 0} s^{-1}\mathbb{P}(U \le sx, V \ge 1 - sy) \le x(1 - \lambda).$$

Let λ increase to zero to find that if $\psi'(1)/\psi'(0) = 0$ then

$$\mathbb{P}(U \le sx, V \ge 1 - sy) = o(s) \text{ as } s \to 0.$$

Second, assume that $\psi'(1)/\psi'(0) > 0$, that is, $-\infty < \psi'(0) \le \psi'(1) < 0$. Then $\psi(0) = -\int_0^1 \psi' < \infty$ and $\psi(u) = \psi(0) + u\psi'(0) + o(u)$ as $u \to 0$ as well as $\psi^{\leftarrow}\{\psi(0) - t\} = -\max(t, 0)/\psi'(0) + o(t)$ as $t \to 0$. We find

$$\begin{aligned}\mathbb{P}(U \le sx, V \ge 1 - sy) &= sx - \psi^{\leftarrow}\{\psi(0) + sx\psi'(0) - sy\psi'(1) + o(s)\} \\ &= sx + \max\{-sx\psi'(0) + sy\psi'(1), 0\}/\psi'(0) + o(s) \\ &= \min[x, \{\psi'(1)/\psi'(0)\}y]s + o(s) \text{ as } s \to 0,\end{aligned}$$

as required. □

Corollary 5.4.2. *Let ψ be a generator of order $d \ge 3$ and let $\boldsymbol{U}$ be a d-variate random vector with joint distribution given by the d-variate Archimedean copula with generator ψ. For all $\boldsymbol{x} \in (0, \infty)^d$ and all $i, j \in \{1, \ldots, d\}$,*

$$\mathbb{P}(U_i \le sx_i, U_j \ge 1 - sx_j) = o(s), \text{ as } s \to 0.$$

Proof. If i is equal to j then there is nothing to prove, so assume i and j are different. The joint distribution function of the random pair (U_i, U_j) is given by the bivariate Archimedean copula with generator ψ, whence, by Theorem 5.4.1,

$$\mathbb{P}(U_i \le sx_i, U_j \ge 1 - sx_j) = \min\left(x_i, \frac{\psi'(1)}{\psi'(0)}x_j\right) s + o(s), \text{ as } s \to 0.$$

The function $\psi^{\leftarrow}$ is continuously differentiable, and its derivative, given by $D\psi^{\leftarrow} = 1/(\psi' \circ \psi^{\leftarrow})$, converges at infinity to zero. Hence $\psi'(0) = \infty$, making the first term on the right-hand side of the previous display vanish, as required. □

5.4.2 Asymptotic independence

For positive x and nonnegative integer k, denote the rising factorial by the Pochhammer symbol

$$(x)_k = \frac{\Gamma(x+k)}{\Gamma(x)} = \begin{cases} 1 & \text{if } k=0, \\ x(x+1)\cdots(x+k-1) & \text{if } k \geq 1. \end{cases}$$

Further, for two positive functions a and b defined in some neighbourhood of $c \in [-\infty, \infty]$, we say that $a(t) \sim b(t)$ as $t \to c$ if $a(t)/b(t) \to 1$ as $t \to c$.

Lemma 5.4.3. *If ψ is a generator of order d and if the function $x \mapsto \psi(1/x)$ is regularly varying at infinity of positive index θ, then for all $k = 0, \ldots, d-1$, the function $(-D)^k \psi^{\leftarrow}$ is regularly varying of index $-\theta^{-1} - k$ and*

$$(-D)^k \psi^{\leftarrow}(t) \sim (\theta^{-1})_k t^{-k} \psi^{\leftarrow}(t), \ \text{ as } t \to \infty. \tag{5.17}$$

Proof. The function $x \mapsto \psi(1/x)$, defined for $x \in [1, \infty)$, is increasing with inverse function $y \mapsto 1/\psi^{\leftarrow}(y)$. Since the former function is regularly varying of positive index θ, the latter must be regularly varying of positive index θ^{-1}. Hence, the function $\psi^{\leftarrow}$ itself is regularly varying of index $-\theta^{-1}$.

We now proceed by induction on d. Note that (5.17) is trivially fulfilled for $k = 0$.

First let $d = 2$. Since $\psi^{\leftarrow}$ is absolutely continuous with Radon-Nikodym derivative $D\psi^{\leftarrow}$ and since $\psi^{\leftarrow}(t) \to 0$ as $t \to \infty$, we have $\psi^{\leftarrow}(t) = -\int_t^\infty D\psi^{\leftarrow}$ for all positive t. Further, since $x \mapsto \psi(1/x)$ is regularly varying with some positive index, necessarily $\psi(u) \to \infty$ as $u \to 0$, that is, ψ is strict. Hence, $\psi^{\leftarrow}(t) > 0$ for all $t \in [0, \infty)$. As, moreover, the function $-D\psi^{\leftarrow}$ is nonnegative and nonincreasing, it must actually be positive. By a version of the monotone density theorem, the function $-D\psi^{\leftarrow}$ must be regularly varying of index $-\theta^{-1} - 1$ and $-D\psi^{\leftarrow}(t) \sim \theta^{-1} t^{-1} \psi^{\leftarrow}(t)$ as $t \to \infty$.

Second, let $d \geq 3$. By Lemma 5.2.2, the function ψ is also a generator of order $d-1$, and thus, by the induction hypothesis, for every $k = 0, \ldots, d-2$, the function $(-D)^k \psi^{\leftarrow}$ is regularly varying at infinity with index $-\theta^{-1} - k$ and the asymptotic relation (5.17) holds true. Further, recall from Remark 5.2.4 that the function $D^{d-2}\psi^{\leftarrow}$ is absolutely continuous and that there exists a version of its Radon-Nikodym derivative, $D^{d-1}\psi^{\leftarrow}$, such that $(-D)^{d-1}\psi^{\leftarrow}$ is nonnegative and nonincreasing. Since $(-D)^{d-2}\psi^{\leftarrow}$ is positive and converges at infinity to zero, we must have $(-D)^{d-2}\psi^{\leftarrow}(t) = \int_t^\infty (-D)^{d-1}\psi^{\leftarrow}$ for all positive t and $(-D)^{d-1}\psi^{\leftarrow}$ must be positive. Moreover, by the induction hypothesis, the function $(-D)^{d-2}\psi^{\leftarrow}$ is regularly varying of index $-\theta^{-1} - (d-2)$ and satisfies (5.17) with $k = d-2$. Apply Lemma 5.1.1 to find that $(-D)^{d-1}\psi^{\leftarrow}$ must be regularly varying of index $-\theta^{-1} - (d-1)$ and

$$\begin{aligned} (-D)^{d-1}\psi^{\leftarrow}(t) &\sim \{\theta^{-1} + (d-2)\} t^{-1} (-D)^{d-2}\psi^{\leftarrow}(t) \\ &\sim (\theta^{-1})_{d-1} t^{-(d-1)} \psi^{\leftarrow}(t) \text{ as } t \to \infty, \end{aligned}$$

as required. □

Theorem 5.4.4. *Let ψ be a generator of order d and let $\boldsymbol{U}$ be a d-variate random vector whose distribution function is the d-variate Archimedean copula C with generator ψ. For every $\boldsymbol{x} \in (0, \infty)^d$ and every subset J of $\{1, \ldots, d\}$ such that both J and its complement J^c are non-empty,*

$$\begin{aligned} &\mathbb{P}(\forall j \in J^c : U_j \leq s x_j; \forall j \in J : U_j \geq 1 - s x_j) \qquad (5.18) \\ &\sim \ (\theta^{-1})_{|J|} \left(\sum_{j \in J^c} x_j^{-\theta} \right)^{-\theta^{-1} - |J|} s\psi(s)^{-|J|} \prod_{j \in J} \psi(1 - s x_j) \end{aligned}$$

as $s \to 0$.

Proof. Let $s > 0$ be small enough such that $sx_j < 1$ for all $j = 1, \dots, d$. The probability on the right-hand side of (5.18) is equal to the C-volume of the hypercube $\prod_1^d I_j$, where $I_j = [0, sx_j]$ if $j \in J^c$ and $I_j = [1 - sx_j, 1]$ if $j \in J$. By (5.3), this volume is equal to

$$\int_{\prod_{j\in J}[0,\psi(1-sx_j)]} (-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J} y_j + \sum_{j\in J^c}\psi(sx_j)\right) d(y_j)_{j\in J}.$$

Since $(-D)^{|J|}\psi^{\leftarrow}$ is nonincreasing, the integral in the previous display is bounded from above by

$$(-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J^c}\psi(sx_j)\right)\prod_{j\in J}\psi(1-sx_j)$$

and is bounded from below by

$$(-D)^{|J|}\psi^{\leftarrow}\left(\sum_{j\in J}\psi(1-sx_j) + \sum_{j\in J^c}\psi(sx_j)\right)\prod_{j\in J}\psi(1-sx_j).$$

Since $\psi(u) \to 0$ as $u \to 1$ and since $x \mapsto \psi(1/x)$ is regularly varying at infinity of index θ, we have

$$\begin{aligned}\sum_{j\in J^c}\psi(sx_j) &\sim \sum_{j\in J}\psi(1-sx_j) + \sum_{j\in J^c}\psi(sx_j) \\ &\sim \psi(s)\sum_{j\in J^c} x_j^{-\theta} \text{ as } s \to 0.\end{aligned}$$

As the function $(-D)^{|J|}\psi^{\leftarrow}$ is regularly varying of index $-\theta^{-1} - |J|$, we find that the probability on the right-hand side of (5.18) is asymptotically equivalent to

$$\left(\sum_{j\in J^c} x_j^{-\theta}\right)^{-\theta^{-1}-|J|} (-D)^{|J|}\psi^{\leftarrow}\{\psi(s)\}\prod_{j\in J}\psi(1-sx_j)$$

as $s \to 0$. Finally, apply equation (5.17) to arrive at equation (5.18). □

Remark 5.4.5. *If in Theorem 5.4.4, it is additionally assumed that the function $x \mapsto \psi(1-1/x)$ is regularly varying at infinity of index $-\alpha$, then the expression on the right-hand side of* (5.18) *can be further simplified by using $\psi(1 - sx_j) \sim x_j^{\alpha}\psi(1-s)$ as $s \to 0$, leading to*

$$\begin{aligned}&\mathbb{P}(\forall j \in J^c : U_j \le sx_j; \forall j \in J : U_j \ge 1 - sx_j) \\ &\sim (\theta^{-1})_{|J|}\left(\sum_{j\in J^c} x_j^{-\theta}\right)^{-\theta^{-1}-|J|}\left(\prod_{j\in J} x_j^{\alpha}\right)\left(\frac{\psi(1-s)}{\psi(s)}\right)^{|J|} s\end{aligned} \tag{5.19}$$

as $s \to 0$. Note that since $x\psi(1-1/x) \to \psi'(1) \in [0,\infty)$ as $x \to \infty$, necessarily $\alpha \ge 1$.

Chapter 6

Extreme and copulae

6.1 Introduction and motivation

Heavy-tailed phenomena have received a lot of attention over the last few years, because of crashes of financial market and some major claims for insurance industry, but also strong deviations of weather from "*usual*" phenomena. But as noticed already in Resnick (1987), "*when* $d = 1$ [univariate case]*, concepts such as extreme values, order statistics and record values have natural definitions, but when* $d > 1$ [multivariate case]*, this is no longer the case as several different concepts of ordering are possible*". Hence, the definition of *extreme* in high dimension is closely related to the "*failure region*" in structural design, defined in Coles and Tawn (1994), or the "*extreme market scenarios*" in Embrechts and Balkema (2004): $\boldsymbol{X}$ is extreme when it belongs to some failure region $\mathcal{A}_u \subset \mathbb{R}^d$, with $\mathbb{P}(\boldsymbol{X} \in \mathcal{A}_u)$ small. Several examples can be considered, e.g. $\mathcal{A}_u$ can be either $\{(x_1 \ldots, x_n) | \max\{x_1 \ldots, x_n\} > u\}$, $\{(x_1 \ldots, x_n) | \min\{x_1 \ldots, x_n\} > u\}$ or $\{\boldsymbol{x} | \boldsymbol{\alpha}'\boldsymbol{x} > u\}$. Tawn (1994) observed, in a survey on applications of multivariate extremes, that many problems that involve extremes are "*inherently multivariate by nature*". As pointed out in Embrechts (2003), hydrology and environmental sciences have initiated the research on extreme value theory (see also de Haan (1985, 1990), Smith (1989)), when seeking how high a sea dyke had to be guard against a $1,000$ year storm. Extreme value theory has now implications in any area in risk management, e.g. determining loss claims for major natural disasters, flood levels of rivers, large downward movements in financial markets, wave heights during storms, minimal performances of financial assets, modeling loss distribution in the context of reinsurance pricing... etc.

In a multivariate context, Gumbel and Goldstein (1964) studied river flows of the Ocmulgee river, and the oldest ages of deaths of men and women in Sweden. Gumbel and Mustafi (1967) also considered multivariate extremes in an hydrological context, on the river flows of the Fox river. In environmental science, Buishand (1984) studied some applications to rainfall, and Walshaw (1991), or Anderson and Turkman (1992), considered some applications to wind data in England. In those contexts, Gumbel structure of dependence (see Chapter 1) appeared as a natural framework.

Similarly, Coles and Tawn (1990) used the logistic model, i.e. Gumbel's model (see Chapter 1) on several environmental series at coastal sites. But as noticed one year after in Coles and Tawn (1991) a rather different model can be obtained based on exceedances, i.e. distribution over high thresholds. Among applications where two components have to exceed given thresholds, Smith (1994) mention the study of daily maximum ozone levels, in Chicago, where high levels of ozone (difficult to model studying only ozone levels) can be characterized by high levels of temperature, and low windspeed. Actually, as pointed out in Dias and Embrechts (2004), when studying dependence in the context of joint exceedances of financial returns, survival Clayton performs better than Gumbel's.

Let X denote a random variable. For univariate extremes, two approaches can be considered to study extremal events. The first one is based on the limiting distribution of an affine transformation of the maxima $X_{n:n} = \max\{X_1, ..., X_n\}$, where $X_1, ..., X_n$ are i.i.d. random variables. Fisher-Tippett Theorem (Theorem 3.2.3 in Embrechts, Klüppelberg and Mikosch (1997)) states that the Generalized Extreme Value (GEV) distribution, defined as

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right) & \xi \neq 0 \\ \exp(-\exp(-x)) & \xi = 0, \end{cases} \tag{6.1}$$

where $1+\xi x > 0$, appears as the limiting distribution of the normalized maxima, for some appropriate normalizing constants. This is also the so-called Von Mises parameterization. Since the limiting distribution does not depend on the affine transformation, it is possible to introduce the notion of *max domain of attraction*: F_X is in the max-domain of attraction of the GEV distribution with parameter ξ if the limiting distribution of maximum of any sample from distribution F_X converges towards the GEV distribution with parameter ξ.

The second approach is based on the study of possible limits for the exceedance distribution, i.e. $X - u$ given $X > u$ when u tends to infinity. Pickands-Balkhema-de Haan Theorem (Theorem 3.4.5 in Embrechts, Klüppelberg and Mikosch (1997)) states that the Generalized Pareto Distribution (GPD), defined as

$$G_{\xi,\beta}(x) = \begin{cases} 1-(1+\xi x/\beta)^{-1/\xi} & \xi \neq 0 \\ 1-\exp(-x/\beta) & \xi = 0, \end{cases} \tag{6.2}$$

where $\beta > 0$ and $x \geq 0$ for $\xi \geq 0$, or $0 \leq x \leq -\beta/\xi$ for $\xi < 0$, appears as the limiting distribution of conditional excess over high thresholds. More precisely, for a large class of random variables X, there exists a function $\beta(\cdot)$ such that

$$\lim_{u\to\infty}\left\{\sup_{0\leq x}\left|\mathbb{P}(X-u \leq x|X>u) - G_{\xi,\beta(u)}(x)\right|\right\} = 0. \tag{6.3}$$

If the two possible limiting distributions for tail events are different, there are related through the same tail index ξ. More precisely, for all $\xi \in \mathbb{R}$, the following assertions are equivalent,

1. $F_X \in MDA(H_\xi)$, i.e. there are (a_n) and (b_n) such that
$$\lim_{n\to\infty}\mathbb{P}(X_{n:n} \leq a_n x + b_n) = H_\xi(x),\ x \in \mathbb{R}.$$
2. There exists a positive, measurable function $a(\cdot)$ such that for $1+\xi x > 0$,
$$\lim_{u\to\infty}\frac{\overline{F}_X(u+xa(u))}{\overline{F}_X(u)} = \lim_{u\to\infty}\mathbb{P}\left(\frac{X-u}{a(u)} > x\,|X>u\right) = \begin{cases} (1+\xi x)^{-1/\xi} & \text{if } \xi \neq 0, \\ \exp(-x) & \text{if } \xi = 0. \end{cases}$$

In the bivariate case, we will see that the two analogous approaches (maximum compononentwise and the threshold approach) are not equivalent.

6.2 Upper Tail Dependence Copulae, hidden regular variation and Ledford and Tawn's approach

6.2.1 Notations

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a d-variate random vector with distribution function F and marginal distribution functions F_j for $j = 1, \ldots, d$. Define the random vector $\boldsymbol{U} = (U_1, \ldots, U_d)$ by $U_j = 1 - F_j(X_j)$ for $j = 1, \ldots, d$. Let C be the distribution function of $\boldsymbol{U}$.

Throughout, we make the assumption that marginal distribution functions $F_1, \dots, F_d$ are continuous and (as in Chapter 2), that $C(\boldsymbol{u})$ is positive for every $\boldsymbol{u} > \mathbf{0}$. Under this assumption each of the random variables U_j is uniformly distributed on the unit interval. As a consequence, C is a copula (although it is not the copula of F), called the *survival copula* of F. Moreover, it means that if $\boldsymbol{x}$ is such that $\mathbb{P}(X_j > x_j) > 0$ for every $j = 1, \dots, d$, then also $\mathbb{P}(\boldsymbol{X} > \boldsymbol{x}) > 0$.

Set $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ with $Y_j = 1/U_j$ for $j = 1, \dots, d$. The marginal distributions of the vector $\boldsymbol{Y}$ are standard Pareto, that is, $\mathbb{P}(Y_j \le y) = 1 - 1/y$ for all $y \ge 1$ and all $j = 1, \dots, d$. Notice that (relatively) large values of X_j correspond to small values of U_j and to large values of Y_j.

Definition 6.2.1. *The* lower tail dependence copula *of C at $\boldsymbol{u} > \mathbf{0}$ is the copula of the conditional distribution of the vector $\boldsymbol{U}$ given $\{\boldsymbol{U} \le \boldsymbol{u}\}$. Such a copula will be noted $C_{\boldsymbol{u}}$.*

Our interest lies in the asymptotic behavior of $C_{\boldsymbol{u}}$ as $\boldsymbol{u}$ tends to $\mathbf{0}$ along some curve in the positive orthant. The class of limit copulae are natural candidates for modelling the dependence structure of a random vector given that all of its components are extreme.

The general theory for the bivariate case has been written down in Charpentier (2004), building upon earlier work for Archimedean random pairs in Juri and Wüthrich (2003) or general symmetric random pairs in Juri and Wüthrich (2004). In these papers, the link between lower tail dependence copulae and extreme value theory is recognized but not fully worked out.

Our aim, then, is to explore this link in depth. Expected benefits are, firstly, a coherent theory of lower tail dependence copulae in arbitrary dimensions and, secondly, an avenue towards non-trivial examples and ensuing statistical methodology.

6.2.2 Multivariate extreme value theory

Let $\mathbb{E} = [0, \infty]^d \setminus \{\mathbf{0}\}$ be the compactified d-dimensional positive orthant punctured at the origin. Complements of sets are to be interpreted with respect to $\mathbb{E}$; for instance $[\mathbf{0}, \boldsymbol{x}]^c = \{\boldsymbol{y} \in [0, \infty]^d \mid \boldsymbol{y} \not\le \boldsymbol{x}\}$ for $\boldsymbol{x} \ge \mathbf{0}$.

A *Radon measure* ν on $\mathbb{E}$ is a non-negative Borel measure such that $\nu(K)$ is finite for every compact $K \subset \mathbb{E}$. Note that a subset K of $\mathbb{E}$ is compact if and only if it is closed in $[0, \infty]^d$ and does not contain the origin. A function $f : \mathbb{E} \to \mathbb{R}$ is said to have compact support if and only if the closure of the set $\{\boldsymbol{x} \in \mathbb{E} \mid f(\boldsymbol{x}) \neq 0\}$ is compact, or equivalently, if there exists $\boldsymbol{y} > \mathbf{0}$ such that $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{E}$ such that $\boldsymbol{x} \le \boldsymbol{y}$.

A sequence of Radon measures ν_n on $\mathbb{E}$ is said to converge *vaguely* to a Radon measure ν on $\mathbb{E}$ if and only if $\int_{\mathbb{E}} f(\boldsymbol{x}) \nu_n(\mathrm{d}\boldsymbol{x}) \to \int_{\mathbb{E}} f(\boldsymbol{x}) \nu_n(\mathrm{d}\boldsymbol{x})$ for every continuous function $f : \mathbb{E} \to \mathbb{R}$ with compact support; notation $\nu_n \xrightarrow{v} \nu$. An equivalent condition is that $\nu_n(B) \to \nu(B)$ for every Borel subset B of $\mathbb{E}$ for which $\nu(\partial B) = 0$, with ∂B denoting the topological boundary of B.

A full treatment of the theory of vague convergence of measures can be found Resnick (1987), chapter 3.

Exponent measure

Recall that $\boldsymbol{Y}$ is a d-variate random vector defined by $Y_j = 1/U_j$ and $U_j = 1 - F_j(X_j)$ for $j = 1, \dots, d$, where $\boldsymbol{X}$ is a d-variate random vector with continuous marginal distribution functions F_j for $j = 1, \dots, d$. Throughout this section, we make the following assumption.

Assumption 6.2.2. *The distribution of $\boldsymbol{Y}$ is multivariate regularly varying with exponent measure ν, that is, there exists a Radon measure ν on $\mathbb{E}$ such that*

$$t\mathbb{P}(t^{-1}\boldsymbol{Y} \in \cdot\,) \xrightarrow{v} \nu(\,\cdot\,), \ \ \text{as } t \to \infty,$$

in $\mathbb{E}$.

The measure ν in Assumption 6.2.2 is necessarily homogeneous of order 1, that is,

$$\nu(s\,\cdot\,) = s^{-1}\nu(\,\cdot\,), \text{ for all } 0 < s < \infty. \tag{6.4}$$

In particular, the measure ν does not put any mass on the rays through infinity,

$$\nu([\mathbf{0},\boldsymbol{\infty})^c) = 0. \tag{6.5}$$

Moreover, the marginal measures must be continuous: $\nu(\{\boldsymbol{x} \in \mathbb{E} \mid x_j = y\}) = 0$, for all $y \in (0,\infty)$ and $j = 1,\ldots,d$. Since the margins of $\boldsymbol{Y}$ are standard Pareto,

$$\nu(\{\boldsymbol{x} \in \mathbb{E} \mid x_j > y\}) = \lim_{t\to\infty} t\mathbb{P}(Y_j > ty) = y^{-1} \tag{6.6}$$

for all $0 < y < \infty$ and all $j = 1,\ldots,d$.

According to Resnick (1987), chapter 5, the distribution function F of $\boldsymbol{X}$ is in the domain of attraction of a d-variate extreme value distribution function G if and only if (i) the marginal distribution functions of F are in the respective domains of attractions of the marginal distribution functions of G, and (ii) Assumption 6.2.2 holds. In that sense, Assumption 6.2.2 is fairly natural. The limiting measure ν in Assumption 6.2.2 is called the *exponent measure* because for $\boldsymbol{y} \in (\mathbf{0},\boldsymbol{\infty}]$,

$$t\mathbb{P}(t^{-1}\boldsymbol{Y} \in [\mathbf{0},\boldsymbol{y}]^c) \to \nu([\mathbf{0},\boldsymbol{y}]^c), \text{ as } t \to \infty, \tag{6.7}$$

and thus

$$\mathbb{P}(\boldsymbol{Y} \le n\boldsymbol{y})^n \to \exp\{-\nu([\mathbf{0},\boldsymbol{y}]^c)\}, \text{ as } n \to \infty,$$

that is, the rescaled component-wise maximum of n independent copies of $\boldsymbol{Y}$ converges in distribution to the d-variate extreme value distribution function defined by the right-hand side of the previous display.

Stable tail dependence function

A convenient way of representing an exponent measure ν is by the following function. For a vector $\boldsymbol{x}$, denote $\boldsymbol{x}^{-1} = (x_1^{-1},\ldots,x_d^{-1})$.

Definition 6.2.3. *The* stable tail dependence function *of the exponent measure ν is defined by*

$$l(\boldsymbol{x}) = \nu([\mathbf{0},\boldsymbol{x}^{-1}]^c) = \nu(\{\boldsymbol{y} \in \mathbb{E} \mid \max(x_1y_1,\ldots,x_dy_d) \ge 1\})$$

for $\boldsymbol{x} \in [\mathbf{0},\boldsymbol{\infty})$.

The stable tail dependence function admits a convenient interpretation in terms of the random vector $\boldsymbol{U}$ with components $U_j = 1/Y_j = 1 - F_j(X_j)$ for $j = 1,\ldots,d$. Under Assumption 6.2.2,

$$\lim_{s\downarrow 0} s^{-1}\mathbb{P}(\exists j = 1,\ldots,d : U_j \le sx_j) = l(\boldsymbol{x}), \tag{6.8}$$

for all $\boldsymbol{x} \in [\mathbf{0},\boldsymbol{\infty})$. The stable tail dependence function satisfies the following properties:

$$\max(x_1,\ldots,x_d) \le l(\boldsymbol{x}) \le x_1 + \cdots + x_d, \text{ for all } \boldsymbol{x} \in [\mathbf{0},\boldsymbol{\infty}); \tag{6.9}$$

$$l(c\boldsymbol{x}) = cl(\boldsymbol{x}) \text{ for all } 0 < c < \infty,\ \boldsymbol{x} \in [\mathbf{0},\boldsymbol{\infty}); \tag{6.10}$$

$$l \text{ is convex.} \tag{6.11}$$

The first two properties follow directly from (6.8); convexity of l is an immediate consequence of equation (6.23) below. In the bivariate case, the above three properties are also sufficient for a function $l : [0, \infty)^2 \to \mathbb{R}$ to be the stable tail dependence function of an exponent measure; in the higher-dimensional case, this is not the case. Note that property (6.9) implies that

$$l(x\boldsymbol{e}_j) = x \text{ for all } 0 \leq x < \infty,\ j = 1, \dots, d, \tag{6.12}$$

where $\boldsymbol{e}_j$ denotes the d-dimensional jth unit vector.

A function that is related to the stable tail dependence function is the function

$$r(\boldsymbol{x}) = \nu([\boldsymbol{x}^{-1}, \boldsymbol{\infty}]) = \nu(\{\boldsymbol{y} \in \mathbb{E} \mid \min(x_1 y_1, \dots, x_d y_d) \geq 1\}), \tag{6.13}$$

for all $\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\infty})$. It is related to the random vector $\boldsymbol{U}$ by the formula

$$\lim_{s \downarrow 0} s^{-1} \mathbb{P}[(U_1 \leq s x_1, \dots, U_d \leq s x_d) = r(\boldsymbol{x}) \tag{6.14}$$

for all $\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\infty})$. Note that $r(\boldsymbol{x})$ is equal to zero as soon as one of the x_j is equal zero. The function r and the stable tail dependence function l are related through the inclusion-exclusion formula. In particular, in two and three dimensions, we have

$$r(x, y) = x + y - l(x, y), \tag{6.15}$$

$$r(x, y, z) = x + y + z - l(x, y, 0) - l(0, y, z) - l(x, 0, z) + l(x, y, z), \tag{6.16}$$

for $(x, y, z) \in [0, \infty)^3$. In general dimension d, using the formula

$$\min(A) = \sum_{B \subset A : |B| \geq 1} (-1)^{|B|-1} \max(B),$$

valid for finite subsets, A, of the real line, we have from equations (6.23) and (6.24) below,

$$r(\boldsymbol{x}) = \sum_{I \subset \{1, \dots, d\} : |I| \geq 1} (-1)^{|I|-1} l(\boldsymbol{x}_I), \tag{6.17}$$

where $\boldsymbol{x}_I$ is the d-dimensional vector such that $(\boldsymbol{x}_I)_i = x_i \mathbf{1}(i \in I)$ for all $i = 1, \dots, d$.

Pickands dependence function

Because of the homogeneity property, the stable tail dependence function l is completely determined by its values on the unit simplex $\{\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\infty}) \mid x_1 + \cdots + x_d = 1\}$. In the bivariate case, the unit simplex can be identified with the unit interval. This leads to the following definition.

Definition 6.2.4. *The* Pickands dependence function *of a bivariate exponent measure ν with stable tail dependence function l is defined as*

$$A(w) = l(1 - w, w), \textit{ for all } 0 \leq w \leq 1.$$

The corresponding stable tail dependence function can be recovered from

$$l(x, y) = (x + y) A\left(\frac{y}{x + y}\right), \text{ for all } (x, y) \in (0, \infty)^2. \tag{6.18}$$

Necessary and sufficient conditions for a function $A : [0, 1] \to \mathbb{R}$ to be a genuine Pickands dependence function are the following:

$$\max(1 - w, w) \leq A(w) \leq 1, \text{ for all } 0 \leq w \leq 1; \tag{6.19}$$

$$A \text{ is convex.} \tag{6.20}$$

This yields a convenient way to generate parametric bivariate extreme value models.

Sometimes, the function $w \mapsto A(1-w)$ is called the Pickands dependence function as well.

In general dimension d, the Pickands dependence function A corresponding to the stable tail dependence function l is defined by

$$A(w_2, \ldots, w_d) = l(w_1, \ldots, w_d), \qquad \text{with } w_1 = 1 - (w_2 + \cdots + w_d)$$

for $(w_2, \ldots, w_d) \in [0, \infty)^{d-1}$ such that $w_2 + \cdots + w_d \leq 1$. Properties of A can be derived from properties of l: in particular, A is convex and

$$\max(w_1, \ldots, w_d) \leq A(w_2, \ldots, w_d) \leq 1.$$

However, except for the case $d = 2$, these conditions are not sufficient to guarantee that A is a proper Pickands dependence function.

Spectral or angular measure

Let $\| \cdot \|$ denote an arbitrary norm on $\mathbb{R}^d$. Denote

$$\aleph = \{\boldsymbol{x} \in \mathbb{E} \mid \|\boldsymbol{x}\| = 1\}, \tag{6.21}$$

the intersection of the $\|\cdot\|$-unit sphere with $\mathbb{E}$. Define the mapping $T : [\mathbf{0}, \boldsymbol{\infty}) \setminus \{\mathbf{0}\} \mapsto (0, \infty) \times \aleph$ by $T(\boldsymbol{x}) = (\|\boldsymbol{x}\|, \boldsymbol{x}/\|\boldsymbol{x}\|)$. Think of T as a transformation to polar coordinates, with $\|\boldsymbol{x}\|$ the radial component and $\boldsymbol{x}/\|\boldsymbol{x}\|$ the angular component of a vector $\boldsymbol{x}$. Note that T is a homeomorphism of the spaces $[\mathbf{0}, \boldsymbol{\infty}) \setminus \{\mathbf{0}\}$ and $(0, \infty) \times \aleph$.

Definition 6.2.5. *The* spectral *or* angular *measure of the exponent measure ν with respect to the norm $\| \cdot \|$ is defined as the measure S on $\aleph$ given by*

$$S(B) = \nu\left(\{\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty}) \mid \|\boldsymbol{x}\| \geq 1, \, \boldsymbol{x}/\|\boldsymbol{x}\| \in B\}\right),$$

for all Borel subsets B of $\aleph$.

Since $\{\boldsymbol{x} \in \mathbb{E} \mid \|\boldsymbol{x}\| \geq 1\}$ is compact, the total mass $S(\aleph)$ is finite. By the homogeneity property (6.4),

$$\nu\left(\{\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty}) \mid \|\boldsymbol{x}\| \geq r, \, \boldsymbol{x}/\|\boldsymbol{x}\| \in B\}\right) = r^{-1} S(B)$$

for every $0 < r < \infty$ and every Borel subset B of $\aleph$. Hence, the measure ν is completely determined by its angular measure S and the homogeneity property (6.4): for every ν-integrable function $f : \mathbb{E} \to \mathbb{R}$,

$$\int_{\mathbb{E}} f(\boldsymbol{x}) \nu(\mathrm{d}\boldsymbol{x}) = \int_{\aleph} \int_0^\infty f(r\boldsymbol{w}) r^{-2} \, \mathrm{d}r \, S(\mathrm{d}\boldsymbol{w}) \tag{6.22}$$

Apply (6.22) to the indicator functions of the sets $[0, \boldsymbol{x}^{-1}]^c$ and $[\boldsymbol{x}^{-1}, \infty]$ for $\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty})$ to get, by (6.2.3) and (6.13),

$$l(\boldsymbol{x}) = \int_{\aleph} \max(w_1 x_1, \ldots, w_d x_d) \, S(\mathrm{d}\boldsymbol{w}), \tag{6.23}$$

$$r(\boldsymbol{x}) = \int_{\aleph} \min(w_1 x_1, \ldots, w_d x_d) \, S(\mathrm{d}\boldsymbol{w}). \tag{6.24}$$

Not every finite Borel measure S on $\aleph$ can arise as the angular measure of a limiting measure ν in Assumption 6.2.2. By equations (6.12) and (6.23),

$$\int_{\aleph} w_j \, S(\mathrm{d}\boldsymbol{w}) = 1, \text{ for all } j = 1, \ldots, d. \tag{6.25}$$

Conversely, every finite Borel measure S on $\aleph$ satisfying (6.25) is the angular measure of some exponent measure ν.

A popular choice for the norm in the definition of the angular measure is the L^1 norm, $\|\boldsymbol{x}\| = |x_1| + \cdots + |x_d|$. In that case, the set $\aleph$ coincides with the unit simplex $\{\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty}) \mid x_1 + \cdots + x_d = 1\}$. In the bivariate case, we can identify $\aleph$ with the unit interval by the correspondance of $(w, 1-w) \in \aleph$ with $w \in [0,1]$. Equations (6.23) and (6.24) then simplify to

$$\begin{aligned} l(x,y) &= \int_{[0,1]} \max\{wx, (1-w)y\}\, S(\mathrm{d}w), \\ r(x,y) &= \int_{[0,1]} \min\{wx, (1-w)y\}\, S(\mathrm{d}w), \end{aligned}$$

for all $(x,y) \in [0,\infty)^2$.

Example 6.2.6. *The random vector* $\boldsymbol{X}$ *is said to be* asymptotically independent *if*

$$\mathbb{P}(F_i(X_i) > 1-p, F_j(X_j) > 1-p) = o(p), \text{ as } p \to 0$$

for every integer pair $1 \le i < j \le d$. *This is equivalent to*

$$\begin{aligned} t\mathbb{P}(t^{-1}\boldsymbol{Y} \in [\mathbf{0}, \boldsymbol{x}]^c) &= t\mathbb{P}(\exists j = 1, \ldots, d : Y_j > tx_j) \\ &\to x_1^{-1} + \cdots + x_d^{-1} = \nu([\mathbf{0}, \boldsymbol{x}]^c), \text{ as } t \to \infty \end{aligned}$$

for every $\boldsymbol{x} > \mathbf{0}$. *The spectral measure* S *of the corresponding exponent measure* ν *with respect to the* L^1*-norm is the measure consisting of point masses of size unity at the* d *vertices of the* d*-dimensional unit simplex* $\aleph = \{\boldsymbol{x} \in [\mathbf{0}, \boldsymbol{\infty}) \mid x_1 + \cdots + x_d = 1\}$. *In particular, the exponent measure* ν *does not put any mass on the interior of the positive orthant, that is,* $\nu((\mathbf{0}, \boldsymbol{\infty}]) = 0$. *In the bivariate case, the corresponding Pickands dependence function is equal to* $A = 1$.

6.2.3 Limits of lower tail dependence copulae

Note that under the assumption that $C(\boldsymbol{u}) > 0$ if all components are strictly positive,

$$\lim_{p \downarrow 0} p^{-1}\mathbb{P}(U_1 \le p, \ldots, U_d \le p) = \nu([\mathbf{1}, \boldsymbol{\infty}]) = r(\mathbf{1}), \tag{6.26}$$

with r as in (6.13). Note that in the bivariate case, the quantity in (6.26) is equal to the coefficient of tail dependence. We now make the following additional assumption.

Assumption 6.2.7. *The quantity in equation* (6.26) *is positive.*

Notice that Assumption 6.2.7 is equivalent to each of the following conditions:

- $r(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty})$;
- $S(\aleph_0) > 0$ where $\aleph_0 = \{\boldsymbol{w} \in \aleph \mid w_j > 0 \text{ for all } j = 1, \ldots, d\}$.

We arrive at the following lemma. For vectors $\boldsymbol{x}$ en $\boldsymbol{y}$, let $\boldsymbol{xy}$ be the vector with components $x_j y_j$.

Lemma 6.2.8. *Under Assumptions 6.2.2 and 6.2.7,*

$$\lim_{s \downarrow 0} \mathbb{P}(\boldsymbol{U} \le s\boldsymbol{xu} \mid \boldsymbol{U} \le s\boldsymbol{x}) = \frac{r(\boldsymbol{ux})}{r(\boldsymbol{x})} = H(\boldsymbol{u}; \boldsymbol{x})$$

for every $\boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty})$ *and every* $\boldsymbol{u} \in [\mathbf{0}, \mathbf{1}]$.

In Lemma 6.2.8, we approach the origin along the direction $\boldsymbol{x}$. We condition on the vector $\boldsymbol{U}$ being contained in the hyperrectangle $[\mathbf{0}, s\boldsymbol{x}]$, rescale the distribution to the d-dimensional hypercube, $[\mathbf{0}, \mathbf{1}]$, and find that as s decreases to zero the resulting distribution function converges to the distribution function $\boldsymbol{u} \mapsto H(\boldsymbol{u}; \boldsymbol{x})$. Notice that

$$H(\,\cdot\,; c\boldsymbol{x}) = H(\,\cdot\,; \boldsymbol{x}), \text{ for all } 0 < c < \infty,\ \boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty}), \tag{6.27}$$

so it is really only the direction of $\boldsymbol{x}$ which determines the limit. The marginal distribution functions of $H(\,\cdot\,; \boldsymbol{x})$ are given by

$$H_j(u; \boldsymbol{x}) = H(\boldsymbol{u}; \boldsymbol{x}) \quad \text{with} \quad u_j = \begin{cases} u & \text{if } i = j, \\ 1 & \text{if } i \in \{1, \dots, d\} \setminus \{j\}, \end{cases}$$

for $u \in [0, 1]$ and $j = 1, \dots, d$; in particular, they are continuous.

Now, let $C(\,\cdot\,; \boldsymbol{x})$ be the copula of the distribution function $H(\,\cdot\,; \boldsymbol{x})$. Recall that $C_{\boldsymbol{u}}$ is the copula of the conditional distribution of $\boldsymbol{U}$ given $\{\boldsymbol{U} \le \boldsymbol{u}\}$ for $\boldsymbol{u} > \mathbf{0}$. Since the margins of $H(\,\cdot\,; \boldsymbol{x})$ are continuous, the following result is a corrollary to Lemma 6.2.8.

Lemma 6.2.9. *Under Assumptions 6.2.2 and 6.2.7,*

$$\lim_{s \downarrow 0} C_{s\boldsymbol{x}}(\boldsymbol{p}) = C(\boldsymbol{p}; \boldsymbol{x})$$

for every $\boldsymbol{x} \in (\mathbf{0}, \boldsymbol{\infty})$ and every $\boldsymbol{p} \in [\mathbf{0}, \mathbf{1}]$.

In words, if we approach the origin along the direction determined by $\boldsymbol{x}$, then the corresponding family of lower tail dependence copulae $C_{s\boldsymbol{x}}$ converges to the copula $C(\,\cdot\,; \boldsymbol{x})$ of the distribution function $H(\,\cdot\,; \boldsymbol{x})$ in Lemma 6.2.8.

Invariance

Definition 6.2.10. *A copula C is invariant if for every $\boldsymbol{u} > 0$, the lower tail dependence copula $C_{\boldsymbol{u}}$ is the same. Hence $C_{\boldsymbol{u}} = C$ for every $\boldsymbol{u} > 0$.*

Example 6.2.11. *Let $C(u_1, ..., u_d, \theta)$ denote Clayton copula,*

$$C(u_1, ..., u_d, \theta) = \left[u_1^{-\theta} + ... + u_d^{-\theta} - (d-1)\right]^{-1/\theta}, \ \theta \ge 0, \tag{6.28}$$

If $\boldsymbol{U}$ has distribution function C, then

$$\begin{aligned}
\mathbb{P}[\boldsymbol{U} \le s\boldsymbol{x}\boldsymbol{u} | \boldsymbol{U} \le s\boldsymbol{x}] &= \frac{\mathbb{P}[\boldsymbol{U} \le s\boldsymbol{x}\boldsymbol{u}]}{\mathbb{P}[\boldsymbol{U} \le s\boldsymbol{x}]} \\
&= \frac{\left[(sx_1u_1)^{-\theta} + ... + (sx_du_d)^{-\theta} - (d-1)\right]^{-1/\theta}}{\left[(sx_1)^{-\theta} + ... + (sx_d)^{-\theta} - (d-1)\right]^{-1/\theta}} \\
&= \frac{\left[(x_1u_1)^{-\theta} + ... + (x_du_d)^{-\theta} - (d-1)\, s^{\theta}\right]^{-1/\theta}}{\left[(x_1)^{-\theta} + ... + (x_d)^{-\theta} - (d-1)\, s^{\theta}\right]^{-1/\theta}} \\
&\to \frac{\left[(x_1u_1)^{-\theta} + ... + (x_du_d)^{-\theta}\right]^{-1/\theta}}{\left[(x_1)^{-\theta} + ... + (x_d)^{-\theta}\right]^{-1/\theta}} \quad \textit{as } s \to 0, \\
&= H(\boldsymbol{u}; \boldsymbol{x}).
\end{aligned}$$

Marginal distributions are then

$$H_i\left(u_i;\boldsymbol{x}\right)=\frac{\left[(x_1)^{-\theta}+...+(x_iu_i)^{-\theta}+...+(x_d)^{-\theta}\right]^{-1/\theta}}{\left[(x_1)^{-\theta}+...+(x_d)^{-\theta}+...+(x_d)^{-\theta}\right]^{-1/\theta}}$$

which can be inverted in

$$H_i^{-1}\left(p_i;\boldsymbol{x}\right)=\left(\frac{\left[(x_1)^{-\theta}+...+(x_i)^{-\theta}+...+(x_d)^{-\theta}\right]p_i^{-\theta}-\left((x_1)^{-\theta}+...+(x_d)^{-\theta}\right)}{x_{d=i}}\right)^{-1/\theta}$$

The associated copula denoted $C\left(\boldsymbol{u};\boldsymbol{x}\right)$ *is then*

$$\begin{aligned}C\left(\boldsymbol{u};\boldsymbol{x}\right) &= \frac{\left[\left[(x_1)^{-\theta}+...+(x_d)^{-\theta}\right]\left(u_1^{-\theta}+...+u_d^{-\theta}-(d-1)\right)\right]^{-1/\theta}}{\left[(x_1)^{-\theta}+...+(x_d)^{-\theta}\right]^{-1/\theta}}\\ &= C\left(u_1,...,u_d,\theta\right)=C\left(\boldsymbol{u}\right).\end{aligned}$$

Some more examples

Example 6.2.12. *If the random vector* $\boldsymbol{X}$ *is comonotone, then by* (6.14)

$$r(\boldsymbol{x})=\min(x_1,\ldots,x_d),\ \textit{for all } \boldsymbol{x}\in[\boldsymbol{0},\boldsymbol{\infty}),$$

whence

$$H(\boldsymbol{u};\boldsymbol{x})=\frac{\min(u_1x_1,\ldots,u_dx_d)}{\min(x_1,\ldots,x_c)},\ \textit{for all } \boldsymbol{u}\in[\boldsymbol{0},\boldsymbol{1}],\ \boldsymbol{x}\in(\boldsymbol{0},\boldsymbol{\infty}).$$

The copula of this distribution function is easily seen to be the comonotone copula

$$C(\boldsymbol{p})=\min(p_1,\ldots,p_d),\ \textit{for all } \boldsymbol{p}\in[\boldsymbol{0},\boldsymbol{1}].$$

The Clayton copula is known to arise naturally in the theory of lower tail dependence copulae. According to Example 6.2.13, it corresponds to the negative logistic model.

Example 6.2.13. *The* negative logistic *dependence structure (Coles and Tawn (1991), or Joe (1990)) has stable tail dependence function*

$$l(\boldsymbol{x})=\sum_{i=1}^{d}x_i-\sum_{I\subset\{1,\ldots,d\}:|I|\geq 2}(-1)^{|I|}\left(\sum_{i\in I}(\psi_{I,i}x_i)^{-\theta_I}\right)^{-1/\theta_I} \tag{6.29}$$

for $\boldsymbol{x}\in[\boldsymbol{0},\boldsymbol{\infty})$, *with the convention that* $0^{-1}=\infty$ *and* $\infty^{-1}=0$. *The parameter vector* $(\theta_I,\psi_{I,i}|I\subset\{1,\ldots,d\},i\in I)$ *should be such that*

(i) $\theta_I>0$ *for all* $I\subset\{1,\ldots,d\}$,

(ii) $\psi_{I,i}\geq 0$ *for all* $I\subset\{1,\ldots,d\}$ *and* $i\in I$,

(iii) $\sum_I(-1)^{|I|}\psi_{I,i}\leq 1$ *for all* $i=1,\ldots,d$, *the sum ranging over all* $I\subset\{1,\ldots,d\}$ *such that* $i\in I$. *Let us prove that*

$$r(\boldsymbol{x})=\left(\sum_{i=1}^{d}\left(\psi_ix_i\right)^{-\theta}\right)^{-1/\theta}, \tag{6.30}$$

where we used the abbreviations $\psi_i = \psi_{\{1,\dots,d\},i}$ *and* $\theta = \theta_{\{1,\dots,d\}}$.

Let $\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\infty})$. *For* $J \subset \{1,\dots,d\}$, *let* $\boldsymbol{x}_J$ *be the* d*-dimensional vector such that* $(\boldsymbol{x}_J)_j = x_j$ *if* $j \in J$ *and* $(\boldsymbol{x}_J)_j = 0$ *otherwise. Then for* $J \subset \{1,\dots,d\}$,

$$l(\boldsymbol{x}_J) = \sum_{j\in J} x_j - \sum_{I\subset J:|I|\geq 2} (-1)^{|I|} \left(\sum_{i\in I} (\psi_{I,i}x_i)^{-\theta_I} \right)^{-1/\theta_I}.$$

By (6.17) *and* (6.29),

$$\begin{aligned} r(\boldsymbol{x}) &= \sum_{J\subset\{1,\dots,d\}:|J|\geq 1} (-1)^{|J|-1} l(\boldsymbol{x}_J) \\ &= \sum_{j=1}^{d} \left(\sum_{J\subset\{1,\dots,d\}:j\in J} (-1)^{|J|-1} \right) x_j \\ &\quad + \sum_{I\subset\{1,\dots,d\}:|I|\geq 2} \left(\sum_{J\subset\{1,\dots,d\}:I\subset J} (-1)^{|J|-|I|} \right) \left(\sum_{i\in I} (\psi_{I,i}x_i)^{-\theta_I} \right)^{-1/\theta_I}. \end{aligned}$$

Now it is not hard to see that

$$\sum_{J\subset\{1,\dots,d\}:j\in J} (-1)^{|J|-1} = 0, \text{ for all } j = 1,\dots,d,$$

as well as

$$\sum_{J\subset\{1,\dots,d\}:I\subset J} (-1)^{|J|-|I|} = \begin{cases} 0 & \text{if } I \neq \{1,\dots,d\}, \\ 1 & \text{if } I = \{1,\dots,d\}. \end{cases}$$

Combine the three previous displays to arrive at equation (6.30).

If ψ_i *is positive for all* $i = 1,\dots,d$, *then Assumption 6.2.7 is satisfied, and*

$$H(\boldsymbol{u};\boldsymbol{x}) = \left(\frac{\sum_{i=1}^d (\psi_i u_i x_i)^{-\theta}}{\sum_{i=1}^d (\psi_i x_i)^{-\theta}} \right)^{-1/\theta}$$

for all $\boldsymbol{u} \in [0,1]^d$ *and* $\boldsymbol{x} \in (\boldsymbol{0}, \boldsymbol{\infty})$. *The copula of this distribution is easily seen to be the Clayton copula* (6.28).

In the bivariate case, we can write $H(\,\cdot,\cdot\,;w) = H(\,\cdot,\cdot\,;1-w,w)$ for $0 < w < 1$ and write everything in terms of the Pickands dependence function A in Definition 6.2.4, yielding

$$H(u,v;w) = \frac{\{u(1-w)+vw\}\left\{1 - A\left(\frac{vw}{u(1-w)+vw}\right)\right\}}{1-A(w)}. \tag{6.31}$$

for $0 < w < 1$ and $(u,v) \in [0,1]^2$. Note that Assumption 6.2.7 is equivalent to $A(w) < 1$ for $0 < w < 1$. As every convex function $A : [0,1] \to \mathbb{R}$ such that $\max(w, 1-w) \leq A(w) \leq 1$ for all $w \in [0,1]$ is a Pickands dependence function, equation (6.31) constitutes a convenient mechanism to generate valid models for limits of lower tail dependence copulae.

Example 6.2.14. *For* $0 < \psi \leq 1/2$, *the function*

$$A(w) = 1 - \psi w + \psi w^3, \text{ for all } 0 \leq w \leq 1,$$

a special case of the so-called asymmetric mixed model, *is the Pickands dependence function corresponding to a bivariate exponent measure satisfying Assumption 6.2.7. By* (6.15) *and* (6.18),

$$r(x,y) = (x+y)\left\{1 - A\left(\frac{y}{x+y}\right)\right\} = \psi\frac{2x^{-1}+y^{-1}}{(x^{-1}+y^{-1})^2}$$

for all $(x,y) \in (0,\infty)^2$. *The corresponding asymptotic distribution function appearing in Lemma 6.2.8 is given by*

$$H(u,v;x,y) = \frac{(x^{-1}+y^{-1})^2}{2x^{-1}+y^{-1}} \cdot \frac{2(ux)^{-1}+(vy)^{-1}}{\{(ux)^{-1}+(vy)^{-1}\}^2}$$

for $(u,v) \in (0,1]^2$ *and* $(x,y) \in (0,\infty)^2$. *In terms of* $w = y/(x+y) \in (0,1)$, *we can simplify the expression in the previous display to*

$$H(u,v;w) = H(u,v;1-w,w) = \frac{1}{1+w} \cdot \frac{2u^{-1}w + v^{-1}(1-w)}{\{u^{-1}w+v^{-1}(1-w)\}^2}$$

for $(u,v) \in (0,1]^2$ *and* $0 < w < 1$. *For* $0 < p < 1$ *and* $0 < q < 1$, *the solutions* $u = u(p;w)$ *and* $v = v(q;w)$ *in the interval* $(0,1)$ *to the equations* $H(u,1;w) = p$ *and* $H(1,v;w) = q$ *are given by*

$$\begin{aligned} u(p;w) &= \frac{w}{1-w}[\{1-(1-w^2)p\}^{-1/2} - 1], \\ v(q;w) &= \frac{2(1-w^2)q}{\{4w(1+w)q+1\}^{1/2} + 2w(1+w)q - 1} \end{aligned}$$

Hence, for $0 < w < 1$, *the copula of the distribution function* $H(\,\cdot,\cdot\,;w)$ *is given by*

$$C(p,q;w) = H(u(p;w), v(q;w); w), \text{ for all } 0<p<1,\ 0<q<1.$$

Note that the copula is different for different directions w.

Example 6.2.15. *The bivariate Gumbel or logistic dependence structure has stable tail dependence function*

$$l(x_1,x_2) = (x_1^\theta + x_2^\theta)^{1/\theta}, \text{ for all } (x_1,x_2) \in [0,\infty)^2.$$

The parameter θ *can range in the interval* $[1,\infty]$, *with* $\theta = 1$ *corresponding to independence and* $\theta = \infty$ *corresponding to comonotonicity. In case* $\theta > 1$, *Assumption 6.2.7 is satisfied with*

$$\begin{aligned} r(x_1,x_2) &= x_1 + x_2 - (x_1^\theta + x_2^\theta)^{1/\theta}, \\ H(u_1,u_2;x_1,x_2) &= \frac{u_1x_1 + u_2x_2 - \{(u_1x_1)^\theta + (u_2x_2)^\theta\}^{1/\theta}}{x_1+x_2-(x_1^\theta+x_2^\theta)^{1/\theta}}. \end{aligned}$$

The inverses of the marginal distribution functions of $H(\,\cdot\,;\boldsymbol{x})$ *do not admit explicit expressions, so neither does its copula.*

Example 6.2.16. *The* asymmetric logistic *dependence structure in dimension d has stable tail dependence function (Tawn (1990))*

$$l(x_1,...,x_d) = \sum_{c\in\mathcal{C}}\left(\sum_{i\in c}(\psi_{c,i}x_i)^{r_c}\right)^{1/r_c}$$

for $(x_1,...,x_d) \in (0,\infty)^d$, *where* $\mathcal{C}$ *is the ensemble of nonempty subsets of* $\{1,...,d\}$, *with parameters* $r_c \geq 1$ *for all* $c \in \mathcal{C}$. *For* $i = 1,...,d$, $\psi_{c,i} = 0$ *if* $i \notin c$, $\psi_{c,i} \geq 0$ *if* $i \in c$, *and* $\sum_{c\in\mathcal{C}}\psi_{c,i} = 1$. ...

and $0 < \psi_j \leq 1$ *for* $j = 1,...,d$.

6.2.4 Asymptotic independence

Example 6.2.17. *Consider Cuadras-Augé copula (as called in Nelsen (1999), or Marshall & Olkin copula with identical parameter),*

$$C(u_1, ..., u_d, \theta) = \min\left\{u_1^{-\theta}, ..., u_d^{-\theta}\right\} \cdot u_1 ... u_d, \ \theta \in [0,1],$$

If $\boldsymbol{U}$ has distribution function C, then, if $\boldsymbol{u} = u\mathbf{1}$,

$$\begin{aligned} \mathbb{P}[\boldsymbol{U} \le s\boldsymbol{x}u | \boldsymbol{U} \le s\boldsymbol{x}] &= \frac{\mathbb{P}[\boldsymbol{U} \le s\boldsymbol{x}u]}{\mathbb{P}[\boldsymbol{U} \le s\boldsymbol{x}]} \\ &= \frac{\min\left\{(sx_1u_1)^{-\theta}, ..., (sx_du_d)^{-\theta}\right\} \cdot sx_1u ... sx_du}{\min\left\{(sx_1)^{-\theta}, ..., (sx_d)^{-\theta}\right\} \cdot sx_1 ... sx_d} \\ &= \min\left\{u^{-\theta}, ..., u^{-\theta}\right\} u ... u = H(u\mathbf{1}; \boldsymbol{x}). \end{aligned}$$

Note in that case, it is already a copula, hence $H(u\mathbf{1}; \boldsymbol{x}) = C(u\mathbf{1}; \boldsymbol{x}) = C(\boldsymbol{u})$. Hence Cuadras-Augé copula is invariant if $\boldsymbol{u} = u\mathbf{1}$: this was called "invariance on the diagonal" in Charpentier (2004).

6.3 Pickands-Balkema-de Haan in dimension 2

In this following section, we will extend Pickands-Balkema-de Hann Theorem, which gives the expression of the exceeding distributions. Section 6.3.2 will focus on quantile based thresholds, i.e. we will be interested in the limiting distribution of

$$(X, Y) \text{ given } \{X > F_X^{\leftarrow}(p) \text{ and } Y > F_X^{\leftarrow}(p)\} \text{ as } p \to 1,$$

while Section 6.3.3 will focus on level based thresholds, i.e. we will be interested in the limiting distribution of

$$(X, Y) \text{ given } \{X > z \text{ and } Y > z\} \text{ as } z \to \infty.$$

6.3.1 A short word on regular variation for $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ functions

The regular variation notion considered so far in this chapter is *orthant based.* More specifically, given $\boldsymbol{x} \in \mathbb{E} = [0, \infty)^2 \backslash \{\mathbf{0}\}$, the limit is obtained under the ray $(\mathbf{0}, \boldsymbol{x})$ in $\mathbb{R}^2$, considering the evolution of $F(t \cdot \boldsymbol{x})$ when $t \to \infty$.

Meerschaert (1998) and Meerschaert and Scheffler (2001), following Balkema (1973), defined a general concept of regular variation on Lie groups, given general results for a general theory of regular variation in $\mathbb{R}^d$. As we will see, there are several ways to define regularly varying functions $\mathbb{R}^2 \to \mathbb{R}$: f is said to be regularly varying at infinity is there are $h : \mathbb{R} \to \mathbb{R}$ and $\lambda : \mathbb{R}^2 \to \mathbb{R}$ such that one of the following relationship holds,

1. *ray convergence*
$$\lim_{t\to\infty} f(tx, ty) \cdot h(t)^{-1} = \lambda(x, y),$$
2. *directional convergence*, given $r, s : \mathbb{R} \to \mathbb{R}$ such that $r(t), s(t) \to \infty$ as $t \to \infty$, both regularly varying at infinity (with index α and β respectively), then
$$\lim_{t\to\infty} f(r(t)x, s(t)y) \cdot h(t)^{-1} = \lambda(x, y),$$

3. there is a sequence (A_t) of $GL(\mathbb{R}^2)$ operators, regularly varying with index E (see Definition 6.3.1 below) such that

$$\lim_{t\to\infty} f\left(A_t^{\leftarrow}\left(\begin{array}{c} x \\ y \end{array}\right)\right)\cdot h(t)^{-1} = \lambda(x,y).$$

For those three cases, different limiting behavior can be observed,

1. there is $\theta \in \mathbb{R}$ such that
$$\lambda(tx,ty) = t^{\theta}\lambda(x,y),$$
for all $x,y,t>0$, i.e. λ is an *homogeneous function*,

2. there is $\theta \in \mathbb{R}$ such that
$$\lambda(t^{\alpha}x,t^{\beta}y) = t^{\theta}\lambda(x,y),$$
for all $x,y,t>0$, i.e. λ is a *generalized homogeneous function*,

3. there is $\theta \in \mathbb{R}$ such that
$$\lambda(f\left(t^{-E}\left(\begin{array}{c} x \\ y \end{array}\right)\right)) = t^{\theta}\lambda(x,y),$$
for all $x,y,t>0$, i.e. λ is a *generalized homogeneous function.*

Consider a linear operator $A \in L\left(\mathbb{R}^2\right)$, the associated exponential operator and the family of power operators as

$$\exp\left(A\right) = \sum_{k=0}^{\infty}\frac{A^k}{k!} \text{ and } t^A = \exp\left(A\log t\right),$$

for all $t>0$. Suppose $f:\mathbb{R}^+\to GL\left(\mathbb{R}^2\right)$ is Borel measurable, where $GL\left(\mathbb{R}^2\right)$ denotes the group of invertible linear operators on $\mathbb{R}^2$. Following Balkema (1973), define regular variation as follows,

Definition 6.3.1. *$f:\mathbb{R}^+\to GL\left(\mathbb{R}^2\right)$ is said to be regularly varying at infinity with index E if*

$$\lim_{t\to\infty} f\left(\lambda t\right) f\left(t\right)^{\leftarrow} = \lambda^{E}, \tag{6.32}$$

for all $\lambda>0$. If $E=0$, f is said to be slowly varying.

Suppose $f:\mathbb{R}^+\to GL\left(\mathbb{R}^2\right)$ is Borel measurable and

$$\lim_{t\to\infty} f\left(\lambda t\right) f\left(t\right)^{\leftarrow} = \phi\left(\lambda\right)\in GL\left(\mathbb{R}^d\right),$$

for all $\lambda>0$. Then, there are some linear operator E such that $\phi\left(\lambda\right)=\lambda^{E}$ for all $\lambda>0$.

Based on the previous notions, it is possible to introduce a general definition for multivariate regular variation.

Definition 6.3.2. *A Borel measurable function $g:\Gamma\subset\mathbb{R}^2\to\mathbb{R}^+$ is said to be regularly varying if there exists $f:\mathbb{R}^+\to GL\left(\mathbb{R}^2\right)$ regularly varying with index $-E$ and $h:\mathbb{R}^+\to\mathbb{R}^+$ regularly varying with index $\beta\neq 0$ such that*

$$\lim_{t\to\infty}\frac{g\left(f\left(t\right)^{\leftarrow}x_t\right)}{h\left(t\right)} = \phi\left(x\right)>0 \tag{6.33}$$

where $x_t\to x$ in Γ. If all eigenvalues of E have positive real part, g is said to vary regularly at infinity.

As shown in Meerschaert and Scheffler (2001), limiting functions (again) satisfy some functional equation: if $g:\Gamma\subset\mathbb{R}^2\to\mathbb{R}^+$ is regularly varying and if Equation (6.33) holds for some $f:\mathbb{R}^+\to GL\left(\mathbb{R}^2\right)$ regularly varying with index $-E$ and $h:\mathbb{R}^+\to\mathbb{R}^+$ regularly varying with index $\beta\neq 0$ then

$$\lambda^{\beta}\cdot\phi\left(x\right)=\phi\left(\lambda^{E}x\right),$$

for all $\lambda>0$ and all $x\in\Gamma$. The idea is to observe that

$$\frac{g\left(f\left(\lambda t\right)^{\leftarrow}x\right)}{h\left(\lambda t\right)}=\frac{f\left(\lambda\right)^{\leftarrow}x_t}{h(t)}\cdot\frac{h(t)}{h\left(\lambda t\right)},$$

where $x_t=f\left(\lambda t\right)\cdot f\left(\lambda\right)^{\leftarrow}\cdot x$. Since f is regularly varying, with $f()\cdot f()^{\leftarrow}\to\lambda^E$, and therefore, $x_t\to\lambda^E\cdot x$. Taking the limit of both sides of, as $t\to\infty$, we get $\phi\left(x\right)=\phi\left(\lambda^E x\right)\cdot\lambda^{-\beta}$.

The approach considered in de Haan, Omey and Resnick (1984) is a particular case of the previous one (the *directional convergence*): given $r,s:\mathbb{R}\to\mathbb{R}$ such that $r(t),s(t)\to\infty$ as $t\to\infty$, both regularly varying at infinity (with index α and β respectively), then

$$\lim_{t\to\infty}f(r(t)x,s(t)y)\cdot h(t)^{\leftarrow}=\lambda(x,y).$$

The particularity is that de Haan, Omey and Resnick (1984) defined a kind of regular variation optimized for nonnegative joint distributions. For signed joint distributions, the approach of the previous section is more general. Actually, for nonnegative joint distributions, the general definition does not add very much, since one would presumably assume that f fixes the set $\{x>0,y>0\}$, or at least maps this set into itself. Note that in that case, E is diagonal with respect to the standard basis vector.

6.3.2 A first bivariate extension of Pickands-Balkema-de Haan Theorem: thresholds as quantiles

Assume that X has a distribution F_X in the max-domain of attraction of the Fréchet distribution, with parameter $\alpha>0$. From Pickands-Balkema-de Haan Theorem, there is a function $a(\cdot)$ such that, for $1+\alpha x>0$,

$$\lim_{u\to\infty}1-\frac{1-F_X\left(u+xa\left(u\right)\right)}{1-F_X\left(u\right)}=\lim_{u\to\infty}\mathbb{P}\left(X\leq u+a\left(u\right)|X>u\right)=G_{\alpha}\left(x\right),$$

where $G_\alpha(\cdot)$ denotes the Generalized Pareto distribution with parameter α, i.e. $G_\alpha\left(x\right)=1-\left(1+\alpha x\right)^{-1/\alpha}$ (see Equation (6.2)). And analogously, assume that Y has tail index β, such that

$$\lim_{v\to\infty}1-\frac{1-F_Y\left(v+yb\left(v\right)\right)}{1-F_Y\left(v\right)}=\lim_{v\to\infty}\mathbb{P}\left(Y\leq v+b\left(v\right)|Y>v\right)=G_{\beta}\left(y\right),$$

for some function $b(\cdot)$.

Hence, from those relationships, some equivalent for marginal tail distributions can be obtained, e.g.

$$\begin{cases}F_X\left(u+xa\left(u\right)\right)\sim 1-\left[1-F_X\left(u\right)\right]\cdot\left[1-G_\alpha\left(x\right)\right]\\ F_Y\left(v+yb\left(v\right)\right)\sim 1-\left[1-F_Y\left(v\right)\right]\cdot\left[1-G_\beta\left(y\right)\right]\end{cases}$$

Using the expression of joint survival probabilities, expressed through the survival copula and marginal survival probabilities, $\mathbb{P}(X > x, Y > y) = C^* (1 - F_X(x), 1 - F_Y(y))$, we get

$$\begin{aligned}
&\lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \,\middle|\, X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p) \right) \\
= &\lim_{p\to 0} \frac{C^* \left(1 - F_X\left(F_X^{\leftarrow}(1-p) + xa\left(F_X^{\leftarrow}(1-p)\right)\right), 1 - F_Y\left(F_Y^{\leftarrow}(1-p) + yb\left(F_Y^{\leftarrow}(1-p)\right)\right)\right)}{C^*(p,p)} \\
= &\lim_{p\to 0} \frac{C^* \left(\left[1 - F_X\left(F_X^{\leftarrow}(1-p)\right)\right] \cdot \left[1 - G_\alpha(x)\right], \left[1 - F_Y\left(F_Y^{\leftarrow}(1-p)\right)\right] \cdot \left[1 - G_\beta(y)\right]\right)}{C^*(p,p)} \\
= &\lim_{p\to 0} \frac{C^* \left(p \cdot [1 - G_\alpha(x)], p \cdot [1 - G_\beta(y)]\right)}{C^*(p,p)}.
\end{aligned}$$

In that case, note that results on extended regular variation (de Haan, Omey and Resnick (1984)) are not necessary, since we are focusing here on the asymptotic behavior of $C^*(p \cdot x^*, p \cdot y^*)/C^*(p,p)$ as $p \to 0$, where $x^* = 1 - G_\alpha(x)$ and $y^* = 1 - G_\beta(y)$.

In the case where C^* is assumed to be symmetric, the approach of Juri and Wüthrich (2004) and Wüthrich (2004) can be extended easily, without the assumption that X and Y should have identical distributions.

Proposition 6.3.3. *Let C^* denote a symmetric copula, such that there exists a continuous function $h : \mathbb{R}^+ \to \mathbb{R}^+$ with $h(x) > 0$ for $x > 0$, and such that*

$$\lim_{u\to 0} \frac{C^*(xu, u)}{C^*(u,u)} = h(x) \text{ for all } x \geq 0. \tag{6.34}$$

Then $h(0) = 0$, $h(1) = 1$, and there exists $\theta \in \mathbb{R}$ such that

$$h(x) = x^\theta h\left(\frac{1}{x}\right) \text{ for all } x > 0,$$

Further, if $H(x,y) = y^\theta h(x/y)$ for all $x, y > 0$, with $H(x,y) = 0$ if either $x = 0$ or $y = 0$, then

$$\lim_{p\to 0} \left\| C^*_{(p,p)} - C_0 \right\|_\infty = 0, \text{ where } C_0(x,y) = H(h^{\leftarrow}(x), h^{\leftarrow}(y)).$$

Proof. The convergence for the infinite norm is a particular case of Charpentier and Juri (2004), but the limiting behavior was initially obtained in Juri and Wüthrich (2004). □

From this relationship, a first bivariate extension of Pickands-Balkema-de Haan Theorem can be obtained, under the assumption that C^* is a symmetric copula.

Theorem 6.3.4. *Let X and Y be two random variables in the Fréchet domain of attraction, with tail indices $\alpha > 0$ and $\beta > 0$ respectively. From Pickands-Balkema-de Haan Theorem, there are functions $a(\cdot)$ and $b(\cdot)$ such that*

$$\lim_{u\to\infty} 1 - \frac{1 - F_X(u + xa(u))}{1 - F_X(u)} = \lim_{u\to\infty} \mathbb{P}(X \leq u + a(u) \,|X > u) = G_\alpha(x),$$

$$\lim_{v\to\infty} 1 - \frac{1 - F_Y(v + yb(v))}{1 - F_Y(v)} = \lim_{v\to\infty} \mathbb{P}(Y \leq v + b(v) \,|Y > v) = G_\beta(y),$$

where $G_\alpha(\cdot)$ denotes the Generalized Pareto distribution with parameter α. If the survival copula of $(X,Y)^t$ is a symmetric copula, satisfying assumptions of Proposition 6.3.3 (i.e. Equation

(6.34)) for some function h, and some parameter θ. Set $\gamma = \theta/\beta$, then

$$\begin{aligned}
&\lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \middle| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p) \right) \\
= \;& (1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\beta}} \right).
\end{aligned}$$

Proof. From calculations of the previous paragraph, note that

$$\begin{aligned}
\kappa(x,y) &= \lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \middle| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p) \right) \\
&= \lim_{p\to 0} \frac{C^*\left(p\cdot[1-G_\alpha(x)], p\cdot[1-G_\beta(y)]\right)}{C^*(p,p)}.
\end{aligned}$$

Since C^* satisfies assumptions of Proposition 6.3.3 (i.e. Equation (6.34)), for all $x, y \in [0,1]$,

$$\lim_{p\to 0} C^* - (p,p)(x,y) = H(h^{\leftarrow}(x), h^{\leftarrow}(y)) = C_h(x,y),$$

were $C_h(x,y) = H(h^{\leftarrow}(x), h^{\leftarrow}(y))$ where $H(x,y) = y^\theta h(x/y)$. Further,

$$\begin{aligned}
\kappa(x,y) &= H([1-G_\alpha(x)],[1-G_\beta(y)]) = [1-G_\beta(y)]^\theta h\left(\frac{[1-G_\alpha(x)]}{[1-G_\beta(y)]}\right) \\
&= (1+y)^{-\theta/\beta} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\beta}} \right).
\end{aligned}$$

This finishes the proof of Theorem 6.3.4. □

Corollary 6.3.5. *Under the hypothesis of Theorem 6.3.4, the limiting copula is C_h^* where $C_h(x,y) = H(h^{\leftarrow}(x), h^{\leftarrow}(y))$ where $H(x,y) = y^\theta h(x/y)$.*

Note further that the limiting behavior can be written in terms of uniform convergence, with notions of Equation (6.2), i.e.

$$G_{\alpha,a(\cdot)}(x) = 1 - (1 + \alpha x/a(\cdot))^{-1/\alpha}.$$

Proposition 6.3.6. *Under the assumptions of Theorem 6.3.4, the convergence is uniform, i.e.,*

$$\begin{aligned}
&\lim_{p\to 0} \sup_{x,y} \left| \mathbb{P}\left(X - F_X^{\leftarrow}(1-p) > x, Y - F_Y^{\leftarrow}(1-p) > y \middle| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p) \right) \right. \\
&\quad \left. -(1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\beta}} \right) \right| = 0.
\end{aligned}$$

Proof. Since the limit is a continuous function, the convergence is uniform (see e.g. Embrechts, Klüppelberg and Mikosh (1997)). □

Example 6.3.7. *In the case where the copula of $(X,Y)^t$ is the survival Clayton copula, with parameter $\theta > 0$,*

$$C^*(x,y) = \left(x^{-\theta} + y^{-\theta} - 1\right)^{-1/\theta},$$

In that case, since this copula is invariant by truncature, and symmetric, note that $C_h = C^*$, with $g(x) = ([x^{-\theta}+1]/2)^{-1/\theta}$, and therefore

$$\begin{aligned}
\kappa(x,y) &= \lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \middle| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p) \right) \\
&= \left(\left(\frac{(1+x)^{\theta/\alpha}+1}{2} \right) + \left(\frac{(1+y)^{\theta/\beta}+1}{2} \right) - 1 \right)^{-1/\theta} \\
&= \left((1+x)^{\theta/\alpha} + (1+y)^{\theta/\beta} \right)^{-1/\theta}.
\end{aligned}$$

Extension when margins are in the Gumbel domain of attraction

Recall that in the case where X is in the Gumbel domain of attraction, there exists function $a(\cdot)$ such that

$$\lim_{u\to\infty} \frac{1 - F_X(u + xa(u))}{1 - F(u)} = \exp(x), \text{ for all } x \in \mathbb{R}.$$

In that case,

$$F_X(u + xa(u)) \sim 1 - [1 - F_X(u)] \cdot \exp(x),$$

and substituting in relationships obtained above, one gets, similarly that

$$\begin{aligned}
&\lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \middle| X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p) \right) \\
= &\lim_{p\to 0} \frac{C^*\left(p \cdot [\exp(x)], p \cdot [1 - G_\beta(y)]\right)}{C^*(p,p)}.
\end{aligned}$$

From the regular variation property, one gets that this probability (denoted $\kappa(x,y)$) can be written

$$\kappa(x,y) = H(\exp(x), [1 - G_\beta(y)]) = (1+y)^{-\theta/\beta} h\left(\frac{\exp(x)}{(1+y)^{-1/\beta}} \right).$$

In the case where both have light tails, this probability becomes

$$\kappa(x,y) = \exp(\theta y) h\left(\exp(x-y)\right).$$

Hence, there is no need to assume that X and Y have identical distributions to obtain limiting results (as in Juri and Wüthrich (2004)).

6.3.3 A second bivariate extension of Pickands-Balkema-de Haan Theorem: thresholds as level

As pointed out in Section 6.3.1, the notion of regular variation considered here can be extended in a more general context. Consider here the case where threshold are levels. Assume again that X and Y have tail indices $\alpha, \beta > 0$, so that

$$\begin{cases} F_X(z + xa(z)) \sim 1 - [1 - F_X(u)] \cdot [1 - G_\alpha(z)] \\ F_Y(z + yb(z)) \sim 1 - [1 - F_Y(z)] \cdot [1 - G_\beta(y)] \end{cases}$$

Using the expression of joint survival probabilities, expressed through the survival copula and marginal survival probabilities, $\mathbb{P}(X > x, Y > y) = C^*(1 - F_X(x), 1 - F_Y(y))$, we get

$$\begin{aligned}
&\lim_{z\to\infty} \mathbb{P}\left(\frac{X-z}{a(z)} > x, \frac{Y-z}{b(z)} > y \,\middle|\, X > z, Y > z\right) \\
= &\lim_{z\to\infty} \frac{C^*(1 - F_X(z + xa(z)), 1 - F_Y(z + yb(z)))}{C^*(1 - F_X(z), 1 - F_Y(z))} \\
= &\lim_{z\to\infty} \frac{C^*([1 - F_X(z)] \cdot [1 - H_\alpha(x)], [1 - F_Y(z)] \cdot [1 - H_\beta(y)])}{C^*(1 - F_X(z), 1 - F_Y(z))} \\
= &\lim_{z\to\infty} \frac{C^*([1 - F_X(z)] \cdot [1 - G_\alpha(x)], [1 - F_Y(z)] \cdot [1 - G_\beta(y)])}{C^*(1 - F_X(z), 1 - F_Y(z))}.
\end{aligned}$$

In that case, setting $x^* = 1 - G_\alpha(x)$, $y^* = 1 - G_\beta(y)$, $r(z) = 1 - F_X(z)$ and $s(z) = 1 - F_Y(z)$, one is interested in the limit

$$\lim_{z\to\infty} \frac{C^*(r(z) \cdot x^*, s(z) \cdot y^*)}{C^*(r(z), s(z))},$$

where both r and s are regularly varying at infinity (since X and Y have tail indices $\alpha, \beta > 0$). In that case, assuming some regular variation properties for C^*, some limiting properties can be obtained.

Theorem 6.3.8. *Let X and Y be two random variables in the Fréchet domain of attraction, with tail indices $\alpha > 0$ and $\beta > 0$ respectively. From Pickands-Balkema-de Haan Theorem, there are functions $a(\cdot)$ and $b(\cdot)$ such that*

$$\lim_{u\to\infty} 1 - \frac{1 - F_X(u + xa(u))}{1 - F_X(u)} = \lim_{u\to\infty} \mathbb{P}(X \le u + a(u) \,|X > u) = G_\alpha(x),$$

$$\lim_{v\to\infty} 1 - \frac{1 - F_Y(v + yb(v))}{1 - F_Y(v)} = \lim_{v\to\infty} \mathbb{P}(Y \le v + b(v) \,|Y > v) = G_\beta(y),$$

where $G_\alpha(\cdot)$ denotes the Generalized Pareto distribution with parameter α. If the survival copula of $(X,Y)^t$, c^* is regularly varying under direction $(1 - F_X, 1 - F_Y)$, i.e. there exists λ such that

$$\lim_{z\to\infty} \frac{C^*([1 - F_X(z)] \cdot x, [1 - F_Y(z)] \cdot y)}{C^*([1 - F_X(z)], [1 - F_Y(z)]} = \lambda(x, y),$$

then there exists a function h and a parameter $\gamma \in \mathbb{R}$ such that $\lambda(x, y) = y^\gamma h(xy^{-\beta/\alpha})$, and therefore

$$\begin{aligned}
&\lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \,\middle|\, X > z, Y > z\right) \\
= &(1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\alpha}}\right).
\end{aligned}$$

Proof. From calculations of the previous paragraph, note that

$$\begin{aligned}
\kappa(x, y) &= \lim_{p\to 0} \mathbb{P}\left(\frac{X - F_X^{\leftarrow}(1-p)}{a\left(F_X^{\leftarrow}(1-p)\right)} > x, \frac{Y - F_X^{\leftarrow}(1-p)}{b\left(F_Y^{\leftarrow}(1-p)\right)} > y \,\middle|\, X > F_X^{\leftarrow}(1-p), Y > F_Y^{\leftarrow}(1-p)\right) \\
&= \lim_{z\to\infty} \frac{C^*([1 - F_X(z)] \cdot [1 - G_\alpha(x)], [1 - F_Y(z)] \cdot [1 - G_\beta(y)])}{C^*(1 - F_X(z), 1 - F_Y(z))}.
\end{aligned}$$

Since C^* satisfies assumptions directional regular variation (Theorem **??** and Proposition **??**), for all $x, y \in [0,1]$,

$$\begin{aligned}\kappa(x,y) &= \lambda([1-G_\alpha(x)],[1-G_\beta(y)]) = [1-G_\beta(y)]^{\gamma\beta} h\left([1-G_\alpha(x)][1-G_\beta(y)]^{-\beta/\alpha}\right)\\ &= (1+y)^{-\gamma} h\left(\frac{(1+x)^{-1/\alpha}}{(1+y)^{-1/\alpha}}\right).\end{aligned}$$

This finishes the proof of Theorem 6.3.8.

□

Remark 6.3.9. *In the case where X and Y have identical distributions (hence $\alpha = \beta$), then the two approaches are equivalent, and the limiting distribution is the same.*

6.4 UTDC and tail dependence measures

In this section, we focus on dependence measures, "*one of the most widely studied subjects in probability and statistics*", as mentioned in Jogdeo (1982). More specifically, we will introduce some "*scale-invariant*" dependence measures, following the ideas of Hoeffding (1940, 1941), and the axiomatic proposed by Scarsini (1984). Dependence measures are here considered in the context of dependence between random variables(2-dimensional dependence).

6.4.1 Dependence and concordance: an axiomatic approach

Dependence measures

Following Rényi (1959), define dependence measures δ as follows

Definition 6.4.1. *δ is measure of dependence if and only if δ satisfies*

1. $0 \leq \delta(X,Y) \leq +1$, $\delta(X, \pm X) = +1$,
2. $\delta(X,Y) = \delta(Y,X)$,
3. *if X and Y are independent, then* $\delta(X,Y) = 0$,
4. $\delta(aX+b, cY+d) = \delta(X,Y)$ *for all* $a, c > 0$ *and any* b, d.

Note that δ is supposed to be a *measure*, and hence, it should be positive. As a consequence, such a functional cannot be used to distinguish *positive* and *negative* dependence. In his initial paper, Rényi also added an additional assumption related to the normal distribution, i.e. if (X,Y) is a Gaussian vector with correlation r, then $\delta(X,Y) = |r|$. But as mention in Schweizer and Wolff (1976), those assumptions are much too constraining. Therefore several of the axioms were modified to obtain monotone dependence (see Schweizer and Wolff (1976, 1981)).

Concordance measures

Hoeffding (1942) considered three fundamental conditions that should satisfy a "*measure of the degree of relationship between two random variables*",

1. δ should lie between two fixed finite bounds ("*say* 0 *and* 1"),

2. δ should be equal to the lower bound if and only if X and Y are independent,

3. δ should be equal to the upper bound if and only if X and Y are functionally dependent (in the sense mentioned at the beginning of Section 1.3.).

Scarsini (1984) suggested the following fundamental properties that a measure κ of concordance should satisfy. The idea is to define a total order on the set of bivariate distributions, which should be consistent with the partial order $\preceq_{PQD}$ (also called the *concordance ordering*), defined as $(X_1, Y_1) \preceq_{PQD} (X_2, Y_2)$ if and only if

$$\mathbb{P}(F_{X_1}(X_1) \leq u, F_{Y_1}(Y_1) \leq v) \leq \mathbb{P}(F_{X_2}(X_2) \leq u, F_{Y_2}(Y_2) \leq v),$$

for all $0 \leq u, v \leq 1$.

Definition 6.4.2. *κ is measure of concordance if and only if κ satisfies*

1. *κ is defined for every pair (X, Y) of continuous random variables,*

2. *$-1 \leq \kappa(X, Y) \leq +1$, $\kappa(X, X) = +1$ and $\kappa(X, -X) = -1$,*

3. *$\kappa(X, Y) = \kappa(Y, X)$,*

4. *if X and Y are independent, then $\kappa(X, Y) = 0$,*

5. *$\kappa(-X, Y) = \kappa(X, -Y) = -\kappa(X, Y)$,*

6. *if $(X_1, Y_1) \preceq_{PQD} (X_2, Y_2)$, then $\kappa(X_1, Y_1) \leq \kappa(X_2, Y_2)$,*

7. *if $(X_1, Y_1), (X_2, Y_2), \ldots$ is a sequence of continuous random vectors that converge to a pair (X, Y) then $\kappa(X_n, Y_n) \to \kappa(X, Y)$ as $n \to \infty$.*

As pointed out in Scarsini (1984), most of the axioms are "*self-evident*". Note simply that the first one is necessary in order to build some total order (which is not the case for the covariance for instance, since vectors have to be in L^2). Note that the second one is related to the normalizing property of the correlation. Added with the fifth one, it helps to define some concepts of negative dependence.

Nevertheless, note that a stronger concept can be considered when the fourth item is replaced by

4′ X and Y are independent if and only if $\kappa(X, Y) = 0$.

If κ is measure of concordance, then, if f and g are both strictly increasing, then $\kappa(f(X), g(Y)) = \kappa(X, Y)$. Further, $\kappa(X, Y) = 1$ if $Y = f(X)$ with f almost surely strictly increasing, and analogously $\kappa(X, Y) = -1$ if $Y = f(X)$ with f almost surely strictly decreasing (see Scarsini (1984) or). From this result, it comes that those measures of dependence are then copula-based, in the sense that if (X_1, Y_1) and (X_2, Y_2) have the same copula (denoted C), then $\kappa(X_1, Y_1) = \kappa(X_2, Y_2) = \kappa(C)$.

Example 6.4.3. *Several copula-based measures of dependence can be considered (see Schweizer and Wolff (1981)), for example*

$$\sigma(X, Y) = 12 \int_0^1 \int_0^1 |C(u, v) - C^{\perp}(u, v)| du dv,$$

which represents the volume between the surfaces C and $C^{\perp}$. Observe that $\sigma(X, Y) = 0$ if and only if X and Y are independent.

6.4.2 Kendall's tau and Spearman's rho

We discuss in this section two measures of concordance, Kendall's tau and Spearman's rho. They shall provide interesting alternatives to Pearson's (linear) when this measure may be inappropriate (see Embrechts McNeil and Straumann (2002)).

Spearman's rho can be defined as Pearson's (linear) correlation between $U = F_X(X)$ and $V = F_Y(Y)$ (also called "*ranks*"). Since U and V are uniform, $\mathbb{E}(U) = \mathbb{E}(V) = 1/2$, and $Var(U) = Var(V) = 1/12$, and therefore,

$$\rho(X,Y) = corr(U,V) = \frac{\mathbb{E}(UV) - 1/4}{1/12} = 12\mathbb{E}(UV) - 3,$$

Definition 6.4.4. *Let (X,Y) denote a random pair of continuous random variables, with copula C, then Spearman's rho is defined as $\rho(X,Y) = 12\mathbb{E}(F_X(X)F_Y(Y)) - 3$, or equivalently*

$$\rho(X,Y) = 12\int_0^1\int_0^1 C(u,v)dudv - 3 = 12\int_{\mathbb{R}}\int_{\mathbb{R}}[F_{XY}(x,y) - F_X(x)F_Y(y)]dxdy.$$

As mentioned in Drouet-Mari and Kotz (2001), the expression on the right corresponds to the quantification of the strength of the PQD dependence at point (x,y). Spearman's rho can be seen as an average measure of PQD dependence.

This measure was initially introduced through its empirical version in Spearman (1904). It could be defined as

$$\rho(X,Y) = 3[\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)],$$

where (X_1,Y_1), (X_2,Y_2) and (X_3,Y_3) denote three independent versions of (X,Y) (see Nelsen (1999)).

Definition 6.4.5. *Let (X,Y) denote a random pair of continuous random variables, with copula C, then Kendall's tau is defined as*

$$\tau(X,Y) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1 = 4\mathbb{E}(C(F_X(X),F_Y(Y))) - 1.$$

Again, initially, Kendall's tau was not defined using copulae, but as the probability of concordance, minus the probability of discordance, i.e.

$$\tau(X,Y) = 3[\mathbb{P}((X_1 - X_2)(Y_1 - Y_2) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_2) < 0)],$$

where (X_1,Y_1) and (X_2,Y_2) denote two independent versions of (X,Y) (see Nelsen (1999)).

If X and Y are continuous, then Kendall's tau and Spearman's rho are measures of concordance (satisfying all the axioms of Definition 6.4.2) (see Scarsini (1984)).

Concordance measures might be interesting since they satisfy property 2 of Definition 6.4.2, i.e. if X and Y are comonotonic $\rho(X,Y) = 1$ and $\tau(X,Y) = 1$ and if X and Y are counter-comonotonic $\rho(X,Y) = -1$ and $\tau(X,Y) = -1$. But further, note that the converse also holds here (see Embrechts, Hoeing and Juri (2002)). □

Example 6.4.6. *Let X and Y be two random variable with an Archimedean copula C, with generator ϕ. Kendall's tau is given by*

$$\tau(X,Y) = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}dt.$$

Example 6.4.7. *As in Joe (1997), consider for $\theta \in [0,1]$ the mixture of the independent copula, and the upper Fréchet-Hoeffding bound,*

$$C(u,v) = (1-\theta)C^{\perp}(u,v) + \theta C^{+}(u,v), (u,v) \in [0,1] \times [0,1].$$

Then if (X,Y) has copula C, $\rho(X,Y) = \theta$. More generally, consider as in Fréchet (1958) the mixture of the independent copula, and the Fréchet-Hoeffding bounds, with $\alpha, \beta > 0$ and $\alpha+\beta \leq 1$,

$$C(u,v) = \alpha C^{-}(u,v) + (1-\alpha-\beta)C^{\perp}(u,v) + \beta C^{+}(u,v), (u,v) \in [0,1] \times [0,1].$$

Then $\rho(X,Y) = \beta - \alpha$.

To conclude this section, note that if those dependence measures do not characterize independence (in the sense that $\rho(X,Y)$ or $\tau(X,Y)$ do not imply that X and Y are equivalent), those notions have been intensively used to perform some nonparametric tests of independence.

6.5 Measures of tail dependence

6.5.1 Quantifying tail dependence

Joe, Smith and Weismann (1992) and conditional Kendall's tau

In order to assess whether their is independence or dependence in the extremes, Joe, Smith and Weissman (1992) introduce the following idea (keeping their notations): let u_1 and u_2 denote upper π-quantiles of the marginal distributions, i.e.

$$u_1 = F_X^{\leftarrow}(1-\pi) \text{ and } u_2 = F_Y^{\leftarrow}(1-\pi),$$

Set $I_X = \mathbb{I}(X > u_1)$ and $I_Y = \mathbb{I}(Y > u_2)$, the indicator that thresholds are exceeded. Then, if $\tau(\pi)$ denotes Kendall's tau (so called tau-b) for the associated 2×2 table for indicator variables,

$$\tau(\pi) = \frac{p(\pi) - \pi^2}{\pi - \pi^2},$$

where $p(\pi)$ is the probability of exceeding both thresholds. This results simply derives from the definition of Kendall's tau, based on the concordance definition.

Joe, Smith and Weissman (1992) mention then that "*by a condition in Sibuya (1960)*", if the joint distribution of (X,Y) is in the max-domain of attraction of some generalized extreme value distribution, then "*the limiting distribution has independent margins if $\tau(\pi) \to 0$ as $\pi \to 0$, and the limiting distribution has dependence if $\tau(\pi) \to \tau > 0$ as $\pi \to 0$* ". This is obtained from Sibuya (1960), since asymptotic independence (in the usual sense) is equivalent to

$$\mathbb{P}(F(X) > 1-s, F(Y) > 1-s) = o(s).$$

From this property, $p(\pi) = o(p)$ and therefore $\tau(\pi) \to 0$ as $\pi \to 0$. As we shall see, this condition is closely related to λ introduced in Joe (1993).

Example 6.5.1. *Figure 6.1 shows the evolution of $\tau(\pi)$ for those 2×2 tables, obtained using indicator variate of exceeding quantile thresholds, for Frank copula, Gumbel and survival Clayton.*

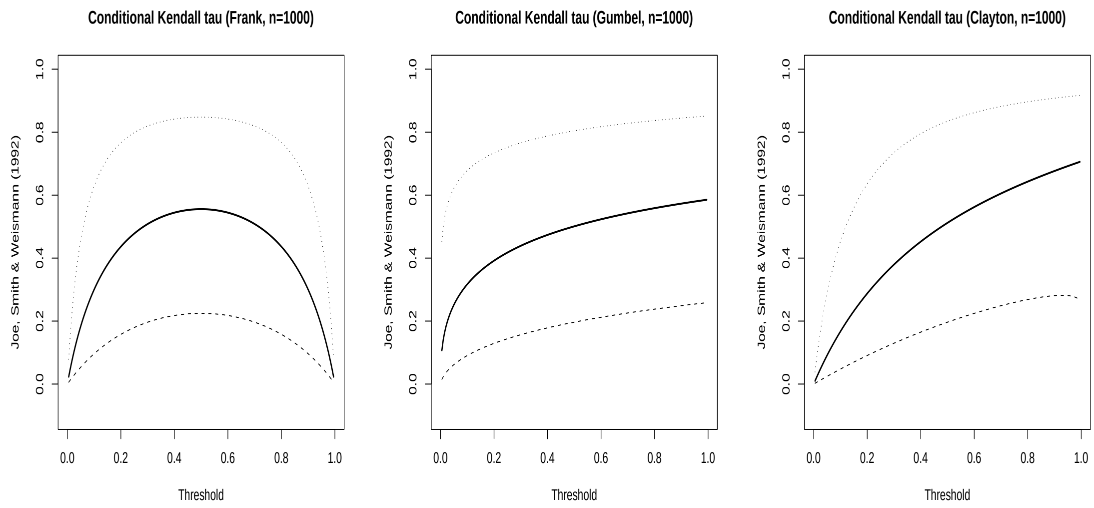


Figure 6.1: Joe, Smith and Weissmann (1992) conditional Kendall's tau, for upper tail dependence (2×2 tables for indicator variate of exceeding quantile thresholds), for Frank copula, Gumbel and survival Clayton, respectively when the overall Kendall's tau is 0.2 (dotted), 0.5 (plain) and 0.8 (dashed).

Ledford and Tawn (1996)'s η, weak tail dependence

Ledford and Tawn (1996), Ledford and Tawn (1997) or Ledford and Tawn (1998)) focused on bivariate survival probabilities to obtain asymptotic results. X and Y are assumed to have unit Fréchet margins. An heuristic approach is the following: if X and Y are *asymptotically* independent, the following equality should hold

$$\mathbb{P}(X > z, Y > z) \sim \mathbb{P}(X > z) \cdot \mathbb{P}(Y > z) = \mathbb{P}(X > z)^2, z \to \infty, \tag{6.35}$$

since X and Y have identical distributions. Thus, if

$$\mathbb{P}(X > z, Y > z) \sim \mathbb{P}(X > z)^{1/\eta}, z \to \infty, \tag{6.36}$$

$\eta = 1/2$ should mean that X and Y are asymptotically independent, if $\eta > 1/2$, there is more weight that in the independent case i.e. X and Y are asymptotically dependent (in the PQD sense). More formally,

$$\mathbb{P}(X > z, Y > z) \sim \mathcal{L}(z)\, \mathbb{P}(X > z)^{1/\eta}, z \to \infty, \tag{6.37}$$

where $\mathcal{L}$ is a slowly varying function at infinity, $\mathcal{L} \in \mathcal{R}_0^\infty$. In this approach, $\mathcal{L}$ is a slowly varying function as $z \to \infty$, and η is the coefficient of tail dependence (and lies in $(0, 1]$). $\eta \to 0$ and $\eta = 1$, with $\mathcal{L} = 1$, correspond respectively to the anti-comonotonic and comonotonic cases (i.e. lower and upper Fréchet-Hoeffding bounds),. Independence is obtained when $\eta = 1/2$ and $\mathcal{L} = 1$. Four types of tail behavior can then be obtained, depending on the value of η (provided that $\mathcal{L}(t) \to c > 0$ when $t \to \infty$).

An estimation of this parameter has been proposed in Draisma, Drees, Ferreira and de Haan (2004), based on some Pareto transformation of margins, using then some Hill estimator.

Joe (1993)'s λ, strong tail dependence

Joe (1993) defined, in the bivariate case a tail dependence measure: let (X, Y) denote a random pair, the upper and lower tail dependence parameters are defined, if the limit exist, as

$$\lambda_L = \lim_{u \to 0} \mathbb{P}\left(X \leq F_X^{\leftarrow}(u) \,|Y \leq F_Y^{\leftarrow}(u)\right),$$

and

$$\lambda_U = \lim_{u \to 1} \mathbb{P}\left(X > F_X^{\leftarrow}(u) \,|Y > F_Y^{\leftarrow}(u)\right).$$

Note that this coefficient can be obtained differently, as in Buishand (1994): consider a pair (X, Y) from an extreme value distribution, and assume that X and Y have unit Fréchet distribution, so that, using Pickands representation, the joint distribution function is

$$\mathbb{P}(X \leq x, Y \leq y) = \exp\left[-\left(\frac{1}{x} + \frac{1}{y}\right) A\left(\frac{y}{x+y}\right)\right].$$

Notice that the distribution of the maximum of X and Y is

$$\mathbb{P}(\max\{X, Y\} \leq x) = \mathbb{P}(X \leq x, Y \leq x) = \exp\left[-\frac{2}{x} A\left(\frac{1}{2}\right)\right] = [\exp(-1/x)]^{\theta},$$

where $\theta = 2A(1/2)$. More generally, if (X, Y) is a pair from an extreme value distribution, with identical marginal distribution function F, the distribution of the maximum is

$$\mathbb{P}(\max\{X, Y\} \leq x) = [F(x)]^{\theta},$$

Thus, $\theta = 2A(1/2)$ can be seen as a dependence measure, and can be used for testing independence (see Tawn (1988)). Buishand (1994) considered random pairs (X, Y) identically distributed, and defined

$$\theta(x) = \frac{\log \mathbb{P}(\max\{X, Y\} \leq x)}{\log \mathbb{P}(X \leq x)}.$$

In the case where (X, Y) has a bivariate extreme value distribution, $\theta(x)$ is constant.

Note that θ belongs to $[1, 2]$, that independence is obtained when $\theta = 2$ and comonotonicity when $\theta = 1$: Joe suggested to consider $2 - \theta$, defined on $[0, 1]$. θ and λ_U are related Fby

$$\lambda_U = 2 - \lim_{x \to \infty} \theta(x),$$

since when $x \to \infty$

$$2 - \frac{\log \mathbb{P}(\max\{X, Y\} \leq x)}{\log \mathbb{P}(X \leq x)} \sim \frac{\mathbb{P}(X > x, Y > x)}{1 - \mathbb{P}(X > x)} = \mathbb{P}(Y > x | X > x).$$

Joe (1993) noticed also, as mentioned earlier, that such a measure does not depend on the marginal distribution and thus, can be expressed using the copula of (X, Y): let (X, Y) denote a random pair with copula C, the upper and lower tail dependence parameters are defined, if the limit exist, as

$$\lambda_L = \lim_{u \to 0} \frac{C(u, u)}{u} \text{ and } \lambda_U = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u} = \lim_{u \to 0} \frac{C^*(u, u)}{u}.$$

6.5.2 Tail rank correlation function

As mentioned in the introduction of this thesis, several tail correlation coefficient have been introduced recently in the literature, e.g. Patton (2004), who considered some so-called "*exceedences correlations*", defined as

$$\rho(u) = \begin{cases} corr(X,Y|X \leq F_X^{\leftarrow}(p) \text{ and } Y \leq F_Y^{\leftarrow}(p)) & \text{if } p \leq 0.5, \\ corr(X,Y|X > F_X^{\leftarrow}(p) \text{ and } Y > F_Y^{\leftarrow}(p)) & \text{if } p > 0.5. \end{cases}$$

The split between low and high quantile is motivate in financial applications (see .e.g. Longin and Solnik (2001) or Ang and Chen (2002)): both high and low return are important. But in insurance or in environmental science, the main interest is either on high or low quantile (e.g. in insurance if the variable of interest is a loss or a gain amount).

Similarly, Boyer, Gibson and Loretan (1997) introduced the following conditional correlation, defined only to focus on upper tails,

$$r_p = \frac{cov\left(X,Y|X \geq F_X^{\leftarrow}(p), Y \geq F_Y^{\leftarrow}(p)\right)}{\sqrt{Var\left(X|X \geq F_X^{\leftarrow}(p), Y \geq F_Y^{\leftarrow}(p)\right)}\sqrt{Var\left(Y|X \geq F_X^{\leftarrow}(p), Y \geq F_Y^{\leftarrow}(p)\right)}}.$$

Boyer, Gibson and Loretan (1997) used this coefficient to quantify changes in correlations of exchange rates, as well as Hauksson et al. (2001). As pointed out, for most exchange rates, r_p is increasing in p. As mentioned in Chapter 1, Pearson's linear correlation might not be an appropriate tool to quantify the strength of the dependence since it depends (strongly) on marginal behavior. Hence, since (X,Y) and $(X,Y|X \geq F_X(p), Y \geq F_Y(p))$ do not have the same marginal distributions, it becomes difficult to compare r and r_p. Does $r_{p'} > r_p$ when $p' > p$ means that there is "*more dependence*" in upper tails ? Does positive dependence rise in up (and down) markets ? In which sense ? As shown on Figure 6.2 the monotonicity of r_p as a function of p does not depend only on the copula, but also marginal distribution. Note that in the case where (X,Y) has survival Clayton's copula (which remains unchanged by truncature, i.e. the copula of $X,Y|X \geq F_X^{\leftarrow}(p), Y \geq F_Y^{\leftarrow}(p)$ is always the same, whatever $p \in [0,1)$), r_p can be either decreasing or increasing.

Note that Malevergne and Sornette (2002) introduced a so-called "*conditional rank correlation*" (focusing on $(U,V)|V > v$, where $U = F_X(X)$ and $V = F_Y(Y)$), defined as,

$$\rho(v) = \frac{cov(U,V|V>v)}{\sqrt{Var(U|V>v)Var(V|V>v)}}.$$

But as pointed out earlier when defining tail conditional copulae, this should not be the way to define properly this coefficient: since margins of $(U,V)|V > v$ are not uniformly distributed, it cannot be a "*rank*" correlation.

6.5.3 Local measures of dependence and functional measures

As in Drouet-Mari and Kotz (2001), a function will be said to be locally PQD in the neighborhood $\mathcal{V}(x_0,y_0)$ of (x_0,y_0) if

$$\mathbb{P}(X > x, Y > y) \geq \mathbb{P}(X > x) \cdot \mathbb{P}(Y > y), \text{ for all } (x,y) \in \mathcal{V}(x_0,y_0).$$

Hence, define restriction of Spearman's rho and Kendall's tau to an open neighborhood of (x_0,y_0) as

$$\rho(x_0,y_0) = \frac{12\int_{\mathcal{V}(x_0,y_0)}[C(u,v) - C^{\perp}(u,v)]dudv}{\int_{\mathcal{V}(x_0,y_0)} dudv}$$

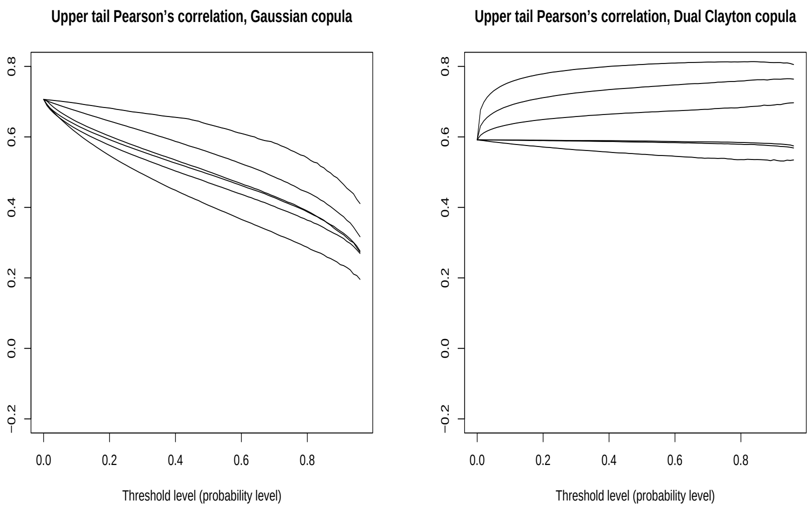


Figure 6.2: Upper tail correlation ($corr(X, Y|X > F_X^{\leftarrow}(u), F_Y^{\leftarrow}(u))$ with Gaussian copula (on the right) and dual Clayton (on the left), and different margins (Gaussian, Student, Lognormal, Pareto) .

and

$$\tau(x_0, y_0) = \frac{4\int_{\mathcal{V}(x_0,y_0)} C(u,v)dC(u,v) - 1}{\int_{\mathcal{V}(x_0,y_0)} dC(u,v)}.$$

Recall that for concordance measures, Proposition **??** implies an invariance by increasing transformation of the margins, i.e. concordance measures are copula based. Analogously, it might be interesting to define some copula based local measures of dependence, i.e. $\mathcal{V}$ should be defined using ranks. Instead of studying the conditional distribution on some neighborhood, it is possible to extend those notions on positive cones, $\mathcal{V}(x_0, y_0) = (x_0, \infty) \times (y_0, \infty)$.

Based on this setting, one can define a "*local*" measure of concordance, similarly to Definition 6.4.2.

Definition 6.5.2. *Consider a subset $\mathcal{V}$ of $\mathbb{R}^2$ with non-null interior, then $\kappa_{\mathcal{V}}$ is a local measure of concordance if and only if κ satisfies*

1. *$\kappa_{\mathcal{V}}$ is defined for every pair (X, Y) of continuous random variables, such that $\mathbb{P}((X, Y) \in \mathcal{V}) \neq 0$,*

2. *$-1 \leq \kappa_{\mathcal{V}}(X, Y) \leq +1$, $\kappa_{\mathcal{V}}(X, X) = +1$,*

3. *$\kappa_{\mathcal{V}}(X, Y) = \kappa_{\mathcal{V}}(Y, X)$,*

4. *if X and Y are independent on $\mathcal{V}$ (i.e. (X, Y) given $(X, Y) \in \mathcal{V}$ is independent), then $\kappa_{\mathcal{V}}(X, Y) = 0$,*

5. *if $((X_1, Y_1)|(X_1, Y_1) \in \mathcal{V}) \preceq_{PQD} ((X_2, Y_2)|(X_2, Y_2) \in \mathcal{V})$, then $\kappa_{\mathcal{V}}(X_1, Y_1) \leq \kappa_{\mathcal{V}}(X_2, Y_2)$,*

6. *if $(X_1, Y_1), (X_2, Y_2), \ldots$ is a sequence of continuous random vectors that converge to a pair (X, Y) then $\kappa_{\mathcal{V}}(X_n, Y_n) \to \kappa_{\mathcal{V}}(X, Y)$ as $n \to \infty$.*

This notion is based on similar axioms as the ones of Definition 6.4.2. Note only that conditions based on $(X, -Y)$ properties do not hold here since $\mathcal{V}$ is not necessarily symmetric.

Natural tail measures of concordance can then be defined, when κ is a measure of concordance (e.g. Spearman's rho or Kendall'tau) and when $\mathcal{V}$ is an upper-quadrant, $[F_X-1(u),\infty)\times[F_Y-1(v),\infty)$, or lower-quadrant, $(-\infty,F_X-1(u))\times(-\infty,F_Y-1(v))$. As mentioned in Chapter 1, since κ can be expressed through the copula C of (X,Y), in the case of lower-quadrant measures of concordance $\kappa_{\mathcal{V}}=\kappa_{(-\infty,F_X-1(u))\times(-\infty,F_Y-1(v))}$ is then a function of the conditional copula $C_{(u,v)}$.

6.5.4 The upper conditional rank correlation

As well as the standard rank correlation is a function of C (or C^*), the upper conditional rank correlation should be defined as a function of the associated conditional copula $C_{(\cdot,\cdot)}$ or $C^*_{(\cdot,\cdot)}$:

Definition 6.5.3. *The upper tail conditional rank correlation of (X,Y) given $\{X>F_X^{\leftarrow}(u),X>F_Y^{\leftarrow}(v)\}$ is defined as*

$$\rho(u,v)=12\int\int C^*_{(1-u,1-v)}(x,y)dC^*_{(1-u,1-v)}*(x,y)-3.$$

Example 6.5.4. *Figure 6.3 shows the evolution of the upper tail conditional rank correlation for Gaussian and Gumbel copulae, for different parameters. Note that they do not cross, i.e. if $C_1\preceq C_2$ then $\rho_1(u,v)\leq\rho_2(u,v)$ for all $u,v\in[0,1)$. Further, note that if the conditional rank correlation if always decreasing for the Gaussian copulae (in some sense, there is less and less dependence in upper tails), it is not the case for Gumbel copulae: if θ is not too high, there is $u^*\in(0,1)$ such that $u^*=\arg\min\{\rho(u,u),u\in[0,1)\}$.*

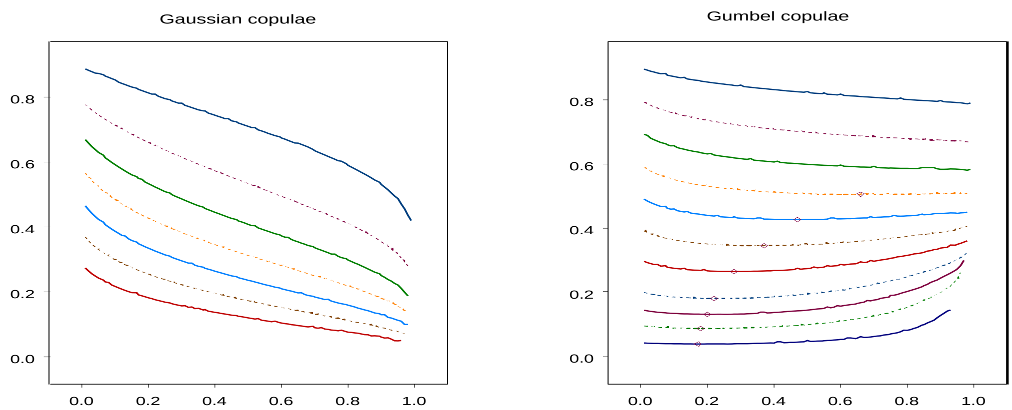


Figure 6.3: Evolution of the upper tail conditional rank correlation (associated to $(X,Y)|X>F_X^{\leftarrow}(u),Y>F_Y^{\leftarrow}(u))$, for Gaussian and Gumbel copulae, for different parameters.

Proposition 6.5.5. *Given u in $[0,1)$, $\rho(u,u)$ is a local concordance measure in the sense of Definition 6.5.2, where $\mathcal{V}$ is defined as $\mathcal{V}=[F_X^{\leftarrow}(u),\infty)\times[F_Y^{\leftarrow}(u),\infty)$.*

Proof. Consider (X,Y) a random pair with survival copula C, and set $U=1-F_X(X)$ and $V=1-F_Y(Y)$. Then (U,V) has distribution function C.

1. If there is no mass in $[u,1)\times[u,1)$, the conditional copula is not defined, and so $\rho(\cdot,u,u)$ cannot be defined. But under the assumption that
$$C(1-u,1-u)=\mathbb{P}((U,V)\in[u,1)\times[u,1))>0$$
the conditional copula of (U,V) given $U>u,V>u$ is defined, and so, $\rho(\cdot,u,u)$ is defined.

2. $\rho(\cdot,u,u)$ belongs to $[-1,+1]$, as Spearman rank correlation. Further, the upper bound can be reached, e.g. if X and Y are comonotonic.

3. The rank correlation of (X,Y) given $\{X>F_X^{\leftarrow}(u)$ and $Y>F_Y^{\leftarrow}(u)\}$ is also the rank correlation of (Y,X) given $\{Y>F_Y^{\leftarrow}(u)\}$ and $\{X>F_X^{\leftarrow}(u)\}$ (the probability rank being the same for both components).

4. If X and Y are independent, $\Psi(C^*,u,u)=C^{\perp}$ for all u.

5. If $(X_1,Y_1)\leq_{PQD}(X_2,Y_2)$ then $C^1\leq C^2$ (for the usual pointwise order) then $C^{1,*}_{(u,u)}\leq C^{2*}_{(u,u)}$ for all u, and thus
$$\rho\left(C^{1,*}_{(u,u)}\right)\leq\rho\left(C^{2,*}_{(u,u)}\right)\text{ for all }u\in[0,1).$$

6. If (X_i,Y_i) is a sequence of continuous vectors that converges towards a pair (X,Y), then for all $\mathcal{V}$, $((X_i,Y_i)\,|(X_i,Y_i)\in\mathcal{V})$ is a sequence of continuous vectors that converges towards a pair $((X,Y)\,|(X,Y)\in\mathcal{V})$.

This finishes the proof of Proposition 6.5.5. □

From Point 3 in the proof, note that $\rho(\cdot,u,v)$ is not a concordance measure if $u\neq v$.

6.5.5 Conditional Kendall's tau

Analogously, based on conditional copulae, it is possible to define an upper tail conditional Kendall's tau. But as well as the conditional rank correlation is not the one introduced in Boyer, Gibson and Loretan (1997) or Malevergne and Sornette (2002), the conditional Kendall's tau defined in this section is not the one defined in Oakes (1989) and Drouet-Mari and Kotz (2001).

Definition 6.5.6. *The upper tail conditional Kendall's tau of (X,Y) given $\{X>F_X^{\leftarrow}(u),X>F_Y^{\leftarrow}(v)\}$ is defined as*
$$\tau(u,v)=4\int\int C^*_{(1-u,1-v)}(x,y)dxdy-1.$$

Proposition 6.5.7. *Given u in $[0,1)$, $\tau(u,u)$ is a local concordance measure in the sense of Definition 6.5.2, where given X and Y, $\mathcal{V}$ is defined as $\mathcal{V}=[F_X^{\leftarrow}(u),\infty)\times[F_Y^{\leftarrow}(u),\infty)$.*

Proof. A proof analogous to the proof of Proposition 6.5.5 holds. □

Example 6.5.8. *In the case of Archimedean copulae, recall that Kendall's tau can be derived directly from the generator (see Example 6.4.6) as*
$$\tau=1+4\int_0^1\frac{\phi(t)}{\phi'(t)}dt.$$

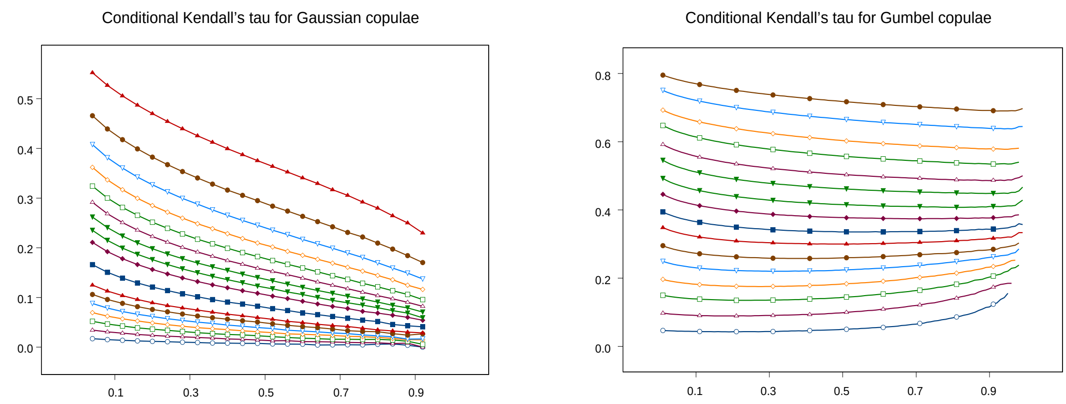


Figure 6.4: Evolution of the upper tail conditional Kendall's tau (associated to $(X,Y)|X > F_X^{\leftarrow}(u), Y > F_Y^{\leftarrow}(u))$, for Gaussian and Gumbel copulae, for different parameters.

Thus, since generators of conditional copula can easily be obtained as

$$\phi_{u,v} = \phi\left(tC\left(u,v\right)\right) - \phi\left(C\left(u,v\right)\right) = \phi\left(t \cdot \phi^{\leftarrow}\left(\phi\left(x\right) + \phi\left(y\right)\right)\right) - \left[\phi\left(u\right) + \phi\left(v\right)\right]$$

conditional Kendall's tau can be expressed as

$$\begin{aligned} \tau\left(u,v\right) &= 1 + 4\int_0^1 \frac{\phi\left(tC\left(u,v\right)\right) - \phi\left(C\left(u,v\right)\right)}{C\left(u,v\right)\cdot\phi'\left(tC\left(u,v\right)\right)}dt \\ &= 1 + 4\int_0^1 \frac{\phi\left(tC\left(u,v\right)\right)}{C\left(u,v\right)\cdot\phi'\left(tC\left(u,v\right)\right)}dt - \int_0^1 \frac{\phi\left(C\left(u,v\right)\right)}{C\left(u,v\right)\cdot\phi'\left(tC\left(u,v\right)\right)}dt \\ &= 1 + \frac{4}{C\left(u,v\right)^2}\int_0^1 \frac{\phi\left(s\right)}{\phi'\left(s\right)}ds - \int_0^1 \frac{\phi\left(C\left(u,v\right)\right)}{C\left(u,v\right)\cdot\phi'\left(tC\left(u,v\right)\right)}dt \\ &= 1 + \frac{\tau - 1}{C\left(u,v\right)^2} - \frac{\phi\left(C\left(u,v\right)\right)}{C\left(u,v\right)}\int_0^1 \frac{dt}{\phi'\left(tC\left(u,v\right)\right)}. \end{aligned}$$

6.5.6 Testing for tail independence

Using conditional Kendall's tau, or conditional Spearman's rho, one can define some testing procedures for tail independence (since distribution of those dependence measure can be obtained only under the assumption of independence).

Testing for independence

Hájek and Sidák (1967), Capéraà and Van Cutsen (1988), and Behnen and Neuhaus (1989) introduced several testing procedures for independence, based on ranks. Given a n sample $(X_1, Y_1), ..., (X_n, Y_n)$ from $F_{X,Y}$, and copula C, let $(R_1, S_1), ..., (R_n, S_n)$ denote the associated pairs of ranks. Those pairs are an appropriate statistics to base a test of $H_0 : C = C^{\perp}$. A standard procedure is based on the empirical version of Kendall's tau,

$$\widehat{\tau}_n = \frac{2}{n(n-1)}\sum_{i<j}\text{sign}(R_i - R_j)\cdot\text{sign}(S_i - S_j).$$

Note that it is an unbiased U statistic (Hoeffding (1948)), and therefore, $\widehat{\tau}_n$ is asymptotically Gaussian under the null hypothesis (and hence asymptotically unbiased). Since its asymptotic variance is $Var(\widehat{\tau}_n) = 2(2n+5)/9n(n-1)$, the null hypothesis should be rejected if, for large enough sample size n, $3\sqrt{n}|\widehat{\tau}_n|/2 > 1.96$. Similarly, another procedure is based on the empirical version of Spearman's rho,

$$\widehat{\rho}_n = \frac{12}{n^3-n}\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right) = -3\frac{n+1}{n-1} + \frac{12}{n(n^2-1)}\sum_i R_i \cdot S_i,$$

which can also be related to some U statistics. Here, variance, under the null hypothesis is simply $Var(\widehat{\rho}_n) = 1/(n-1)$, and therefore, the null hypothesis should be rejected if, for large enough sample size n, $\sqrt{n}|\widehat{\rho}_n| > 1.96$. The closely related U statistic is here $\widehat{u}_n$ defined by

$$\widehat{u}_n = \frac{3}{n(n-1)(n-2)}\sum_{i,j,k}\text{sign}(X_i - X_j)\cdot\text{sign}(Y_i - Y_k) \tag{6.38}$$

where the summation extends not over the n^3 possible triples, but only over the $n(n-1)(n-2)$ triples of distinct subscripts. Further (see Hoeffding (1948)), $\widehat{u}_n$ is an unbiased estimate of

$$u(X,Y) = 3\int\int F_X(x)\,F_Y(y)\,dF_{XY}(x,y) = \rho(X,Y)$$

Hence, the natural estimate of Spearman's ρ, $\widehat{\rho}_n$, can then be written

$$\widehat{\rho}_n = \frac{n-2}{n+1}\widehat{u}_n + \frac{6}{n+1}\widehat{\tau}_n. \tag{6.39}$$

Note that this estimator is biased :

$$\mathbb{E}\left[\widehat{\rho}_n(X,Y)\right] = \frac{n-2}{n+1}\rho(X,Y) + \frac{3}{n+1}\tau(X,Y)$$

But the estimator as asymptotically unbiased, and asymptotically normally distributed.

More generally, several testing procedures have been considered, of the form

$$\widehat{t}_n = \frac{1}{n}\sum_i \mathcal{J}\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right),$$

with specific score functions $\mathcal{J}$ (e.g. Bhuchongkul (1964) suggested $\mathcal{J}(u,v) = \Phi^{\leftarrow}(u)\cdot\Phi^{\leftarrow}(v)$, also called van der Waerden test).

Example 6.5.9. *Consider a class of copulae C_θ, with $C_{\theta_0} = C^\perp$, such that $\theta < \theta'$ implies $C_\theta \preceq C_{\theta'}$, with continuous density c_θ so that $\partial c_\theta = \partial c_\theta/\partial\theta$ is square integrable, and exists at θ_0. Shirahata (1975) and Genest and Verret (2004) show that the locally most powerful test, for random pairs with copula C_θ, is*

$$T_n^* = \frac{1}{n}\sum_i \mathcal{T}(R_i, S_i),\ \text{where } \mathcal{T}(x,y) = \mathbb{E}\left(\frac{\partial c_{\theta_0}}{c_{\theta_0}}(U_{x:n}, V_{y:n})\right),$$

for all $x, y \in \{1, ..., n\}$, and $U_1, ..., U_n, V_1, ..., V_n$ are uniformly distributed over $[0,1]$. Hence, this test is asymptotically equivalent to

$$T_n^* = \frac{1}{n}\sum_i \partial c_{\theta_0}\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right).$$

As noticed by Genest and Verret (2004), for several families of copulae, the independence test based on Spearman's rho is locally most powerful. In the case of Clayton or Gumbel copulae, the locally most powerful test is based on

$$T_n^*(\textit{Clayton}) = \frac{1}{n}\sum_i \log\frac{R_i}{n+1}\cdot\log\frac{S_i}{n+1}.$$

Estimating conditional dependence measures

Those conditional dependence measures can be estimated using natural estimators introduced in the previous paragraph, with notations of Section 6.5.3.

Definition 6.5.10. *Let $\mathcal{V}$ denote a subset of $\mathbb{R}^2$, and $\mathcal{I}$ the subset of indexes $i \in \{1, ..., n\}$ such that $(X_i, Y_i) \in \mathcal{V}$. Let R_i^* and S_i^* denote the ranks of the X_i and Y_i respectively, among $\{X_i, i \in \mathcal{I}\}$ and $\{Y_i, i \in \mathcal{I}\}$. Let n^* denote the cardinal of $\mathcal{I}$. Then*

$$\widehat{\tau}_n(\mathcal{V}) = \frac{2}{n^*(n^*-1)} \sum_{i<j\in\mathcal{I}} sign(R_i^* - R_j^*) \cdot sign(S_i^* - S_j^*),$$

and

$$\widehat{\rho}_n(\mathcal{V}) = -3\frac{n^*+1}{n^*-1} + \frac{12}{n^*(n^{*2}-1)} \sum_{i\in\mathcal{I}} R_i^* \cdot S_i^*,$$

are natural estimators of $\tau(\mathcal{V}) = \tau((X,Y)|(X,Y) \in \mathcal{V})$ and $\rho(\mathcal{V}) = \rho((X,Y)|(X,Y) \in \mathcal{V})$.

Using standard properties of those estimators for non-conditional samples, note that several properties can be derived.

Proposition 6.5.11. *$\widehat{\tau}_n(\mathcal{V})$ is an unbiased estimate of $\tau(\mathcal{V})$, and $\widehat{\rho}_n(\mathcal{V})$ is a biased estimate of $\rho(\mathcal{V})$.*

Proof. n^* denotes the number of points in $\mathcal{V}$ (e.g. the upper quadrant). Note that the expected value of $\widehat{\tau}_n(\mathcal{V})$ is $\mathbb{E}(\widehat{\tau}_n(\mathcal{V})) = \mathbb{E}(\mathbb{E}(\widehat{\tau}_n(\mathcal{V})|n^*))$. Here $\widehat{\tau}_n(\mathcal{V})|n^*$ is the natural estimate of Kendall's tau for some n^*-sample. Hence, it is unbiased, and so $\mathbb{E}(\widehat{\tau}_n(\mathcal{V})|n^*) = \tau(\mathcal{V})$, and thus, $\mathbb{E}(\widehat{\tau}_n\mathcal{V})) = \tau(\mathcal{V})$.

For Spearman's rho $\mathbb{E}(\widehat{\rho}_n(\mathcal{V})) = \mathbb{E}(\mathbb{E}(\widehat{\rho}_n(\mathcal{V})|n^*))$. But here, $\widehat{\rho}_n(\mathcal{V})|m$ is a biased estimate of $\rho(\mathcal{V})$, and more precisely,

$$\mathbb{E}(\widehat{\rho}_n(\mathcal{V})|n^*) = \frac{n^*-2}{n^*+1}\rho(\mathcal{V}) + \frac{3}{n^*+1}\tau(\mathcal{V}).$$

Hence,

$$\mathbb{E}(\widehat{\rho}_n(\mathcal{V})) = \rho(\mathcal{V})\sum_{k=0}^{n}\frac{k-2}{k+1}\mathbb{P}(n^*=k) + \tau(\mathcal{V})\sum_{k=0}^{n}\frac{3}{k+1}\mathbb{P}(n^*=k). \tag{6.40}$$

This finishes the proof of Proposition 6.5.11. □

In order to get a better understanding of the behavior of this estimator (and therefore asymptotic approximations), we should get the (exact) distribution of n^*.

On the (exact) distribution of n^* when $\mathcal{V} = [F_X^{\leftarrow}(u), +\infty) \times [F_Y^{\leftarrow}(u), +\infty)$

Let $(X_1, Y_1), ..., (X_n, Y_n)$ denote a n-i.i.d. sample with copula C, and joint distribution function $F_{X,Y}$. Set $U_i = R_i/(n+1)$ and $V_i = S_i/(n+1)$ where R_i and S_i denote respectively the ranks of X_i and Y_i. Note that the number n^* of observations (X_i, Y_i) which belong to region $\mathcal{V} = [F_X^{\leftarrow}(u), +\infty) \times [F_Y^{\leftarrow}(u), +\infty)$ is a random variable with joint distribution

$$n^* \sim \mathcal{B}(n, \pi)$$

where

$$\pi = \mathbb{P}(X > F_X^{\leftarrow}(u), Y > F_Y^{\leftarrow}(u)) = \overline{F}_{X,Y}(F_X^{\leftarrow}(u), F_Y^{\leftarrow}(u)) = C^*(1-u, 1-u).$$

But, since the (U_i, V_i)'s are rank based, note that the distribution of n^* (number of observations within the upper quadrant) is slightly more difficult to get, because of the constraints

$$\mathrm{Card}\{i \in \{1, ..., n\}, U_i > u\} = \mathrm{Card}\{j \in \{1, ..., n\}, V_j > u\} = [(n+1)(1-u)],$$

where $[\cdot]$ stands for the integer part. An idea can be to use Bayes formula, i.e. $\mathbb{P}(U > u, V > u) = \mathbb{P}(U > u | V > u) \cdot \mathbb{P}(V > u)$. Hence, the following result holds,

Proposition 6.5.12. *The cardinal n^* of $\{i \in \{1, ..., n\}, U_i > u$ and $V_i > u\}$ is a random variable with distribution*

$$n^* \sim \mathcal{B}(m, \pi) \text{ where } m = [(n+1)(1-u)] \text{ and } \pi = \frac{C^*(1-u, 1-u)}{1-u}.$$

Proof. Since the R_i's and the S_i's are the ranks, there are necessarily exactly $m = [(n+1)(1-u)]$ observations (U_i, V_i) in the regions $[1-u, 1] \times [0, 1]$. Consider the subset of m observations, $\{(U_i, V_i), U_i > u\}$ (see Figure 6.5). In that case, there are no other constraint, and therefore, within those m observations, the number of observations such that both component exceed u is (using the result mentioned at the beginning of this section) a Binomial distribution, with parameters m and

$$\pi = \mathbb{P}(V > u | U > u) = \frac{\mathbb{P}(V > u, U > u)}{\mathbb{P}(U > u)} = \frac{C^*(1-u, 1-u)}{1-u}.$$

This finishes the proof of Proposition 6.5.12. □

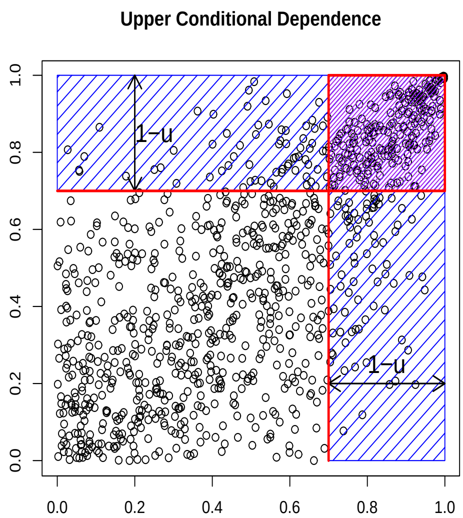


Figure 6.5: Scatterplot of some sample $\{(U_i, V_i), i = 1, ..., n\}$ with regions $\{(U_i, V_i), U_i > u\}$, $\{(U_i, V_i), U_i > u\}$ and $\{(U_i, V_i), U_i > u, V_i > u\}$. Note that any hatched area should include $m = [(n+1)(1-u)]$ observations.

Proposition 6.5.13. *$\widehat{\rho}_n(u, u)$ is asymptotically an unbiased estimator of $\rho(u, u)$.*

Proof. n^* denotes the number of observations in the upper quadrant, for n observations $(X_1, Y_1), ..., (X_n, Y_n)$. Hence, n^* is a Binomial distribution, $\mathcal{B}(m, \pi)$ where $m = [(n+1)(1-u)]$ and

$$\pi = \mathbb{P}(V > u | U > u) = \frac{\mathbb{P}(V > u, U > u)}{\mathbb{P}(U > u)} = \frac{C^*(1-u, 1-u)}{1-u}.$$

From Equation (6.40), recall that

$$\mathbb{E}\left(\widehat{\rho}_n\left(\mathcal{V}\right)\right) = A_n \rho\left(\mathcal{V}\right) + B_n.$$

Therefore, on the one hand, we set

$$A_n = \sum_{k=0}^{n} \frac{\mathbb{P}(n^* = k)}{k+1} = \sum_{k=0}^{m} \frac{1}{k+1} \binom{m}{k} \pi^k (1-\pi)^{m-k},$$

and therefore

$$\begin{aligned} A_n &= \sum_{k=0}^{m} \frac{m!}{(k+1)!(m-k)!} \pi^k (1-\pi)^{m-k} = \frac{1}{mp} \sum_{k=0}^{m} \frac{(m+1)!}{(k+1)!(m-k)!} \pi^{k+1} (1-\pi)^{m-k} \\ &= \frac{1}{mp} \sum_{k=0}^{m} \binom{m+1}{k+1} \pi^{k+1} (1-\pi)^{(m+1)-(k+1)} = \frac{1}{mp} \sum_{k=1}^{m+1} \binom{m+1}{k} \pi^k (1-\pi)^{(m+1)-(k)} \\ &= \frac{1-(1-\pi)^{m+1}}{mp} \sim \frac{1-(1-\pi)^{(n+1)(1-u)+1}}{(n+1)(1-u)p} \to 0, \end{aligned}$$

as $n \to \infty$. On the other hand, one gets that, similarly

$$\begin{aligned} B_n &= \sum_{k=0}^{n} \frac{k-2}{k+1} \mathbb{P}(n^* = k) = \sum_{k=2}^{n} \frac{k+1}{k+1} \mathbb{P}(n^* = k) + \sum_{k=0}^{n} \frac{-3}{k+1} \mathbb{P}(n^* = k) \\ &= 1 + \sum_{k=0}^{n} \frac{-3}{k+1} \mathbb{P}(n^* = k) \to 1 \text{ as } n \to \infty. \end{aligned}$$

So, finally, $\mathbb{E}(\widehat{\rho}_n\,(u, v)) \to \rho(u, v)$ as $n \to \infty$. This finishes the proof of Proposition 6.5.13. □

Remark 6.5.14. *Note that an elegant proof can also be written, using the relationship $\mathbb{P}(n^* = k) = \mathbb{P}(n^* > k) - \mathbb{P}(n^* > (k-1))$, and calling upon results on Abel series.*

Example 6.5.15. *Figure 6.6 shows the distribution of $\widehat{\rho}_n(u, u)$ and $\widehat{\tau}_n(u, u)$, for $u = 0.5$ and $n = 1,000$ for the survival Clayton copula ($\tau = 0.5$ and $\rho = 0.68$), obtain using $1,000,000$ Monte Carlo simulations (and smoothing the distributions). From the Gaussian shape of the distribution, it is intuitive to look for approximations, that might be useful to perform easily some tail independence tests.*

Approximations for the distribution of n^*, and

Using standard properties of the limiting behavior of the Binomial distribution, note that, given $u \in [0, 1)$ as $n \to \infty$,

$$n^* \sim \mathcal{N}\left(m(n, u, C^*), \sqrt{V(n, u, C^*)}\right),$$

where

$$\begin{cases} m(n, u, C^*) = nC^*(1-u, 1-u) \\ V(n, u, C^*) = nC^*(1-u, 1-u)\left(1 - \dfrac{C^*(1-u, 1-u)}{1-u}\right). \end{cases}$$

In the case where $C = C^{\perp}$ (H_0), this asymptotic distribution can be simplified:

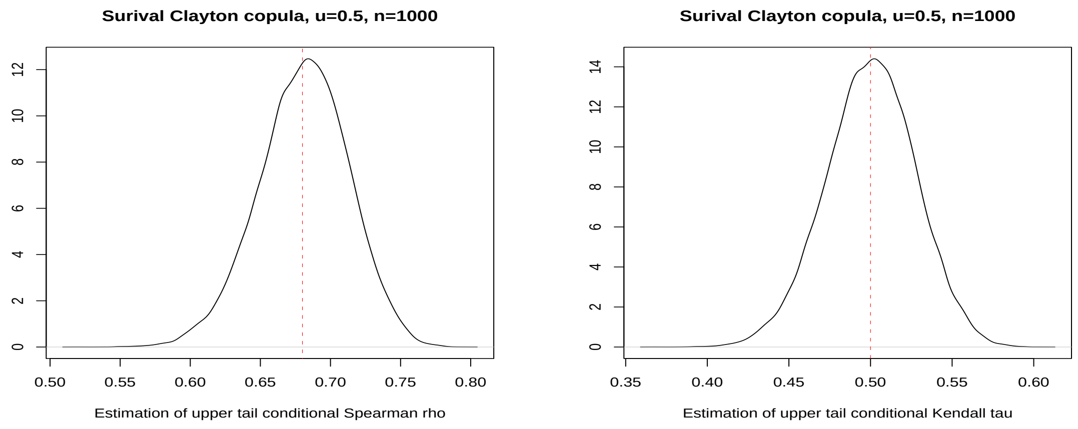


Figure 6.6: Distribution of $\widehat{\rho}_n(u,u)$ and $\widehat{\tau}_n(u,u)$, for $u = 0.5$ and $n = 1000$ for the survival Clayton copula.

Proposition 6.5.16. *Under the hypothesis of independence ($H_0 : C = C^* = C^\perp$), the distribution of n^*, as $n \to \infty$ can be approximated by*

$$n^* \overset{\perp}{\sim} \mathcal{N}\left(n(1-u)^2, \sqrt{nu(1-u)}\right).$$

Example 6.5.17. *This approximation (for the distribution of n^*) can be observe on Figure 6.5.6 where for 50,000 simulated n-samples, the real-distribution is plotted (a Binomial distribution) versus the asymptotic approximation (the Gaussian distribution).*

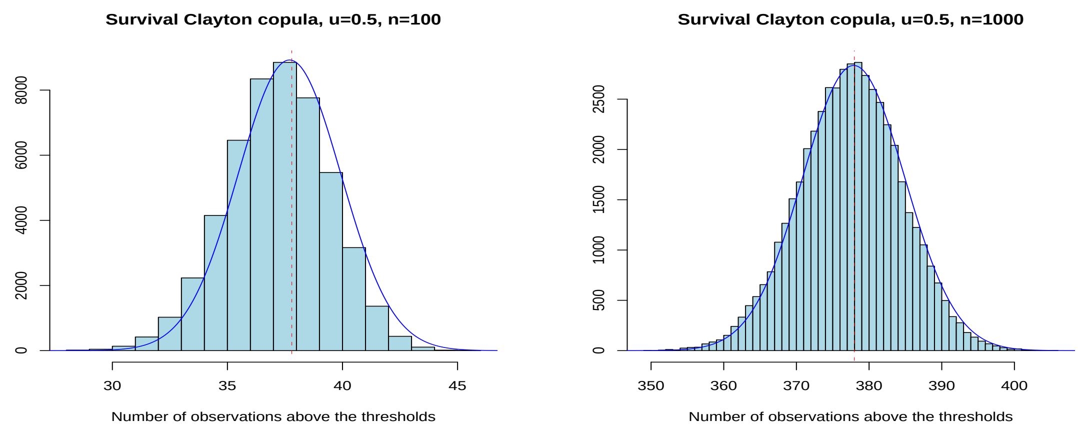


Figure 6.7: Distribution of the number of observations in $[0.5, 1]^2$, for samples of size 100 (on the left) and 1,000 (on the right). The histogram is the "*real*" distribution obtained from 50,000 simulated n-sample, while the line is the Gaussian approximation.

Remark 6.5.18. *Recall that the Gaussian approximation for a binomial distribution $\mathcal{B}(n,p)$ is a "good" approximation when $np \geq 30$ and $n(1-p) \geq 30$ (see e.g. Hollander and Wolfe (1999)). Under the independence hypothesis, one can use the Gaussian approximation when $n(1-u)^2 \geq 30$. Table 6.5.6 gives some ideas on the meaning of this inequality. Note that to use a Gaussian approximation when $u = 90\%$, on needs, at least, 3,000 observations.*

u	$n \geq 30/(1-u)^2$	n	$u \geq \sqrt{30/n}$
50%	120	100	45%
80%	750	1,000	83%
90%	3,000	5,000	92%
95%	12,000	10,000	95%
99%	300,000	100,000	98%

Table 6.1: Link between the total size of the sample and the maximal value of u so that the Gaussian approximation (of the distribution of the number of exceeding observations) is valid.

On the distribution of $\widehat{\tau}_n(u,u)$ and $\widehat{\rho}_n(u,u)$

Given an n-sample, Kendall's tau and Spearman's rho can be estimated using procedure mentioned at the beginning of this section, but no non-null exact distribution can be obtained (some tables can be consulted, see e.g. Hollander and Wolfe (1999)).

On asymptotic distributions of the estimators under independence

Under independence, given an n-sample, $\mathbb{E}(\widehat{\rho}_n) \to 0$ and $\mathbb{E}(\widehat{\tau}_n) = 0$, with an asymptotic Gaussian distribution. Hence, if n is large enough

$$\widehat{\rho}_n \overset{\perp}{\sim} \mathcal{N}\left(0, \frac{1}{\sqrt{n}}\right) \text{ and } \widehat{\tau}_n \overset{\perp}{\sim} \mathcal{N}\left(0, \frac{2}{3\sqrt{n}}\right).$$

Since n^* is itself asymptotically normally distribution, the asymptotic behavior of $\widehat{\rho}_n(u,u)$ and $\widehat{\tau}_n(u,u)$ can be obtained, as a mixture of Gaussian distributions.

Proposition 6.5.19. *Given $u \in (0,1)$, as $n \to \infty$,*

$$\mathbb{P}(\widehat{\rho}_n(u,u) \leq r) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} \phi\left(\sqrt{z-1}y\right) \phi\left(\frac{z-n(1-u)^2}{2\sqrt{nu(1-u)^2}}\right) dzdy,$$

and

$$\mathbb{P}(\widehat{\tau}_n(u,u) \leq r) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} \phi\left(\sqrt{z-1}y\right) \phi\left(\frac{(z-n(1-u)^2)}{3\sqrt{nu(1-u)^2}}\right) dzdy,$$

where ϕ denotes the density of the $\mathcal{N}(0,1)$ distribution.

Proof. This result is obtained through substitution. □

Example 6.5.20. *This asymptotic distribution (for $\widehat{\rho}_n(u,u)$ and $\widehat{\tau}_n(u,u)$) can be observe on Figure 6.5.6 where for* 50,000 *simulated n-samples of independent variables X and Y, when* $u = 0.5$.

Note that a more convenient approximation can be obtained, assuming that that the limiting distribution is simply a Gaussian distribution, assuming that uncertainty arises from the estimation of dependence measures, not the number of observations n^*.

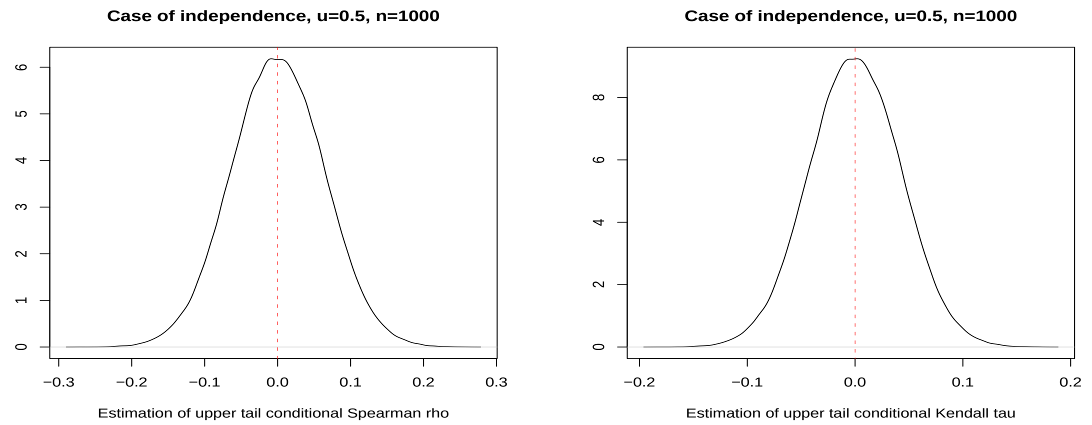


Figure 6.8: Distribution of $\widehat{\rho}_n(u)$ and $\widehat{\tau}_n(u)$, for $u = 0.5$ and $n = 1000$ in the independent case.

Proposition 6.5.21. *Given $u \in (0,1)$, as $n \to \infty$,*

$$\mathbb{P}(\widehat{\rho}_n(u,u) \leq r) \stackrel{\perp}{=} \int_{-\infty}^{x} \phi\left(\sqrt{[n(1-u)^2]-1} \cdot y\right) dy$$

and

$$\mathbb{P}(\widehat{\tau}_n(u,u) \leq r) \stackrel{\perp}{=} \int_{-\infty}^{x} \phi\left(\frac{3}{2}\sqrt{[n(1-u)^2]-1} \cdot y\right) dy$$

where ϕ denotes the density of the $\mathcal{N}(0,1)$ distribution.

Proof. This results is obtained substituting a Dirac at point $\mathbb{E}(n^*) = n(1-u)^2$ to the Gaussian distribution, with mean $n(1-u)^2$. $\square$

Example 6.5.22. *Distribution of estimation those estimators can be seen on Figure 6.5.6, for $\widehat{\rho}_n(u)$ and $\widehat{\tau}_n(u)$, when $u = 0.5$ or $u = 0.9$ and a sample size $n = 250$. The plain line is the "exact" distribution, obtained using $1,000,000$ Monte Carlo simulations, and the dotted line the asymptotic Gaussian approximation (and the mixture of Gaussian, which is the "exact asymptotic" distribution). Not that this if this approximation is good for small u, the closer to 1 is u, the worst is the approximation. When u is close to 1, n^* is more and more volatile (compared with its average).*

Note on those graphs that tails are underestimated, and thus, confidence intervals are too tight. an alternative is to use the two first moments of $\widehat{\rho}_n(u,u)$ and $\widehat{\tau}_n(u,u)$. Respectively,

$$\mathbb{E}(\widehat{\rho}_n(u,u)) = \mathbb{E}(\mathbb{E}(\widehat{\rho}_n(u,u)|n^*)) \to 0 \text{ as } n \to \infty$$

$$\mathbb{E}(\widehat{\tau}_n(u,u)) = \mathbb{E}(\mathbb{E}(\widehat{\tau}_n(u,u)|n^*)) \to 0 \text{ as } n \to \infty.$$

For the variance,

$$\begin{aligned} Var(\widehat{\rho}_n(u,u)) &= Var(\mathbb{E}(\widehat{\rho}_n(u,u)|n^*)) + \mathbb{E}(Var(\widehat{\rho}_n(u,u)|n^*)) \\ &\to Var(0) + \mathbb{E}\left(\frac{1}{\sqrt{n^*}}\right), \end{aligned}$$

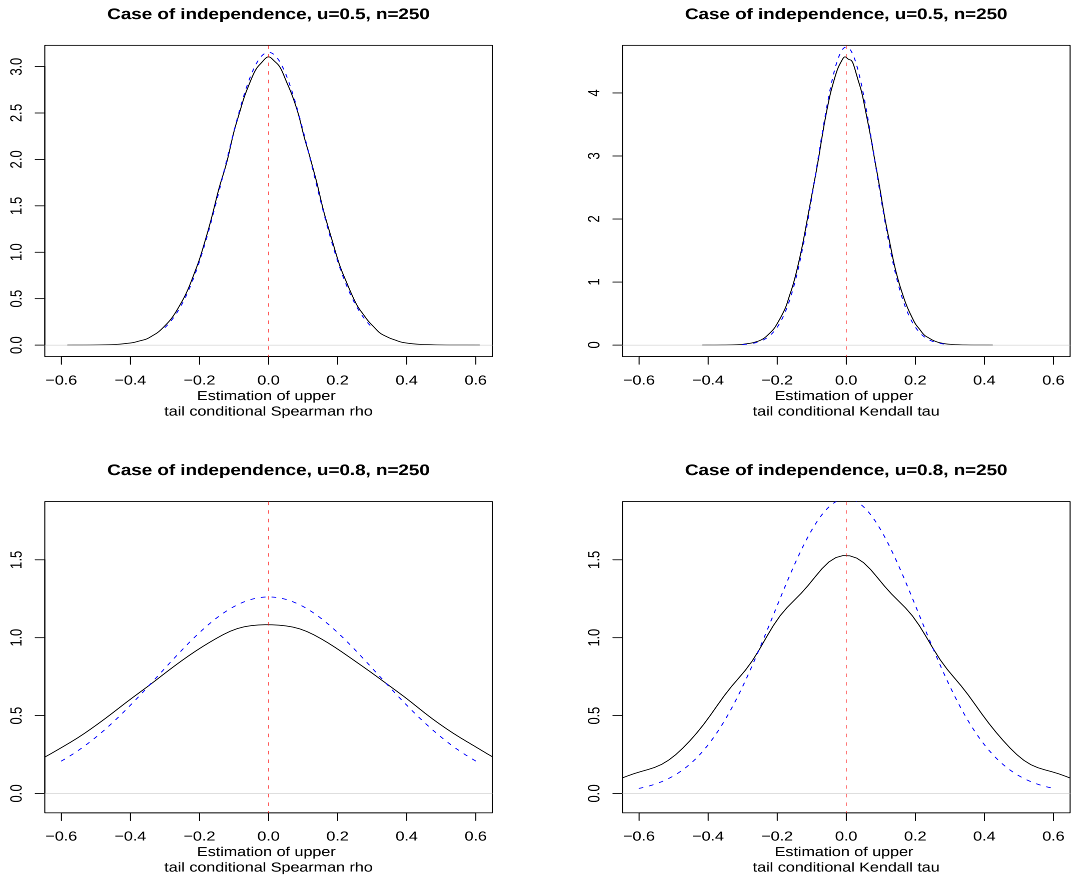


Figure 6.9: Distribution of estimation of $\widehat{\rho}_n(u)$, on the right, and $\widehat{\tau}_n(u)$, on the left, when $u = 0.5$ on top and $u = 0.8$ below, when $n = 250$. The plain line is the "*exact*" distribution, and the dotted line the asymptotic Gaussian approximation.

$$\begin{aligned} Var(\widehat{\tau}_n(u,u)) &= Var(\mathbb{E}(\widehat{\tau}_n(u,u)|n^*)) + \mathbb{E}(Var(\widehat{\tau}_n(u,u)|n^*)) \\ &\to Var(0) + \mathbb{E}\left(\frac{2}{3\sqrt(n^*)}\right). \end{aligned}$$

Further, from Proposition **??**, this expected value can be approximated as follows,

$$\mathbb{E}\left(\sqrt{n^*}^{-1}\right) = \sum_{k\in\mathbb{N}} k^{-1/2}\mathbb{P}(n^*=k) = \sum_{k=0}^{m} k^{-1/2}\binom{m}{k}(1-u)^k u^{m-k}, \tag{6.42}$$

for any $u \in [0,1)$ and $m = [(n+1)(1-u)]$.

Testing for tail independence

An idea to test tail independence is to test whether , given an n sample $(X_1, Y_1), ..., (X_n, Y_n)$, $\widehat{\rho}_n(u)$ or $\widehat{\tau}_n(u)$ are "*closed*" to 0. From the Gaussian approximation, this closeness notion can be explicitly written in terms of confidence interval.

A natural test of tail independence is derived as follows. X and Y are tail independent if and only if

$$-u_{1-\alpha/2}\sqrt{n(1-u)^2} \leq \widehat{\rho}_n(u) \leq +u_{1-\alpha/2}\sqrt{n(1-u)^2}, \text{ for } u \text{ close to } 1,$$

or

$$-u_{1-\alpha/2}\frac{3}{2}\sqrt{n(1-u)^2} \leq \widehat{\tau}_n(u) \leq +u_{1-\alpha/2}\frac{3}{2}\sqrt{n(1-u)^2}, \text{ for } u \text{ close to } 1,$$

where $u_{1-\alpha/2}$ denotes the quantile of order $1-\alpha/2$ of the $\mathcal{N}(0,1)$ distribution.

Example 6.5.23. *Figure 6.5.6 shows graphically the results of this testing procedure, for simulated samples of size $n = 10,000$, based on Spearman's rho and Kendall's tau (respectively on the left, and on the right), with an independent sample, with survival Clayton, and two Gaussian samples, $r = 0.2$ and $r = 0.7$ (from top to bottom).*

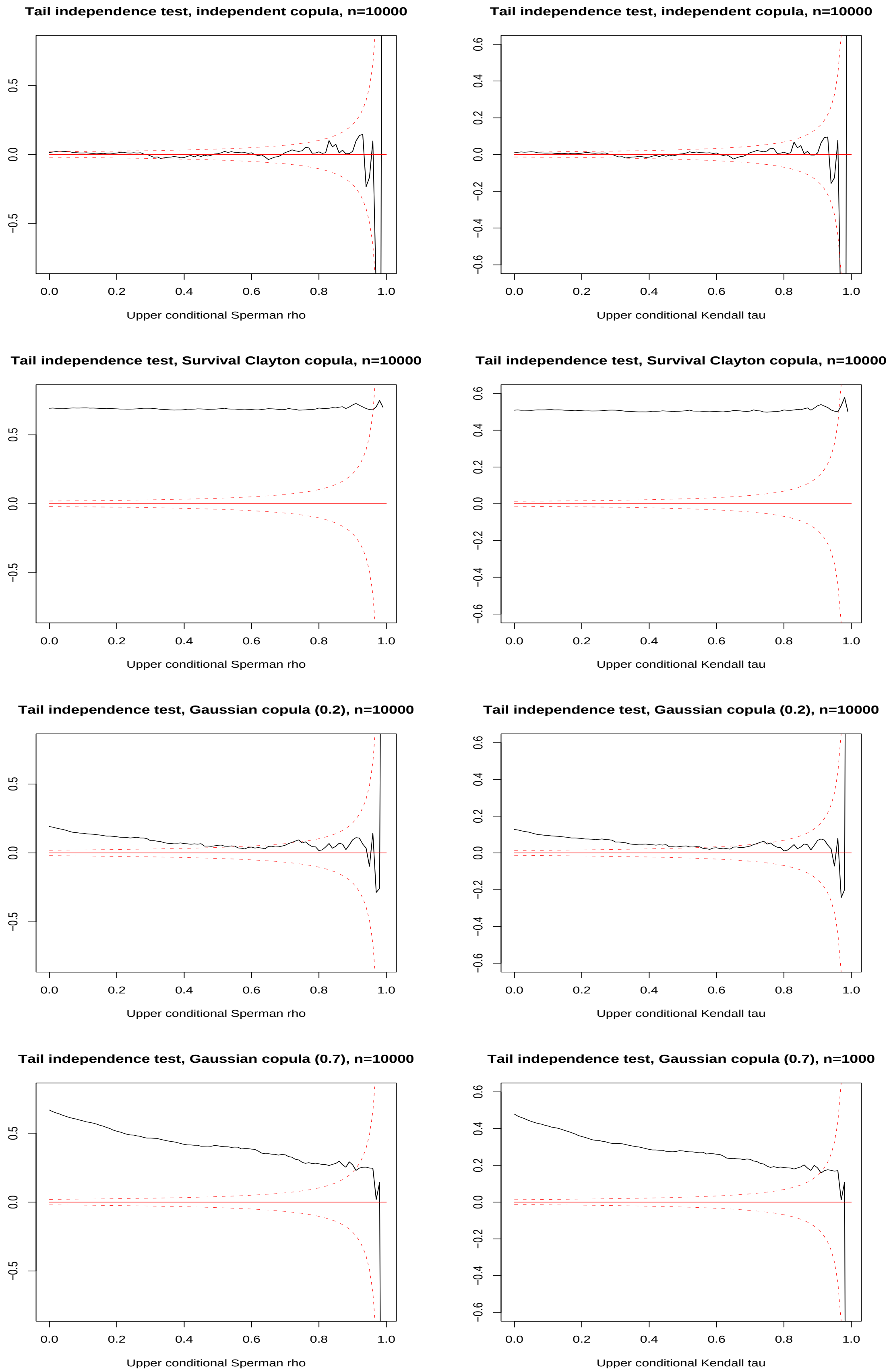


Figure 6.10: Evolution of $\widehat{\rho}_n(u)$, on the right, and $\widehat{\tau}_n(u)$ as functions of u, for simulated samples of size $n = 10,000$, with confidence intervals for $\widehat{\tau}_n(u)$ and $\widehat{\tau}_n(u)$ under the assumption of independence. From top to bottom are simulated an independent sample, survival Clayton, and two Gaussian samples, $r = 0.2$ and $r = 0.7$

Chapter 7

Nonparametric estimation of copulae density

7.1 Introduction

copulae are a way of formalizing dependence structures of random vectors. Although they are known for a long time (Sklar (1959)), they have been rediscovered relatively recently in applied sciences (biostatistics, reliability, biology etc). In finance, they have become a standard tool with broad applications: multi-asset pricing (especially complex credit derivatives), credit portfolio modeling, risk management, etc. For instance, see Li (1999), Patton (2001), or Longin and Solnik (2001), among others.

Although the concept of copulae is well understood, it is now recognized that their empirical estimation is a harder and trickier task. Lots of traps and technical difficulties are present, and these are most of the time ignored or underestimated by practitioners. The problem is that the estimation of copulae implies usually that every marginal distribution of the underlying random vectors must be evaluated and plugged into an estimated multivariate distribution. Such a procedure produces unexpected and unusual effects w.r.t. the usual statistical procedures: non-standard limiting behaviors, noisy estimations, etc (see the discussion in Fermanian and Scaillet (2005), e.g.).

In this chapter, we focus on the practical issues practitioners are faced with, especially concerning estimation and visualization. In the first section, we expose a general setting for the estimation of copulae. Such a framework nests most of the available techniques. In the second section, we deal with the estimation of the copula density itself, with a particular focus on estimation near the boundaries of the unit square.

7.2 A general approach for the estimation of copula functions

copulae involve several underlying functions: the marginal cumulative distribution functions and a joint cdf. To estimate copula functions, the first issue consists in specifying how to estimate separately the margins and the joint law. Moreover, some of these functions can be fully known. Depending on the assumptions some quantities have to be estimated parametrically, semi- or even non-parametrically. In the latter case, the practitioner has to choose between the usual methodology of using "empirical counterparts" and invoking smoothing methods well-known in statistics: kernels, wavelets, orthogonal polynomials, nearest neighbors, etc.

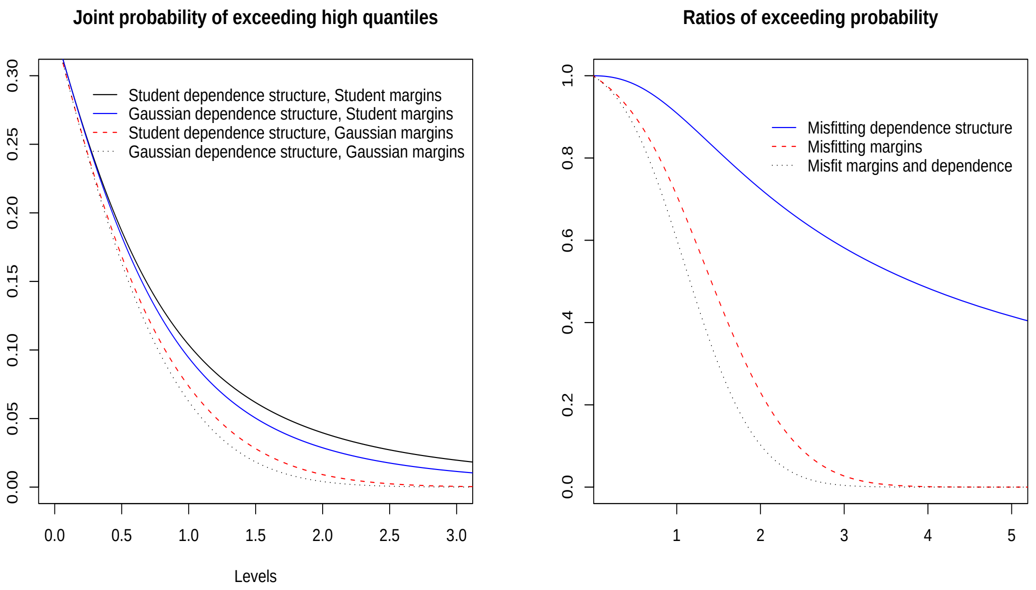


Figure 7.1: Function χ when (X, Y) is a Student random vector, and when either margins or the dependence structure are misspecified. The associated ratios of exceeding probability correspond to the χ function obtained for the misspecified model vs. the true χ (for the true Student model).

Obviously, the estimation precision and the graphical results are functions of all these choices. A true known marginal can improve a lot the results under well-specification, but the reverse is true under misspecification (even under a light one). Without any valuable prior information, non-parametric estimation should be favored, especially for margin estimation.

To illustrate this point Figure 7.1 shows the graphical behaviour of the exceeding probability function

$$\chi : p \mapsto \mathbb{P}\left(X > F_X^{\leftarrow}(p), Y > F_Y^{\leftarrow}(p)\right)$$

. If the true underlying model is a multivariate Student vector (X, Y), the associated probability is the upper line. If either marginal distributions are misspecified (e.g. Gaussian marginal distributions), or the dependence structure is misspecified (e.g. joint Gaussian distribution), these probabilities are always underestimated, especially in the tails.

Now, let us introduce our framework formally. Consider the estimation of a d-dimensional copula C, that can be written

$$C(\boldsymbol{u}) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right).$$

Obviously, all the marginal cdfs have been denoted by F_k, $k = 1, \ldots, d$, when the joint cdf is F. Along this chapter, the inverse operator $^{\leftarrow}$ should be understood as a generalized inverse, namely that for every function G,

$$G^{\leftarrow}(x) = \inf\{y | G(y) \geq x\}.$$

Assume we have observed a T-sample $(\boldsymbol{X}_t)_{t=1,\ldots,T}$. They are some realizations of the d-random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$. Note that we do not assume that $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{dt})$ are mutually independent (at least for the moment).

Every marginal cdf, say the k-th can be estimated empirically by

$$F_k^{(1)}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(X_{kt} \leq x),$$

and $[F_k^{(1)}]^{\leftarrow}(u_k)$ is simply the empirical quantile corresponding to $u_k \in [0,1]$. Another way is to smooth such cdfs, and the simplest way is to invoke the kernel method (see, e.g., Härdle and Linton (1994) or Pagan and Ullah (1999) for an introduction): consider a univariate kernel function $K : \mathbb{R} \longrightarrow \mathbb{R}$, $\int K = 1$, and a bandwidth sequence h_T (or simpler h hereafter), $h_T > 0$ and $h_T \longrightarrow 0$ when $T \to \infty$. Then, $F_k(x)$ can be estimated by

$$F_k^{(2)}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{K}\left(\frac{x - X_{kt}}{h}\right),$$

for every real number x, by denoting $\mathbb{K}$ the primitive function of K: $\mathbb{K}(x) = \int_{-\infty}^{x} K$.

There exists another common case: assume that an underlying parametric model has been fitted previously for the k-th margin. Then, the natural estimator for $F_k(x)$ is some cdf $F_k^{(3)}(x, \hat{\theta}_k)$ that depends on the relevant estimated parameter $\hat{\theta}_k$. When such a model is well-specified, $\hat{\theta}_k$ is tending almost surely to a value θ_k such that $F_k(\cdot) = F_k^{(3)}(\cdot, \theta_k)$. The last limiting case is the knowledge of the true cdf F_k. Formally, we will set $F_k^{(0)} = F_k$.

Similarly, the joint cdf F can be estimated empirically by

$$F^{(1)}(\boldsymbol{x}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(\boldsymbol{X}_t \leq \boldsymbol{x}),$$

or by the kernel method

$$F^{(2)}(\boldsymbol{x}) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{K}\left(\frac{x - \boldsymbol{X}_t}{h}\right),$$

with a d-dimensional kernel $\mathbb{K}$, so that

$$\mathbb{K}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} K,$$

for every $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. Besides, there may exist an underlying parametric model for $\boldsymbol{X}$: F is assumed to belong to a set of multivariate cdfs indexed by a parameter τ. A consistent estimation $\hat{\tau}$ for the "true" value τ allows setting $F^{(3)}(\cdot) = F(\cdot, \hat{\tau})$. Finally, we can denote $F^{(0)} = F$.

Therefore, generally speaking, a d-dimensional copula C can be estimated by

$$\hat{C}(\boldsymbol{u}) = F^{(j)}\left([F_1^{(j_1)}]^{\leftarrow}(u_1), \dots, [F_d^{(j_d)}]^{\leftarrow}(u_d)\right), \tag{7.1}$$

for every indices $j, j_1, j_2, \dots, j_d$ that belong to $\{0, 1, 2, 3\}$. Thus, it is not so obvious to discriminate between all these competitors, especially without any parametric assumption.

Every estimation method has its own advantages and drawbacks. The full empirical method ($j = j_1 = \dots = j_d = 1$ with the notations of Equation (7.1)) has been introduced in Deheuvels (1979, 1981a, 1981b) and studied more recently by Fermanian et al. (2004), in the independent case, and by Doukhan et al. (2004) in a dependent framework. It provides a robust and universal way for estimation purposes. Nonetheless, its discontinuous feature induces some difficulties: the graphical representations of the copula can be unpleasant and not intuitive. Moreover, there is no unique choice for building the inverse function of $F_k^{(1)}$. In particular, if $X_{k1} \leq \ldots \leq X_{kT}$ is the ordered sample on the k-axis, then the inverse function of $F_k^{(1)}$ at some point t/T may be chosen arbitrarily between X_{kt} and $X_{k(t+1)}$. Finally, since the copula estimator is not differentiable when only one empirical cdf is involved in Definition (7.1), it cannot be used straightforwardly to get an estimate of the associated copula density (by differentiation of $\hat{C}(\boldsymbol{u})$ with respects to all its arguments) or for optimization purposes, for instance.

Smooth estimators are nicer for graphical purposes, and can provide more easily the intuition for getting the "true" underlying parametric distribution. However, they depend on an auxiliary smoothing parameter (h in the case of the kernel method, e.g.), and suffer from the well-known "curse of dimensionality": the higher the dimension (d with our notations), the worse the performance in terms of convergence rates. In other words, as the dimension increases, the complexity of the problem increases exponentially [1]. Such methods can be invoked safely in practice when $d \leq 3$ and for samples sizes larger than a couple of hundreds of observations (which is usual in finance). The theory of fully smoothed copulae ($j = j_1 = \ldots = j_d = 2$ with the notations of Definition (7.1)) can be found in Fermanian and Scaillet (2003) in a strongly dependent framework.

A more comfortable situation exists when "good" parametric assumptions are put into (7.1), for the marginal cdfs and/or the joint cdf F. The former case is relatively usual because there exist a lot of univariate models for financial variables (see Alexander (2002), for instance). Nevertheless, for a lot of dynamic models (stochastic volatility models, e.g.), their (unconditional) marginal cdfs cannot be written explicitly. And, obviously, we are under the threat of a misspecification, that can have disastrous effects (see Fermanian and Scaillet, 2005). Concerning a parametric assumption for F itself, our opinion is balanced. At first glance, we are absolutely free for choosing an "interesting" parametric family $\mathcal{F}$ of d-dimensional cdfs and that would contain the true law F. But, by setting for every real number x and every $k = 1, \ldots, d$

$$F_k^{(3)}(x) = F(+\infty, \ldots, +\infty, x, +\infty, \ldots |\hat{\tau}),$$

where x is the k-th argument of F, we should have found the "right" marginal distributions too, to be coherent with ourselves. Indeed, the joint law contains the marginal ones. Then, the estimated copula should be

$$\hat{C}(\boldsymbol{u}) = F^{(3)}\left([F_1^{(3)}]^{\leftarrow}(u_1), [F_2^{(3)}]^{\leftarrow}(u_2), \ldots, [F_d^{(3)}]^{\leftarrow}(u_d)\right).$$

Actually, the problem is really to find a sufficiently rich family $\mathcal{F}$ ex ante, that might generate all empirical features. What people do is more clever. They choose a parametric family $\mathcal{F}^*$ and other marginal parametric families $\mathcal{F}_k^*$, $k = 1, \ldots, d$, and set

$$C(\boldsymbol{u}) = \hat{F}^*\left([\hat{F}_1^*]^{\leftarrow}(u_1), \ldots, [\hat{F}_d^*]^{\leftarrow}(u_d)\right)$$

for some $\hat{F}^* \in \mathcal{F}$, and $\hat{F}_k^* \in \mathcal{F}_k$ for every $k = 1, \ldots, d$. Note that the choice of all the parametric families is absolutely free of constraints, and these families are not related to each others (they can be arbitrary and independently chosen). It is the usual way of generating new copula families. The price to be paid is that the true joint law F does not belong to $\mathcal{F}^*$ generally speaking. Similarly, the true marginal laws F_k do not belong to the sets $\mathcal{F}_k^*$ in general.

[1]For instance, the number of histogram grid cells increases exponentially. This effect cannot be avoided, even by other estimation methods. Under smoothness assumptions on the density, the amount of training data required for nonparametric estimators increases exponentially with the dimension (see e.g. Stone (1980)).

When a parametric assumption is done in such a case, the standard estimation procedure is semi-parametric: the copula is a function of some parameter $\theta = (\tau, \theta_1, \ldots, \theta_d)$. Recall that the copula density c is the derivative of C w.r.t. each of its arguments:

$$c_\theta(\boldsymbol{u}) = \frac{\partial^d}{\partial_1 \ldots \partial_d} C(\boldsymbol{u}).$$

Here, the copula density c_θ itself can be calculated under a full parametric assumption. Thus, we get an estimator of θ by maximizing the log-likelihood

$$\sum_{t=1}^{T} \log c_\theta \left(\widehat{F}_1(X_{1t}), \ldots, \widehat{F}_d(X_{dt}) \right),$$

for some $\sqrt{T}$-convergent estimates $\widehat{F}_k(X_{kt})$ of the marginal cdfs. Obviously, we may choose $\widehat{F}_k = F_k^{(1)}$ or $F_k^{(2)}$.

Note that such an estimator is called an "omnibus estimator", and it can be seen as a maximum likelihood estimator of θ after replacing the unobservable ranks $F_k(X_{kt})$ by the pseudo observations. The asymptotic distribution of the estimator was studied in Genest et al. (1995) and Shi and Louis (1995). The main interest of semi-parametric estimation is to avoid possible misspecification of marginal distributions, which may over-estimate the degree of dependence in the data (see e.g. Silvapulle, Kim and Silvapulle (2004)). Note finally that Chen and Fan (2004a, 2004b) have developed the theory of this semi-parametric estimator in a time-series context.

Thus, depending on the degree of assumptions about the joint and marginal models, there exists a wide range of possibilities for estimating copula functions as provided by Equation (7.1). The only trap is to be sure that the assumptions done for margins are consistent with those done for the joint law. The statistical properties of all these estimators are the usual ones, namely consistency and asymptotic normality.

7.3 The estimation of copula densities

After the estimation of C by $\hat{C}$ as in Equation (7.1), it is tempting to define an estimate of the copula density c at every $\boldsymbol{u} \in [0,1]^d$ by

$$\hat{c}(\boldsymbol{u}) = \frac{\partial^d}{\partial_1 \ldots \partial_d} \hat{C}(\boldsymbol{u}).$$

Unfortunately, this works only when $\hat{C}$ is differentiable. Most of the time, this is the case when the marginal and joint cdfs are parametric or nonparametrically smoothed (by the kernel method, for instance). In the latter case and when d is "large" (more than 3), the estimation of c can be relatively poor because of the curse of dimensionality.

Nonparametric estimation procedures for the density of a copula function have already been proposed by Behnen, Huskova, and Neuhaus (1985) or Gijbels and Mielniczuk (1990). These procedures rely on symmetric kernels, and have been detailed in the context of uncensored data. Unfortunately, such techniques are not consistent on the boundaries of $[0,1]^d$. They suffer from the so-called "boundary bias". Such bias can be significant in the neighborhood of the boundaries too, depending on the size of the bandwidth. Hereafter, we will propose some solutions to cope with such an issue. To ease notations and without a lack of generality, we will restrict ourselves to the bivariate case ($d = 2$). Thus, our random vector will be denoted by (X, Y) instead of (X_1, X_2).

In the following sections, we will study some properties of some kernel-based estimators, and illustrate some of these by simulations. The benchmark will be a simulated sample, whose size is $T = 1,000$ and that will be generated by a Frank copula with copula density

$$c^{Fr}(u, v, \theta) = \frac{\theta[1 - e^{-\theta}]e^{-\theta(u+v)}}{([1 - e^{-\theta}] - (1 - e^{-\theta u})(1 - e^{-\theta v}))^2},$$

and Kendall's tau equal to 0.5. Hence, the copula parameter is $\theta = 5.74$. This density can be seen on Figure 7.2 together with its contour plot on the right.

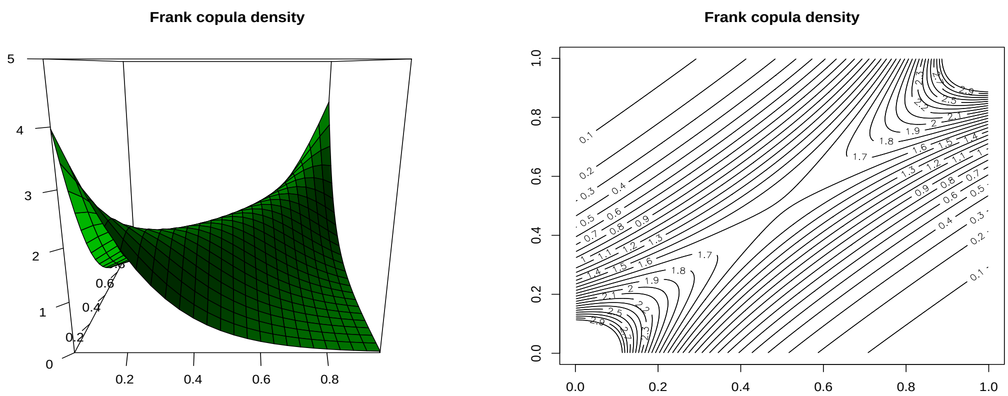


Figure 7.2: Density of the Frank copula with a Kendall tau equal to 0.5.

7.3.1 Nonparametric density estimation for distributions with finite support

A first approach relies on a kernel based estimation of the density based on the pseudo-observations $(F_{X,T}(X_i), F_{Y,T}(Y_i))$, where $F_{X,T}$ and $F_{Y,T}$ are the empirical distribution functions,

$$F_{X,T}(x) = \frac{1}{T+1}\sum_{i=1}^{T} \mathbf{11}(X_i \leq x) \text{ and } F_{Y,T}(y) = \frac{1}{T+1}\sum_{i=1}^{T} \mathbf{1}(Y_i \leq y),$$

where the factor $T+1$ (instead of standard T, as in Deheuvels (1979) for instance) allows avoiding boundary problems: the quantities $F_{X,T}(X_i)$ and $F_{Y,T}(Y_i)$ are the ranks of the X_i's and the Y_i's divided by $T+1$, and therefore take values

$$\left\{\frac{1}{T+1}, \frac{2}{T+1}, ..., \frac{T}{T+1}\right\}.$$

Standard kernel-based estimators of the density of pseudo-observations yield, using diagonal bandwidth (see Wand and Jones (1995))

$$\widehat{c}_h(u, v) = \frac{1}{Th^2}\sum_{i=1}^{T} K\left(\frac{u - F_{X,T}(X_i)}{h}, \frac{v - F_{Y,T}(Y_i)}{h}\right),$$

for a bivariate kernel $K : \mathbb{R}^2 \longrightarrow \mathbb{R}$, $\int K = 1$.

The variance of the estimator can be derived, and is $\mathcal{O}((Th^2)^{\leftarrow})$. Moreover, it is asymptotically normal at every point $(u,v)\in(0,1)$:

$$\frac{\widehat{c}_h(u,v)-\mathbb{E}(\widehat{c}_h(u,v))}{\sqrt{Var(\widehat{c}_h(u,v))}}\xrightarrow{\mathcal{L}}\mathcal{N}(0,1).$$

As a benchmark, Figure 7.2 shows the theoretical density of a Frank copula. We plot on Figure 7.3 the standard Gaussian kernel estimator based on the sample of pseudo-observations $(\widehat{U}_i,\widehat{V}_i)\equiv(F_{X,T}(X_i),F_{Y,T}(Y_i))$.

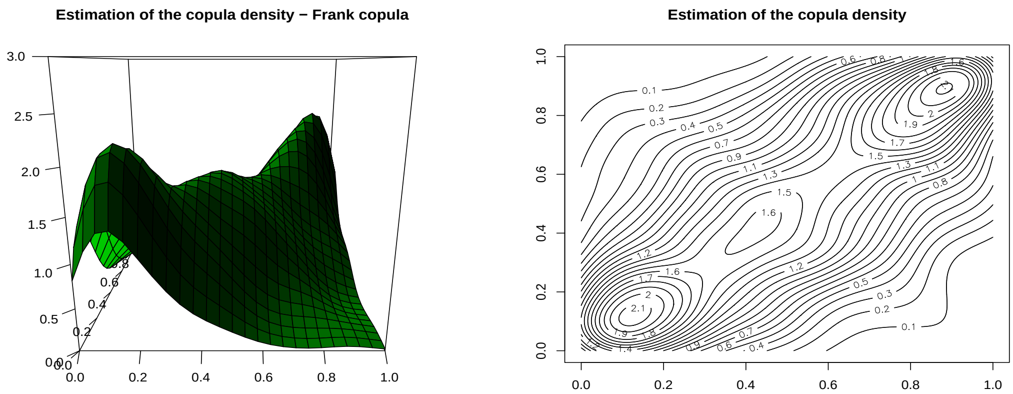


Figure 7.3: Estimation of the copula density using a Gaussian kernel based with 1,000 observations drawn from a Frank copula.

Recall that even if kernel estimates are consistent for distributions with unbounded support, the boundary bias when support is bounded can yield some ill underestimation (even if the distribution is twice differentiable in the interior of its support).

We can explain this phenomenon easily in the univariate case. Consider a T sample $X_1,...,X_T$ of a positive random variable with density f. The support of their density is then $\mathbb{R}^+$. Let K denote a symmetric kernel, whose support is $[-1,+1]$. Then, for all $x\geq 0$, using a Taylor expansion, we get

$$\begin{aligned}\mathbb{E}(\widehat{f}_h(x)) &= \int_{-1}^{x/h} K_h(y)f(x-hy)dy\\ &= f(x)\cdot\int_{-1}^{x/h} K_h(y)dy\\ &\qquad -h\cdot f'(x)\cdot\int_{-1}^{x/h} zK_h(y)dy+O(h^2).\end{aligned}$$

Hence, since the kernel is symmetric, $\int_{-1}^{x/h}K_h(y)dy\xrightarrow{h\to 0}1/2$ when $x=0$, and therefore,

$$\mathbb{E}(\widehat{f}_h(0))=\frac{1}{2}f(0)+O(h).$$

Note that, if $x>0$, the expression $\int_{-1}^{x/h}K_h(y)dy$ is 1 when h is sufficiently small (when $x>h$ to be specific). Thus, this integral cannot be one uniformly w.r.t. every $x\in(0,1]$. And for more general kernels, it has no reason to be equal to 1. In the latter case, since this expression can be calculated, normalizing $\widehat{f}_h(x)$ by dividing by $\int_{-1}^{x/h} K(z)dz$ (at each x) achieves consistency. Nonetheless, it remains a bias that is of order $O(h)$. Using some boundary kernels (see Gasser and Müller (1979)), it is possible to achieve $O(h^2)$ everywhere in the interior of the support.

Consider the case of variables uniformly distributed on $[0,1]$, $U_1,...,U_n$. Figure 7.4 shows kernel-based estimators of the uniform density, with Gaussian kernel and different bandwidth s, with $n=100$ simulated variables on the left, $n=1,000$ simulated variables on the right. In that case, for any $h>0$,

$$\mathbb{E}(\widehat{f}_h(0)) = \int_0^1 K_h(y)dy = \frac{1}{h\sqrt{2\pi}}\int_0^1 \exp\left(-\frac{y^2}{2h^2}\right)dy \xrightarrow{h\to 0} \frac{1}{2} = \frac{f(0)}{2},$$

and in the interior, i.e. $x\in(0,1)$

$$\begin{aligned}\mathbb{E}(\widehat{f}_h(x)) &= \int_0^1 K_h(y-x)dy \\ &= \frac{1}{h\sqrt{2\pi}}\int_0^1 \exp\left(-\frac{(y-x)^2}{2h^2}\right)dy \xrightarrow{h\to 0} 1 = f(x).\end{aligned}$$

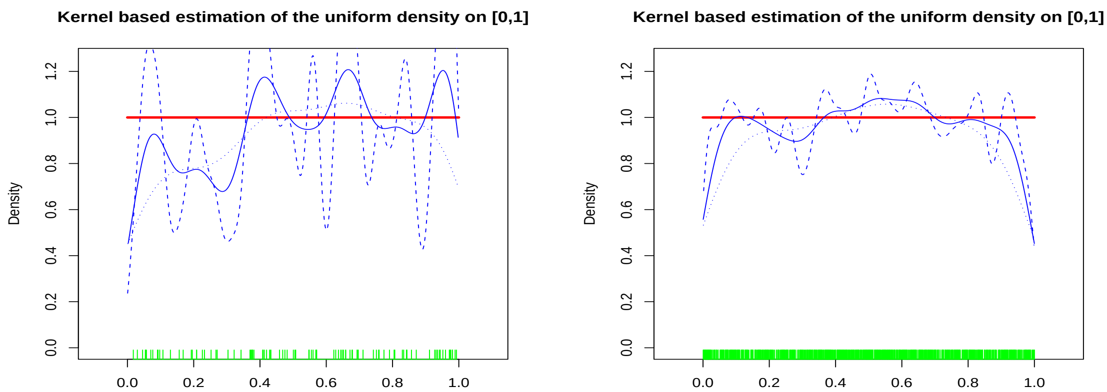


Figure 7.4: Estimation of the uniform density using a Gaussian kernel and different bandwidth with $n=100$ and $1,000$ observations.

Dealing with bivariate copula densities, we observe the same phenomenon. On boundaries, we obtain some "multiplicative bias", $1/4$ in corners and $1/2$ in the interior of borders. The additional bias is of order $O(h)$ on the frontier, and standard $O(h^2)$ in the interior. More precisely, in any corners (e.g. $(0,0)$)

$$\mathbb{E}(\widehat{c}_h(0,0)) = \frac{1}{4}c(u,v)+O(h).$$

on the interior of the borders (e.g. $u=0$ and $v\in(0,1)$),

$$\mathbb{E}(\widehat{c}_h(0,v)) = \frac{1}{2}c(u,v)+O(h).$$

and in the interior ($(u,v)\in(0,1)\times(0,1)$),

$$\mathbb{E}(\widehat{c}_h(u,v)) = c(u,v)+O(h^2).$$

The bias can be observed on Figure 7.5 which represents the diagonal of the estimated density for several samples.

Several techniques have been introduced to get a better estimation on the borders, for univariate densities:

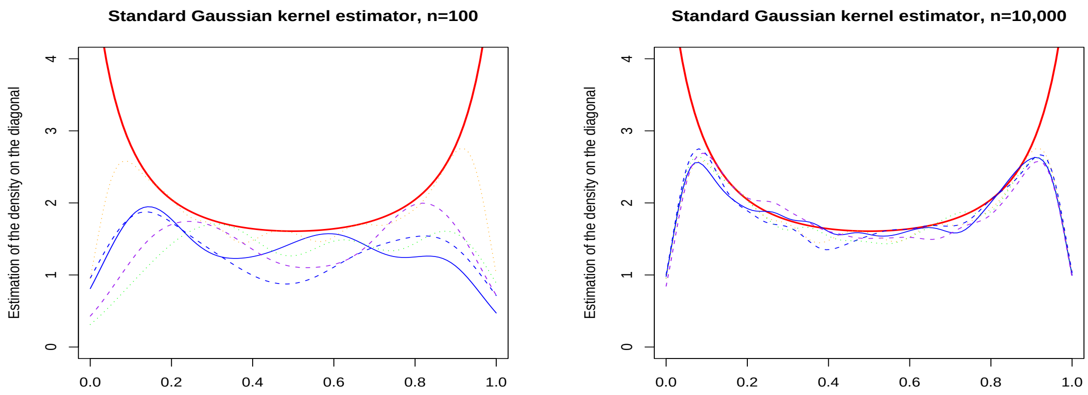


Figure 7.5: Estimation of the copula density on the diagonal using a (standard) Gaussian kernel with 100 and 10,000 observations drawn from a Frank copula.

- mirror image modification (Deheuvels and Hominal (1989), Schuster (1985)), where artificial data are obtained, using symmetric (mirror) transformations on the borders;
- transformed kernels (Devroye and Györfi (1985), Wand, Marron and Ruppert (1991)), where the idea is to transform the data X_i using a bijective mapping ϕ so that $\phi(X_i)$'s have support $\mathbb{R}$. Efficient kernel based estimation of the density of the $\phi(X_i)$'s can be derived, and, by the inverse transformation, we get back the density estimation of the X_i's themselves;
- boundary kernels (Gasser and Müller (1979), Rice (1984), Müller (1991)), where a smooth distortion is considered near the border, so that the bandwidth and the kernel shape can be modified (the closer to the border, the smaller).

Finally, the last section will briefly mention the impact of pseudo-observations, i.e. working on samples

$$\{(F_{X,T}(X_1), F_{Y,T}(Y_1)), ..., (F_{X,T}(X_T), F_{Y,T}(Y_T))\},$$

instead of

$$\{(F_X(X_1), F_Y(Y_1)), ..., (F_X(X_T), F_Y(Y_T))\},$$

as if we know the true marginal distributions.

7.3.2 Mirror image

The idea of this method, developed by Deheuvels and Hominal (1979) and Schuster (1985), is to add some "*missing mass*" by reflecting the sample w.r.t. the boundaries. They focus on the case where variables are positive, i.e. whose support is $[0, \infty)$. Formally and in its simplest form, it means replacing $K_h(X_i - x)$ by $K_h(x + X_i) + K_h(X_i - x)$. The estimator of the density is then

$$\widehat{f}_h(x) = \frac{1}{Th} \sum_{i=1}^{T} \left\{ K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + X_i}{h}\right) \right\}.$$

In the case of densities whose support is $[0, 1] \times [0, 1]$, the non-consistency can be corrected on the boundaries, but the convergence rate of the bias will remain $O(h)$ on the boundaries which is larger than the usual rate $O(h^2)$ obtained in the interior if $h \to 0$. The only case where the usual rate of convergence is obtained on boundaries is when the derivative of the density is zero on such subsets. Note that the variance is 4 times higher in corners and 2 times higher in the interior of borders.

For copulae, instead of using only the "pseudo-observations" $(\widehat{U}_i, \widehat{V}_i) \equiv (F_{X,T}(X_i), F_{Y,T}(Y_i))$, the mirror image consists in reflecting each data point with respect to all edges and corners of the unit square $[0,1] \times [0,1]$. Hence, additional observations can be considered, i.e. the $(\pm\widehat{U}_i, \pm\widehat{V}_i)$'s, the $(\pm\widehat{U}_i, 2 - \widehat{V}_i)$'s, the $(2 - \widehat{U}_i, \pm\widehat{V}_i)$'s and the $(2 - \widehat{U}_i, 2 - \widehat{V}_i)$'s. Hence, consider

$$
\begin{aligned}
&\hat{c}_h(u,v) \\
= \; & \frac{1}{Th^2} \sum_{i=1}^{T} \left\{ K\left(\frac{u - \widehat{U}_i}{h}\right) K\left(\frac{v - \widehat{V}_i}{h}\right) + K\left(\frac{u + \widehat{U}_i}{h}\right) K\left(\frac{v - \widehat{V}_i}{h}\right) + \right. \\
& K\left(\frac{u - \widehat{U}_i}{h}\right) K\left(\frac{v + \widehat{V}_i}{h}\right) + K\left(\frac{u + \widehat{U}_i}{h}\right) K\left(\frac{v + \widehat{V}_i}{h}\right) + \\
& K\left(\frac{u - \widehat{U}_i}{h}\right) K\left(\frac{v - 2 + \widehat{V}_i}{h}\right) + K\left(\frac{u + \widehat{U}_i}{h}\right) K\left(\frac{v - 2 + \widehat{V}_i}{h}\right) + \\
& K\left(\frac{u - 2 + \widehat{U}_i}{h}\right) K\left(\frac{v - \widehat{V}_i}{h}\right) + K\left(\frac{u - 2 + \widehat{U}_i}{h}\right) K\left(\frac{v + \widehat{V}_i}{h}\right) + \\
& \left. K\left(\frac{u - 2 + \widehat{U}_i}{h}\right) K\left(\frac{v - 2 + \widehat{V}_i}{h}\right) \right\}.
\end{aligned}
$$

Figure 7.6 has been obtained using the reflection principle. We can check that the fit is far better than in Figure 7.3.

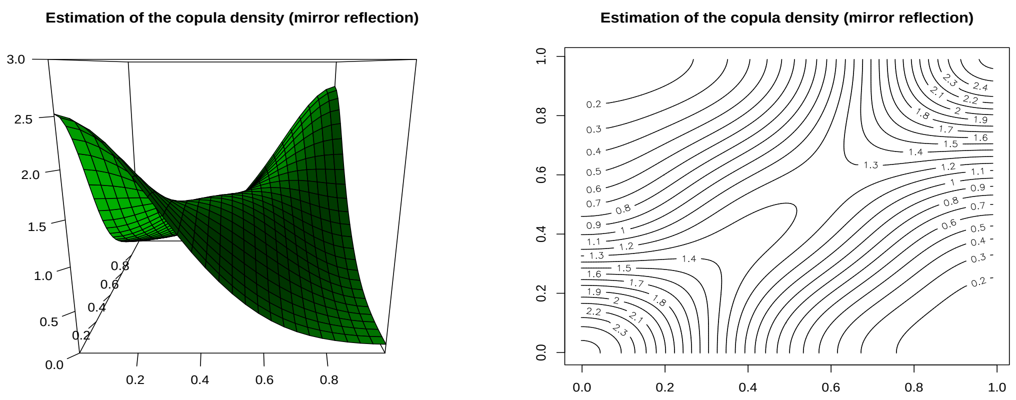


Figure 7.6: Estimation of the copula density using a Gaussian kernel and the mirror reflection principle with 1,000 observations from the Frank copula.

7.3.3 Transformed kernels

Recall that c is the density of (U,V), $U = F_X(X)$ and $V = F_Y(Y)$. The two latter rvs follow uniform distributions (marginally). Consider a distribution function G of a continuous distribution on $\mathbb{R}$, with differentiable strictly positive density g. We build new rvs $\tilde{X} = G^{\leftarrow}(U)$ and

$\tilde{Y} = G^{\leftarrow}(V)$. Then, the density of $(\tilde{X}, \tilde{Y})$ is

$$f(x, y) = g(x)g(y)c[G(x), G(y)]. \tag{7.2}$$

This density is twice continuously differentiable on $\mathbb{R}^2$, and the standard kernel approach applies.

Since we do not observe a sample of (U, V) but instead pseudo-observations $(\hat{U}_i, \hat{V}_i)$, we build an "approximated sample" of the transformed variables $(\tilde{X}_1, \tilde{Y}_1), ..., (\tilde{X}_T, \tilde{Y}_T)$, by setting $\tilde{X}_i = G^{\leftarrow}(\hat{U}_i)$ and $\tilde{Y}_i = G^{\leftarrow}(\hat{V}_i)$. Thus, the kernel estimator of f is

$$\hat{f}(x, y) = \frac{1}{Th^2} \sum_{i=1}^{T} K\left(\frac{x - \tilde{X}_i}{h}, \frac{y - \tilde{Y}_i}{h}\right). \tag{7.3}$$

The associated estimator of c is then deduced by inverting (7.2),

$$c(u, v) \quad = \quad \frac{f(G^{\leftarrow}(u), G^{\leftarrow}(v))}{g(G^{\leftarrow}(u))g(G^{\leftarrow}(v))}, \qquad (u, v) \in [0, 1] \times [0, 1],$$

and therefore we get

$$\begin{aligned} \widehat{c}_h(u, v) \quad = \quad & \frac{1}{Th^2 g(G^{\leftarrow}(u)) \cdot g(G^{\leftarrow}(v))} \\ & \sum_{i=1}^{T} K\left(\frac{G^{\leftarrow}(u) - G^{\leftarrow}(\hat{U}_i)}{h}, \frac{G^{\leftarrow}(v) - G^{\leftarrow}(\hat{V}_i)}{h}\right), \end{aligned}$$

Note that this approach can be extended by considering different transformations G_X and G_Y, different kernels K_X and K_Y, or different bandwidths h_X and h_Y, for the two marginal random variables.

Figure 7.7 has been obtained using the transformed kernel, where K was a Gaussian kernel and G was respectively the cdf of the $\mathcal{N}(0, 1)$ distribution.

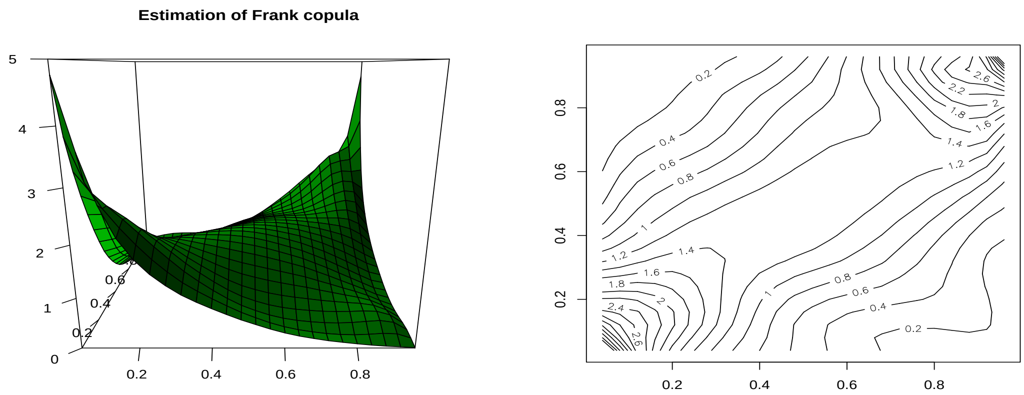


Figure 7.7: Estimation of the copula density using a Gaussian kernel and Gaussian transformations with 1,000 observations drawn from the Frank copula.

The absence of a multiplicative bias on the borders can be observed in Figure 7.8, where the diagonal of the copula density is plotted, based on several samples. The copula density estimator obtained with transformed samples has no bias, is asymptotically normal, etc. Actually, we get all the usual properties of the multivariate kernel density estimators.

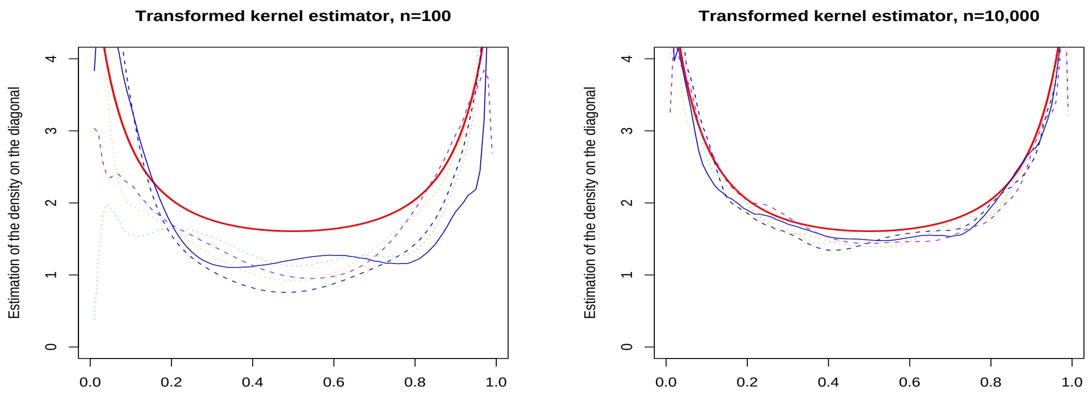


Figure 7.8: Estimation of the copula density on the diagonal using a Gaussian kernel and Gaussian transformations with 100 and 10,000 observations drawn from a Frank copula.

7.3.4 Beta kernels

In this section we examine the use of the beta kernel introduced by Brown and Chen (1999), and Chen (1999, 2000) for nonparametric estimation of regression curves and univariate densities with compact support, respectively.

Following an idea by Harrell and Davis (1982), Chen (1999, 2000) introduced the Beta kernel estimator as an estimator of a density function with known compact support $[0,1]$, to remove the boundary bias of the standard kernel estimator:

$$\widehat{f}_h(x) = \frac{1}{T}\sum_{i=1}^{T} K\left(X_i, \frac{x}{h}+1, \frac{1-x}{h}+1\right),$$

where $K(\cdot,\alpha,\beta)$ denotes the density of the Beta distribution with parameters α and β,

$$K(x,\alpha,\beta) = \frac{x^{\alpha}(1-x)^{\beta}}{B(\alpha,\beta)}, x \in [0,1], \text{ where } B(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

The main difficulty when working with this estimator is the lack of a simple "rule of thumb" for choosing the smoothing parameter h.

The beta kernel has two leading advantages. First it can match the compact support of the object to be estimated. Secondly it has a flexible form, and changes the smoothness in a natural way as we move away from the boundaries. As a consequence beta kernel estimators are naturally free of boundary bias, and can produce estimates with a smaller variance. Indeed we can benefit from a larger effective sample size since we can pool more data. Monte Carlo results available in these papers show that they have better performance compared to other estimators which are free of boundary bias, such as local linear (Jones (1993)) or boundary kernel (Müller (1991)) estimators. Renault and Scaillet (2004) also report better performance compared to transformation kernel estimators (Silverman (1986)). Besides Bouezmarni and Rolin (2001, 2003) show that the beta kernel density estimator is consistent even if the true density is unbounded at the boundaries. This feature may arise in our situation as well. For example the density of a bivariate Gaussian copula is unbounded at the corners $(0,0)$ and $(1,1)$. Therefore beta kernels are appropriate candidates to build well-behaved nonparametric estimators of the density of a copula function.

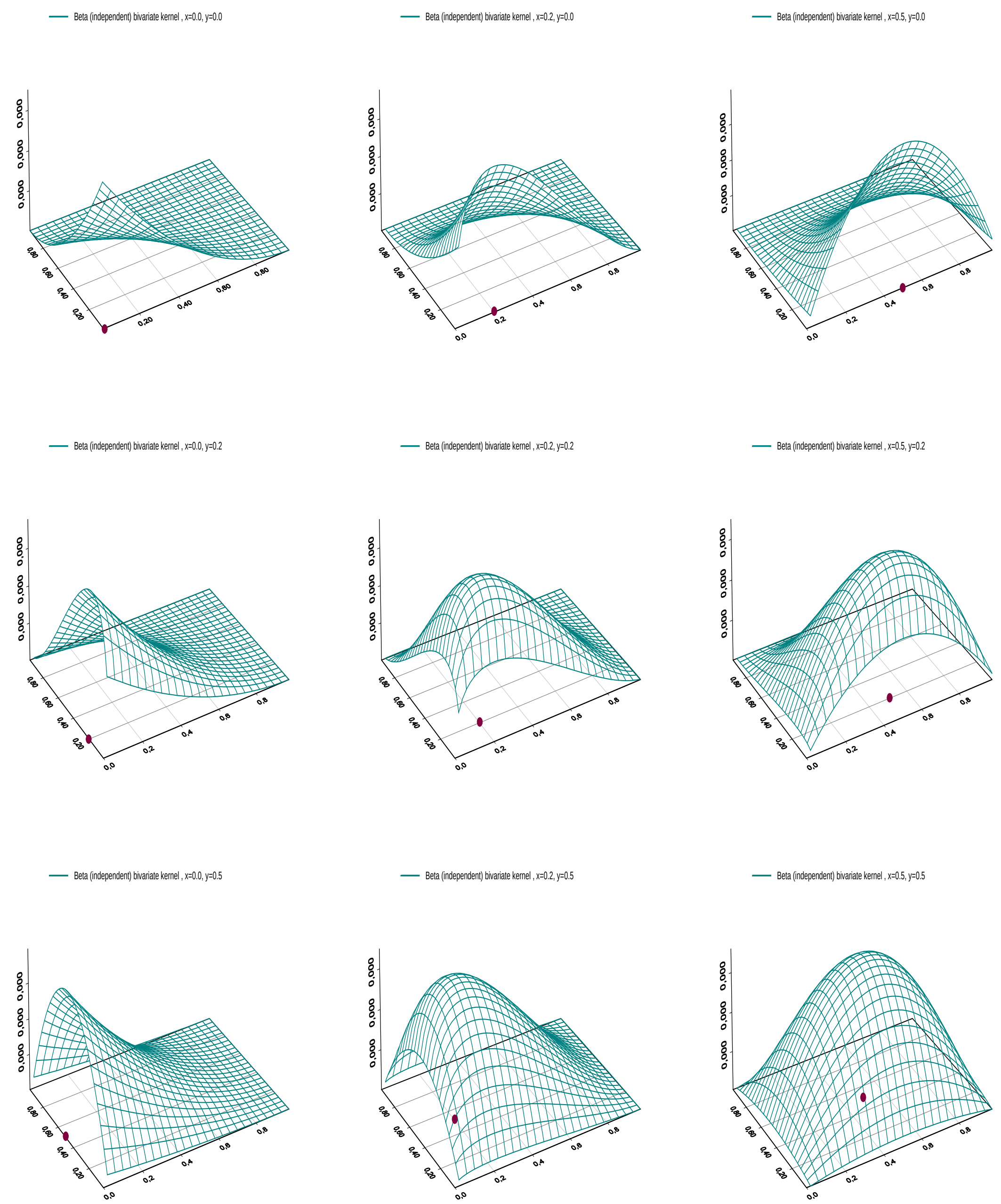


Figure 7.9: Shape of bivariate Beta kernels for different values of u and v.

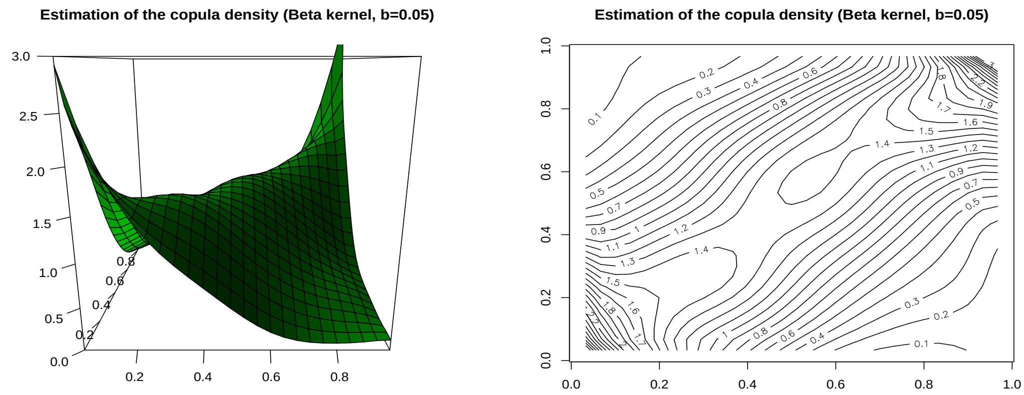


Figure 7.10: Estimation of the copula density using Beta kernels with 1,000 observations drawn from a Frank copula.

The Beta-kernel based estimator of the copula density at point (u, v), is obtained using product beta kernels, which yield

$$\widehat{c}_h(u,v) = \frac{1}{Th^2}\sum_{i=1}^{T} K\left(X_i, \frac{u}{h}+1, \frac{1-u}{h}+1\right)\cdot K\left(Y_i, \frac{v}{h}+1, \frac{1-v}{h}+1\right).$$

Figure 7.9 shows that the shape of the product beta kernels for different values of u and v is clearly adaptive.

For convenience, the bandwidths are here assumed to be equal, but more generally, one can consider one bandwidth per component. See Figure 7.10 for an example of an estimation based on Beta kernels and a bandwidth $h = 0.05$.

Let $(u, v) \in [0, 1] \times [0, 1]$. The bias of $\widehat{c}(u, v)$ is of order h, $\widehat{c}_h(u, v) = c(u, v) + \mathcal{O}(h)$. The absence of a multiplicative bias on the boundaries can be observed on Figure 7.11, where the diagonal of the copula density is plotted, based on several samples.

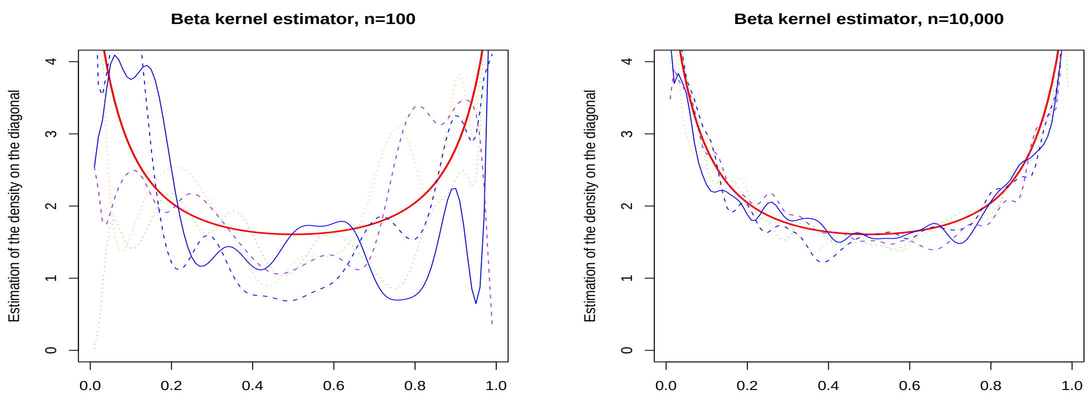


Figure 7.11: Estimation of the copula density on the diagonal using Beta kernels with 100 and 1,000 observations drawn from a Frank copula.

On the other hand, note that the variance depends on the location. More precisely, $Var(\widehat{c}_h(u,v))$ is $\mathcal{O}((Th^{\kappa})^{\leftarrow})$, where $\kappa = 2$ in corners, $\kappa = 3/2$ in borders, and $\kappa = 1$ in the interior of $[0,1]^2$. Moreover, as well as "*standard*" kernel estimates, $\widehat{c}_h(u,v)$ is asymptotically normally distributed,

$$\sqrt{Th\kappa'}[\widehat{c}_h(u,v) - c(u,v)] \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma(u,v)^2\right), \text{ as } nb^{\kappa'} \to \infty \text{ and } h \to 0,$$

where κ' depends on the location, and where $\sigma(u,v)^2$ is proportional with $c(u,v)$.

7.3.5 Working with pseudo-observations

As we know, most of the time, the marginal distributions of random vectors are unknown, as recalled in the first section. Hence, the associated copula density should be estimated, not on samples $(F_X(X_i), F_Y(Y_i))$, but on pseudo samples $(F_{X,T}(X_i), F_{Y,T}(Y_i))$.

Figure 7.12 shows some scatterplots when the margins are known (i.e. we know $(F_X(X_i), F_Y(Y_i)))$, and when margins are estimated (i.e. $(F_{X,T}(X_i), F_{Y,T}(Y_i))$. Note that the pseudo sample is more "*uniform*", in the sense of a lower discrepancy (as in Quasi Monte Carlo techniques, see e.g. Niederreiter (1992)). Here, by mapping every point of the sample on the marginal axis, we get uniform grids, which is a type of "Latin hypercube" property (see Jaeckel (2002), for instance).

Because samples are more "*uniform*" using ranks and pseudo-observations, the variance of the estimator of the density, at some given point $(u,v) \in (0,1) \times (0,1)$ is usually smaller. For instance, Figure 7.13 shows the impact of considering pseudo observations, i.e. substituting $F_{X,T}$ and $F_{Y,T}$ to unknown marginal distributions F_X and F_Y. The dotted line shows the density of $\hat{c}(u,v)$ from 100 observations (U_i, V_i) (drawn from the same Frank copula), and the straight line shows the density of $\hat{c}(u,v)$ from the sample of pseudo-observations (i.e. the ranks of the observations).

A heuristic interpretation can be obtained from Figure 7.12. Consider the standard kernel based estimator of the density, with a rectangular kernel. Consider a point (u,v) in the interior, a bandwidth h such that the square $[u-h, u+h] \times [v-h, v+h]$ lies in the interior of the unit square. Given a T sample, an estimation of the density at point (u,v) involves the number of points located in the small square around (u,v). Such a number will be denoted by N, and it is a random variable. Larger N provide more precise estimations.

Assume that the margins are known, or equivalently, let (U_1,V_1),...,(U_T,V_T) denote a sample with distribution function C. The number of points in the small square, say N_1, is random and follows a binomial law with size T and some parameter p_1. Thus, we have $N_1 \sim \mathcal{B}(T, p_1)$ with

$$\begin{aligned} p_1 &= \mathbb{P}\left((U,V) \in [u-h, u+h] \times [v-h, v+h]\right) \\ &= C(u+h, v+h) + C(u-h, v-h) \\ &\quad -C(u-h, v+h) - C(u+h, v-h) \end{aligned}$$

and therefore

$$Var(N_1) = Tp_1(1-p_1).$$

On the other hand, assume that margins are unknown, or equivalently that we are dealing with a sample of pseudo-observations $\left(\widehat{U}_1, \widehat{V}_1\right), \ldots, \left(\widehat{U}_T, \widehat{V}_T\right)$. By construction of pseudo-observations, we have is

$$\#\left\{\widehat{U}_i \in [u-h, u+h]\right\} = \lfloor 2hT \rfloor$$

where $\lfloor \cdot \rfloor$ denotes the integer part. As previously, the number of points in the small square N_2 satisfies $N_2 \sim \mathcal{B}(\lfloor 2hT \rfloor, p_2)$ where

$$p_2 = \mathbb{P}\left(\left(\widehat{U}, \widehat{V}\right) \in [u-h, u+h] \times [v-h, v+h] \middle| \widehat{U} \in [u-h, u+h]\right)$$

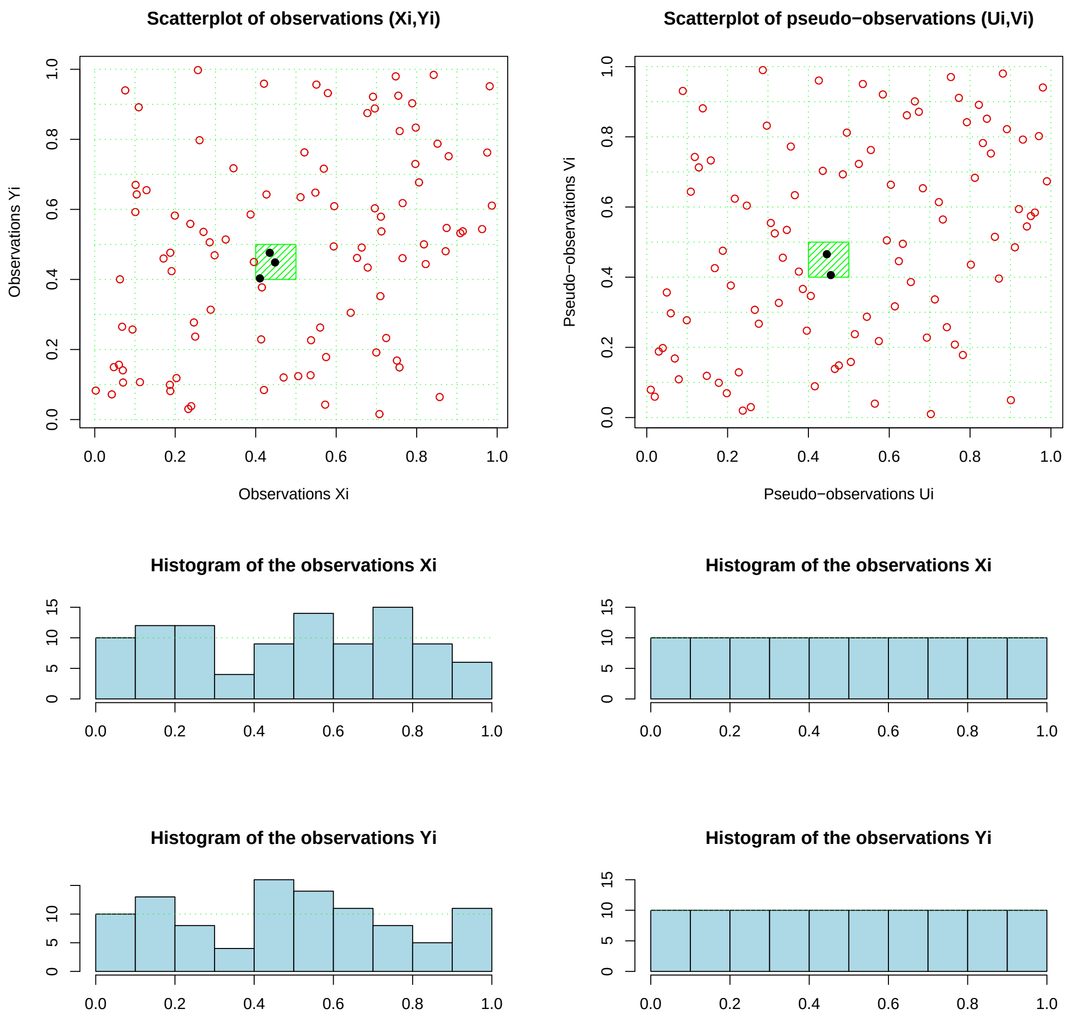


Figure 7.12: Observations and pseudo-observation: 100 observations (X_i, Y_i) drawn from a Frank copula, associated pseudo-sample $(U_i, V_i) = (\hat{F}_X(X_i), \hat{F}_Y(Y_i))$, and histograms of margins.

$$= \frac{\mathbb{P}\left(\left(\widehat{U}, \widehat{V}\right) \in [u-h, u+h] \times [v-h, v+h]\right)}{\mathbb{P}\left(\widehat{U} \in [u-h, u+h]\right)}$$

$$\approx \frac{C(u+h, v+h) + C(u-h, v-h) - C(u-h, v+h) - C(u+h, v-h)}{2h}$$

$$= \frac{p_1}{2h}$$

And therefore the expected number of observations is the same for both methods ($\mathbb{E}[N_1] \simeq \mathbb{E}[N_2] \simeq Tp_1$), but

$$Var(N_2) \approx 2hTp_2(1-p_2) = 2hT\frac{p_1}{2h}\left(1 - \frac{p_1}{2h}\right) = \frac{T}{2h}p_1(2h - p_1).$$

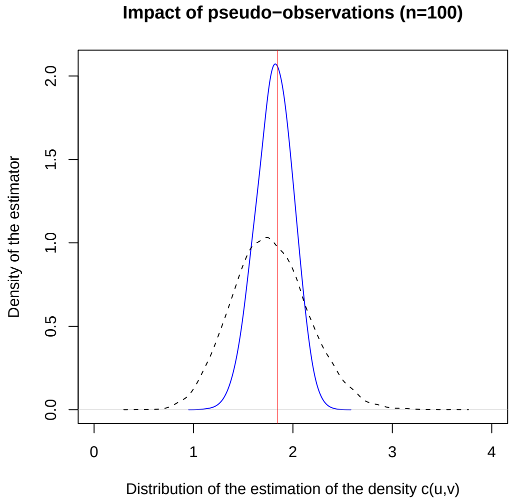


Figure 7.13: The impact of estimating from pseudo-observations. The dotted line the the distribution of $\hat{c}(u,v)$ from sample $(F_X(X_i), F_Y(Y_i))$, and the plain line from pseudo sample $(F_{X,T}(X_i), F_{Y,T}(Y_i))$.

Thus,

$$\frac{Var\left(N_2\right)}{Var\left(N_1\right)} = \frac{Tp_1\left(2h-p_1\right)}{2hTp_1\left(1-p_1\right)} = \frac{2h-p_1}{2h-2hp_1} \leq 1$$

since $h \leq 1/2$ and thus $2hp_1 \leq p_1$.

So finally, the variance of the number of observations in the small square around (u,v) is larger than the variance of the number of pseudo-observations in the same square. Therefore, this larger uncertainty concerning the relevant sub-sample used in the neighborhood of (u,v) in the former case implies a loss of efficiency. The consequence of this result is largely counterintuitive. By working with pseudo-observations instead of "true" ones, we would expect an additional noise, what should induce more noisy estimated copula densities. This is not the case, actually, as we have just shown.

7.4 Estimation of copula density for censored data

Consider the case of right-censoring, i.e. the value observed is $Y = \min\left\{Y^0, C\right\}$, where C is a censoring value (e.g. a policy limit) and Y_0 is known only when $Y_0 \leq C$. The data are pairs (Y,δ) where δ is the censoring indicator (equal to one if Y_0 is observed, and zero if not). Assume that Y^0 and C have respectively distribution function F_Y and G, and that C is independent of Y^0. Note that $\mathbb{P}\left(Y > y\right) = \left(1 - F^0\left(y\right)\right)\left(1 - G\left(y\right)\right)$.

Consider some i.i.d. sample $\left(Y_1, \delta_1\right), ... \left(Y_n, \delta_n\right)$. Under censoring, the analog of the empirical distribution function is the product-limit estimator proposed by Kaplan and Meier (1958)

$$1 - \widehat{F}^0\left(x\right) = \prod_{i, Y_i \leq x} \left(\frac{n-i}{n-i+1}\right)^{\delta_i}.$$

Note that since $1-\delta$ is the indicator of censoring, an analogous expression can be obtained to estimate the distribution of the censoring value,

$$1-\widehat{G}(x)=\prod_{i,Y_i\leq x}\left(\frac{n-i}{n-i+1}\right)^{1-\delta_i}.$$

7.4.1 Stock sampling and bias due to censoring

Example 7.4.1. *Figure 7.14 shows some density estimation on the diagonal, with no censoring on the left, 5% censored data in the middle and 25% censored data on the left, with 8 sets of $n = 1,0000$ simulated data from Frank copula. Stock sampling can be observe as the bias increases as u goes to 1, and as the proportion of censored data increases. Figure on bottom-left shows the estimated density $\widehat{c}(12/13, 12/13)$ with no censoring, and 25% censored observations.*

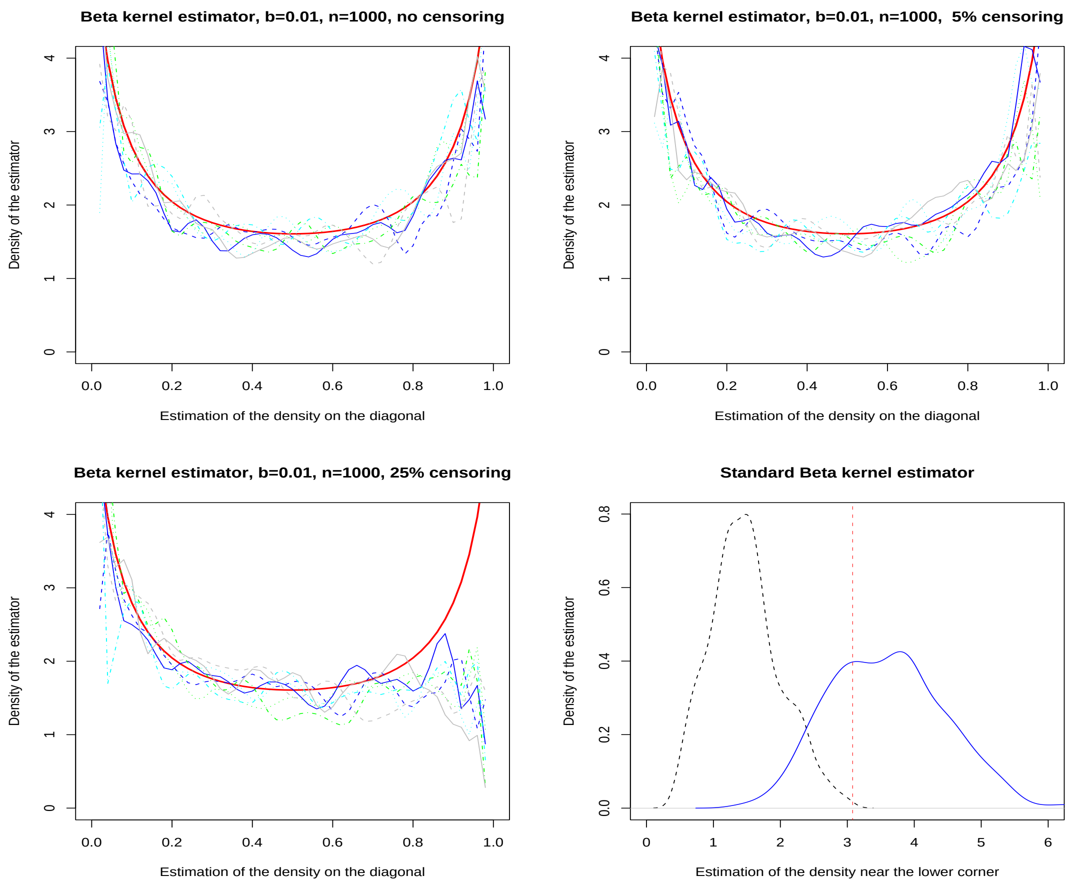


Figure 7.14: Impact of censoring on the estimation of the density: estimation of the copula density on the diagonal, with non censored data (top-left), 5% censored data (top-right) and 25% censored (bottom-left); distribution of $\widehat{c}(12/13, 12/13)$ with no censoring (plain) and 25% censoring (dotted) (bottom-right), from 1,000 simulated samples.

7.4.2 Bootstrap bias correction

Recall that if $\widehat{\theta} = \widehat{\theta}(X_1, ..., X_n)$ is an estimator of $\theta = \theta(F_X)$, the bias is simply $b_{F_X} = \mathbb{E}_{F_X}\left(\widehat{\theta}(X_1, ..., X_n)\right) - \theta(F_X)$. As pointed out in Efron and Tibshirani (1993), bootstrap techniques can be used to estimate the bias. The bootstrap estimate of the bias is the estimated bias obtained by substituting $\widehat{F}_X$ for F_X, where $\theta\left(\widehat{F}_X\right)$ is the plug-in estimate of θ, which may differ from $\widehat{\theta}(X_1,...,X_n)$. In practice, generate B bootstrap samples from $\{X_1,...,X_n\}$, denoted $\{X_1^{*b},...,X_n^{*b}\}$, $b=1,...,B$. Evaluate $\widehat{\theta}^{*b}=\widehat{\theta}\left(X_1^{*b},...,X_n^{*b}\right)$ and approximate the bootstrap approximation of $\mathbb{E}_{\widehat{F}_X}\left(\widehat{\theta}(X_1,...,X_n)\right)$ by the average of the bootstrap statistics, and therefore

$$\widehat{b}_{\widehat{F}_X}=\frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}\left(X_1^{*b},...,X_n^{*b}\right)-\theta\left(\widehat{F}_X\right)$$

From this estimator, the idea is to correct $\widehat{\theta}$ to obtain a less biased estimate. Here the bootstrap sample is obtained classically.

Following the idea developed in Efron (1981), consider the following resampling scheme, when data are censored, adjusted for bivariate pairs,

1. Generate $U_1^\star,U_2^\star,...,U_n^\star,V_1^\star,V_2^\star,...,V_n^\star$ independently, uniformly on $[0,1]$,
2. Generate Y_i^{0*} independently from $\widehat{F}_Y^0(x)$, $Y_i^{0\star}=\widehat{F}_Y^{0-1}(U_i^*)$,
3. Generate C_i^* independently from $\widehat{G}(x)$, $C_i^\star=\widehat{F}_Y^{0-1}(V_i^*)$,
4. Let $j\in\{1,...,n\}$ such that $Y_i^{0*}=Y_j$, set $X_i^*=X_j$
5. Set $Y_i^*=\min\left\{Y_i^{0*},C_i^*\right\}$ and $\delta_i^*=\mathbb{I}\left\{Y_i^{0*}\leq C_i^*\right\}$
6. Estimate the marginal Kaplan-Meier distribution functions $\widehat{F}_Y^{0*}(\cdot)$ and $\widehat{G}^*(\cdot)$, as
$$1-\widehat{F}^{0*}(x)=\prod_{i,Y_i^*\leq x}\left(\frac{n-i}{n-i+1}\right)^{\delta_i^*}\text{ and }1-\widehat{G}^*(x)=\prod_{i,Y_i^*\leq x}\left(\frac{n-i}{n-i+1}\right)^{1-\delta_i^*}.$$
7. Set $(U_i^\star,V_i^\star)=\left(\widehat{F}^*(X_i^*),\widehat{F}^{0*}(Y_i^*)\right)$ and estimate the density of the sample $(U_1^\star,V_1^\star),...,(U_n^\star,V_n^\star)$

Example 7.4.2. *Figure 7.15 shows some Beta kernel density estimation on the diagonal, with* 5% *censored data in the middle and* 25% *censored data on the left, without correction (plain) and with a bootstrap correction (dotted).*

This technique can be applied also on the Loss-ALAE dataset, where Loss amonts are censored data (upper limit for policies). Figure 7.16 compares the density of the copula on the diagonal and the theoretical density of Gumbel copula.

7.5 Concluding remarks

We have discussed how various estimation procedures impact the estimation of tail probabilities in a copula framework. Parametric estimation may lead to severe underestimation when the parametric model of the margins and/or the copula is misspecified. Nonparametric estimation may also lead to severe underestimation when the smoothing method does not take into account potential boundary biases in the corner of the density support. Since the primary focus of most risk management procedures is to gauge these tail probabilities, we think that the methods analyzed in the previous lines might help to better understand the occurrence of extreme risks in

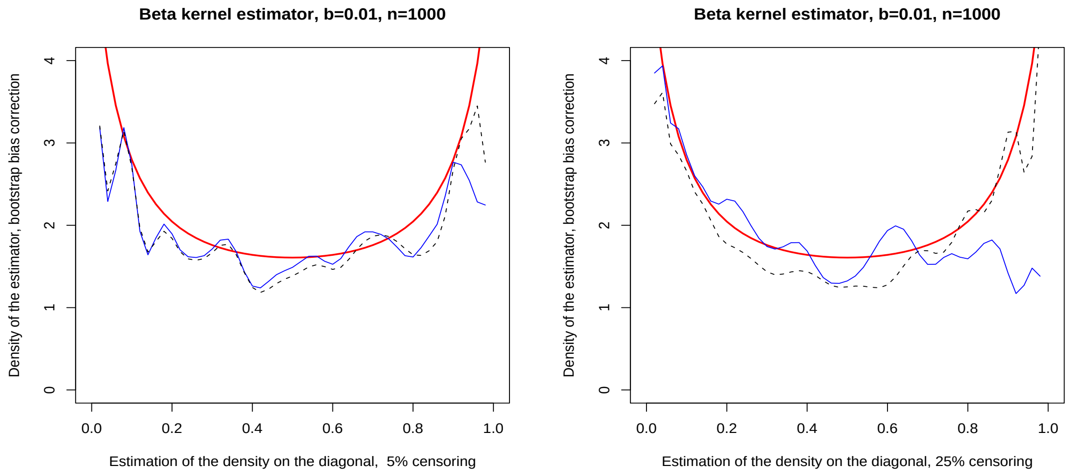


Figure 7.15: Bootstrap correction of the censoring bias on the estimation of the density (using Beta kernel estimation), with 5% censored data on the left, 25% censored data on the right.

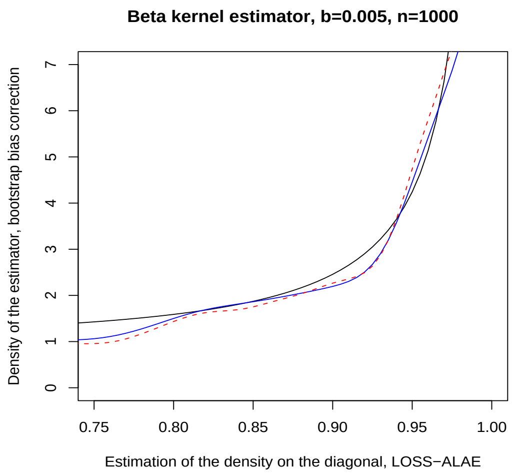


Figure 7.16: Beta kernel estimator for Loss-ALAE, compare with Gumbel copula.

stand alone positions (single asset) or inside a portfolio (multiple assets). In particular we have shown that nonparametric methods are simple powerful visualization tools to spot dependencies among various risks. A clear assessment of these dependencies should help to design better risk measurement tools within a VaR or expected shortfall framework.

Chapter 8

Temporal dependencies for natural events

8.1 Introduction and motivations

In February 2005, opening the conference on *Climate change: a global, national and regional challenge*, chairman Dennis Tirpak pointed out that "*there is no longer any doubt that the Earth's climate is changing* [...] *globally, nine of the past 10 years have been the warmest since records began in* 1861". He singled out the heatwave that gripped western Europe in 2003 as an example: Europe's worst natural disaster in 50 years killed as many as $30,000$ people and inflicted an estimated 30 billion dollars in damage.

One of the major issue is to get an accurate estimate of those risks, e.g. August 2003' heat wave, but also storms as in December 1999, or floods in July 2002 or August 2005, in (Central) Europe. Costs of natural events may be huge for insurers.

8.1.1 The return period as a risk measure

From this observation, it becomes crucial to rethink the concept of *return period*, which is a key notion in risk management. The return period is a key notion in hydrology (see e.g. Gumbel (1941) for an introduction of the concept, Martins and Clarke (1993) for maximum likelihood estimation, or Davis , Duckstein and Fogel (1976) for a Bayesian approach). As in Gumbel (1941), "*we suppose that the events are independent of one another: the occurrence of a high or low value for x has no influence on the value of any succeeding observation*": the X_i's have to be independent. Further, as pointed out in the conclusions "*we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made*": X_i's have to be identically distributed.

Annualized maxima (called *flood* events in Gumbel (1941)). Formally, if F_X is the distribution function of the yearly maxima of a random variable X, and if $X_1, ..., X_n$ are i.i.d. annualized maxima, the number of required Bernoulli experiments for the event $\{X > u\}$ to occur for the first time is a geometric random variable with mean $\mathbb{P}(X > u)$: if $N(u) = \inf\{i \geq 1, X_i > u\}$, $\mathbb{P}(N(u) = k+1) = p(1-p)^k$ where $p = \mathbb{P}(X > u)$. Thus, the period of return is the associated expected value,

$$\mathbb{E}(N(u)) = \frac{1}{p} = \frac{1}{\mathbb{P}(X > u)}.$$

The following definition is also used: the *centennial event* is u^* such that $\mathbb{E}(N) = 100$ years,

Insured Loss	Victims	Date	Event - *Country*
45,000	1,326	24.08.**2005**	Katrina, Hurricane - *USA*
22,274	43	23.08.1992	Andrew, Hurricane - *USA, Bahamas*
18,450	60	17.01.1994	Earthquake, Northridge - *USA*
10,000	34	20.09.**2005**	Rita, Hurricane - *USA, Cuba*
10,000	35	16.10.**2005**	Wilma, Hurricane - *Mexico, USA, Cuba*
8,272	24	11.08.**2004**	Charley, Hurricane - *USA, Cuba*
8,097	51	27.09.1991	Typhoon Mireille - *Japan*
6,864	95	25.01.1990	Winterstorm Daria - *France, Europe*
6,802	110	25.12.1999	Winterstorm Lothar - *France, Europe*
6,610	71	15.09.1989	Hugo, Hurricane - *Puerto Rico, USA*
5,170	38	26.08.**2004**	Frances, Hurricane - *USA, Bahamas*
5,157	22	15.10.1987	Storms and floods - *France, Europe*
4,770	64	25.02.1990	Winterstorm Vivian - *Central Europe*
4,737	26	22.09.1999	Typhoon Bart - *Japan*
4,230	600	20.09.**2004**	Georges, Hurricane - *USA, Caribbean*
4,136	3,034	13.09.1998	Jeanne, Hurricane - *USA, Caribbean*
3,707	45	06.09.**2004**	Typhoon Songda - *Japan*
3,475	41	05.06.2001	Tropical storm Allison - *USA*
3,403	45	02.05.2003	Tornadoes, hail, thunderstorms - *USA*
3,169	6,425	17.01.1995	Earthquake Kobe - *Japan*

Table 8.1: Major natural catastrophes, over 3 billion US Dollars (index to 2005), since 1950 (source: Swiss Re, with estimations as at December 2005 for hurricanes occurred in 2005).

i.e.

$$\overline{F}_X(u^*) = \frac{1}{100} \text{ or equivalently } u^* = F_X^{\leftarrow}\left(1 - \frac{1}{100}\right) = VaR(X, 0.99)$$

A dual approach is to determine the level $U(n)$ such that one may expect a single event larger than $U(n)$ occurring in n years. Equivalently, solve

$$\mathbb{E}\left(\sum_{i=1}^{n} \mathbf{1}(X_i > U(n))\right) = 1,$$

given $n \in \mathbb{N}$, i.e.

$$\mathbb{P}(X > U(n)) = \frac{1}{n}.$$

Remark 8.1.1. *Note that this geometric distribution for the time between occurrence is closely related to the* lack of memory *property (see Chukova, Dimitrov & Khalil (1992), or Chukova & Dimitrov (1993)). The underlying idea is that the geometric distribution is the only discrete distribution satisfying the recurrence property* $\mathbb{P}(X = n + k) = \mathbb{P}(X = n) \cdot \mathbb{P}(X = k)$*, for all* $k, n \in \mathbb{N}$*, while, similarly the exponential distribution is the only continuous distribution satisfying the analogous of the recurrence property,* $f(x + y) = f(x) \cdot f(y)$*, where* f *denotes the density.*

If most of the results are known in the i.i.d. context, note that (**?**) was probably the first one to highlight that this notion is more difficult to interpret when temporal dependence can be exhibited.

8.1.2 Outline of the *een bijgevoegde stelling*

In the first section of this *een bijgevoegde stelling*, we should start with the study of windstorms. After windstorms of December 1999 in France and Belgium, it appeared that consecutive occurrence of major windstorms (the first storm *Lothar* occured 48 hours before the second one, *Martin*). This persistence phenomena of storms was first analyzed by Haslett and Raftery (1989). Unfortunate, when working on the same dataset, it appeared that all their motivations for modeling windspeed using long range dependence models were not relevant: the slow deacrease of the autocorrelation function was simply a seasonal effect. Nevertheless, when studying series carefully, a seasonal trend cannot be used to model the series, and their is still persistence in residuals. Hence, an idea was then to use GARMA processes (*Gegenbauer ARMA*) instead of ARFIMA processus (*Fractional ARIMA*). As we will see using simulations, skipping the long range effect will underestimate risk estimations of two consecutive big storms.

Section 2 will extend this approach to temperature time series, in order to estimate properly the return period of August 2003' heat wave. As for the heat wave in Chicago in 1995 (which also killed hundreds of people), the reason why those events were dramatic is not that very high temperature have been reached, but because the temperature has been high during several days. More specifically, the minimal temperature (nighttime temperature) remained very high during several nights, and this had an important health impact. Here again, modeling persistence is the main issue when estimating the period of return of such an event. But if the seasonal cycle was very small for windspeed, it will be easily identifiable for temperature in Paris. The main problem for temperature is the increasing trend due to global warming (+3 degree in one century). After modeling and removing the trend, GARMA processes will also be used. But as mentioned in Dacunha-Castelle (2004) instead of focusing on fractional processes to model persistence, it is possible to obtain almost similar results with heavy tailed noise. Hence, the end of this section will focus on the comparison of two models: a Gaussian GARMA process, with long range dependence, and a Student ARMA process, with short range dependence, but heavier tails. As we will see using simulations, those two concepts yield different estimations of the period of return of this event, depending on the definition of the heat wave (11 consecutive days exceeding 19° C, or 3 consecutive days exceeding 24° C).

And finally, in section, we will focus on the study of flood events. Floods and river flows have been intensively studied between 1940 and 1955. Two approaches have been considered, both on annualized maxima of river levels. Gumbel (1941) noticed that annualized maxima were i.i.d. and that the so called Gumbel distribution could be used to model this series of maxima. On the other hand, Hurst (1951) observed, on longer series (700 years, versus 100 in Gumbel's work) that annualized maxima were not independently distributed, but strongly dependent: fractional processes were perfect to model the dynamics of those series. Hence, from independence to strong dependence, the good approach is maybe in between. And furthermore, getting a better understanding of the dynamics is crucial in order to get a proper estimate of the period of return. Since there is no chance to obtain long range dependence from only 100 years of data, the idea will be, here, to used financial models, used to get a more adequate model for transactions, when high frequency data are available.

Remark 8.1.2. *This* een bijgevoegde stelling *focuses on statistical applications, while the thesis was more theoretical. Hence, there will be no theorem in the following sections, but rather a presentation of statistical methods, applied to environmental series. Dataset with daily windspeed is the same as the one used in Haslett and Raftery (1989)*[1]*. Dataset with daily temperature in Paris Montsouris have been provided by the European Climate Assessment & Dataset project* [2]*.*

[1] http://lib.stat.cmu.edu/datasets/

[2] http://eca.knmi.nl/dailydata/index.php

And finally, the dataset with daily river level of the Saugeen river was kindly provided by Anne Catherine Favre (INRS Québec).

Acknowledgment I would like to thank David Sibaï, co-author for some results, Dider Dacunha-Castelle for a stimulating discussion on long range with light tails, versus short range with heavy tails, and Anne Catherine Favre for discussions we had on modeling hydrological series.

8.2 Modeling windspeed and windstorms

In December 1999, France and most of western Europe countries have been hit by two major windstorms, Lothar and Martin, the second one starting 48 hours after the first one. This occurrence of several windstorms within a few days is not the first one. And it appears that modeling windstorms occurrence as an homogeneous Poisson process was not relevant.

Haslett and Raftery (1989) suggested to use long-memory processes to model daily windspeed. The motivation was that autocorrelations were slowly decreasing (Figure 8.1 on the left, with 100 lags, i.e. 3 months). But considering more lags (10 years), it appears that the slow decrease was simply the beginning of a seasonal effect (Figure 8.1 on the right).

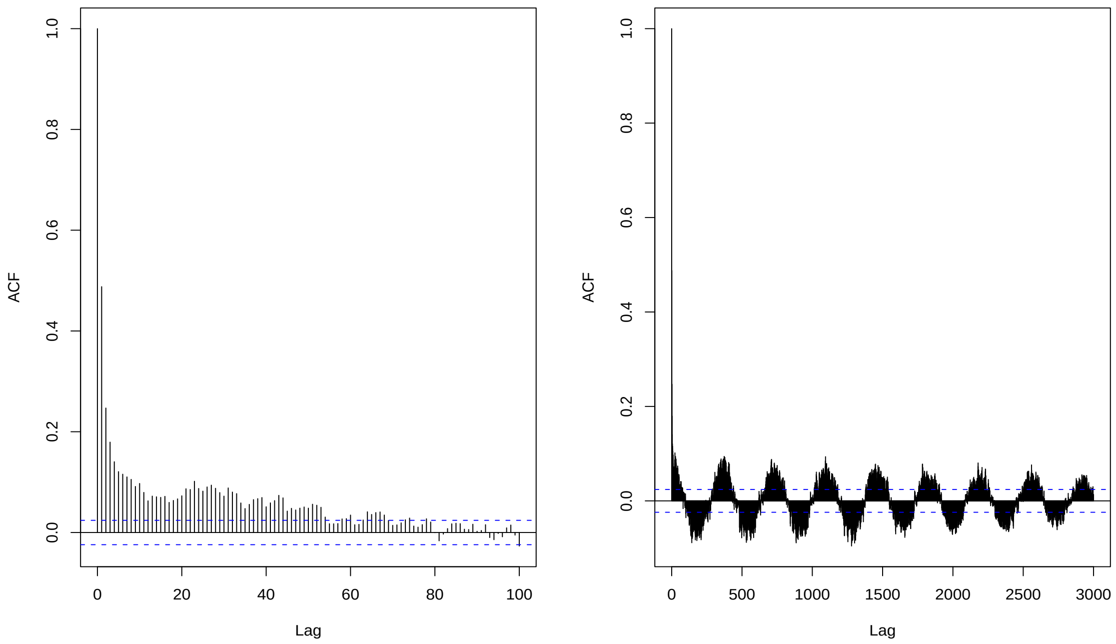


Figure 8.1: Autocorrelations, daily windspeed in Ireland.

This seasonal behavior, that can be observed on the autocorrelogramm can also be visualized on Figure 8.2, where daily windspeed is plotted, from January till December. Hence, windspeed in higher in winter than in summer. But note also that the cycle is very small, with a noise having a large variance.

8.2.1 Modeling stationary time series

Let $(Y_t)_{t\in\mathbb{Z}}$ denote a stochastic process, i.e. a sequence of random variables. $(Y_t)_{t\in\mathbb{Z}}$ is said to be strongly stationary if, for any $n \in \mathbb{N}$, $t_1 < ... < t_n$ and $h \in \mathbb{Z}$, the following joint distributions are

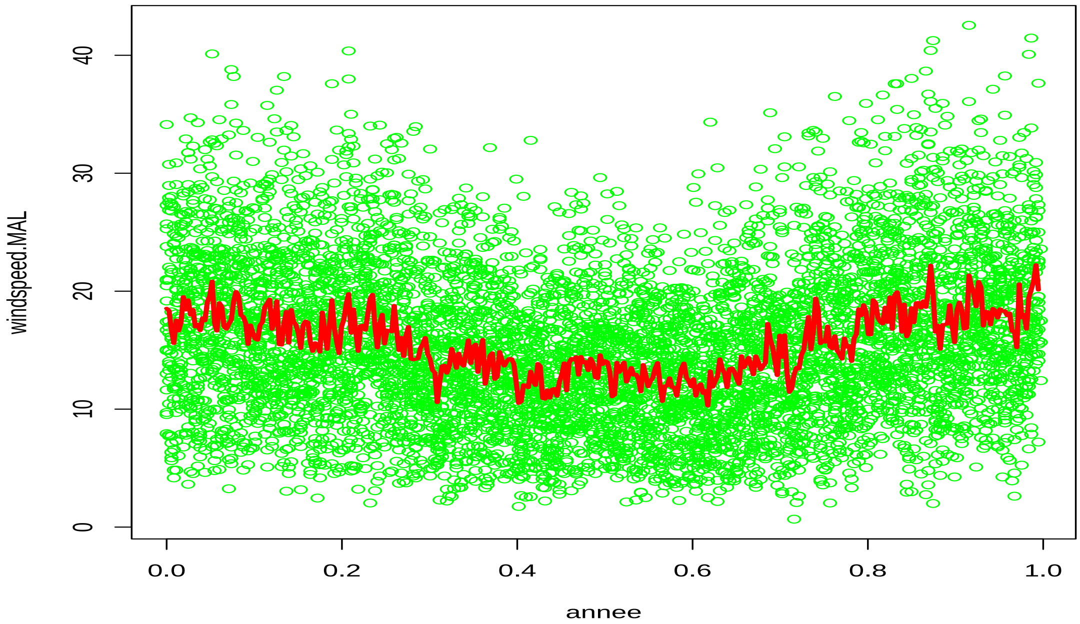


Figure 8.2: Daily windspeed, small cycle, large noise.

equal $(Y_{t_1}, ..., Y_{t_n}) \overset{\mathcal{L}}{=} (Y_{t_1+h}, ..., Y_{t_n+h})$. If variables Y_t are squared integrable, a weaker condition can be considered: $(Y_t)_{t\in\mathbb{Z}}$ is weakly stationary, or stationary in the L^2 sense if $\mathbb{E}(Y_t) = m$ for any $t \in \mathbb{Z}$, and $cov(X_t, X_s) = \gamma(|t-s|)$, for any s, t. In that case, define the autocovariance function, for all $h \in \mathbb{Z}$, as

$$h \mapsto \gamma_X(h) = cov(X_t, X_{t-h}) = \mathbb{E}(X_t X_{t-h}) - \mathbb{E}(X_t) \cdot \mathbb{E}(X_{t-h}).$$

Another point of view when studying weakly stationary processes is to study, instead of the autocovariance function, its Fourier transform, i.e.

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma_X(h) \exp(i\omega h)$$

for all $\omega \in [0, 2\pi]$. Those two notions are equivalent, since

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma_X(h) \cos(\omega h)$$

and

$$\gamma_X(h) = \int_0^\pi \cos(\omega h) f_X(\omega) \, d\omega, \text{ où } \gamma_X(h) = cov(X_t, X_{t-h}),$$

(the *spectral representation theorem*, see Theorem 4.3.1. in Brockwell and Davis (1991), or Theorem 8.31 in Gouriéroux and Monfort (1997)).

Set finally $\rho_X(h)$ the autocorrelation of order h, defined as $\rho_X(h) = \gamma_X(h)/\gamma_X(0)$.

8.2.2 ARIMA processes

Amongst the most widely used models of stationary processes are the ARMA processes, *autoregressive moving average* (see e.g. Chapter 3 in Brockwell and Davis (1991) or section 8.1 in

Gouriéroux and Monfort (1997)). An *autoregressive* (AR) model is a linear difference equation, with constant coefficient. A *moving average* (MA) model is one which expresses a process as a linear combination of a white noise, and a finite number of lagged value.

From Wold theorem (Section 2.6 in Brockwell and Davis (1991)), it is possible to represent every stationary process with a moving average representation (perhaps of infinite order). But in practice, the interest is limited. Similarly, adequate approximation of stochastic processes can be be obtain with high-order AR models, but it is limited for low-order AR models. Hence, a large number of data is needed to find robust estimators.

When combining AR and MA components in a mixed models, the approximation with a limited number of parameters is greatly increased. Thus, ARMA processes are interesting to model stationary series.

A qth-order moving average process, MA(q), $(X_t)_{t\in\mathbb{Z}}$ satisfies

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise process ($\mathbb{E}(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$ and $\gamma_\varepsilon(h) = 0$ for any $h \neq 0$), and where $\Theta(L)$ is a polynomial of lag operator L ($LX_t = X_{t-1}$). Note that is can also be expressed as as an AR(∞) process, $\Theta^{-1}(L)X_t = \varepsilon_t$ if all the roots of the polynomial Θ lie outside the unit circle (see section 8.1 in Gouriéroux and Monfort (1997)). Note that a moving average process is always stationary, when $q < \infty$.

A pth-order autoregressive process, AR(p), $(X_t)_{t\in\mathbb{Z}}$ satisfies

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = \varepsilon_t = \Phi(L)X_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise process and where $\Phi(L)$ is a polynomial of lag operator L. Note that is can always be expressed as as an MA(∞) process (Wold theorem), written $X_t = \Phi^{-1}(L)\varepsilon_t$. Further, $(X_t)_{t\in\mathbb{Z}}$ is stationary if all the roots of the polynomial Φ lie outside the unit circle (see section 8.1 in Gouriéroux and Monfort (1997)).

A ARMA(p, q) process, $(X_t)_{t\in\mathbb{Z}}$ satisfies

$$\Phi(L)X_t = \Theta(L)\varepsilon_t$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise process, where all the roots of the polynomial Φ lie outside the unit circle. If 1 is a root of Φ, then $(X_t)_{t\in\mathbb{Z}}$ is no longer stationary, and it is called integrated ARMA process, satisfying

$$(1 - L)^d\Phi'(L)X_t = \Theta(L)\varepsilon_t.$$

For instance, if $(1 - L)X_t = \varepsilon_t$, $(X_t)_{t\in\mathbb{Z}}$ is a random walk.

8.2.3 Long range dependence, or long memory processes

A stationary process $(Y_t)_{t\in\mathbb{Z}}$ is said to have long range dependence if

$$\sum_{h=1}^{\infty} |\rho_X(h)| = \infty,$$

and short range dependence if not (from McLeod and Hippel (1978)).

Remark 8.2.1. *As we will see more deeply in section 8.4, long range dependence has been introduced in Hurst (1951), in hydrology (on the Nile series), and more deeply studied by Mandelbrot (1965, 1972). Note that those processes have been introduced in Finance in the 90's (see e.g. Cheung and Lai (1993), Cheung (1993) or Baillie and Bollerslev (1994) for some applications on exchange rate dynamics). A wide survey can be found in Baillie (1996) or Jasiak (1999).*

Recall that stationary ARMA processes have autocorrelations that are quickly decreasing, i.e.

$$|\rho(h)| \leq C \cdot r^h, \text{ for all } h = 1, 2, ...$$

where $r \in]0, 1[$ (see Section 3.6 in Brockwell and Davis (1991)). This is the main reason why those processes are said to have *short* range dependence: for small values of h, $corr(X_t, X_{t-h})$ can be relatively small (and non-significant). A wide class of long memory processes is obtained when autocorrelations are slowly decreasing, at a power rate,

$$\rho(h) \sim C \cdot h^{2d-1} \text{ as } h \to \infty, \tag{8.1}$$

where $d \in]0, 1/2[$. Such a relationship can be obtained when considering stochastic processes defined as

$$(1 - L)^d X_t = \varepsilon_t,$$

where (ε_t) is a (weak) white noise ($\mathbb{E}(\varepsilon_t) = 0$ and $\rho(h) = 0$ when $h \neq 0$), and where $(1 - L)^d$ is defined as

$$(1 - L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 + ... = \sum_{j=0}^{\infty} \phi_j L^j,$$

where

$$\phi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<k\leq j} \left(\frac{k-1-d}{k}\right) \text{ for all } j = 0, 1, 2, ...$$

if L denotes the lag operator (see Section 13.2 in Brockwell and Davis (1991)).

When $-1/2 < d < 1/2$, this process is stationary, and it has the following moving-average representation,

$$X_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i} \text{ were } \theta_i = \frac{\Gamma(i+d)}{\Gamma(j+1)\Gamma(d)}.$$

Note further that it is invertible, and if further $Var(\varepsilon_t) = 1$, the autocovariance function is

$$\gamma_X(h) = \frac{\Gamma(1-2d)\Gamma(h+d)}{\Gamma(d)\Gamma(1-d)\Gamma(h+1-d)} \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \cdot h^{2d-1},$$

as $h \to \infty$. Finally, its spectral density satisfies

$$f_X(\omega) = \left(2 \sin \frac{\omega}{2}\right)^{-2d} \sim \omega^{-2d},$$

as $\omega \to 0$.

Remark 8.2.2. *Those processes have been defined also in continuous time (e.g. Mandelbrot and Van Ness (1968)), under the name of fractionary brownian process.*

Let $(X_t)_{t\in\mathbb{N}}$ denote a long memory process, and set, $\overline{X}_n = [X_1 + ... + X_n]/n$ the empirical average over the first n observations, then, since $\rho(h) \to 0$ as $h \to \infty$,

$$\mathbb{E}(\overline{X}_n - \mathbb{E}(X))^2 \to 0 \text{ as } n \to \infty,$$

and for any $-1/2 < d < 1/2$,

$$n^{1-2d}\mathbb{E}(\overline{X}_n - \mathbb{E}(X))^2 \to C \text{ as } n \to \infty,$$

i.e. the empirical average is not asymptotically distributed anymore (see Taqqu (1975) and Fox and Taqqu (1986)). Furthermore, when $0 < d < 1/2$, Bartlett's formula on asymptotic normality of empirical autocorrelations (see Propositions 7.3.1-7.3.4 in Brockwell and Davis (1991)) does not hold anymore.

In order to detect long range dependence, from a theoritical point of view, a naive approach would be to substitute empirical autocorrelation and empirical spectral density, and check that

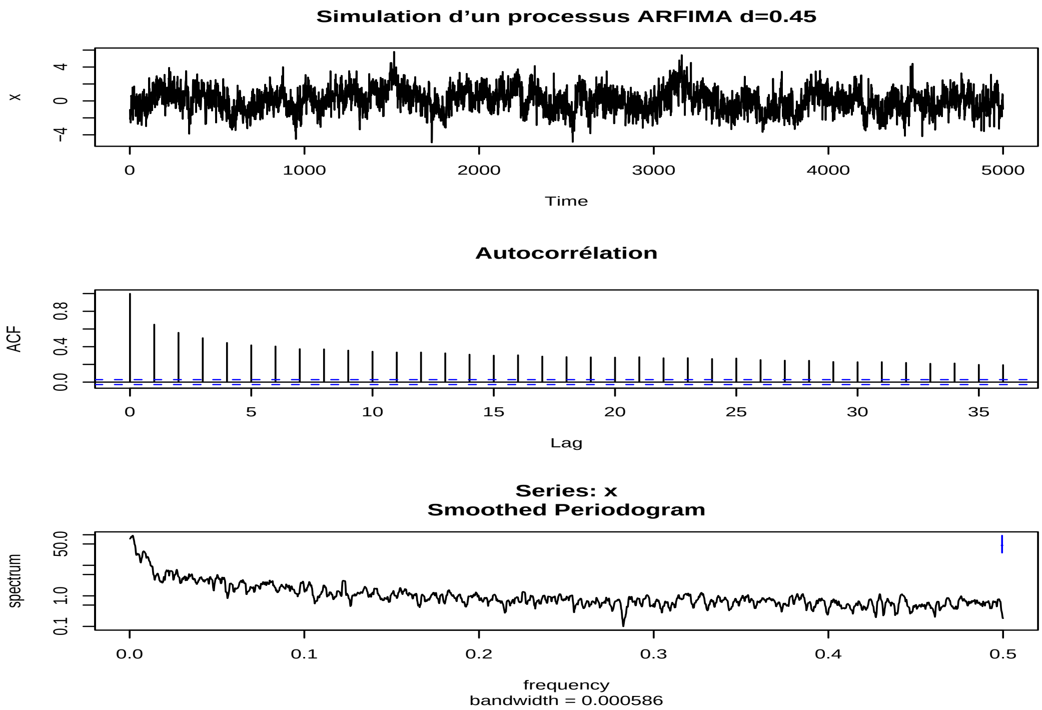


Figure 8.3: Simulation of an ARFIMA process, $d = 0.45$.

- the autocorrelations tend to 0 with a power decrease,
- the spectral density tends to infinity in 0,

where

$$\widehat{\rho}_n(h) = \frac{\widehat{\gamma}_n(h)}{\widehat{\gamma}_n(0)} \to \rho(h) \text{ where } \widehat{\gamma}_n(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \overline{X}_n)(X_{t+h} - \overline{X}_n),$$

as $n \to \infty$ (under ergodic additional assumptions on $(X_t)_{t\in\mathbb{Z}}$, see Hannen (1973) for exact assumptions in order to obtain asymptotic properties), and

$$I_n(\omega) = \frac{1}{n}\left\|\sum_{t=1}^{n}\exp(-i\omega t)X_t\right\|^2, \omega \in [0,\pi],$$

is not a consistent estimator of $f(\omega)$, but satisfies $\mathbb{E}(I_n(\omega)) \to f_X(\omega)$ as $n \to \infty$ (under very general assumptions, see ...).

Note that if Equation (8.1) holds, then

$$\log \widehat{\rho}_n(h) \sim \log C - d \log h.$$

Thus, a natural estimator of d can be obtained from the slope of points $(\log h, \log \widehat{\rho}_n(h))_{h\in\mathbb{N}}$ (see Figure 8.4 with a white noise, $d = 0$, and Figure 8.5 with an ARFIMA process, $d = 0.3$).

Example 8.2.3. *Figure 8.6 is based on the dataset studied in Hurst (1951), on the Nile yearly maxima. The slow decrease of the autocorrelation can be observed on those* 900 *years. But in the case of a shorter series (extracting* 150 *years, Figure 8.7) note that it becomes much more difficult to assess whether the series exhibits long range dependence or not.*

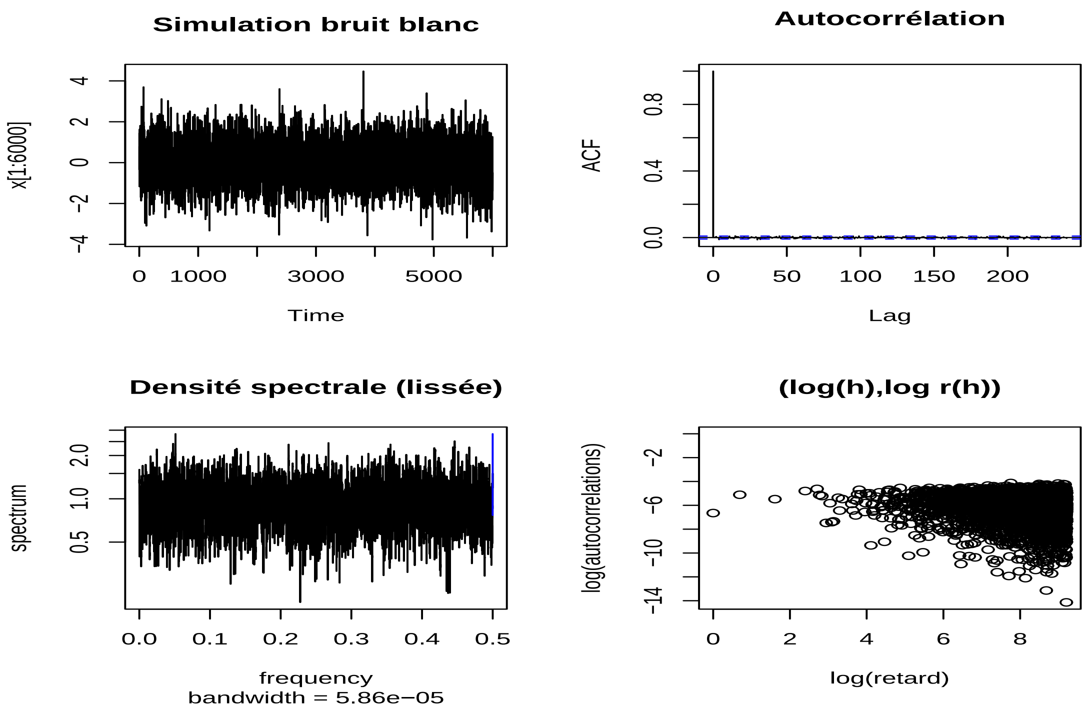


Figure 8.4: Simulated white noise process.

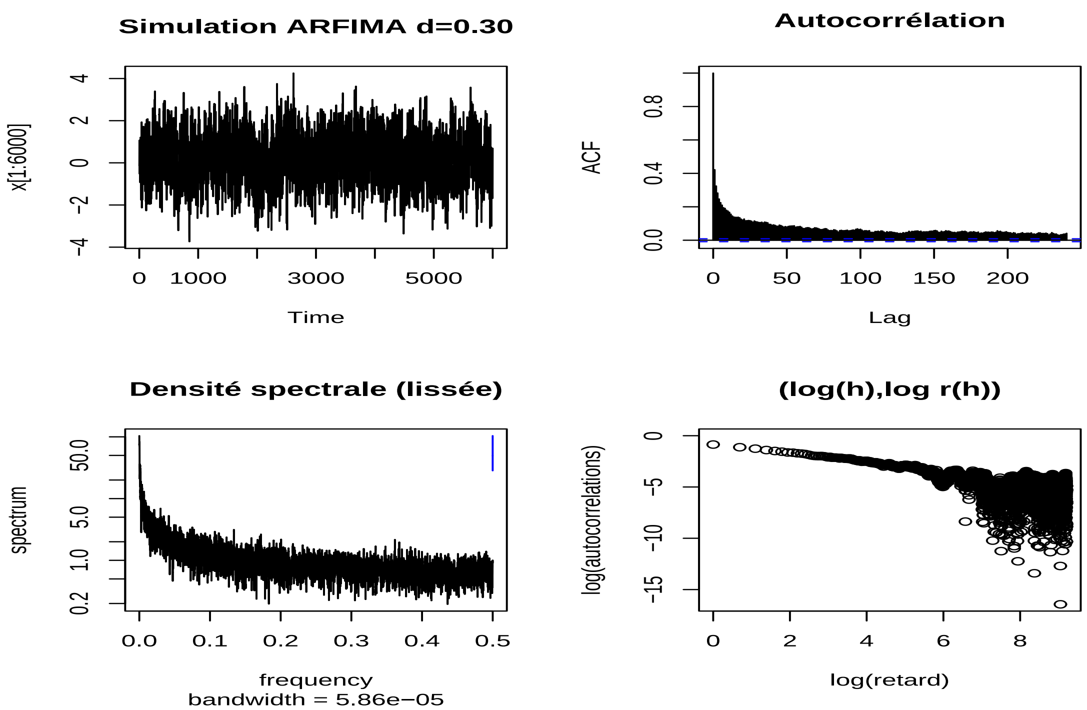


Figure 8.5: Simulated ARFIMA process, $d = 0.30$.

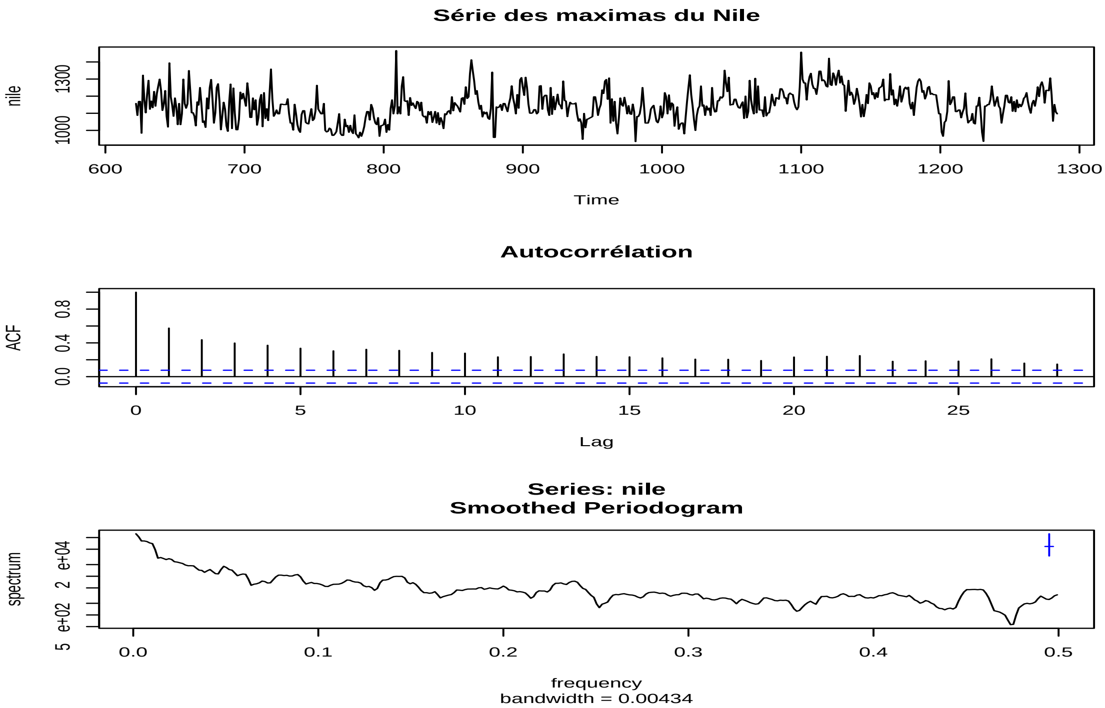


Figure 8.6: Nile series, studied in Hurst (1951).

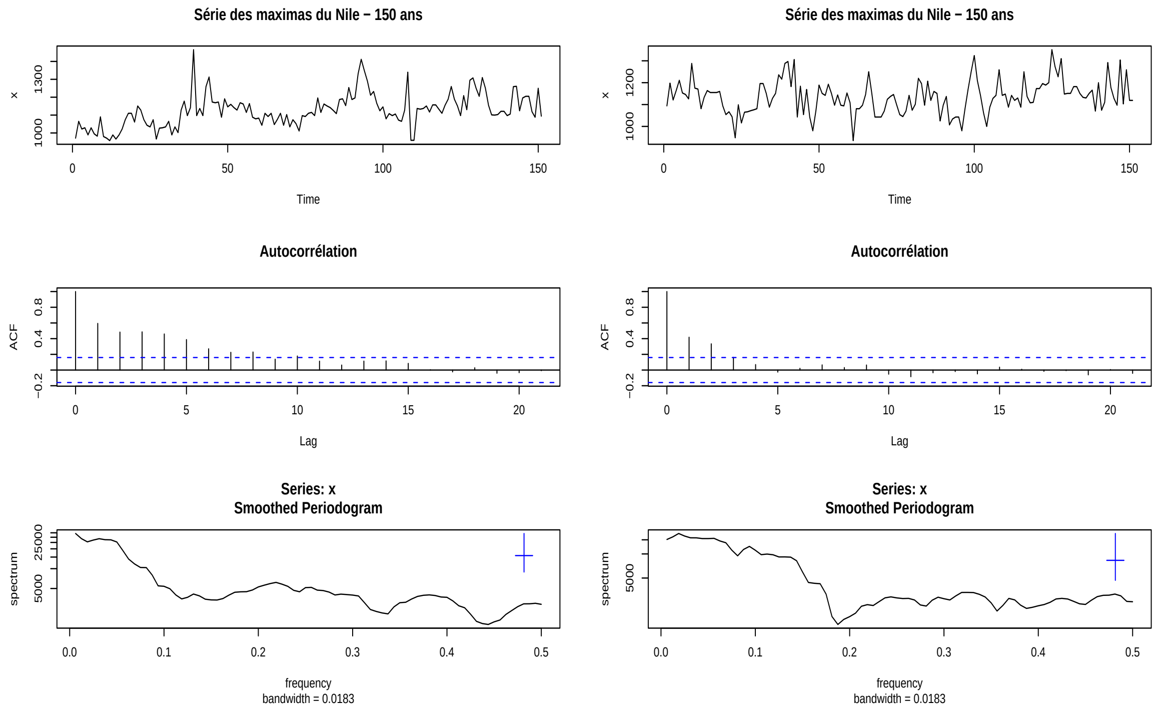


Figure 8.7: Nile series, studied in Hurst (1951) - 150 observations.

8.2.4 A more general model for long range dependence

Long range dependence has been obtained here with a pole in 0 of the spectral density. But a more generally, Gray, Zhang and Zoodward (1989) extended the notion of long memory by noting that the spectral density of a long memory process is unbounded for some frequency between 0 and π:

Definition 8.2.4. *A stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ is said to have long range dependence if its spectral density has a pole, i.e. there exists $\omega_0 \in [0,\pi]$ such that $f_X(\omega) \to \infty$ when $\omega \to \omega_0$.*

Note that this pole is closely related to seasonal effect. Further, the pole may not be unique (this will yield multiple factors models).

Example 8.2.5. *Hosking (1981, 1984) proposed two kinds of models to obtain poles in $\omega_0 > 0$:*

- *considering operators $(1-\phi L + L^2)^d$,*
- *considering operators $(1-L^s)^d$.*

The first class of models were studied in Gray, Zhang and Woodward (1989), which introduced this operator to model seasonal series with long range dependence. Hence, GARMA(p,d,q) processes were actually introduced in Hosking (1981) as

$$\Phi(L)(1-2uL+L^2)^\lambda X_t = \Theta(L)\varepsilon_t,$$

but they were not studied in that paper due to the difficulty to invert operator $(1-2uL+L^2)^\lambda$. It has been done in Gray, Zhang and Woodward (1989), using Gegenbauer polynomial: given $\lambda \neq 0$, $|Z|<1$ and $|u|\leq 1$,

$$(1-2uL+L^2)^{-\lambda} = \sum_{i=0}^{\infty} P_{i,\lambda}(u)L^n,$$

where

$$P_{i,\lambda}(u) = \sum_{k=0}^{[i/2]}(-1)^k \frac{\Gamma(\lambda+n-k)}{\Gamma(\lambda)}\frac{(2u)^{n-2k}}{[k!(n-2k)!]}$$

If $|u|<1$, the limit of $(\omega-\omega_0)^{2\lambda}f(\omega)$ exists when $\omega \to \omega_0$, where ω_0 is Gegenbauer's frequency, defined as $\omega = \cos^{-1}(u)$.

Note further that if $|u|<1$ and $0<\lambda<1/2$, then

$$\rho(h) \sim C \cdot h^{2\lambda-1} \cdot \cos(\omega_0 \cdot h) \text{ when } h \to \infty.$$

The estimation is usually done in two steps (see Ferrara (2000)),

- estimation of u (which should be related to the seasonality of the series),
- estimation of the long range index λ, or equivalenty d.

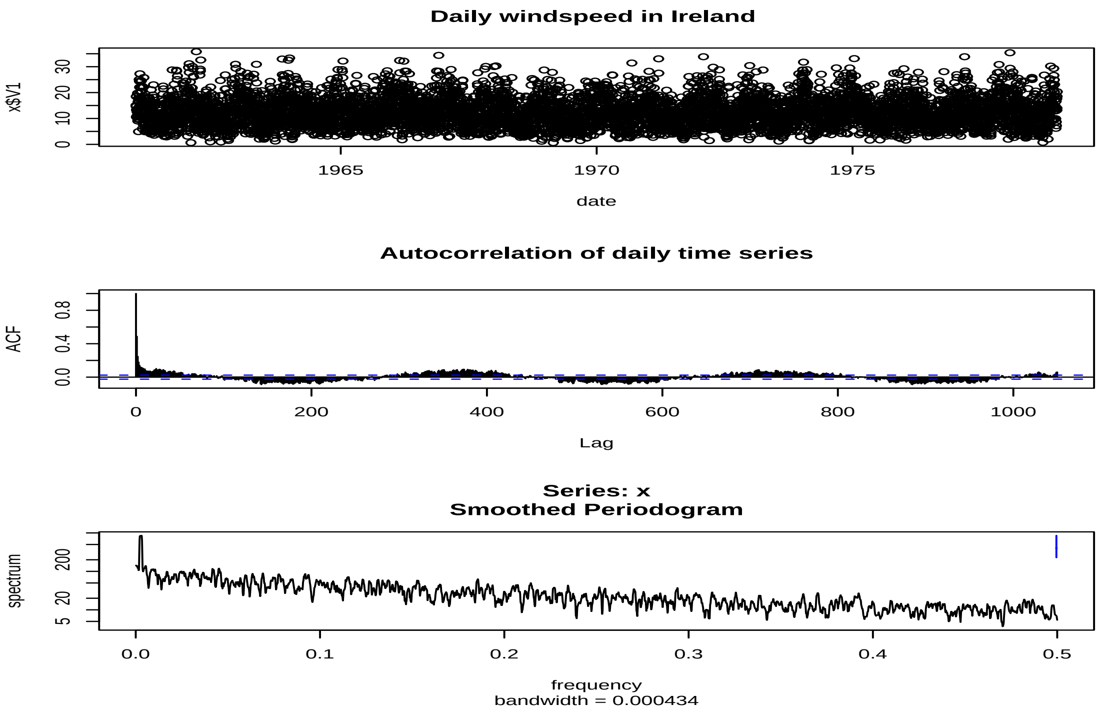


Figure 8.8: Daily windspeed in Ireland.

8.2.5 Estimation of parameters

A natural estimator is based on the R/S statistics (Hurst (1951), and Mandelbrot and Wallis (1969)). Set

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S_n^2 = \frac{1}{n}\sum_{i=1}^{n}[X_i - \overline{X}_n]^2$$

and then define

$$Q_n = \frac{R_n}{S_n} \text{ where } R_n = \max_{k=1,\dots,n}\sum_{i=1}^{k}(X_i - \overline{X}_n) - \min_{k=1,\dots,n}\sum_{i=1}^{k}(X_i - \overline{X}_n).$$

For a wide class of processes, $n^{-d-1/2}\cdot Q_n$ has a nondegenerated limit when $n\to\infty$ (Mandelbrot (1975) or Taqqu (1975)). More specifically,

$$n^{-d-1/2}\cdot\mathbb{E}(Q_n)\sim c \text{ where } \log\mathbb{E}(Q_n)\sim\gamma + H\log n,$$

where $H = d + 1/2$ is Hurst long memory index.

Parameter u is related to the seasonality of the series, or formally

$$\widehat{u} = \cos(\widehat{\omega}_0) \text{ where } \widehat{\omega}_0 \in \text{argmax}_{\omega\in[0,2\pi]}\{I_n(\omega)\}.$$

If I_n is the periodogram of series $(X_t)_{t=1,\dots,n}$, and if

$$\widehat{\omega}_0 = \text{argmax}\{I_n(\omega), \omega\in[0,\pi]\},$$

then for any $\alpha\in[0,1]$,

$$n^{\alpha}(\widehat{\omega}_0 - \omega_0)\to 0, \text{ where } \omega_0 = \cos^{-1}(u),$$

as proved in Yajima (1996) (under normality assumptions for stochastic process $(X_t)_{t\in\mathbb{Z}}$). In the case of daily windspeed in Ireland, $\widehat{\omega}_0 = 2\pi/365$.

In a general context, when estimating the additional parameters of the GARMA process,

$$(\boldsymbol{\alpha}, \sigma, d) = ((\phi_1, ..., \phi_p, \theta_1, ..., \theta_q), \sigma, d),$$

where σ^2 is the variance of the white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$. Whittle's approach can be considered (see Hosoya (1997) or Taqqu and Teverovsky (1997)) using asymptotic approximations for the log-likelihood. In that case, estimators $\widehat{\boldsymbol{\alpha}}$, $\widehat{\sigma^2}$ and $\widehat{d}$ are asymptotically independent, and normally distributed (see Hosoya (1997) or Palma and Chan (2005)). In the case of daily windspeed, $\widehat{\lambda} = 0.126$ and $\widehat{\sigma^2} = 4.47^2$.

Remark 8.2.6. *On hourly series, GARMA with k factors can also be used (see e.g.Bouette et al. (2005) and Figure 8.9).*

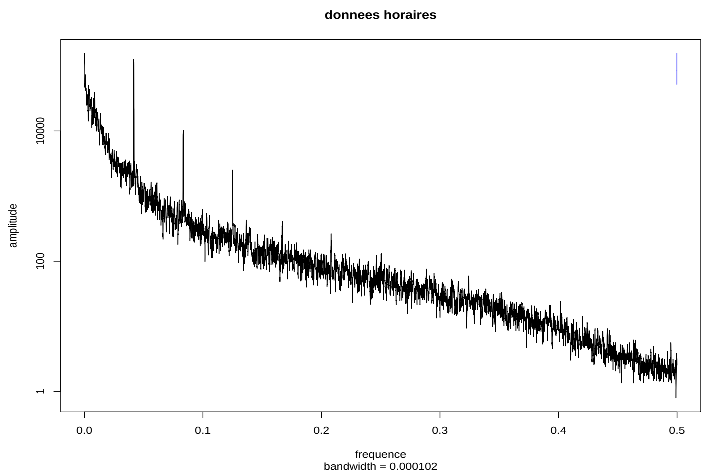


Figure 8.9: Spectral density of hourly series of windspeed in the Netherlands.

8.2.6 Estimation of return periods

It is possible to use a recursive algorithm to generate a GARMA process (see Gray, Zhang and Woodward (1989), using ideas from Rainville (1960)),

$$X_t = \mu + \sum_{j=0}^{M} C_j(d, u) \cdot \varepsilon_{t-j},$$

where M is large (e.g. 290,000) and where $C_j(d, u)$ satisfies a recursive relationship

$$C_j(d, u) = 2\nu \left(\frac{d-1}{j} + 1\right) \cdot C_{j-1}(d, u) - \left(2\frac{d-1}{j} + 1\right) \cdot C_{j-2}(d, u).$$

In order to understand the impact of long range dependence, an idea is to compare probabilities to have a *strong* wind during n consecutive days,

Those probabilities can also be visualized on Figure 8.10.

		15 knots	20 knots	25 knots
2 days	$GARMA$	47.7%	8.5%	0.3%
	seasonal $ARMA$	36.9%	3.6%	0.0%
	ratio	0.775	0.421	0.117
3 days	$GARMA$	46.3%	8.0%	0.3%
	seasonal $ARMA$	27.7%	1.4%	0.0%
	ratio	0.597	0.176	0.010
4 days	$GARMA$	45.3%	7.6%	0.3%
	seasonal $ARMA$	20.6%	0.5%	0.0%
	ratio	0.454	0.062	0.001

Table 8.2: Probabilities to have strong wind during consecutive days.

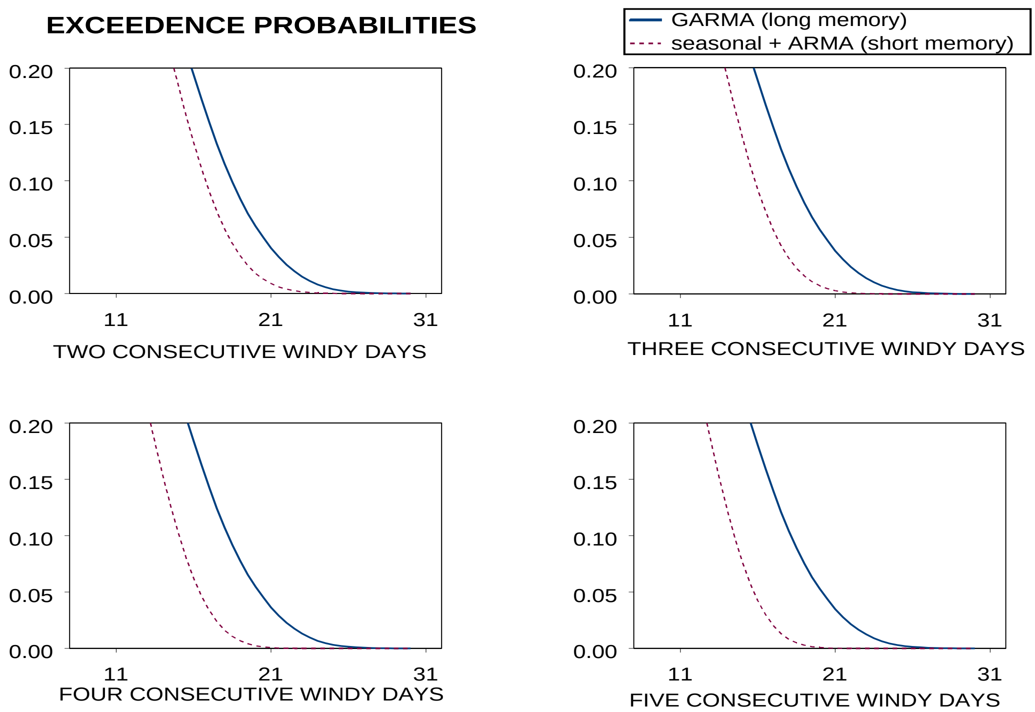


Figure 8.10: Probabilities to have strong wind during n consecutive days.

8.3 2003's heat wave and its return period

The summer of 2003 will be remembered for the extreme heat, and the approximately 40,000 heat-related deaths over western Europe. More specifically, the period 1-15 August 2003 was the most intense heat of the summer. The report of Pirard et al. (2005) states that "*Europe experienced an unprecedented heat wave in the Summer* 2003. *In France, it was the warmest summer recorded for* 53 *years in terms of minimal, maximal and average temperature and in terms of duration*". Luerbacher et al. (2004) even claim that the "*summer of* 2003 *was by far the hottest summer since* 1500". But because of global warming, their estimate of the return period of that event is 250 years. Hence, nobody was expecting such an event, and nothing had been planned in France to face it.

Actually, the underestimation of the probability of occurrence of such an event was already mentioned in the Third IPCC Assessment (Intergovernmental Panel on Climate Change (2001)). More specifically, it is pointed out that treatment of extremes (e.g. trends in extreme high temperature) is "*clearly inadequate*". Karl and Trenberth (2003) noticed that "*the likely outcome is more frequent heat waves*", "*more intense and longer lasting*" added Meehl and Tebaldi (2004). In this section, the goal is to get an accurate estimate of the return period of that event.

8.3.1 Characteristics of 2003's heat wave

One of the characteristics of 2003's heat wave has not been the intensity, but the length during 10 to 20 days in several major cities in France. For instance, in Nîmes, there were more than 30 days with temperatures higher than 35° C (versus 4 in hot summers, and 12 in the previous heat wave, in 1947). Similarly, the average maximum (minimum) temperature in Paris peaked over 35° C (approached 20° C) for 10 consecutive days, 4-13 August. Previous records were 4 days in 1998 (8 to 11 of August), and 5 days in 1911 (8 to 12 of August). Similar conditions were found in London, where maximum temperatures peaked above 30°C during the period 4-13 August (see Burt (2004) and Burt and Eden (2004)).

The use of minimal temperature was initiated by Karl and Knight (1997) when modeling 1995 heatwave in Chicago: they concentrated on the severity of an annual "*worst heat event*", and suggested that several nights with no relief from very warm nighttime minimum temperature should be most important for health impact (see also Kovats and Koppe (2005)).

8.3.2 Modeling temperature

Smith (1993) or Dempster and Liu (1995) suggested that, on a long period, the average annual temperature should be decomposed as follows:

- an increasing *linear* trend,
- a random component, with *long range dependence.*

Hence, modeling daily windspeed, a natural idea would be to consider three components,

- an increasing *linear* trend,
- an annual *seasonal* effect,
- a random component, with *long range dependence.*

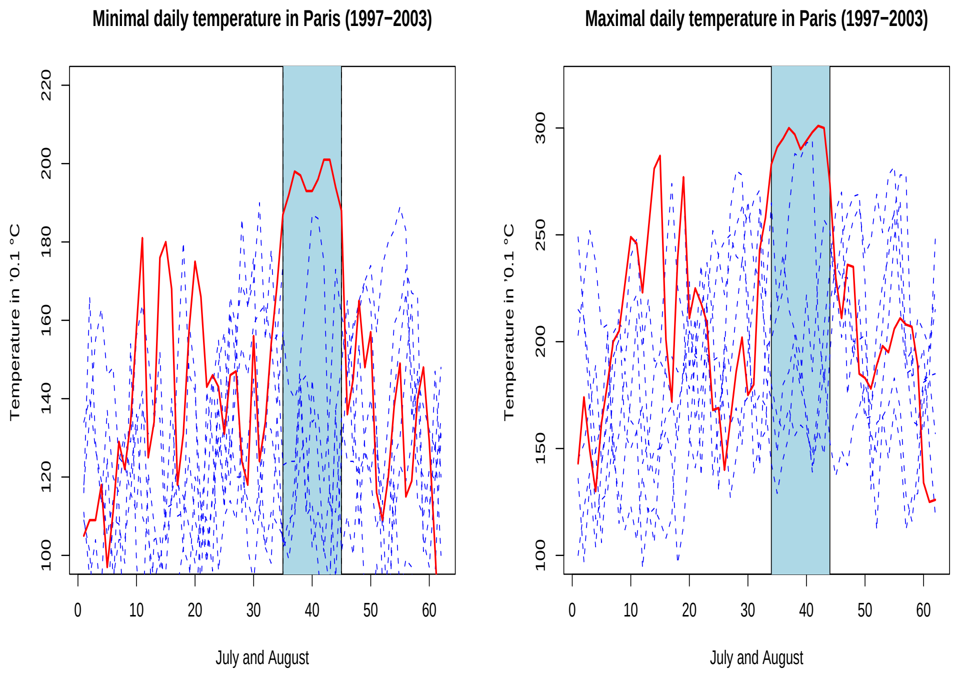


Figure 8.11: Daily temperature in Paris, years 1997 to 2002 in dotted lines, and 2003 and plain line. The plain area is the hottest 10 day period of 2003.

Modeling the increasing trend

The global warming can easily be observed using some nonparametric regression (see e.g. Figure 8.12, based on local regression with lowess functions). At first sight, some linear trend can be assumed.

Estimating the linear trend of a statistical sample is usually a trivial problem, but here, several difficulties may arise, due to the effect of (possible) long range dependence.

Hence, if $\gamma_X(\cdot)$ denotes the autocovariance function of a stationary process $(X_t)_{t\in\mathbb{Z}}$,

$$Var(\overline{X}_n) = \frac{\gamma_X(0)}{n} + \frac{2}{n}\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\gamma(k),$$

where $\overline{X}_n$ is the standard empirical mean of a sample $\{X_1, ..., X_n\}$ (see Brockwell and Davis (1991), or Smith (1993)). Furthermore, if autocovariance function satisfies $\gamma(h) \sim a \cdot h^{2d-1}$ as $h \to \infty$, then

$$Var(\overline{X}_n) \sim \frac{a}{d(2d-1)} \cdot n^{2d-2},$$

as derived in Samarov and Taqqu (1988). And further, the ordinary least squares estimator of the slope β (in the case where the X_i's are regressed on some covariate Y) is still

$$\widehat{\beta} = \frac{\sum X_i(Y_i - \overline{Y}_n)}{\sum(Y_i - \overline{Y}_n)^2}.$$

As shown in Yajima (1988), and more generally in Yajima (1991) in the case of general regressors,

$$Var(\widehat{\beta}) \sim \frac{36a(1-d)}{d(1+d)(2d+1)} \cdot n^{2d-4}.$$

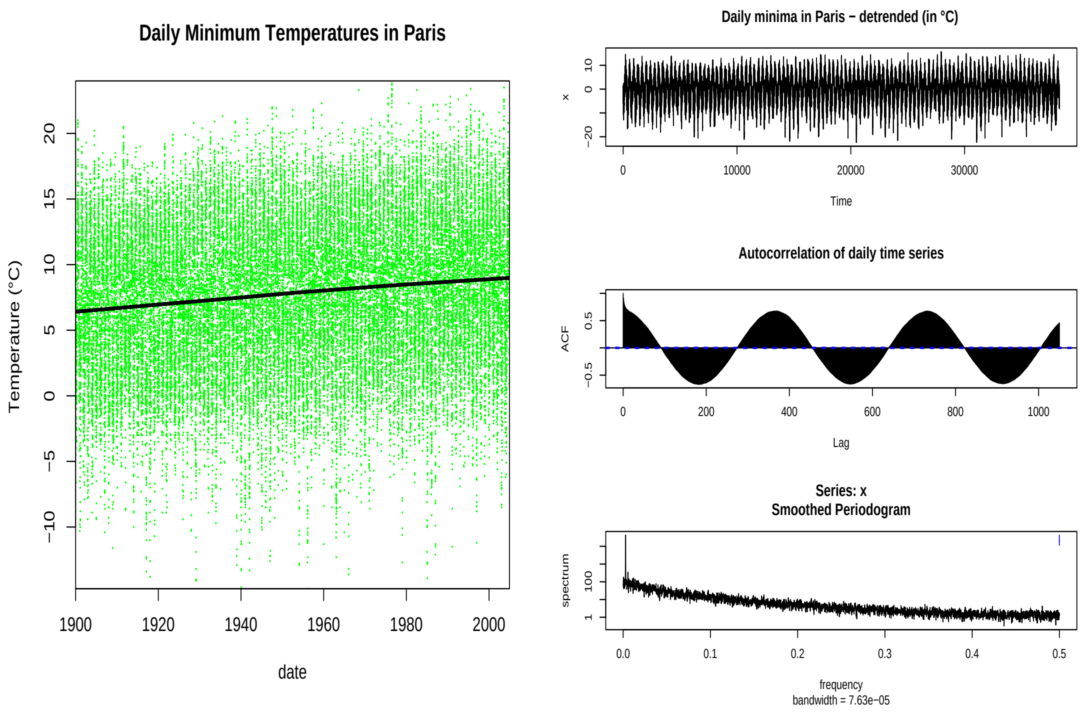


Figure 8.12: Trend of the series, and analysis of the series of residuals.

Remark 8.3.1. *More generally, Deo (1997) considered some nonparametric regression, and Beran et al. (2002) robust local polynomial regression, with long-range dependence errors, that is*

$$X_i = g(t_i) + Y_i \text{ where } t_i = \frac{i}{n}, \text{ for all } i = 1, 2, ..., n,$$

where g is an unknown function (sufficiently smooth), and $(Y_t)_{t\in\mathbb{Z}}$ is a stationary process with long-range dependence (see also Remark 8.3.2).

In the case of the minimal temperature in Paris, the model is

$$X_t = \underset{(0.05878)}{6.33843} + \underset{(0.0000012)}{0.000068371}t + Y_t,$$

where $t = 1, 2, ..., 38330$ (from January 1900 till September 2004). From this point, the residual series $(Y_t)_{t\in\mathbb{Z}}$ will become our series of interest, that we should model.

Modeling the seasonal effect with and long range noise

Here, parameter u is related to the seasonality of the series, or formally

$$\widehat{u} = \cos(\widehat{\omega}_0) \text{ where } \widehat{\omega}_0 = \text{argmax}_{\omega\in[0,2\pi]}\{I_n(\omega)\}.$$

Hence, because of the annual cycle of temperature above the tropic, $\widehat{u} = \cos(2\pi/365)$.

And, from section 8.2.5, recall that Whittle's approach can be considered (see Hosoya (1997) or Taqqu and Teverovsky (1997)) in order to estimate paramaters of the GARMA process

$$(\boldsymbol{\alpha}, \sigma, d) = ((\phi_1, ..., \phi_p, \theta_1, ..., \theta_q), \sigma, d),$$

where σ^2 is the variance of the white noise. In the case of daily temperature in Paris, $\widehat{d} = 0.185$, $\widehat{\phi}_1 = 0.56$, $\widehat{\sigma^2} = 2.22^2$.

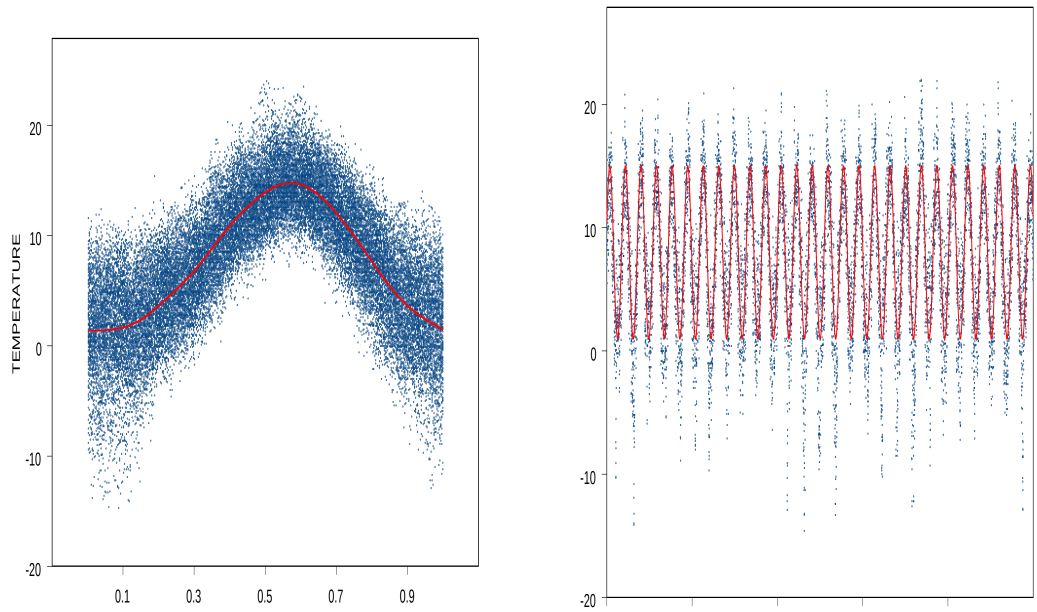


Figure 8.13: Seasonal cycle of the series.

Remark 8.3.2. *An alternative could be to consider SEMIFAR models, introduced by Beran (1999)* - semiparametric fractional autoregressive. *Recall that an ARIMA(p, δ,0) model - $\delta \in \mathbb{N}$ - satisfies*

$$\Phi(L)((1-L)^{\delta}X_t - \mu) = \varepsilon_t.$$

ARFIMA(p, d,0) processes - $d \in (-1/2, 1/2)$ - have been defined as

$$\Phi(L)(1-L)^{d}(X_t - \mu) = \varepsilon_t.$$

By extension, $(X_t)_{t\in\mathbb{Z}}$ will be called SEMIFAR if there are $\delta \in (0,1)$ and $d \in (-1/2, 1/2)$ such that

$$\Phi(L)(1-L)^{d}((1-L)^{\delta}X_t - g(t)) = \varepsilon_t,$$

for some with noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$, and some smoothed function g.

8.3.3 Long range or fat tailed distribution ?

As pointed out in Smith (1993), "*we do not believe that the autoregressive model provides an acceptable method for assessing theses uncertainties*". Based on the optimistic scenario (end of the linear trend), three models will be compared,

- a Gaussian model, with a seasonal effect, and long-range dependence (the previous model),
- an heavy tails model, with a seasonal effect, and short-range dependence,
- an as a benchmark, a Gaussian model, with a seasonal effect, and short-range dependence.

Consider generally the case of a linear process, $(Y_t)_{t\in\mathbb{Z}}$, such that

$$Y_t = \sum_{k\in\mathbb{Z}} \psi_k \varepsilon_{t-k}, t \in \mathbb{Z},$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise, centered. Note that $(Y_t)_{t\in\mathbb{Z}}$ is stationary, and it is stationary in the L^2 sense if

$$Var(Y_t) = Var(\varepsilon_t) \sum_{k\in\mathbb{Z}} \psi_k^2 < \infty.$$

Example 8.3.3. *If $(Y_t)_{t\in\mathbb{Z}}$ is a stationary $AR(1)$ process, $Y_t = \alpha Y_{t-1} + \varepsilon_t$, with $|\alpha| < 1$, can be inverted as follows,*

$$Y_t = \alpha Y_{t-1} + \varepsilon_t = \alpha[\alpha Y_{t-2} + \varepsilon_{t-1}] + \varepsilon_t = \ldots = \sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k}.$$

Following examples of section 8.2.3, consider the following fractionary integrated process, AFIMA$(0, d, 0)$,

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \text{ were } \psi_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)},$$

where $-1/2 < d < 1/2$.

As pointed out in the thesis, the study of extremal events is related to the fatness of tails.

Definition 8.3.4. *A random variable Z is said to have heavy tails if its distribution is in the max-domain of attraction of the Fréchet distribution. Hence, there exists $\alpha > 0$ such that $\mathbb{P}(Z > z) = z^{-\alpha}\mathcal{L}(z)$ where $\mathcal{L}$ is slowly varying.*

Hence, Gaussian distributions do not have fat tails (they belong to the max-domain of attraction of the Gumbel), but among the class of elliptical distributions (see section 1.5 of the thesis), the t distribution has heavy tails. Recall that its density is

$$f_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{2}\right)^{-(\nu+1)/2}.$$

Remark 8.3.5. *Note that the t-distribution can be written as a scale mixture of Gaussians where the mixture factor has an inverse gamma distribution (Kelker (1970))*

Definition 8.3.6. *A random variable Z is said to have a stable distribution if for every $n \in \mathbb{N}$, there exist constant $a_n > 0$ and b_n such that the sum $Z_1 + ... + Z_n$ has the same distribution as $a_n Z + b_n$ for all i.i.d. random variables $Z_1, ..., Z_n$, with the same distribution as Z.*

Heavy tailed time series, case of infinite variance

In previous sections, we have considered the case of of stationary series, with Gaussian noise. More generally, assume that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise with a symmetric α-stable distribution, with generating function

$$\phi(t) = \mathbb{E}(e^{it\varepsilon}) = \exp(-c|t|\alpha), t \in \mathbb{R},$$

where $\alpha \in (0, 2]$ and $c > 0$. Note that $Var(\varepsilon_t)$ is finite if and only if $\alpha = 2$. Further, note that

$$Y_t \stackrel{\mathcal{L}}{=} \varepsilon_t \left(\sum_{k\in\mathbb{Z}} |\psi_j|^\alpha\right)^{1/\alpha},$$

(from the definition of stable distributions, see Feller (1971)). It comes that Y_t is then also α-stable. Further, from Theorem 22.8 in Billingsley (1999), it can be obtained that $(Y_t)_{t\in\mathbb{Z}}$ is well defined if and only if

$$\sum_{k\in\mathbb{Z}} |\psi_j|^\alpha \text{ is finite.}$$

Assuming that time series had finite variance was historically crucial to legitimate regression models (since L^2 is an Hilbert space, and projection techniques can be used). In the case variance is infinite, alternative ideas should be used (see Section 13.3 in Brockwell and Davis (1991), or Bhansali (1996) for a detailed survey).

Further, autocorrelation functions cannot be defined, since variance is not finite. Nevertheless, one can define quantities

$$\rho(h) = \frac{\sum_{k\in\mathbb{Z}} \psi_k \psi_{k+|h|}}{\sum_{k\in\mathbb{Z}} \psi_k^2} \text{ for all } h \in \mathbb{Z}.$$

In the case of finite variance processes, those quantities can be interpreted as autocorrelations, but not anymore in the general case. In the Gaussian case, recall that for each m,

$$\sqrt{n}(\widehat{\rho}_n(h) - \rho(h))_{h=1,\ldots,m},$$

has a non-degenerated limiting Gaussian distribution (Proposition 7.3.4. in Brockwell and Davis (1991)). In the case of α-stable processes,

$$(n \log n)^{1/\alpha}(\widehat{\rho}_n(h) - \rho(h))_{h=1,\ldots,m},$$

has a non-degenerated limiting distribution, related to some α-stable distributions (Theorem 7.3.2 in Embrechts, Klüppelberg and Mikosh (1997)).

Similarly, the spectral density, which was related to the Fourier series of autocorrelation cannot be defined anymore. Nevertheless, the periodogram

$$\widehat{I}_n(\omega) = \frac{1}{n}\left\|\sum_{t=1}^{n} Y_t e^{-i\omega t}\right\|^2 = \sum_{|h|<n} \widehat{\gamma}_n(h) e^{-i\omega h}, \omega \in [-\pi, \pi] \tag{8.2}$$

is still defined. The periodogram is not a consistant estimator of the spectral density, but under mild condition, it is not far away from consistency (see Sections 10.1 and 10.3 in Brockwell and Davis (1991)). Hence, if $Var(\varepsilon_t) < \infty$, for all frequencies $0 < \omega_1 < \ldots < \omega_m < \pi$, for any $m \in \mathbb{N}$

$$(\widehat{I}_n(\omega_k))_{k=1,\ldots,m} \xrightarrow{\mathcal{L}} \frac{Var(\varepsilon_t)}{2} \left(\|\psi(\omega_k)\|^2 (U_k^2 + V_k^2) \right)_{k=1,\ldots,m},$$

where U_i, V_j's are i.i.d. $\mathcal{N}(0,1)$ random variables, where

$$\|\psi(\omega)\|^2 = \left\|\sum_{t\in\mathbb{Z}} \psi_t e^{-i\omega t}\right\|^2.$$

Similarly, as $\widehat{\rho}_n(h)$ and $\rho(h)$ were still defined if $Var(\varepsilon_t) = \infty$, one can defined $\widehat{I}_n(\omega)$ from equation 8.2. And thus, from Theorem 7.4.3. in Embrechts, Klüppelberg and Mikosh (1997), one gets that for all frequencies $0 < \omega_1 < \ldots < \omega_m < \pi$, for any $m \in \mathbb{N}$

$$(\widehat{I}_n(\omega_k))_{k=1,\ldots,m} \xrightarrow{\mathcal{L}} \frac{Var(\varepsilon_t)}{2} \left(\|\psi(\omega_k)\|^2 (U_k^2 + V_k^2) \right)_{k=1,\ldots,m},$$

where $(U_1, \ldots, U_m, V_1, \ldots, V_m)$ is an α-stable random vector (with maybe non independent components).

If $(Y_t)_{t\in\mathbb{Z}}$ is an ARMA process with an α-stable white noise, Whittle estimator (based on the spectral density) can be used. Set $\boldsymbol{\alpha} = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$, where

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, t \in \mathbb{Z}.$$

That process can be inverted under some assumptions (see), and $(Y_t)_{t\in\mathbb{Z}}$ is then a linear process, with parameters $\psi_j(\boldsymbol{\alpha})$, $j \in \mathbb{N}$. If

$$\|\psi(\omega, \boldsymbol{\alpha})\|^2 = \left| \sum_{t\in\mathbb{Z}} \psi_t(\boldsymbol{\alpha}) e^{-i\omega t} \right|^2,$$

Whittle estimator is

$$\widehat{\boldsymbol{\alpha}}_n = \text{argmax}_{\boldsymbol{\alpha}} \left\{ \frac{2\pi}{n} \sum_{|t|<n} \frac{\widehat{I}_n(2\pi t/n)}{\|\psi(2\pi t/n, \boldsymbol{\alpha})\|^2} \right\}.$$

Then $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha})$ has a non-degenerated limiting distribution, if $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a Gaussian white noise. If $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a stable white noise, $(n/\log n)^{1/\alpha}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha})$ has a non-degenerated limiting distribution (see theorem 7.5.3. in Embrechts, Klüppelberg and Mikosh (1997)). Hence, note that the rate of convergence is then $O((n/\log n)^{-1/\alpha})$, which is faster than $O(n^{-1/2})$ is the Gaussian case.

Heavy tailed time series, case of finite variance

An alternative is to consider the case of a linear process where the noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ is supposed to be heavy tailed, i.e. in the maximum domain of attraction of the Fréchet distribution. Hence, assume that $\mathbb{P}(\varepsilon_t > x) = x^{-\alpha}\mathcal{L}(x)$, for some slowly varying function $\mathcal{L}$, with $\alpha > 2$.

If variance is finite, least-squares methods can be motivated (see Brockwell and Davis (1991)). An alternative is to consider maximum likelihood methods (see Section 5.2. in Brockwell and Davis (1991)). In a general context (with finite variance noise), a maximum likelihood procedure for estimating an ARMA process without the Gaussian assumption has been developed by Lii and Rosenblatt (1996). The derivation of the likelihood is given in Appendix.

Remark 8.3.7. *Polasek and Pai (1998) proposed to use Gibbs sampler for Bayesian estimation, with a white noise t-distributed. The idea is to use the fact that the t-distribution can be written as a mixture of Gaussians (see Remark 8.3.5)*

From a quick look at autocorrelations of the series, we have modeled $(Y_t)_{t\in\mathbb{Z}}$ with a Gaussian ARMA(2,2) model,

$$Y_t = \underset{(0.0414)}{1.428}\, Y_{t-1} - \underset{(0.0319)}{0.479}\, Y_{t-2} + \varepsilon_t - \underset{(0.0414)}{0.666}\, \varepsilon_{t-1} - \underset{(0.0072)}{0.103}\, \varepsilon_{t-2}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is stationary with variance $\widehat{\sigma^2} = 5.023$. Figure 8.14 is the analysis of series $(\varepsilon_t)_{t\in\mathbb{Z}}$: the white noise assumption is relevant, but it is far from being Gaussian, as shown on Figures 8.15 and 8.16. Hence, three estimations are performed and outputs are presented in Table 8.3.

Remark 8.3.8. *From Table 8.3, it appears that maximum likelihood estimates with the Gaussian and the t distribution are almost equal. Hence, in Lii and Rosenblatt (1996), simulations are performed when residuals are t-distributed (with 4 degrees of freedom). They observed that "estimates of parameters are quite accurate" [...] especially "when the roots are moved farther away from the unit circle".*

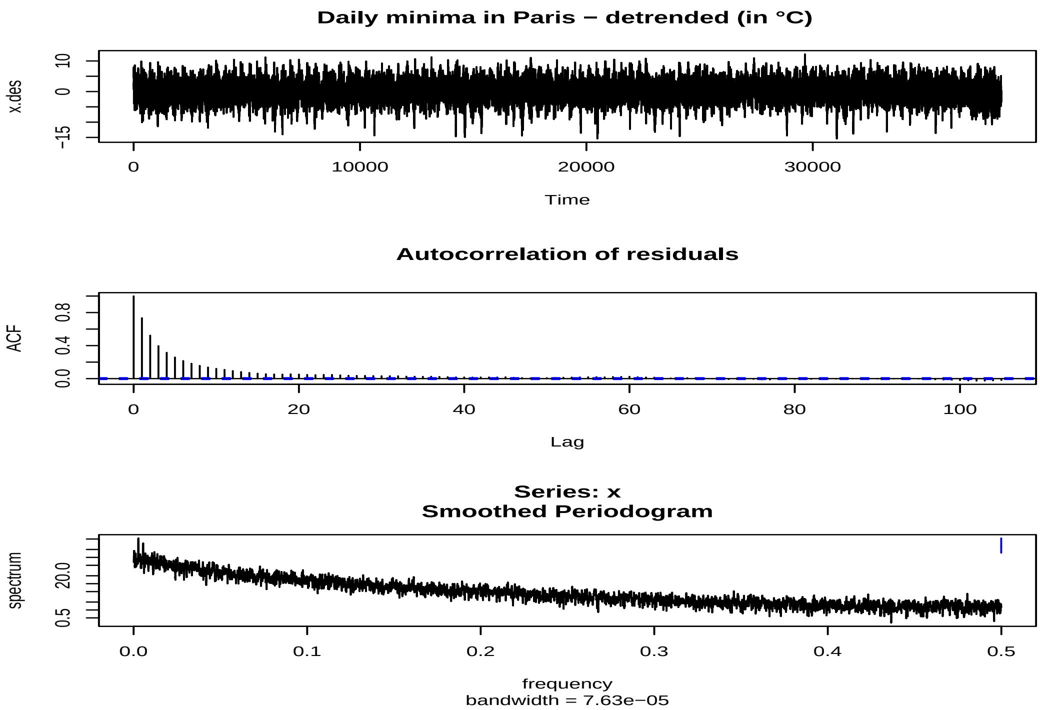


Figure 8.14: Residuals $(\varepsilon_t)_{t\in\mathbb{Z}}$.

		$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\sigma}^2$
Gaussian	ML	1.4196	-0.4733	-0.6581	-0.1032	5.023
		(0.0419)	(0.0322)	(0.0419)	(0.0072)	
t $(\nu = 20)$	ML	1.4191	-0.4738	-0.6571	-0.1032	5.023
		(0.0418)	(0.0323)	(0.0418)	(0.0073)	
t $(\nu = 5)$	ML	1.4134	-0.4725	-0.6551	-0.1035	5.023
		(0.0418)	(0.0324)	(0.0418)	(0.0072)	
	LS	1.4020	-0.4600	-0.6406	-0.1038	5.023
		(0.0418)	(0.0319)	(0.0417)	(0.0071)	

Table 8.3: Parameter estimation for the ARMA process.

8.3.4 Return period for two scenarios

Two scenarios on the future evolution of the linear trend will be considered in this section:

- an *optimistic* scenario, where we assume that there will be no more increasing trend in the future,
- a *pessimistic* scenario, where we assume that the trend will remain, with the same slope.

Definition of the heat wave

In order to compare the two models, two alternative definitions of the heat wave will be considered (both may characterize the phenomena of the beginning of August in Paris),

- during 11 consecutive days, the temperature was higher than 19° C (type (A)),

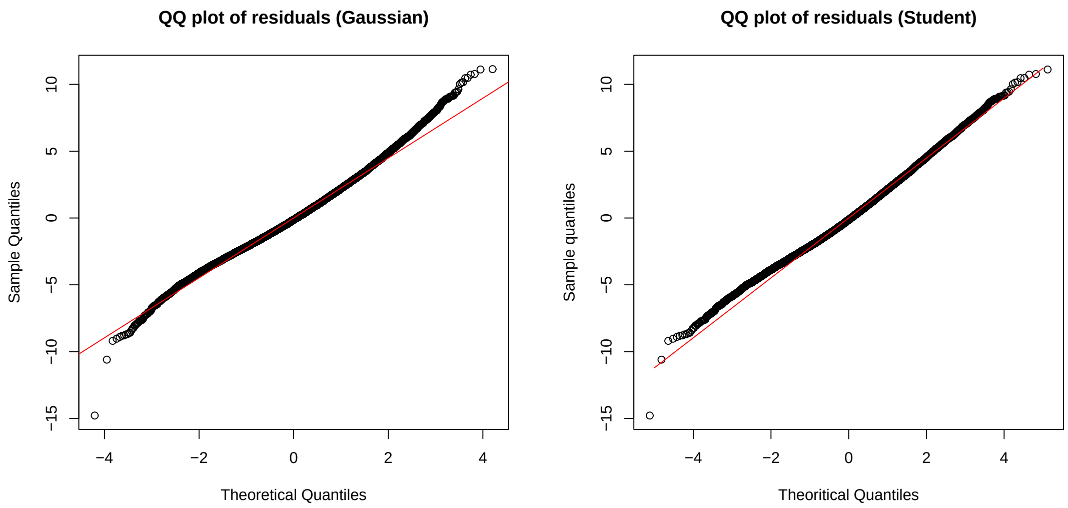


Figure 8.15: Residuals $(\varepsilon_t)_{t\in\mathbb{Z}}$, versus Gaussian and t-distribution.

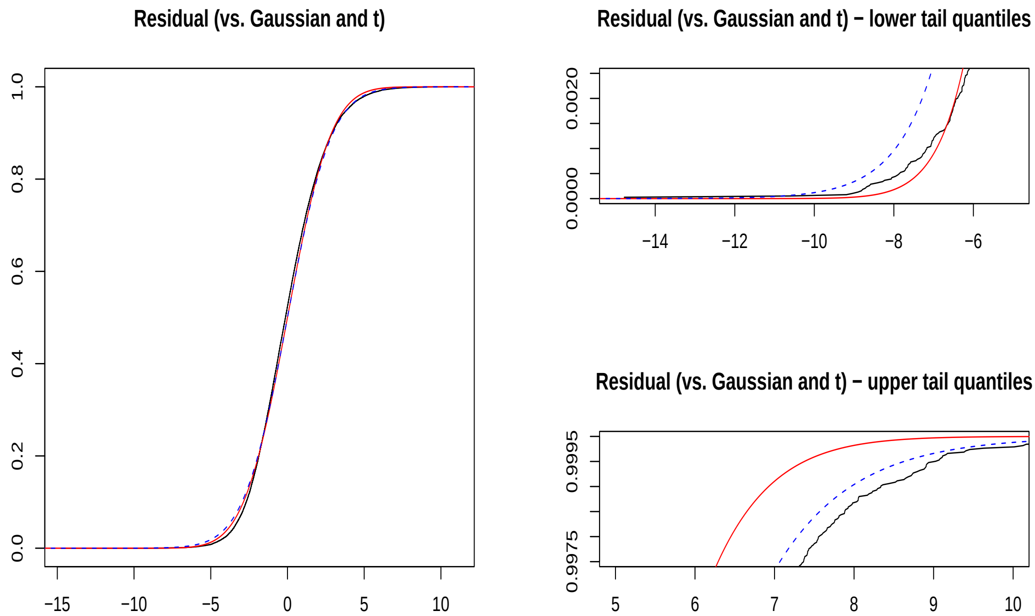


Figure 8.16: Residuals $(\varepsilon_t)_{t\in\mathbb{Z}}$, Gaussian (plain line) or t distributed (dotted line).

- during 3 consecutive days, the temperature was higher than 24° C (type (B)).

The outputs of simulations can be visualized on Figures ?? and ??. Here 10,000 simulations over 300 years have been used. The large plain line is the result of GARMA processes simulations, i.e. long range dependence, and Gaussian noise. The dotted line is the result of ARMA processes simulations, with a Student noise (heavier tails than the Gaussian, plotted with a light gray line).

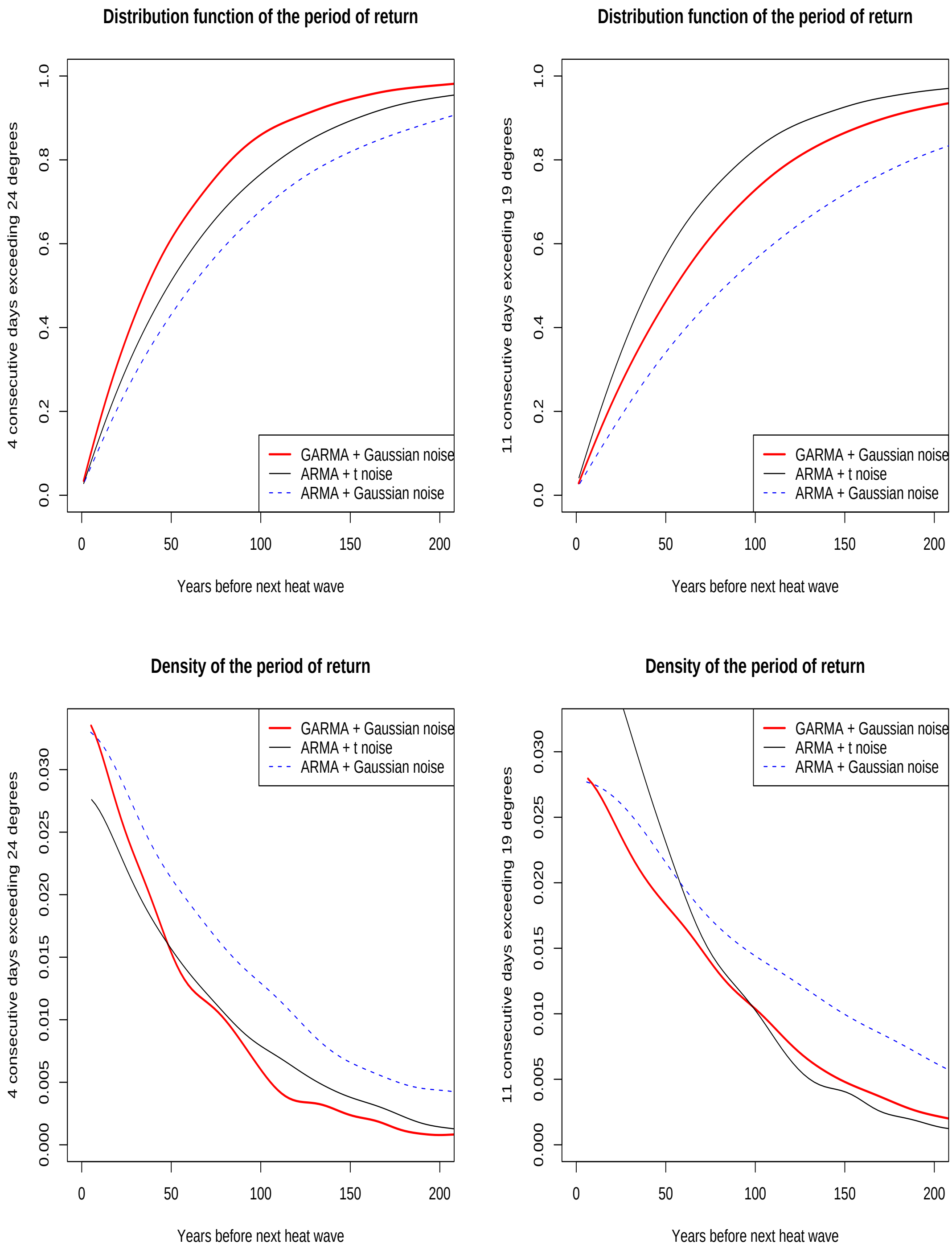


Figure 8.17: Survival distributions and densities of time before the next heat wave event, when heat wave is 11 consecutive days with temperature higher than 19° C, on the left, and 3 consecutive days with temperature higher than 24° C, on the right. We assume no more linear trend in the future (optimistic scenario).

Note that for the two scenarios and the two alternative models, different periods of returns are obtained,

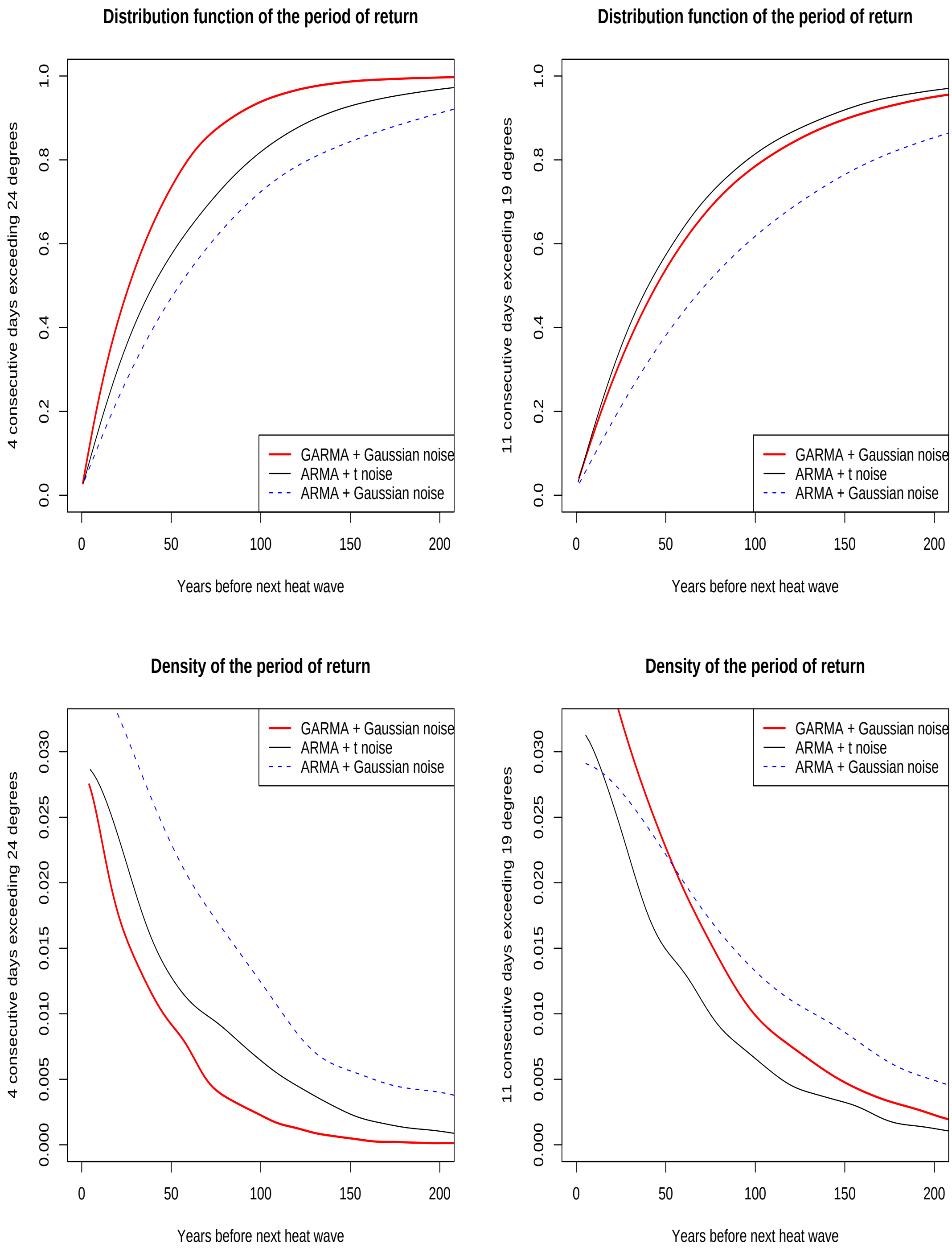


Figure 8.18: Survival distributions and densities of time before the next heat wave event, when heat wave is 11 consecutive days with temperature higher than 19° C, on the left, and 3 consecutive days with temperature higher than 24° C, on the right. We keep the linear trend in the future (pessimistic scenario).

Remark 8.3.9. *In this section, the aim was to estimate the return period using only time series techniques (i.e. endogeneous estimation). In order to build up warnings, many factors should be*

	short memory short tail noise	short memory heavy tail noise	long memory short tail noise
optimistic	88 years	69 years	53 years
pessimistic	79 years	54 years	37 years

Table 8.4: Periods of return (expected value, in years) before the next heat wave similar with August 2003 (type (A)).

	short memory short tail noise	short memory heavy tail noise	long memory short tail noise
optimistic	115 years	59 years	76 years
pessimistic	102 years	51 years	64 years

Table 8.5: Periods of return (expected value, in years) before the next heat wave similar with August 2003 (type (B)).

included (see e.g. Black et al. (2004)). Hence, the probability to have a very hot summer was quite large even in May, based on anomalies study.

8.4 Floods, how to chose between Hurst and Gumbel ?

As recalled in Mandelbrot and Wallis (1968), models in hydrology initally assumed that "*precipitation were random and Gaussian* [....] *with successive year'precipitation either mutually independent or with a short memory*". Those two assumptions have been treated separately, when dealing with annualized maxima, Gumbel assuming that observations were independent but with a non-Gaussian distribution (the so-called Gumbel distribution), while Hurst assumed that yearly maxima were Gaussian but with long memory. We shall recall briefly in the two next paragraphs the main ideas of those two approach, from independence (Gumbel) of annualized maxima to range memory dependence (Hurst). As we will see those two approaches yield completely different results in terms of risk measurement (e.g. return periods). The last paragraph of this section will briefly highlight that this problem of two approaches also holds in finance, and that some alternative have arisen.

8.4.1 Gumbel's independence and extreme value models: a static model

Consider $(X_i)_{i\in\mathbb{N}}$ an i.i.d. sequence of random variables. Define, for all $n \in \mathbb{N}^*$ the associated i-th statistic order $X_{i:n}$. One gets easily that $X_{n:n} \overset{a.s.}{\to} x_F = \sup\{x \in \mathbb{R}, F_X(x) < 1\}$ (x_F is then infinite for non-bounded random variables). This limiting results does not provide much information in terms of limiting distribution. Following the construction of "*stable*" distribution, in the context of extremes, the idea will be to consider normalizing sequences (a_n) and (b_n) so that there exists a limiting distribution for the maxima $X_{n:n}$. Fisher-Tippett Theorem (see Theorem 3.2.3 in Embrechts, Klüppelberg and Mikosch (1997)) states that the Generalized Extreme Value

(GEV) distribution, defined as

$$H_{\xi}(x) = \begin{cases} \exp\left(-\left(1+\xi x\right)^{-1/\xi}\right) & \xi \neq 0 \\ \exp\left(-\exp\left(-x\right)\right) & \xi = 0, \end{cases} \tag{8.3}$$

where $1 + \xi x > 0$, appears as the limiting distribution of the normalized maxima, for some appropriate normalizing constants. This is also the so-called Von Mises parameterization.

Let $Y_1, ..., Y_m$ denote independent and identically distributed variables, say daily discharge, or daily maximum river height. Let X_i denote the annualized maxima for year i. Then, the X_i's are then i.i.d., and moreover, their distribution should be close to some GEV distribution. Notice that the special case $\xi = 0$ (Gumbel law) was obtained by Gumbel when studying hydrological data.

Example 8.4.1. *The data considered here are average daily streamflows of the Saugeen River, measured between 1914 and 2003. The autocorrelation function of the annual maxima series (Figure 8.19, on the left hand side) seems at first to validate the independent sampling hypotheses since no autocorrelation is significant. When fitting a GEV distribution on these annual maxima (see Figure 8.19, on the right hand side), one finds that the Gumbel distribution, with a tail parameter* $\widehat{\xi} = -0.02$*, and a standard deviation* 0.07 *(and* $\widehat{\mu} = 238$ *and* $\widehat{\sigma} = 90.6$*) is adequate. Testing the nullity of the* ξ *coefficient with a modified likelihood ratio test (see Hosking (1984)) confirms this conclusion. (The hypothesis that the annual maxima do not follow a Gumbel distribution is rejected with a threshold of 1%).*

But as pointed out in Gumbel (1958,page 238), "*it must be admitted that the good fit cannot be foreseen from the theory, which is based on three assumptions (1) the distribution of the daily discharges is of the exponential type (2)* $n = 365$ *is sufficiently large (3) the daily observations are independent* [...] *The thirst assumption does not hold.*"

Note that the GEV approach, induced by Gumbel's work, has be in use for practical work for more than 30 years, since it was recommended for hydrological applications in NERC (1975). But again, as mentioned in Gumbel (1958, page 164), "*the exact distribution of extreme values holds only for independent observations*".

8.4.2 Hurst's strong dynamics and fractional processes: a dynamic model

The problem studied by Edwin Hurst, who spent a lifetime studying the Nile and the problems related to water storage, was to determine the design of an ideal reservoir based upon the given record of observed discharges from the lake. He investigated many natural phenomena, such as river discharges, mud sediments and tree rings, and introduced the so-called R/S statistics, where R denotes the "*range*" (the difference between the maximum and the minimum influx through a given period of time τ), and S the "*standard deviation*" on the same period of time. He noticed that the observed rescaled range R/S for many records in time was well described by the empirical relation $R/S = (\tau/2)^H$, where H is called Hurst component (but called K by Hurst) . When observations are independent, R/S should be proportional to $\sqrt{\tau}$ (i.e. $H = 1/2$), but as pointed out by Hurst, in many natural phenomena, we have $H > 1/2$. The explanation given by Hurst of this characteristic was (see e.g. that "*the discharge of a river depends not only on the recent precipitation, but also on earlier rainfalls. The flow in a large river system such as the Nile must depend on the water content in a large drainage area. The amount of water stored in the drainage area will increase in prolonged periods of higher than average precipitation. The excess amount of water stored will then contribute to the discharge in drier years* [...] *for river discharges, the fractal nature* (i.e. the memory effect) *of the drainage area may also contribute*

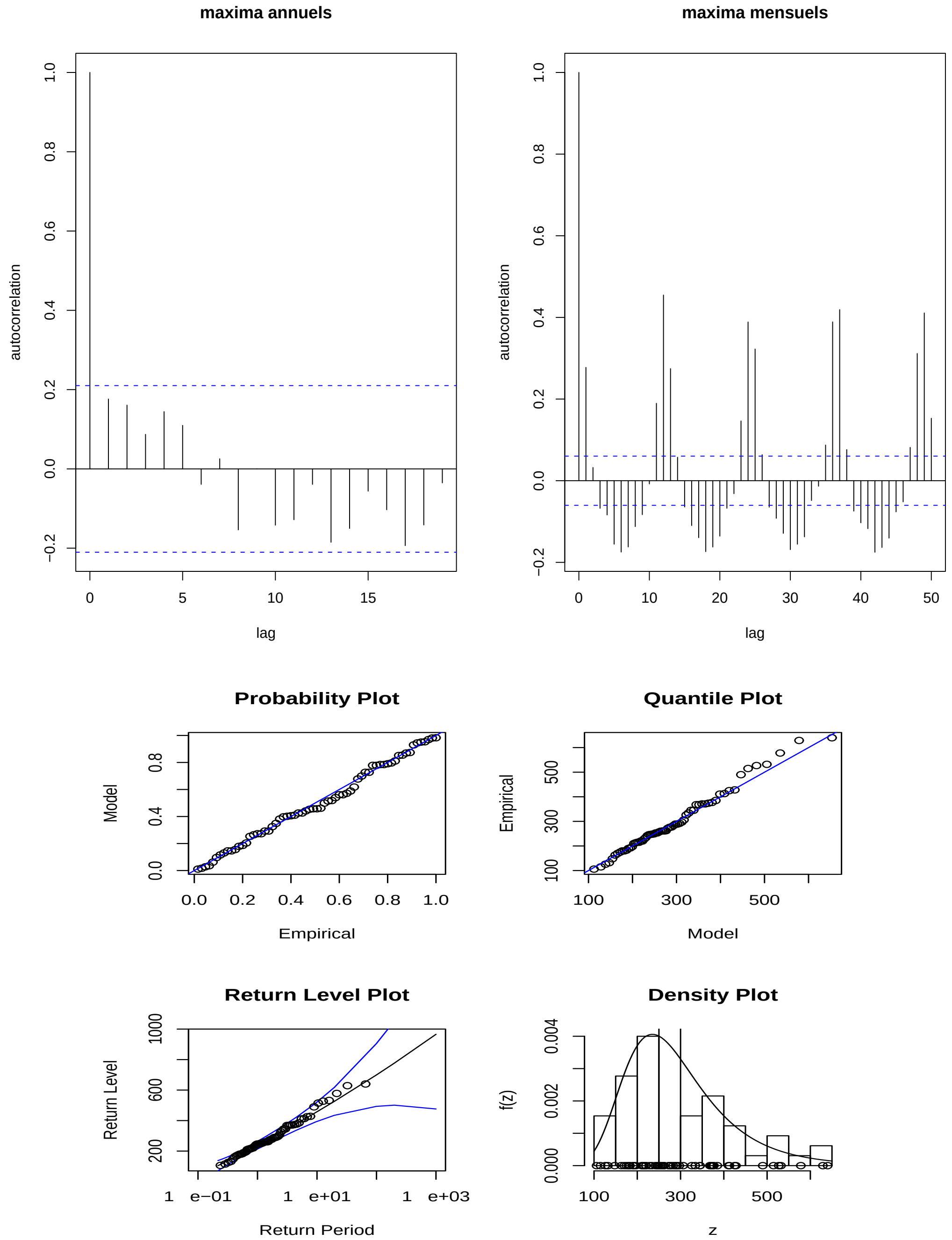


Figure 8.19: Autocorrelation function of the annual streamflow maxima. GEV fit.

to the fractal behavior of river discharges." This phenomena is called "*long-range dependence*" or "*long-term dependence*" (see e.g. Probst and Tardy (1987), Pelletier and Turcotte (1997), or Koscielny-Bunde et al. (2006)), "*fractal behavior*" or (see e.g. Puente (1996), Pandey, Lovejoy and Schertzer (1998), "*scale-invariant*" (see e.g. Foufoula-Georgiou (1999))

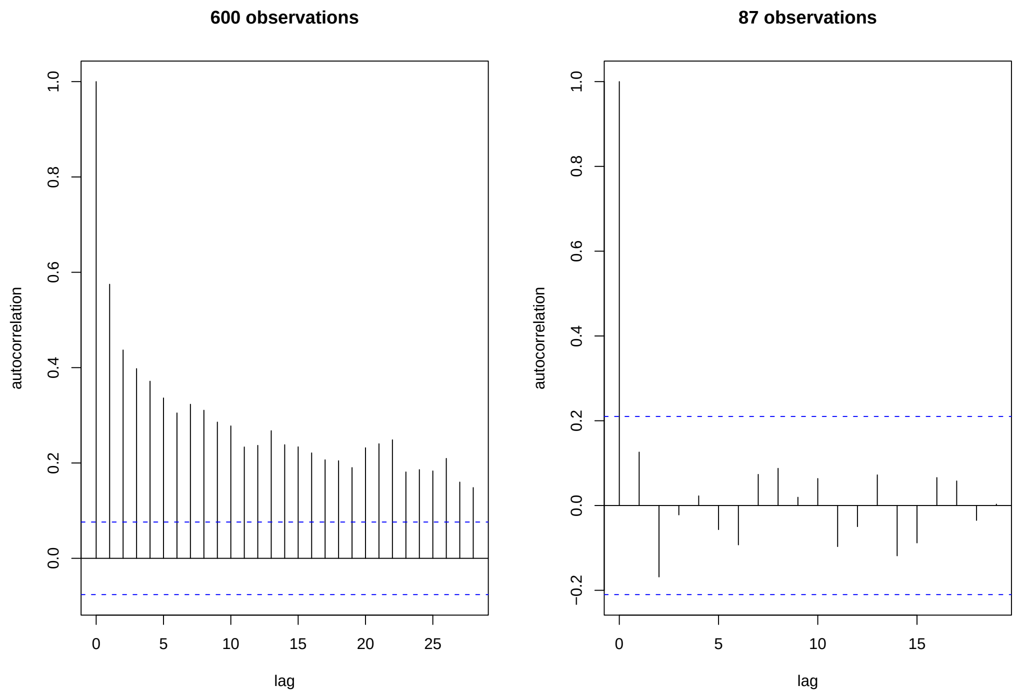


Figure 8.20: Autocorrelation function of the Nile annual maxima series with 600 and 87 observations.

The two approaches presented so far (supposing that annual maxima are independent, as assumed by Gumbel or exhibit long memory as assumed by Hurst) seem to be incompatible. The differences seem to stem from the length of the studied series. Indeed, Hurst had noticed that the *"Joseph effect"* was noticeable only if a very long series was considered. On shorter series, and at different times, the long memory effect vanishes. This would mean, in our case, that we would ignore a dependency that is here, as can be seen on figure 8.20, where 87 is precisely the number of years of data available for the Saugeen river. Hence, it seems reasonable to ask whether the independent annual maxima assumption is relevant : assuming that the data are independent and identically distributed is only possible because hydrologists usually have short series, while the data are actually dependent.

Gumbel's approach lets one model extremal events (i.e. floods), while Hurst's approach does not. How can we model extremal events, while at the same time retaining the underlying dynamic? This leads us to a parallel between high-frequency data finance and hydrology.

8.4.3 The analogy with financial data

The historical approach in hydrological series, mentioned in Mandelbrot and Wallis (1968) also holds in finance: in 1900, Bachelier assume that daily returns of stock prices were independent, and Gaussian random variables. Later one, dynamics was introduced, e.g. based on long range dependence models on stock returns (see e.g. Greene and Fielitz (1977), Aydoyan and Booth (1988) or Lo (1991)), in exchange rates (see e.g. Booth, Kaen and Koros (1987)) or in interest rates (see e.g. Shea (1991)). Moreover, early evidence on periodicities and duration clustering entailed temporal dependence, as discussed in Engle and Russell (1998).

A natural idea to get better models, is to model no only the price at the end of each day, but for all transactions. But all transactions data are inherently irregularly spaced in time. And since most econometric models are specified for *fixed intervals* this poses an immediate complication. Analogously, flood evens are also irregularly spaced in time.

Although empirical evidence has been observed very early in finance - often called the week-end effect, or January effect-, it did not receive much interest until Engle and Russell (1998) introduced the ACD model (Autoregressive Conditional Duration) for modeling the financial duration process. The idea was to consider T_i the time at which the i-th trade occurs, and to define the following set of observations,

$$\{(D_i, (P_i, V_i))\}_{i\in\mathbb{N}} \text{ where } D_i = T_i - T_{i-1},$$

and where P_i and V_i denote respectively the price of the asset when traded, and the volume.

A natural parallel can be obtain for hydrological series, were *flood* events can be characterized through variables of interest T_i's (time of the *flood*, or at least the event {*the river level exceeded some given threshold*}, P_i's (the peak associated to the ith flood event) and V_i's (the volume of surplus water associated to the ith flood event).

Hence, the goal of this paper is to show that instead of focusing on annualized observations for hydrological series, or daily observations for financial series, high frequency data can be considered. As will se in Section 8.4.4 a model, which is slightly different from the one shown previously, can be used to model high frequency data in hydrology. In finance, time is usually treated as one variable, i.e. the time of the transaction. The transaction is a discrete event. However, in hydrology, a flood event cannot be considered as a discrete event, since a flood can last many days. Hence, two time processes will be considered: the beginning of flood process, and the flood duration process.

8.4.4 Dynamic flood modelling

The first dynamic flood models were developed by Todorovic and Zelenhasic (1970), and Todorovic and Rousselle (1970). The idea is to study the duration between floods, as well as for each flood a mark (the river maximum level for example). The dynamic aspect of the phenomenon is therefore not lost. However, these first models still assume that flood events are independent.

Basics on point processes

These first models use results from point process theory, which we recall here briefly.

Definition 8.4.2. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *a probability space. Let* (τ_n) *be an increasing sequence of random variables such that* $\tau_0 = 0$ *and* $\forall i \geq 1, \tau_i > \tau_{i-1}$. *A counting process (or point process) is :*

$$N_t = \sum_{n=1}^{\infty} \mathbb{I}_{(\tau_n \leq t)}$$

N_t is the number of events that occurred before date t, and the sequence τ_n is the sequence of occurring times of the said events. The model used usually in hydrology assumes that the occurring process is a Poisson process, defined as follows :

Definition 8.4.3. *Let's consider the process* (N_t) *defined previously. If for all* $k \geq 1$ *et for all increasing sequence* $0 = t_0 < t_1 < ... < t_k$, *the random variables* $N_{t_1}, N_{t_2} - N_{t_1}, ..., N_{t_k} - N_{t_{k-1}}$

are independent and follow a Poisson distribution with parameter $\lambda(t_i - t_{i-1})$ for $i \in [1, .., k]$ then the point process is called homogeneous Poisson process.

The occurring times sequence has then the following property :

$$\forall n \geq 1, \quad \tau_n - \tau_{n-1} \sim \mathcal{E}(\lambda)$$

The duration between two events follows an exponential law. The parameter λ is called the intensity of the Poisson process.

Non-homogeneous Poisson processes extend the class of homogeneous Poisson process. The λ parameter is not constant anymore and is a deterministic function of time. In both cases, λ has the following property:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\left(N(t + \Delta t) > N(t)\right)}{\Delta t}$$

The $\lambda(t)$ parameter can be interpreted as the instantaneous occurrence probability of an event at date t. If a characteric is associated to a point process, (for example a river level for the flood occurrence process), the point process is then called a marked point process and the characteristic is called a mark.

This model may seem limited in so far as it only has one mark : it could be interesting to integrate both durations (the duration of the flood and the duration between two floods) into the point process driving the occurrence time, and, if a flood occurs, associate two marks (peak and volume) to the same point process.

Two durations model

The basic idea is to consider two point processes, with the first one driving the second one. The driving process here would be the flood occurrence process, and the second one would define the duration between the end of the last flood and the beginning of the next one.

The main advantage of such a model is that each beginning of flood event introduces a waiting time for the next end of flood event, without making any assumption on the type of dependency that may exist between the two duration variables.

From a mathematical standpoint, let t_i be the occurring time of the i-th flood, and t'_i the ending time of the i-th flood. Let's consider the following durations : $X_i = t_{i+1} - t_i$, the duration between two flood beginnings, and $Y_i = t_{i+1} - t'_i$, the duration between the end of a flood and the beginning of the next one. We can then deduce the duration of the flood, $D_i = X_i - Y_i$. A graph showing the relationship between those durations is shown figure 8.21.

Model specification

Without loss of generality, the joint density of the bivariate duration process $\{(X_i, Y_i)\}$ can be written as the product of a conditional density and of the density of one the marginals:

$$p(x_i, y_i | \mathcal{H}_{i-1}, \theta) = g(y_i | X_i = x_i, \mathcal{H}_{i-1}, \theta_2) f(x_i | \mathcal{H}_{i-1}, \theta_1)$$

where g can be seen as the conditional density of flood endings , f the flood beginning density, $(\mathcal{H}_i)$ the filtration generated by $(X_k)_{k \leq i}$ and $\theta = (\theta_1, \theta_2)$ is the model parameter.

A class of models for f et g, which does not make the assumption that observations are independent has been developed by Engle adn Russell (1998), to study the durations between financial transactions. Having noticed that there were many more financial transactions in the morning than at lunch time, they wanted to study the influence of the number of transactions on the stock volatility. They therefore chose to model the duration between transactions as well as

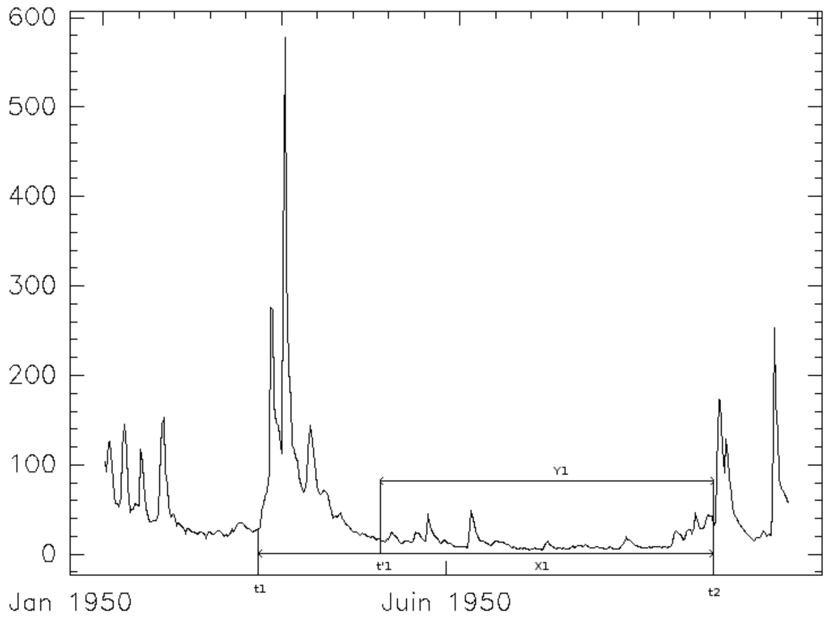


Figure 8.21: Two durations model

the stock price, turning calendar time into economic time to only keep the necessary information. Here, we would like to study floods instead of transactions, but the approach remains the same. These models are called ACD models (*Autoregressive Conditional Duration*).

The ACD(p, q) model

Let $z_i = (X_1, ..., X_{i-1})$ the sequence of past duration realizations. Let $\Psi_i > 0$ be a function of z_i and of θ_1 ($\theta_1 \in \mathbb{R}^s$). The following model is called an ACD(p, q) model:

$$\begin{cases} X_i = \Psi_i \epsilon_i \text{where } (\epsilon_i) \text{ are i.i.d and independent from } z_i \\ \mathbb{E}[X_i | z_i] = \Psi_i \\ \Psi_i = \theta_1 + \sum_{k=1}^p \alpha_k X_{i-k} + \sum_{k=1}^q \beta_k \Psi_{i-k} \end{cases}$$

This model can be rewritten as

$$X_i = \theta_1 + \sum_{k=1}^{\max(p,q)} (\alpha_k + \beta_k) X_{i-k} + \sum_{k=1}^{q} \beta_k \eta_{i-k} + \eta_i$$

where $\eta_i = X_i - \Psi_i$. This is equivalent to saying that the conditional expectation of the duration between transactions follows an $ARMA$ process with highly non-gaussian innovations.

We can link the Poisson process to the ACD model. Indeed, if f_{ϵ,θ_1} denotes the density of ϵ, ($\theta_1 \in \Theta_1$), the density of X_i can be written as :

$$f(x|z_i, \theta_1) = \frac{1}{\Psi_i} f_{\epsilon,\theta_1}\left(\frac{x}{\Psi_i}, \theta_1\right)$$

In the case of the Poisson process, the process intensity is deterministic whereas in the case of of an ACD(p, q) model, if f_{ϵ,θ_1} is the density of an exponentially distributed random variable with parameter 1, $f(x|z_i, \theta_1)$ follows a exponential law whose parameter is randomly distributed.

For practical reasons (especially to avoid the need of constraining the parameters), one can introduce exponential functions in the previous model and build a model called the exponential ACD (following roughly the building of the EGARCH model, see Nelson (1991)). If $\Psi_i(\mathcal{H}_{i-1}, \theta_1) = \mathbb{E}(X_i|\mathcal{H}_{i-1}, \theta_1)$. The conditional duration of floods is then:

$$f(x_i|\mathcal{H}_{i-1}, \theta_1) = \frac{1}{\Psi_i(H_{i-1}, \theta_1)} \exp\left(\frac{-x_i}{\Psi_i(H_{i-1}, \theta_1)}\right),$$

and

$$\Psi_i(\mathcal{H}_{i-1}, \theta_1) = \exp\left(\alpha + \delta \log(\Psi_{i-1}) + \gamma \frac{x_{i-1}}{\Psi_{i-1}} + \beta_1 p_{i-1} + \beta_2 v_{i-1}\right).$$

p_i (peak) and v_i (volume) act here as explanatory variables : we assume that the intensity of the flood process is linked with past floods' peaks and volumes.

g is defined in a similar way, taking into account the fact that the density used here is conditional to x_i. Ψ_i becomes Φ_i, and x_i becomes y_i :

$$\Phi_i(x_i, \mathcal{H}_{i-1}, \theta_1) = \exp\left(\mu + \rho \log(\Phi_{i-1}) + \gamma \frac{y_{i-1}}{\Phi_{i-1}} + \tau \frac{x_i}{\Psi_i} + \eta_1 p_{i-1} + \eta_2 v_{i-1}\right)$$

we can then apply this model to floods (or more precisely to threshold exceedances, see Figure 8.22).

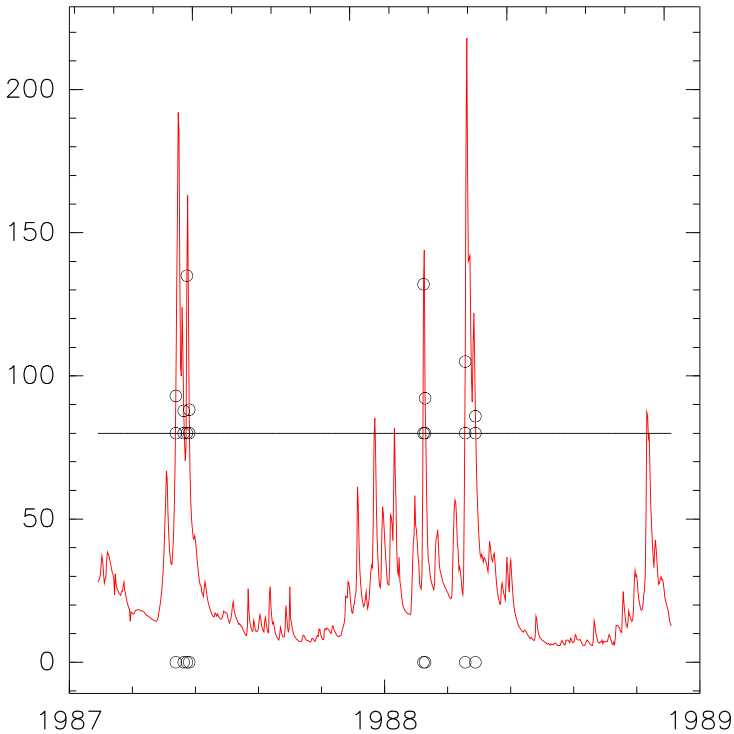


Figure 8.22: Determination of the flood events, on a two year period.

Estimating the model

The log-likelihood of the model for an n-sized sample is:

$$\mathcal{L}(X, Y, \theta_1, \theta_2) = \sum_{i=1}^{n} (\log g(y_i|x_i, \mathcal{H}_{i-1}, \theta_2)) + \sum_{i=1}^{n} (\log f(x_i|\mathcal{H}_{i-1}, \theta_1)$$

One can start by maximizing the second term to get and $\hat{\theta_1}$ then maximize

$$\sum_{i=1}^{N} (\log g(y_i|\hat{x}_i, \mathcal{H}_{i-1}, \theta_2)) .$$

This approach is not equivalent to directly maximizing the log-likelihood when the dependence between X and Y is ill-specified and can result in lesser quality estimators. However, this method, used in Engle and Lunde (2003), is simple.

If the model is truly exponential and if the duration between two floods is well-specified, the residuals associated to the said duration should be independent and be exponentially distributed with parameter 1. In the ACD framework, residuals are defined as:

$$\epsilon_i = \frac{x_i}{\Psi_i(\mathcal{H}_{i-1}, \theta_1)}$$

Mixed distribution for the residuals

The exponential distribution is simple, and often used to model durations, but quite inefficient as we shall see in Section 8.4.5. We will therefore use more general families of distribution, namely a mix of two distributions. Indeed, snow meltdown creates a seasonal effect, which means for the residuals that two effects should be taken into account : one for *regular* excedances (1), and one for the snow meltdown effect (2).

$$\text{distribution of } \varepsilon \begin{array}{ll} \nearrow & \mathcal{L}(\varepsilon, \lambda, \delta) \text{ with probability } \alpha \quad (1) \\ \searrow & \mathcal{L}(\varepsilon, \mu, \gamma) \text{ with probability } 1-\alpha \quad (2) \end{array}$$

We will try modelling the residuals with a mixed exponential distribution and with a mixed Weibull distribution (such a law allows more flexibility than the exponential law when modelling durations, see Barlow and Proschan (1996)).

The mixed exponential model can be written as

$$S(x) = \alpha e^{-\lambda_1 x} + (1-\alpha) e^{-\lambda_2 x}, \quad w_1, w_2 > 0, \quad w_1 + w_2 = 1$$

where S denotes the survival function. Hence, the log-likelihood is:

$$\mathcal{L}(X_1, ..., X_n) = \sum_{k=1}^{n} \log \left(\frac{\alpha}{\lambda_1} e^{-\lambda_1 x} + \frac{1-\alpha}{\lambda_2} e^{-\lambda_2 x} \right)$$

which is then maximized according to $\lambda_1, \lambda_2, w_1, w_2$.

The mixed Weibull distribution has the following density:

$$f(x) = \alpha \left(\lambda \delta^{-\lambda} x^{\lambda-1} e^{-(x/\delta)^{\lambda}} \right) + (1-\alpha) \left(\mu \gamma^{-\mu} x^{\mu-1} e^{-((x-d)/\gamma)^{\mu}} \right)$$

Such a model can seem complex since it has 4 parameters. The parcimonia principle would suggest to focus on a simpler mixture, but, as we shall see in Section 8.4.5 (see e.g. Figure 8.26) the proper distribution can be, in practice, difficult to model with simple laws.

The estimation procedure proposed here is a maximum likelihood estimation procedure for both exponential and Weibull distributions, but recall that a Bayesian approach can also be considered, e.g. using Gibbs' sampler (see Kelly and Krzysztofowicz (2000, 2001) for the use of bayesian models for estimation in rainfall series).

Parametric estimation for a mixture

In reliability theory, when modelling failure-time data, different types of failure are usually considered, e.g. 2. The traditional approach is to assume that there exists latent failure times T_1 and T_2, corresponding to the two causes of failure, and then proceed with the modelling of $T = \min\{T_1, T_2\}$ on the basis that the two causes are independent (see David and Moeschberger (1968)). An alternative approach (see Prentice et al. (1978)) is to consider a two-component mixture model, where

$$\overline{F}(t) = \alpha \overline{F}_1(t) + (1-\alpha)\overline{F}_2(t)$$

where F_i denotes the conditional survival survival function, given that failure is due to the ith cause, and α is the probability that the failure is due to the ith cause. In terms of density, we notice that

$$f(t) = \alpha f_1(t) + (1-\alpha) f_2(t).$$

The underlying model can then be written

$$X_i = \Theta_i Y_{1,i} + (1-\Theta_i) Y_{2,i}$$

where $\Theta_i \sim \mathcal{B}(\alpha)$, and $X_i|\Theta_i$ has density f_1 if $\Theta_i = 1$ and f_2 if $\Theta_i = 0$.

Maximum likelihood Data are obtained from a mixture, and they are said to be *incomplete* since we do not know if one observation X_i is of type (1) or (2). Hence, the X_i's are observed, but not the Θ_i's (also called component labels). Assume that the f_i's belong to the same parametric family, and that $f_i(\cdot) = g(\cdot|\beta_i)$, where β_i can be a vector of parameters. The log-likelihood for a sample $\boldsymbol{X} = (X_1, ..., X_n)$ is then

$$\mathcal{L}(\alpha, \beta_1, \beta_2 | \boldsymbol{X}) = \sum_{i=1}^{n} \log\left[\alpha g(X_i|\beta_1) + (1-\alpha) g(X_i|\beta_2)\right].$$

The maximum likelihood estimates are obtained considering solutions of

$$\frac{\partial \mathcal{L}(\alpha, \beta_1, \beta_2)}{\partial \alpha} = \frac{\partial \mathcal{L}(\alpha, \beta_1, \beta_2)}{\partial \beta_1} = \frac{\partial \mathcal{L}(\alpha, \beta_1, \beta_2)}{\partial \beta_2} = 0.$$

But recall that α should satisfy the constraint $\alpha \in [0,1]$. As pointed out in Titterington, Smith and Makov (1985) explicit calculation of maximum likelihood estimates is usually not possible. However, iterative techniques can be considered, such as the EM algorithm.

The EM algorithm The *EM* algorithm is based on the idea that the X_i's are actually realizations of some unobserved Y_i's,.which should define the *complete* sample, i.e. when dealing with mixture distributions, $Y_i = (X_i, \Theta_i)$. The *complete* log-likelihood for the Y_i's is $\log \mathcal{L}^c(\alpha, \beta_1, \beta_2 | \boldsymbol{Y})$, but it cannot be observed since it is based on the Θ_i's. It is then replaced by its conditional expected value,

$$Q(\xi|\xi_0, \boldsymbol{X}) = \mathbb{E}_{\xi'}\left(\log \mathcal{L}^c(\alpha, \beta_1, \beta_2 | \boldsymbol{X}, \boldsymbol{\Theta}) | \xi_0, \boldsymbol{X}\right),$$

where $\boldsymbol{\Theta}$ is integrated according to the conditional distribution $f(\theta|\xi_0, \boldsymbol{X})$.

The iterative algorithm to estimate parameter ξ is the following: set ξ_0 and at step $k+1$:

1. Step E (estimation): calculate $Q\left(\xi|\xi^{(k)}, \boldsymbol{X}\right)$, based on the expected value with respect to $f\left(\theta|\xi^{(k)}, \boldsymbol{X}\right)$,

2. Step M (maximisation): maximize $Q\left(\xi|\xi^{(k)},\boldsymbol{X}\right)$ and set

$$\xi^{(k+1)}=\arg\max_{\xi} Q\left(\xi|\xi^{(i)},\boldsymbol{X}\right).$$

Example 8.4.4. *In the case of exponential components, $\overline{F}_j(t)=\exp(-\lambda_j t)$ for $t\geq 0$, observe that*

$$Q\left(\xi|\xi^{(k)},\mathbf{X}\right)=\sum_{i=1}^{n}\left[\log\lambda_1-\lambda_1 X_i\right]\mathbb{I}(\Theta_i=1)+\left[\log\lambda_2-\lambda_2 X_i\right]\mathbb{I}(\Theta_i=0),$$

and

$$\lambda_j^{(k+1)}=\frac{1}{X_1+...+X_n}\sum_{i=1}^{n}\phi_{i,j}^{(k)},$$

where

$$\phi_{i,j}^{(k)}=\frac{\lambda_j^{(k)}}{\sum_{u\in\mathcal{S}_j}\lambda_u^{(k)}} \text{and} \left\{\begin{array}{l}\mathcal{S}_1=\{i,\Theta_i=0\}\\ \mathcal{S}_2=\{i,\Theta_i=1\}\end{array}\right.$$

(see Albert and Baxter (1995)).

Example 8.4.5. *In the case of Weibull components, $\overline{F}_j(t)=\exp\left(-[t/\lambda_j]^{\gamma_j}\right)$ for $t\geq 0$, the approach is rather similar but the EM algorithm is not easy to implement.*

Albert and Baxter (1995) considered a pseudo-alternating EM algorithm (see Liu and Rubin (1994)). The equations to solve at the $k+1$th iteration are

$$\lambda_j=\left(\frac{X_1^{\gamma_j}+...+X_n^{\gamma_j}}{\phi_{1,j}^{(k)}+...\phi_{n,j}^{(k)}}\right)^{1/\gamma_j}$$

and

$$\sum_{i=1}^{n}\left[\phi_{i,j}^{(k)}\left(\frac{1}{\gamma_j}+\log\left(\frac{X_i}{\lambda_j}\right)\right)-\left(\frac{X_i}{\lambda_j}\right)^{\gamma_j}\log\left(\frac{X_i}{\lambda_j}\right)\right]=0,$$

where

$$\phi_{i,j}^{(k)}=\frac{f\left(X_i,\lambda_j^{(k)},\alpha_j^{(k)}\right)}{\sum_{u\in\mathcal{S}_j}f\left(X_i,\lambda_u^{(k)},\alpha_u^{(k)}\right)} \text{ where } f(x,\lambda,\alpha)=\frac{\alpha x^{\alpha-1}}{\lambda^{\alpha}}.$$

Example 8.4.6. *Based on the Saugeen data, the previous algorithm have been implemented in order to fit the residuals, see Table 8.6.*

	Mixture	Weibull 1		Weibull 2	
	$\widehat{\alpha}$	$\widehat{\lambda}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\gamma}$
ML	0.70	0.36	0.89	1.01	1.60
(PA)EM	0.68	0.34	0.92	0.98	1.57

Table 8.6: Estimation of the mixed Weibull distribution.

Since the model is now specified, the key issue now is to model the marks properly.

Two durations marked model

We will here take 4 variables into account instead of two : two duration variables (the duration between two flood beginnings as well as the duration between the end of a flood and the beginning of the next one) and two marks (peak streamflows and volumes). In this paragraph and the ones that follow, $\mathcal{H}_i$ will denote the filtration generated by the $(X_k, Y_k, P_k, V_k)_{k<i}$.

By following a similar approach to the previous one, we can write, with f denoting again the conditional density of (P_i, V_i, X_i, Y_i) :

$$f(p_i, v_i, x_i, y_i|\mathcal{H}_{i-1}) = \underbrace{g(p_i, v_i|\mathcal{H}_{i-1}, x_i, y_i)}_{\text{marks}}\underbrace{q(x_i, y_i|\mathcal{H}_{i-1})}_{\text{durations}}$$

By writing the density in such a way, we can separate the roles played by the marks and by the durations, while at the same time retaining the possibility to reuse the previous results. As done previously, we can then separate q in two different parts, which will each follow an E-ACD models.

This model can be simplified by imposing a simple relationship between P and V, and have a single mark two durations model. The idea (albeit simplistic) is to consider that the flood volume can be written as

$$V = P\frac{X-Y}{2}$$

A more accurate model could integrate some error terms in the approach. A first idea could be to assume that $V = P\frac{X-Y}{2} + \varepsilon$ where ε is some Gaussian error term. But as observed on Figure **??**, homoscedasticity of the error cannot be assumed, since the higher the volume, the stronger the variance of the error term. Another idea might be to use a Generalized Linear Models (see McCullagh and Nelder (1988)). For instance, assume that V is Poisson distributed, with expected value $V = P\frac{X-Y}{2}$. Since the Poisson distribution has for variance function identity, the variance term will also be $V = P\frac{X-Y}{2}$, which is more consistent with Figure **??** (more than the *linear* Gaussian model mentioned previously).

Example 8.4.7. *The $(V_i, P_i(X_i - Y_i)/2)$ cloud can be seen Figure 8.23, where the linear model is the dotted line and we can validate this hypotheses.*

One of these two marks being necessary, we chose to keep the peak variable. Using a Pareto law seems right, since it is used to model (Balkema-de Haan theorem, see Embrechts, Klüppelberg and Mikosch (1997)) laws over a threshold). The (conditional) density can then be written as:

$$h(p_i|\mathcal{H}_{i-1}, x_i, y_i, \alpha, \sigma) = \alpha\left(\frac{p_i + b(x_i - y_i) + d}{\sigma}\right)^{-(1+\alpha)}$$

The density of y_i conditionally to x_i being the one given in the previous section, the likelihood of the model is then:

$$\mathcal{L}(P, X, Y, \theta_1, \theta_2, \alpha, \sigma) = \sum_{i=1}^{N}(h(p_i|\mathcal{H}_{i-1}, x_i, y_i, \alpha, \sigma)) + \sum_{i=1}^{N}(\log g(y_i|x_i, \mathcal{H}_{i-1}, \theta_2)) + \sum_{i=1}^{N}(\log f(x_i|\mathcal{H}_{i-1}, \theta_1)$$

8.4.5 Modelling exceedances for the Saugeen river

When modeling extremal events (see Embrechts, Klüppelberg and Mikosh (1997) or Beirlant, Goegebeur, Segers and Teugels (2004)) defined as events for which the variable of interest exceeds a certain level, one usually does the following : define a lower threshold, model events exceeding this threshold, then focus on tails. When modeling floods, the idea remains be the same. We define floods as exceedances over some very high thresholds. In order to study them from a statistical point of view, we shall study exceedances for lower thresholds.

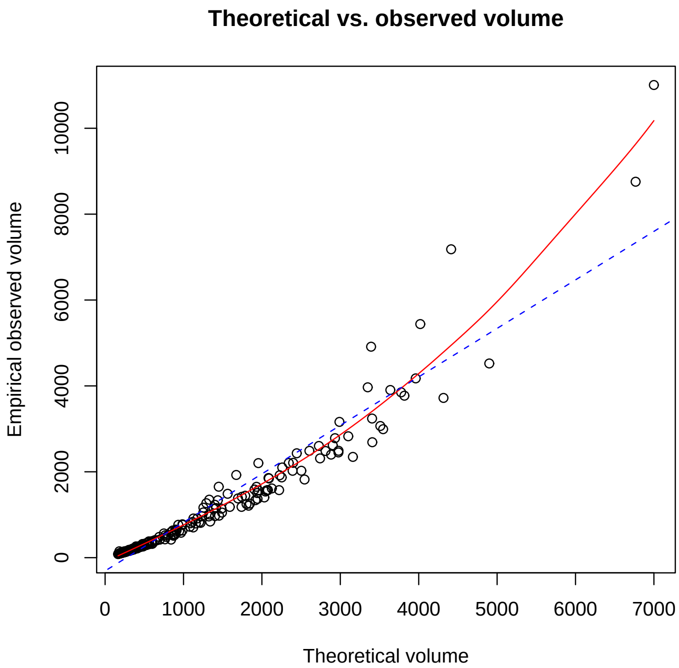


Figure 8.23: Relationship between peak, volume and duration of a flood.

How to define a *flood* (or exceedance of some threshold) events

In order to define the point processes (beginning and ending of the *flood*, or exceedance), the threshold is chosen so that if the variable exceeds the threshold, it is much higher than the threshold. Such a method allows to keep only "interesting" floods, and to limit the number of events considered. The idea is to maximize

$$\mathbb{P}(X_i > f(threshold)|X_i > threshold) = \frac{\#\{X_i > f(threshold)\}}{\#\{X_i > threshold\}}$$

where f is some given function. Assume that f is affine. The constant terms helps to exclude noise due to the very low level of the river, which may artificially bring the threshold down. The multiplicative coefficient has to be chosen higher than 1. Based on the Saugeen's level, we have assumed that the level was 1.5. Higher values would have made the number of exceedances decrease drastically, while lower values would not be associated to flood events. Figure 8.24 represents the probability as a function of the threshold. Note that Figure 8.22 shows exceedances and flood events over a two year period of time.

Estimating the model

Estimating the two-duration model yields the results shown in table 8.7. One can notice the similarities between parameters which play the same role in the two functions. We can notice that past observations do not play a major role, probably because of a scale factor. Simulating such a process yields for X the QQ-plot shown Figure 8.25, which rejects the EACD model : this model overestimates very long durations between floods. Besides, studying the residuals invalidates the estimated model, since the residuals are not exponentially distributed, and have two very apparent modes (see figure 8.26). The second mode in particular has a major influence

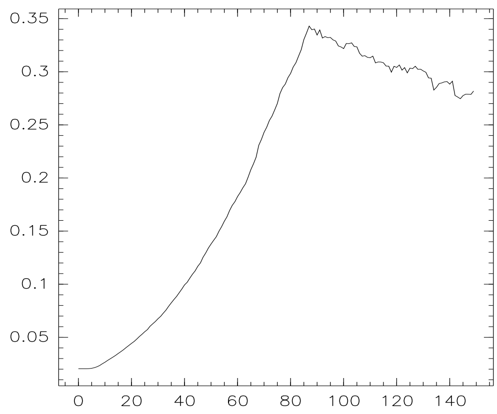


Figure 8.24: Determination of the threshold, which define exceedances.

over the distribution of extremes values. This density may imply that there are two different flood regimes, one being associated with snow meltdown and the other one with regular floods.

$\widehat{\alpha}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\delta}$	$\widehat{\gamma}$	
4.32	$-1.03.10^{-3}$	$2.89.10^{-6}$	0.15	$-9.55.10^{-4}$	
$\widehat{\mu}$	$\widehat{\eta_1}$	$\widehat{\eta_2}$	$\widehat{\rho}$	$\widehat{\gamma}$	$\widehat{\tau}$
4.43	$-1.03.10^{-3}$	$-5.34.10^{-7}$	0.12	$-2.94.10^{-3}$	0.67

Table 8.7: Results of the estimation procedure for the two durations model (X_i) et (Y_i)

As mentioned in the previous section, the exponential assumption does not hold(see Figure 8.25). The Weibull approach being more flexible, we tried to fit such a model on the data. Numerically speaking, the estimates are, for $(\alpha, d, \lambda, \delta, \mu, \gamma)$ give in Table 8.8.

$\widehat{\alpha}$	$\widehat{d}$	$\widehat{\lambda}$	$\widehat{\delta}$	$\widehat{\mu}$	$\widehat{\gamma}$
0.70	1.5	0.36	0.89	1.01	1.60

Table 8.8: Estimation of Exponential model parameters $(\alpha, d, \lambda, \delta, \mu, \gamma)$

Based on this Weibull model, simulated values (using estimated values from the observation) for the fitted model can be compared to the empirical data (see Figure 8.27). The improvement of the results is very noticeable. It seems legitimate to think these results can be improved even more since the estimation procedure was very unprecise.

We have specified a model for the duration, and can now extend it to take marks into account, by following a similar approach to Engle and Rusell (1998) (for a single duration and a single mark).

Figure 8.28 shows the distribution function of mark simulations against empirical data, as well as a qq-plot. We can notice that the qq-plot is misshapen. This stems from the fact that we simulated 17000 floods, and extreme values (99.9% quantile) appeared which are not present in the data.

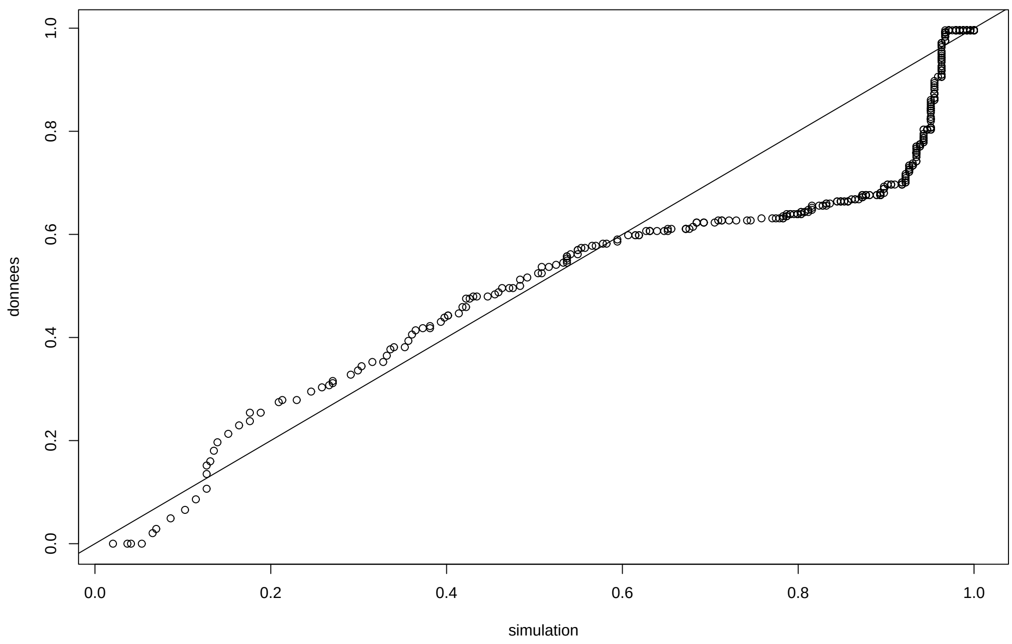


Figure 8.25: Comparison between the empirical distribution function of the data and that of X.

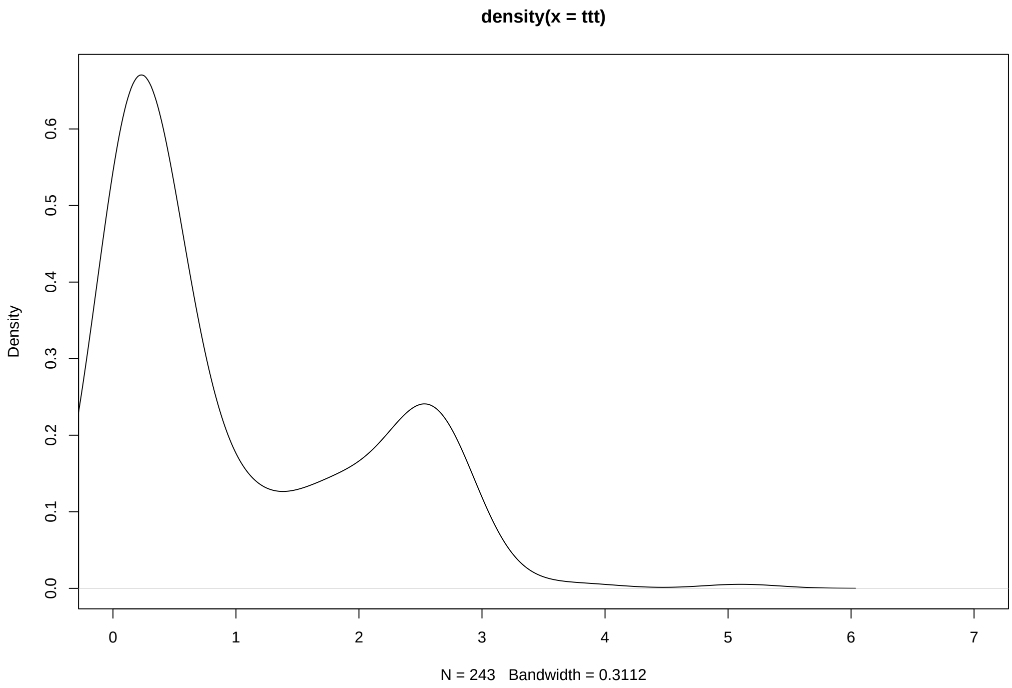


Figure 8.26: Kernel density estimation of residuals

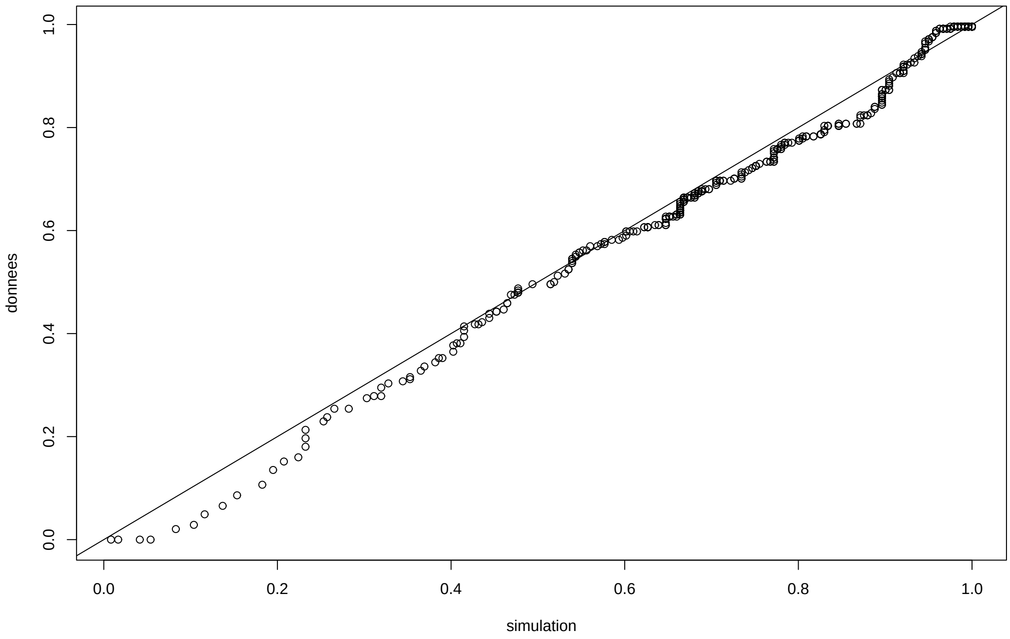


Figure 8.27: Comparison between the empirical distribution function of the data and that of simulations of X for a Weibull mixed law

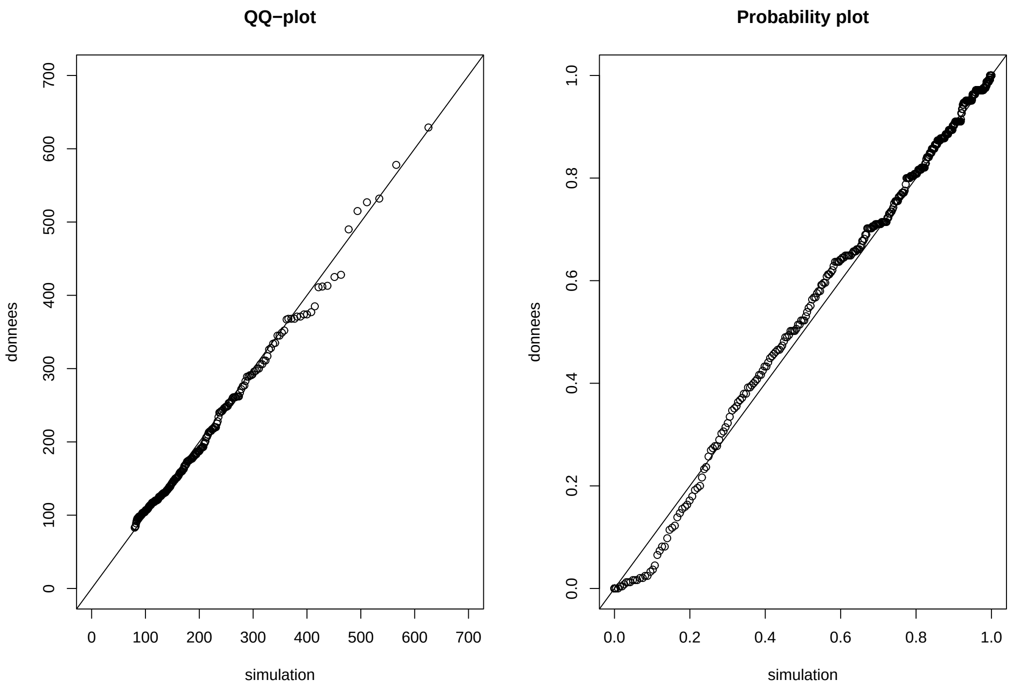


Figure 8.28: QQ-plot and Probability plot.

8.4.6 Back to annual maxima

Since in hydrology the benchmark is the annual maxima, we should now compare this model with the *traditional* approach, studying the aggregated process derived from this detailed approach.

We simulated marks from the estimated model and extracted annual maxima. We then fitted a GEV distribution on these data and found $\widehat{\xi} = 0.03$, $\widehat{\mu} = 96$ and $\widehat{\sigma} = 216$. These results are very close to the static model. Estimating the 90% yields results in accordance with the ones obtained before: The 90% quantile of our dynamic model is 433, against 440 for the static model. The dynamic model we proposed seems therefore "*compatible*" with the usual static one.

Since the models seem *compatible*, it is interesting to study why using a dynamic model could be better. In particular, we can study the duration between two decades floods, i.e. floods which are over the 90% quantile estimated on annual maxima.

Example 8.4.8. *The duration distribution between decennial floods is shown Figure 8.29. It resembles that of a geometric distribution, typical of the return period hypothesis, with more events within a very short period. However, it seems that the return period associated to such a distribution is smaller than the expected ten years : 9 years. This probably comes from the fact that the dynamic model takes into account multiple extreme events in a single year, something which the static model does not allow.*

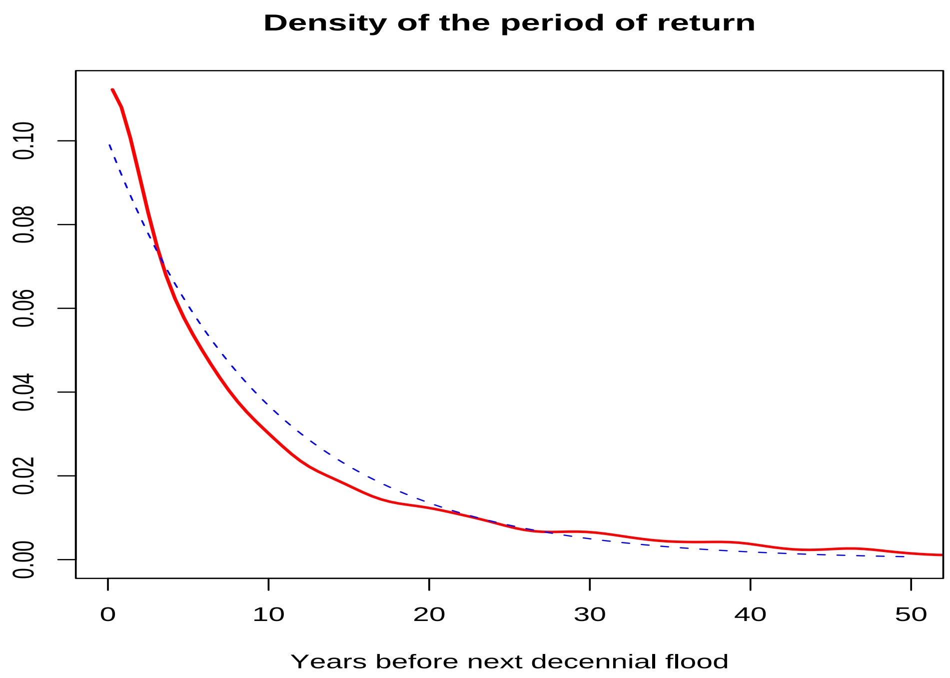


Figure 8.29: Flood duration distribution.

8.4.7 A brief conclusion

The goal of that section, drawing from the field of high-frequency finance, was to work on irregularly spaced observations in order to avoid some inconsistencies observed in several usual models. Indeed, flood events are not necessarily annual : even if there is a strong seasonal component due to snow meltdown (but not regular enough to be integrated in the model, as done in Chavez-Demoulin and Davison (2005)), there is no reason to focus on annual observations.

Empirical evidence suggests that although Gumbel was *almost* right stating that annual maxima are i.i.d., we cannot assume that flood events are independent. More precisely, the *real* return period of flood events (or at least threshold exceedances) is shorter than the one obtained under the i.i.d. assumption. However, although we were able to model time-dependent observations, we still could not reproduce the strong dependence known as the Hurst effect. One of the reason might be the lack of data, for 100 years might not be enough. Since ACD models in finance allow to observe persistence phenomena, it should also be true for hydrological series..

One of the main difficulties was the choice of the threshold defining an exceedance, in short a *flood*event. A low level (annual exceedance) allows to keep a lot of observations but might be useless from a practical standpoint. On the contrary, a high level has practical consequences (in terms of risk management, e.g. assessing the level of a Dam) but there are not enough observations left to model correctly temporal dependence.

Many things remain to be studied, starting with the many numerical problems which arose and need to be solved. Furthermore, it would be interesting to use the same category of models on rainfall series, and try to model jointly rainfall and streamflow.

Appendix: ARMA with t residuals

Assume that $Y_1, \ldots, Y_n$ has the following ARMA(p,q) representation,

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t,$$

where $\Phi(L) = \mathbb{I} - \phi_1 L - \ldots - \phi_p L^p$, and $\Phi(L) = \mathbb{I} + \theta_1 L + \ldots + \theta_q L^q$.

Consider the following matrix notations, $\boldsymbol{y} = (Y_1, ..., Y_n)$ the vector of observations, and $\boldsymbol{e} = (\varepsilon_1, ..., \varepsilon_n)$. Define $\boldsymbol{y}_* = (Y_{p-1}, ..., Y_{-1}, Y_0)$ containing presample elements of $(Y_t)_{t\in\mathbb{Z}}$, and $\boldsymbol{e}_* = (\varepsilon_{q-1}, ..., \varepsilon_{-1}, \varepsilon_0)$ containing presample elements of $(\varepsilon_t)_{t\in\mathbb{Z}}$. A the sample n elements from $(Y_t)_{t\in\mathbb{Z}}$ can be represented in the corresponding vector equation

$$A_*\boldsymbol{y}_* + A\boldsymbol{y} = M\boldsymbol{e} + M_*\boldsymbol{e}_*,$$

where the banded lower-triangular matrix A is of order $n\times n$, while A_* is a $n\times p$ matrix, defined as

$$A = \begin{pmatrix} 1 & & & & & \\ \phi_1 & \ddots & & & (0) & \\ & \ddots & \ddots & & & \\ \phi_p & & \phi_1 & 1 & & \\ & \ddots & & \phi_1 & \ddots & \\ & (0) & \ddots & & \ddots & \ddots \\ & & & \phi_p & & \phi_1 \quad 1 \end{pmatrix} \text{ and } A_* = \begin{pmatrix} \phi_p & & \phi_1 \\ & \ddots & \\ & & \phi_p \\ & & \\ & (0) & \\ & & \end{pmatrix}$$

for the autoregressive representation, and for the moving average component,

$$M = \begin{pmatrix} 1 & & & & & \\ \theta_1 & \ddots & & & (0) & \\ & \ddots & \ddots & & & \\ \theta_q & & \theta_1 & 1 & & \\ & \ddots & & \theta_1 & \ddots & \\ & (0) & \ddots & & \ddots & \ddots \\ & & & \theta_q & & \theta_1 \quad 1 \end{pmatrix} \text{ and } M_* = \begin{pmatrix} \theta_q & & \theta_1 \\ & \ddots & \\ & & \theta_q \\ & & \\ & (0) & \\ & & \end{pmatrix}$$

Hence, an alternative matrix representation is the following,

$$A\boldsymbol{y} = M\boldsymbol{e} + V\boldsymbol{u},$$

where V and $\boldsymbol{u}$ are respectively a $n\times(p+q)$ matrix and a $(p+q)$ vector,

$$V = (-A_* \; M_*) \text{ and } \boldsymbol{u} = (\boldsymbol{y}_* \; \boldsymbol{e}_*)'.$$

So finally, combining those equations, we get

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ A^{-1}V & A^{-1}M \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} \tag{8.4}$$

or, considering the inverted version,

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ A^{-1}V & A^{-1}M \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \mathbb{I} & 0 \\ -M^{-1}V & M^{-1}A \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{pmatrix} \tag{8.5}$$

Since $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise, $Var(\boldsymbol{e}) = \sigma^2\mathbb{I}$. Hence, assume that $\boldsymbol{e} \sim \mathcal{T}(0, \sigma^2\mathbb{I}, \nu)$. Also, $\boldsymbol{y} \sim \mathcal{T}(\sigma^2\Omega)$, for some matrix Ω. So finally, it comes that

$$Var(\boldsymbol{u}) = \sigma^2 \begin{pmatrix} \Omega & 0 \\ 0 & \mathbb{I} \end{pmatrix}$$

From Equation (8.4), the variance of $(\boldsymbol{u}\ \boldsymbol{y})$ is

$$Var \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{y} \end{pmatrix} = \sigma^2 \begin{pmatrix} \Omega & \Omega V' A'^{-1} \\ A'^{-1} V \Omega & A^{-1}(V\Omega V' + MM')A'^{-1} \end{pmatrix}$$

and thus, the variance of $\boldsymbol{y}$ is

$$Var(\boldsymbol{y}) = \sigma^2 A^{-1}(V\Omega V' + MM')A'^{-1},$$

and using properties of triangular matrix,

$$Var(\boldsymbol{y})^{-1} = \frac{1}{\sigma^2} A' M''^{-1}(\mathbb{I} - M^{-1}V(\Omega^{-1} + V'(MM')^{-1}V)^{-1}V'M'^{-1})M^{-1}A.$$

Assume now that elements of $\boldsymbol{e}$ and $\boldsymbol{u}$ are independent, so that the density of $(\boldsymbol{e}, \boldsymbol{u})'$ is the product of the densities of $\boldsymbol{e}$ and $\boldsymbol{u}$,

$$\begin{aligned} f(\boldsymbol{u}, \boldsymbol{e}) &= \frac{\Gamma((\nu + (p+q))/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu\sigma^2)^(p+q)|\Omega|}} \left(1 + \frac{\boldsymbol{u}'\Omega^{-1}\boldsymbol{u}}{\sigma^2\nu}\right)^{-(\nu+(p+q))/2} \\ &\times \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu\sigma^2)^n}} \left(1 + \frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2\nu}\right)^{-(\nu+n)/2} \end{aligned}$$

We assume here that the number of degrees of freedom, ν, is known.

Therefore, joint distribution of $(\boldsymbol{u}, \boldsymbol{y})$ is then obtained from Equations (8.4) and (8.5), since the distribution of $\boldsymbol{y}$ given $\boldsymbol{u}$ is a t distribution, and thus, using $f(\boldsymbol{u}, \boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{u}) \cdot f(\boldsymbol{u})$, we get

$$\begin{aligned} f(\boldsymbol{u}, \boldsymbol{y}) &= \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu\sigma^2)^n}} \left(1 + \frac{(A\boldsymbol{y}_* - V\boldsymbol{u})'(MM')^{-1}(A\boldsymbol{y}_* - V\boldsymbol{u}_*)}{\sigma^2\nu}\right)^{-(\nu+n)/2} \\ &\times \frac{\Gamma((\nu + (p+q))/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu\sigma^2)^(p+q)|\Omega|}} \left(1 + \frac{\boldsymbol{u}'\Omega^{-1}\boldsymbol{u}}{\sigma^2\nu}\right)^{-(\nu+(p+q))/2} \end{aligned}$$

From this expression, the marginal distribution of $\boldsymbol{y}$ can be expressed as

$$f(\boldsymbol{y}) = \frac{\Gamma((\nu+n)/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu\sigma^2)^n|\Sigma|}} \left(1 + \frac{\boldsymbol{y}'\Sigma^{-1}\boldsymbol{y}}{\sigma^2\nu}\right)^{-(\nu+n)/2},$$

where $\Sigma = A^{-1}(V\Omega V' + MM')A'^{-1}$.

Therefore, maximum likelihood estimators of the ARMA components can be obtained by minimizing

$$|\Sigma|^{-1/2} \left(1 + \frac{\boldsymbol{u}'\Omega^{-1}\boldsymbol{u}}{\sigma^2\nu}\right)^{-(\nu+(p+q))/2}$$

as a function of Σ, or the logarithm of that function. Note that the determinant of Σ is

$$|\Sigma| = |A^{-1}(V\Omega V' + MM')A'^{-1}| = |V\Omega V' + MM'|,$$

since $|A| = 1$.

Bibliography

Abdous, B., Fougères, A.L., and Ghoudi, K.(2005). Extreme behavior for bivariate elliptical distributions. *Canadian Journal of Statistics*, **33**, 317-334 .

Abdous, B., Ghoudi, K. and Khoudraji, A. (1999). Nonparametric estimation of the limits dependence function of multivariate extremes. *Extremes*, **2**, 245-268.

Abramson, I.S. (1982). On bandwidth variation in kernel estimates - a square root law. *Annals of Statistics*, **9**, 168-176.

Aczél, J. (1966). Lectures on functional equations and their applications. Academic Press.

Akritas, M. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of the American Statistical Association*, **81**, 1032-1038.

Alaci, V. (1923). Functiuni pseudo-omogenes. *Mathematica Timisoara*, **3**, 3-4.

Albert, J.R.G., & Baxter, L.A. (1995). Applications of the EM Algorithm to the Analysis of Life Length Data. *Applied Statistics*, **44**, 323-341.

Aldous, D.J. (1985). Exchangeability and related topics, in *Ecole d'été de probabilités de Saint-Flour*, **13**, Springer Verlag.

Alsina, C., Nelsen, R.B. and Schweizer, B. (1993). On the characteristic function of a class of binary operations on distribution functions. *Statistic and Probability Letters*, **17**, 85-89.

Alsina, C. and Schweizer, B. (1988). Mixtures are not derivable. *Foundations of Physics Letters*, **1**, 171-174.

Alvo, M., Cabilio, P. and Feigin, P.D. (1982). Asymptotic Theory for Measures of Concordance with Special Reference to Average Kendall Tau. *The Annals of Statistics*, **10**, 1269-1276.

Alzaid, A., Kim, J.S. and Proschan, F. (1991). Laplace ordering and its applications. *Journal of Applied Probability*, **28**, 116-130.

Arjas, E. (1981a). A stochastic process approach to multivariate reliability systems : notions based on conditional stochastic order. *Mathematics of Operations Research*, **6**, 263-276.

Arjas, E. (1981b). The failure and hazard processes in multivariate reliability systems. *Mathematics of Operations Research*, **6**, 551-562.

Arjas, E. & Norros, I. (1984). Life lengths and association : a dynamic approach. *Mathematics of Operations Research*, **9**, 151-158.

Arjas, E. and Norros, I. (1986). A compensator representation of multivariate life length distributions, with applications. *Scandinavian Journal of Statistics*, **13**, 99-112.

ARJAS, E. AND NORROS, I. (1991). Stochastic order and martingale dynamics in multivariate life length models : a review, *in* Stochastic Orders and Decision under Risk, eds. Mosler and Scasini, IMS Lectures Notes, 7-23.

ARNOLD, B.C. (1983). Pareto distributions. International Cooperative Publishing House.

ARNOLD, B.C. (1987). Bivariate distributions with Pareto conditionals. *Statistics and Probability Letters*, **5**, 263-266.

ARNOLD, B.C. (1989). Dependence in conditionally specified distributions. *in* Topics in Statistical Dependence, IMS Lecture Notes, **16**, 13-18.

ARNOLD, B.C., CASTILLO, E., AND SARABIA, J.M. (1992). Conditionally specified distributions. Springer-Verlag.

AVÉROUS, J. (2002). Hazard rate ordering between two life distributions an SI ordering between bivariate vectors and their order statistics. *Working paper.*

AVÉROUS, J. AND, DORTET-BERNARDET, M.B. (2000). Dependence for Archimedean copulas and aging properties of their generating functions. *Indian Statistical Institute*, **66**, 607-620.

AYDOYAN, K., & BOOTH, G.G. (1988) Are there long cycles in common stock returns. *Southern economic journal*, **55**, 141-149.

BAGAI, I. AND PRAKASA RAO, B.L.S. (1995). Kernel based estimates for positive valued random variables. *Sankhya: the Indian Journal of Statistics* (Series A), **57**, 56-67.

BAGGS, G.E. AND NAGARAJA, H.N. (1996). Reliability properties of order statistics from bivariate exponential distributions. *Commun Statist. Stoch. Models*, **12**, 611-631.

BAILLIE, R.T. (1996). Long Memory Processes and Fractional Integration in Econometrics. *Journal of Econometrics*, **73**, 5-59.

BAILLIE, R.T., & BOLLERSLEV, T. (1989). Common Stochastic Trends in a System of Exchange Rates. *Journal of Finance*, **44**, 167-181.

BAILLIE, R.T., & BOLLERSLEV, T. (1994). Cointegration, Fractional Cointegration, and Exchange Rate Dynamics. *Journal of Finance*, **49**, 737-45.

BAILLIE, R.T., CHUNG, C.F., & TIESLAU, M.A. (1996). Analyzing Inflation by the Fractionally Integrated ARMA-GARCH Model. *Journal of Applied Econometrics*, **11**, 23-40.

BALAKRISHNAN, N., CHARALAMBIDES, C. AND PAPADATOS, N.(2003). Bounds on expectations of order statistics from a finite population. *Journal of Statistical Planning and Inference*, **113**, 569-588.

BALKEMA, A.A. (1973). Monotone transformations and limit laws. *Mathematisch Centrum, Amsterdam*, **45**.

BALKEMA, A.A., AND DE HAAN, L. (1974). Residual Life Time at Great Age, *The Annals of Probability* **2**, 792-804.

BALKEMA, G. AND EMBRECHTS, P. (2004). Multivariate excess distributions. *ETHZ Preprint.*

BALL, F. (1985). Deterministic and stochastic epidemics with several kinds of susceptible. *Advanced Applied Probability*, **17**, 1-22.

BANDEEN-ROCHE, K.J. AND LIANG, K.Y. (1996). Modeling Failure-Time Associations in Data with Multiple Levels of Clustering. *Biometrika*, **83**, 29-39.

BARAKAT, H.M. (1999). On moments of bivariate order statistics. *Ann. Inst. Statist. Math.*, **51**, 351-358.

Barbe, Ph., Genest, C., Ghoudi, K. and Rémillard, B. (1996). On Kendall's process. *Journal of Multivariate Analysis*, 58, 197–229.

Barbour, A.D., Lindvall, T. and Rogers, L.C.G. (1991). Stochastic ordering of order statistics. *Journal of Applied Probability*, **28**, 278-286.

Barlow, R.E. and Proschan, F. (1966). Inequalities for linear combinations of order statistics from restricted families, *Annals of Mathematical Statistics*, **37**, 1574-1591.

Barlow, R.E. and Proschan E. (1996). Mathematical theory of reliability. SIAM.

Barlow, R.E. and Mendel, M.B.. (1992). de Finetti-type representations for life distributions. *Journal of the American Statistical Association*, **87**, 1116-1122.

Barlow, R.E. and Spizzichino, F. (1993). Schur-concave survival functions and survival analysis. *Journal of Computational and Applied Mathematics*, **46**, 437-447.

Barlow, R.E. and Proschan, F. (1996). Mathematical theory of reliability. SIAM.

Bartoszewicz, J. (2001). Stochastic comparisons of random minima and maxima from life distributions. *Statistics & Probability Letters*, **55**, 107-112.

Basrak, B. Davis, R.A. and Mikosch, T. (2002). A characterization of multivariate regular variation. *Annals of Applied Probability*, **12**, 908-920.

Bassan, B. and Spizzichino, F. (1999). Stochastic comparison for residual lifetimes and Bayesian notions of multivariate aging. *Advances in Applied Probability*, **31**, 1078-1094.

Bassan, B. and Spizzichino, F. (2000). Stochastic comparison for residual lifetimes and Bayesian notions of multivariate aging. *Advances in Applied Probability*, **31**, 1078-1094.

Bassan, B. and Spizzichino, F. (2004). Relations among univariate aging, bivariate aging and dependence for exchangeable lifetimes. *Journal of Multivariate Analysis*, to appear.

Bassan, B. and Spizzichino, F. (2005). Bivariate survival models with Clayton aging functions. *Insurance: Mathematics and Economics*, **37**, 6-12.

Bäuerle, N. and Müller, A. (1987). Modelling and comparing dependencies in multivariate risk portfolios. *Astin Bulletin*, **28**, 59-76.

Bäuerle, N. and Müller, A. (2004). Stochastic orders and risk measures: consistency and bounds. *Insurance: Mathematics and Economics*, **38**, 132-148.

Behnen, K. (1971). Asymptotic optimality and ARE of certain rank-order tests under contiguity. *The Annals of Mathematical Statistics*, **42**, 325-329.

Behnen, K., Huskova, M. and Neuhaus, G. (1985). Rank Estimators of Scores for Testing Independence. *Statistics and Decision*, **3**, 239-262.

Behnen, K. and Neuhaus, G. (1989). Rank Tests with estimated scores and their application. Teubner.

Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). Statistics of Extremes - Theory and Applications. Wiley.

Bellman, R. E. (1961). Adaptive Control Processes. Princeton University Press.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, University of California, Berkeley.

Berens, H. and DeVore, R. (1980). A characterization of Bernstein polynomial. *Approximation Theory III.*

BERMAN, S.M. (1961). Convergence to bivariate limiting extreme value distributions. *Annals of the Institute of Statistical Mathematics*, **13**, 217-223.

BHANSALI, R.J. (1996). Recent developments in analysis of time series with infinite variance. *Procedings of the Athens International Conference on Applied Probability and Time Series*, Vol **2**,42-54, Springer.

BHUCHONGKUL, S. (1964). A Class of Nonparametric Tests for Independence in Bivariate Populations. *Annals of Mathematical Statistics*, **34**, 138-149.

BICKEL, P.J. AND ROSENBLATT, M. (1973). On Some Global Measures of the Deviations of Density Function Estimates. *Annals of Statistics*, **1**, 71-1095.

BILLINGSLEY, P. (1999), Convergence of Probability Measures. Wiley.

BINGHAM, N.H., GOLDIE, C.M. AND TEUGELS, J.L. (1987). Regular variation. Cambridge University Press, Cambridge.

BLOCK, H.W AND BASU, A.P. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association*, **69**, 1031-1037.

BOLAND, P.J., HOLLANDER, M., JOAG-DEV, K. AND KOCHAR, S.(1994). Bivariate dependence properties of order statistics. *Journal of Multivariate Analysis*, **56**, 75-89.

BOOTH, G.G., KAEN, F.R. & KOROS, P.E. (1987). R/S analysis of foreign exchange rates under two international money regimes. *Journal of monetary economics*, **10**, 407-415.

BOUËTTE, J.C., CHASSAGNEUX, J.F., SIBAÏ, D., TERRON, R. & CHARPENTIER, A. (2006). Windspeed in Ireland: long memory or seasonal effect?. *Stochastic Environmental Research and Risk Assessment*. **20**, 141 - 151.

BOUEZMARNI, T. AND ROLIN, J.M. (2001). Consistency of beta kernel density function estimator. *The Canadian Journal of Statistics*, **31**, 89Ũ98.

BOUEZMARNI, T. AND ROLIN, J.M. (2003). Bernstein estimator for unbounded density function. *Université Catholique de Louvain-la-Neuve*, **0309**.

BOYER, B.H., GIBSON, M.S. AND LORETAN, M. (1997). Pitfalls in Tests for Changes in Correlations. FRB International Finance Discussion Paper **597**.

BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A. AND NESBITT, C.J. (1997). Actuarial mathematics. The Society of Actuaries.

BREIMAN, L., MEISEL, W. AND PURCELL, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, **19**, 135-144.

BREYMANN, W., DIAS, A. AND EMBRECHTS, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, **3**, 1-16.

BROCKWELL, P.J. & DAVIS, R.A. (1991). Time Series: Theory and Methods. Springer-Verlag.

BROWN, B.M. AND CHEN, S.X. (1999). Beta-Bernstein smoothing for regression curves with compact supports. *Scandinavian Journal of Statistics*, **26** 47-59.

BUISHAND, T.A. (1984). Bivariate extreme-value data and the station-year method. *Journal of Hydrology*. **69**. 77-95.

BURT, S. (2004). The August 2003 heatwave in the United Kingdom. Part I: Maximum temperatures and historical precedents. *Weather*, **59**, 199-208.

BURT, S. & EDEN, P. (2004). The August 2003 heatwave in the United Kingdom. Part II: The hottest sites. *Weather*, **59**, 239-246.

Burt, S. (2004). The August 2003 heatwave in the United Kingdom. Part III: Minimum temperatures. *Weather*, **59**, 272-273.

Cacoullos, T. (1966). Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, **18**, 179-190.

Cai, J. and Garrido, J. (1998). Aging properties and bounds for ruin probabilities and stop-loss premiums. *Insurance: Mathematics and Economics*, **23**, 33-43.

Capéraà, P. and Genest, C. (1993). Spearman's ρ is larger than Kendall's τ for positively dependent random variables. *Nonparametric Statistics*. **2**. 183-194.

Capéraà, P., Fougères, A.L. and Genest, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*. **72**. 30-49.

Carriere, J.F. (2000). Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*. **1** 17-32.

Chan, W., Proschan, F. and Sethuraman (1991). Convex ordering among functions, with application to reliability and mathematical statistics. *in* Topics in Statistical Dependence, eds. Block, Sampson and Savits, IMS Lectures Notes. 121-134.

Charpentier, A. (2003). Tail distribution and dependence measures. *Proceedings of the 34th ASTIN Conference*.

Charpentier, A. (2004). Extremes and dependence: a copula based approach. *Proceedings of the 3rd Conference in Actuarial Science & Finance in Samos*.

Charpentier, A. and Juri, A. (2004). Limiting dependence structure for credit defaults. *to appear in Journal of Applied Probability*.

Charpentier, A., Fermanian, J.D. and Scaillet, 0. (2006). Beta kernel estimation of copula density for censored data. *in progress*.

Charpentier, A., Fermanian, J.D. and Scaillet, 0. (2005). Estimating copula densities. *in* Copula Methods in Derivatives and Risk Management: from Credit Risk to Market Risk. *to appear*.

Charpentier, A. (2006). 2003 heat wave and its return period. *submitted*.

Charpentier, A. and Segers, J. (2006a). Convergence of Archimedean Copulas. *CentER Discussion Paper*, 2006-28.

Charpentier, A. and Segers, J. (2006b). Lower tail dependence for Archimedean copulas: characterizations and pitfalls. *CentER Discussion Paper*, 2006-29.

Charpentier, A. and Segers, J. (2006c). Upper tail dependence for Archimedean copulas. *in progress*.

Charpentier, A. and Segers, J. (2006d). Tail dependence copulas and extreme value theory. *in progress*.

Charpentier, A. and Sibaï, D. (2006). Dynamic flood modelling: Combining Hurst and Gumbel's approach. *submitted*.

Chateauneuf, A. (1999). Comonotonicity axioms and rank-dependent expected utility theory for arbitrary consequences. *Journal of Mathematical Economics*, **32**, 21-45 .

Chateauneuf, A. (1991). On the use of capacities in modeling uncertainty aversion and risk aversion. *Journal of Mathematical Economics*, **20**, 343-369.

Chen, S.X. (1999). A beta kernel estimation for density functions. *Computational Statistics and Data analysis*. **31**. 131-135.

CHEN, S.X. (2000). Beta kernel for regression curve. *Statistica Sinica*. **10**. 73-92.

CHEN, S.X. AND FAN, Y. (2004a). Efficient Semiparametric Estimation of Copulas. *Working Paper*.

CHEN, S.X. AND FAN, Y. (2004b). Copula-based Tests for Dynamic Models. *Working Paper*.

CHENG, Q. (1992). On the Unique Representation of Non-Gaussian Linear Processes. *The Annals of Statistics*. **20**, 1143-1145

CHEUNG, Y.W. (1993). Long Memory in Foreign-Exchange Rates. *Journal of Business and Economic Statistics*, **11**, 93-102.

CHOQUET, G. (1955). Theory of capacities. *Annales de l'Institut Fourier (Grenoble)*. **5**. 131-295.

CHOW, Y.S. & TEICHER,H. (1997). Probability theory : independence, interchangeability, martingales. Springer Verlag.

CHRISTOFIDES, T.C. AND VAGGELATOU, E. (2004). A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, **88**, 138-151 .

CLAYTON, D.G.. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika*. **65**, 141-151.

COLANGELO, A., SCARSINI, M. AND SHAKED, M. (2003). Some Positive Dependence Stochastic Orders. *Journal of Multivariate Analysis*, **97**, 46-78.

COLES, S.G. AND TAWN, J.A. (1990). Statistics of coastal flood prevention. *Philosophical Transactions of the Royal Society of London, Series A*. **332**, 457-476.

COLES, S.G. AND TAWN, J.A. (1991). Modelling Extreme Multivariate Events. *Journal of the Royal Statistical Society. Series B (Methodological)*. **53**, 377-392.

COLES, S.G. AND TAWN, J.A. (1994). Statistical Methods for Multivariate Extremes: An Application to Structural Design. *Applied Statistics*. **43**, 1-48.

COLES, S. HEFFERNAN, J. AND TAWN, J.A. (1999). Dependence measures for multivariate extremes. *Extremes*. **2** 339-365.

COOK, R.D. AND JOHNSON, M.E.. (1981). A Family of Distributions for Modelling Non-Elliptically Symmetric Multivariate Data. *Journal of the Royal Statistical Society. Series B (Methodological)*. **43**. 210-218.

COSSETTE, H., DUCHESNE, T. AND MARCEAU, E. (2003). Modeling Catastrophes and their Impact on Insurance Portfolios. *North American Actuarial Journal* **7**, 1-22.

COWLING, A. AND HALL, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society, Series B*, **58**, 551-563.

CREDITMETRIC GROUP. (1997). CreditMetric technical document.

CUADRAS, CM.M. AND AUGÉ, J.. (1981). A continuous general multivariate distribution and its properties. *Comm. in Stat., Theor.Meth., A.* **10**, 339-353, 1981.

DABROWSKA, D. (1989). Kaplan-Meyer estimate on the plane: weak convergence, LIL and the bootstrap. *Journal of Multivariate analysis*, **29**, 308-325.

DACUNHA-CASTELLE, D. (2004). Evènements extrêmes. *Geosciences and Statisics Conference*.

Dall'Aglio, G. (1991). Frechet classes : the beginnings. *in* Advances in Probability Distribution with Given Marginals - Beyond the Copulas *Dall'Aglio,G., Kotz,S. & Salinetti,G. Ed. - Kluwer Academic Publishers* 3-12.

Daniela, K. and Rychlik, T. (2003). Sharp bounds for expectations of spacings from DDA and DFRA families. *Statistics & Probability Letters*, **65**, 303-316.

Daul, S., De Giorgi, E., Lindskog, P. and McNeil, A. (2003). Using the grouped t-copula. *Risk*, 73-76.

David, H.A. and Moeschberger, M.L. (1978). Theory of Competing Risks. Griffin's Statistical Monographs.

David, H.A. (1986). Inequalities for ordered sums. *Annals of the Institute of Statistical Mathematics* **38**. 551-555.

David, H.A. (1993). A note on order statistics for dependent variables. *American Statistician.* **47**. 198-199.

David, H.A. and Nagaraja, H.N. (2003). Order statistics. Wiley.

Davis, M. and Lo, V. (1999b). Modelling default correlation in bond portfolios. *ICBI Report on Credit Risk, edited by Carol Alexander.*

Davis, M. and Lo, V. (1999b). Infectious defaults. *Quantitative Finance*,**1**, 382-387,

Davis, D.R., Duckstein, L., & Fogel, M.M. (1976). Uncertainty in thereturn period of maximum hydrologic events: a Bayesian approach. *Journal of Hydrology*, **31**, 81Ű95.

Davison, A. and Hinkley, D. (1997). Bootstrap Methods and their Applications. Cambridge University Press, Cambridge.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincarré.* **7**, 1-68.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **40**, 317-337.

de Haan, L. and Resnick, S. (1978). Derivatives of regularly varying functions in $\mathbb{R}^d$ and domains of attraction of stable distributions. *Stochastic Processes Appl.* **8**. 349-355.

de Haan, L., Omey, E. and Resnick, S. (1984a). Domains of attraction of regular variation in $\mathbb{R}^d$. *Journal of Multivariate Analysis* **14**, 17-33.

de Haan, L. and Omey, E. (1984b). Integrals and derivatives of regularly varying function in $\mathbb{R}^d$ and domains of attraction of stable distributions II. *Stochastic Processes Appl.* **16**. 157-170.

de Haan, L.(1984) A spectral representation for max-stable processes. *Annals of Probability*, **12**, 1194-1204.

de Haan, L. and Ronde, J.(1998) Sea and wind: multivariate extremes at work. *Extremes*, **1**, 7-45.

Deheuvels, P. (1977). Estimation non paramétrique de la densité par histrogrammes généralisés (II). *Publications de l'Institut de Statistique de l'Université de Paris.* **22**, 1-23.

Deheuvels, P. (1979). Propriétés d'existence et propriétés topologiques des fonctions de dépendance avec applications à la convergence des types pour des lois multivariées. *Comptes Rendus de l'Académie des Sciences de Paris* **288** 145-148.

Deheuvels, P. (1981). A nonparametric test of independence *Publications de l'Institut de Statistique de l'Université de Paris* **26** 29-50.

Deheuvels, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis.* **13**. 257-271.

Deheuvels, P. (1984). Probabilistic aspects of multivariate extremes. *in* Statisical Extremes and Applications, Tiago de Olivera Eds. 117-130.

Deheuvels, P. (1985). Point processes and multivariate extreme values, II. *in* Multivariate Analysis VI, Krishnaiah Eds. 145-164.

Deheuvels, P. and Hominal, P. (1979). Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Revue de Statistique Appliquée*, **27**, 47-68.

Demarta, S. and McNeil, A.J. (2004). The t-Copula and Related Copulas. *International Statististical Review*, **73**, 111Ű129.

Dembo, A and Zeitouni, O. (1998). Large Deviations Techniques and Applications. Springer Verlag.

Dempster, A.P. & Liu, C. (1995).Trend and drift in climatological time series. *Proceedings of the 6th International Meeting on Statistical Climatology, Galway, Ireland*, 21-24.

Denneberg, D. (1990). Premium calculation : why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin* **20**, 181-190.

Denneberg, D. (1994). Non-additive measure and integral. Kluwer Academic Publishers.

Denuit, M. and Cornet, A. (1999). Premium calculation with dependent time-until-death random variables: the widow's pension. *Journal of Actuarial Practice.* **7**. 147-180.

Denuit, M. (2001). Laplace transform ordering of actuarial quantities. *Insurance: Mathematics and Economics.* **29**. 83-102.

Denuit, M. and Dhaene, J. (2003). Simple characterizations of comonotonicity and countercomonotonicity by extremal correlations. *Belgian Actuarial Bulletin.* **3**. 22-27.

Denuit, M. and Charpentier, A. (2004). Mathématiques de l'assurance non-vie, Tome 1: Principes fondamentaux de théorie du risque. Economica.

Denuit, M. and Charpentier, A. (2004). Mathématiques de l'assurance non-vie, Tome 2: Tarification et provisionnement. Economica.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). Actuarial theory for dependent risks. Wiley.

Denuit, M., Purcaru, O. and Van Keilegom, I. (2005). Bivariate archimedean copula modelling for loss-ALAE data in non-life insurance. *Journal of Actuarial Practice* to appear.

Deo, S. (1997). Nonparametric regression with long-memory errors. *Statistics & Probability Letters*, **33**, 89-94.

Durante, F. and Sempi, C. (2004). Semi-copulae. *Kybernetika* **41**, 315Ű328.

Devroye, L. and Györfi, L. (1985). Nonparametric Density Estimation: The L_1 View. John Wiley.

Devroye, L. ,Lugosi, G. and Udina, F. (2000). Inequalities for a new data-based method for selecting nonparametric density estimates. *in* Asymptotics in Statistics and Probability. Papers in Honor of George Gregory Roussas. VSP International Science Publishers, M.L. Puri (editor).

Devroye, L. and Lugosi, G. (2001). Combinatorial Methods in Density Estimation. Springer Verlag.

Devroye, L., Lugosi, G. and Udina, F. (2000). Inequalities for a new data-based method for selecting nonparametric density estimates, in *Asymptotics in Statistics and Probability: Papers in Honor of George Gregory Roussas*, Madan L. Puri ed., 133-154, VSP International Science Publishers.

Dhaene, J., and Denuit, M. (1999). The safest dependence structure among risks. *Insurance: Mathematics and Economics* **25**, 11-21.

Dhaene, J. and Goovaerts, M.J. (1996). Dependency of risks and stop-loss order. *Astin Bulletin* **26**, 201-212.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2001). The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, **31**, 3-33.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: applications. *Insurance: Mathematics and Economics*, **31**, 133-161.

Dias, A., Embrechts, P. (2003). Dynamic copula models for multivariate high-frequency data in finance. *in* Risk Measure for the 21st Century, G.Szegoe ed., Wiley Finance Series, 16, 321 - 335.

Draisma, G., Drees, H., Ferreira, A. and de Haan, L. (2004). Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, **10**, 251-280.

Drouet-Mari, D. and Kotz, S. (2001). Correlation and dependence. World Scientific Publishing Company Inc.

Durrleman, V., Nikeghbali, A. and Roncalli, T. (2000). Copulas approximation and new families. *Groupe de Recherche Opérationel du Crédit Lyonnais, Working Paper.*

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, **76**, 312-319.

Efron, B. and Tishirani, R.J. (1993). An introduction to bootstrap. Chapman & Hall.

Ehrenberg, A.S.C. (1952). On Sampling from a Population of Rankers. *Biometrika*. **39**, 82-87.

Einmahl, J.H.J. (1987) Multivariate empirical processes. Center for Mathematical and Computer Science, Amsterdam.

Einmahl, J.H.J., Mason, D.M. & Ruymgaart, F.H. (1988) Processus empiriques multidimensionnels : aperçu de quelques résultats récents. *International statistical review*, **56**, 139-151.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) Modelling extremal events for insurance and finance. Springer, Berlin.

Embrechts, P., McNeil, A.J. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk management: value at risk and beyond*, edited by Dempster M., Cambridge University Press, 176-223.

Embrechts, P., Hoeing, A. and Juri, A. (2004). Using Copulae to bound the Value-at-Risk for functions of dependent risks. *Finance and Stochastics*. **7**, 145-167.

Embrechts, P. (2004). From Dutch Dykes to Value-at-Risk: Extreme Value Theory and Copulae as Risk Management Tools. *Nomura lectures, University of Oxford.*

Embrechts, P. and Puccetti, G. (2004). Bounds on Value-at-Risk. *ETHZ Preprint.*

ENGLE, R.F. & RUSSELL, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, **66**, 1127–1162.

ENGLE, R.F. (2000). The Econometrics of Ultra High Frequency Data, *Econometrica*, **68**, 1–22.

ENGLE, R.F. & LUNDE, A. (2003). Trades and quotes: A Bivariate Point Process. *Journal of Financial Econometrics* **1**, 159-188.

ESARY, J.D. AND PROSCHAN, F. (1972). Relationships Among Some Concepts of Bivariate Dependence. *The Annals of Mathematical Statistics*, **43**, 651-655.

ESARY, J.D. AND MARSHALL, A.W. (1979). Multivariate distributions with increasing hazard rate average. *Annals of Probability*, **7**, 359-370.

EYRAUD, H. (1934). Sur une représentation nouvelle des corrélations continues. *Comptes Rendus de l'Académie des Sciences de Paris*, **199**, 1356-1358.

FANG, K.T., KOTZ, S. AND NG, K.W. (1990). Symmetric Multivariate and Related Distributions. Chapman and Hall.

FELLER, W.. (1971). An Introduction to Probability Theory and Its Applications, Vol. 2. Wiley.

FERMANIAN, J.D.(2005). Goodness of fit for copulas. *Journal of Multivariate Analysis*, **95**, 119-152.

FERMANIAN, J.D., RADULOVIĆ, D. AND WEGKAMP, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847-860.

. FERMANIAN, J.D. AND SCAILLET, O.(2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, **5**, 25-54.

FERMANIAN, J.D. AND SCAILLET, O. (2004). Some Statistical Pitfalls in Copula Modeling for Financial Applications. *FAME Working Paper*, **108**.

FINDLEY, D.F. (1986). The Uniqueness of Moving Average Representations with Independent and Identically Distributed Random Variables for Non-Gaussian Stationary Time Series. *Biometrika*, **73**, 520-521.

FINDLEY, D.F. (1986). Amendments and Corrections: The Uniqueness of Moving Average Representations with Independent and Identically Distributed Random Variables for Non-Gaussian Stationary Time Series. *Biometrika*, **77**, 235.

FOUGÈRES, A.L. (2003). Multivariate extremes. in *Extreme Values in Finance, Telecommunications, and the Environment*, Finkenstaedt, B. and Rootzen, H. Eds. Chapman and Hall.

FOUFOULA-GEORGIOU, E. (1999). On scaling theories of space-time rainfall: some recent results and open problems. *in* Stochastic methods in hydrology, ed. Barndorff-Nielsen, Gupta, Pérez-Abreu & Waymire Eds. Wolrd Scientific.

FOX, R. & TAQQU, M. (1986). Large sample properties of parameter estimates for strongly dependent stationary time series. *The Annals of Statistics*, **14**, 517–532.

FRANK, M.J., NELSEN, R.B. AND SCHWEIZER, B.. (1987). Best-possible bounds for the distribution of a sumŮa problem of Kolmogorov. *Probabability Theory and Related Fields*, **74**, 199-211.

FRÉCHET, M. (1958). Remarques au sujet de la note précédante *Comptes Rendus de l'Académie des Sciences de Paris*. **249**. 2719-2720.

Fredricks, G.A., Nelsen, R.B. and Rodriguez-Lallena, J.A. (2005). Copulas with fractal support. *Insurance: Mathematics and Economics*, **37**, 42-48.

Frees, E.W., Carriere, J. and Valdez, E. (1996). Annuity valuation with dependent mortality. *The Journal of Risk and Insurance.* **63**. 229-261.

Frees, W.E. and Valdez, E.A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* **2**, 1-25.

Frey, R. and McNeil, A.J. (2004). Dependent defaults in models of portfolio credit risk. *Journal of Risk*, **6**, 59-92.

Gaaralda Guillem, A.I. (2000). Structure de dépendance des lois de valeurs extrêmes bivariées. *Comptes Rendus de l'Académie des Sciences de Paris.* **330**. 351-354.

Gagliardini, P. and Gouriéroux, C. (2006). An efficient nonparametric estimator for models with nonlinear dependence. *Journal of Econometrics*, to appear.

Gaenssler, P. and Stute, W. (1987). Seminar on Empirical Processes. Birkhäuser Verlag, Basel.

Gajek, L. and Rychlik, T. (1996). Projection method for moment bounds on order statistics from restricted families. I. Dependent case. *Journal of. Multivariate Analysis.* **57** 156-174.

Gajek, L. and Rychlik, T. (1998). Projection method for moment bounds on order statistics from restricted families. II. Independent case. *Journal of the American Statistical Association.* **70**. 674-680.

Galambos, J. (1975). Order Statistics of Samples from Multivariate Distributions. *Journal of. Multivariate Analysis.* **64**. 156-182.

Galambos, J. (1978). The asymptotic theory of extreme order statistics. Krieger.

Gasser, T. and Müller, H.G. (1979). Kernel Estimation of Regression Functions. *in* Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics **757**, eds. T. Gasser and M. Rosenblatt, Springer-Verlag, 23-68.

Geluk, J.L. and de Vries, C.G. (2006). Weighted sums of subexponential random variables and asymptotic dependence between returns on reinsurance equities *Insurance: Mathematics and Economics*, **38**, 39-56.

Genest, C. (1986a). The joy of copulas : bivariate distributions with uniform marginals. *The American Statistician*, **40**, 4 280-283.

Genest,C. and MacKay, R.J. (1986b). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *La revue canadienne de statistique* **14**, 145-159.

Genest, C. and Rivest, L.P. (1993). Statistical inference procedures for bivariate Archimedian copulas. *Journal of the American Statistical Association* **88** 423 1034-1043.

Genest, C., Ghoudi, K. and Rivest, L.P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika.* **82** 543-552.

Genest, C., Quesada Molina, J.J and Rodríguez Lallena, J.A. (1995). De l'impossibilité de construire des lois à marges multidimensionelles données à partir de copules. *Comptes Rendus de l'Académie des Sciences de Paris.* **320** 723-726.

Genest, C., Ghoudi, K. and Rivest, L.P. (1998). *Commentaire sur l'article de E. W. Frees et E. A. Valdez* Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 143-149.

Genest, C., Quesada Molina, J.J., Rodríguez Lallena, J.A. and Sempi, C. (1999). A characterization of quasi-copulas. *Journal of Multivariate Analysis*, **69**, 193-205.

Genest, C. and Rivest, L.P. (2001). On the multivariate probability integral transformation. *Statistics and Probability Letters*, **53**, 391-399.

Genest, C. and Werker, J.M. (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. in *Distributions With Given Marginals and Statistical Modelling*, Cuadras Fortiana and Rodríguez Lallena Eds, Kluwer, 103-112.

Genest, C. and Boies, J.C. (2003). Detecting Dependence With Kendall Plots. *The American Statistician*. **57**. 275-284.

Genest, C. and Verret, F. (2005). Verret, Locally most powerful rank tests of independence for copula models. *Journal of Nonparametric Statistics*, **17**, 521-539.

Genest, C., Quéssy, F., and Rémillard, B. (2006). On the joint asymptotic behavior of two rank-based estimators of the association parameter in the gamma frailty model. *Statistics & Probability Letters*, **76**, 10-18.

Geoffroy, J. (1958). Contributions à la théorie des valeurs extrêmes. *Publications de l'Institut de Statistique de l'Université de Paris* **7**/**8** 37-185.

Ghosal (2000). Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics*. **29**, 1264-Ũ1280

Ghoudi, K. Khoudraji, A. and Rivest, L.P. (1998). Propriétés statistiques de copules de valeurs extrêmes bidimensionelles. *Revue canadienne de statistique*, **26**, 187-197.

Ghoudi, K. and Rémillard, B. (1998). Empirical processes based on pseudo-observations. in *Asymptotic methods in probability and statistics: a volume in honor of Miklos Csörgo*, B. Szyskowicz ed., 171-197.

Ghoudi, K. and Rémillard, B. (2004). Empirical processes based on pseudo-observations II: the multivariate case. *Fields Institute Communications Series - AMS* **44**, 341-406.

Ghurye, S.G. and Marshall, A.W. (1984). Shock processes with aftershocks and multivariate lack of memory. *Journal of Applied Probability*, **21**, 786-801.

Gijbels, I. and Mielniczuk, J. (1990). Estimating the Density of a Copula Function. *Communications in Statistics: Theory and Methods*, **19**, 445-464.

Gill, R. D., Vardi, Y. et Wellner, J. (1988). Large sample theory of empirical distributions in biased sampling models. *Annals of Statistics*, **16**, 1069-1112.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics*. **44**. 423-453.

Gneiting, T. (1997). Radial positive definite functions generated by Euclid's hat. *Journal of Multivariate Analysis*, **69**, 88-119.

Gokhale, D.V. (2000). On asymptotic relative efficiencies of a class of rank tests for independence of two variables. *The Annals of the Institute of Statistical Mathematics*. **20**, 255-261.

Goldie, C. and Resnick, S. (1989). Records in a partially ordered set. *The Annals of Probability*. **17**, 678-699.

Gray, H.L., Zhang, N.F., & Woodward, W. (1989). On Generalized Fractional Processes. *Journal of Time Series Analysis*, **10**, 233-57.

Greene, M. & Fielitz, B. (1977). Long term dependence in common stock returns. *Journal of financial econometrics*, **4**, 339-349.

Gumbel, E.J. (1941). The Return Period of Flood Flows. *The Annals of Mathematical Statistics*, **12**, 163-190.

Gumbel E. (1958). Statistics of Extreme. Columbia University Press.

Gumbel, E.J. (1959). Mutltivariate distributions with given margins. *Revista da Faculdade de Ciências de Lisboa (2) A*, **7**, 179-218.

Gumbel, E.J. (1960). Distributions des valeurs extremes en plusieurs dimensions.*Publications de l'Institut de Statistique de l'Université de Paris.* **9** 171-173.

Gumbel, E.J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association* **56** 335-349.

Gumbel, E.J. and Mustafi, C.K. (1967). Some Analytical Properties of Bivariate Extreme Distributions. *Journal of the American Statistical Association* **62** 569-588.

Gupta, R.C On some association measures in bivariate distributions and their relationships. *Journal of Statistical Planning and Inference.* **117**, 83-98.

Hájek, J. and Sidák, Z. (1967). Theory of Rank Tests. Academic Press.

Hájek, J. (1969). A course in nonparametric statistics. Holden-Day.

Hájek, J., Zbyněk, Š. and Sen, P.K.(1999). Theory of rank tests. Academic Press.

Hannan, E.J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability*, **10**, 130Ũ145.

Hall, P. and Park, B.U. (2002). New methods for bias correction at endpoints and boundaries. *Annals of Statistics* **30**, 1460-1479

Hardy, G.H., Littlewood, J.E. and Polya, G. (1934). Inequalities. Cambridge university press.

Harrell, F.E. and Davis, C.E. (1982). A new distribution-free quantile estimator. *Biometrika* **69**, 635-640.

Harris, R.A. (1970). Reliability applications of a bivariate exponential distribution, *Operations Research* **16**, 16-27.

Harris, R.A. (1970). A multivariate definition for increasing hazard rate distribution function, *Annals of Mathematical Statistics* **41**, 713-717.

Haslett, J. & Raftery. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resources. *Journal of the Royal Statistical Society.* **38**. 1-50.

Hauksson, H.A., Dacorogna, M., Domenig, T., Müller, U. & Samorodnisky, G. (2001). Multivariate extremes, aggregation and risk estimation. *Quantitative Finance*, **1**, 79-95.

Heilmann, W.R. (1986). On the impact of independence of risks on stop-loss premiums. *Insurance: Mathematics and Economics* **5**, 197-199.

Heffernan, J.E. (2000). A directory of Coefficients of Tail dependence. *Annals of Mathematical Statistics.* **41**, 713-717.

Heffernan, J. E. and Tawn, J. A. (2004). A Conditional Approach for Multivariate Extreme Values. *Journal of the Royal Statistical Society. Series B (Methodological).* **66**. 497-546 .

Heffernan, J. E. and Resnick S. (2004). Hidden Regular Variation and the Rank Transform. *Cornell Reports.* **1409**.

Hesselager, O., Wang, S. and Willmot, G.E. (1998). Exponential and scale mixtures and equilibrium distributions. *Scandinavian Actuarial Journal,* **98**, 125-142.

Hillali, Y. (1998). Analyse et modélisation des données probabilistes: capacités et lois multidimensionnelles. Thèse de doctorat, Université Paris Dauphine.

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Matematischen Instituts für Angewandte Matematik der Universitat Berlin.* **5**, 181-233, *reprinted as* Scale- Invariant correlation theory *in* The collected works of Wassily Hoeffding, Fisher and Sen eds., Springer Verlag, (1994), 57-107.

Hoeffding, W. (1941). Masstabinvariante Korrelationsmasse für Diskontinuierliche Verteilungen. *Archiv für Mathematische Wirtscafts.* **7**, 49-70, *reprinted as* Scale- Invariant correlations for discontinuous distributions *in* The collected works of Wassily Hoeffding, Fisher and Sen eds., Springer Verlag, (1994), 109-132.

Hoeffding, W. (1942). Stochastische Abhängigkeit und Funktionaler Zusammenhang. *Skandinavisk Aktuarietidskrift.* **25**, 200-227, *reprinted as* Stochastic dependence and functional relationships *in* The collected works of Wassily Hoeffding, Fisher and Sen eds., Springer Verlag, (1994), 109-132.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics.* **19**, 293-325.

Hollander, M., and Wolfe, D.A. (1999). Nonparametric statistical methods. Wiley Interscience.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika*, **68**, 165-76.

Hosking, J.R.M. (1984). Testing whether the shape parameter is zero in the generalized Extreme-Value Distribution. *Biometrika*, **71**, 367-374.

Hosking, J.R.M.(1984). Testing whether the shape parameter is zero in the generalized Extreme-Value Distribution. *Biometrika*, **71**, 367-374.

Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *The Annals of Statistics*, **25**, 105-137.

Hossjer, O. and Ruppert, D. (1995). Asymptotics for the Transformation Kernel Density Estimator (in Curve Estimation). *The Annals of Statistics.* **23**, 1198-1222.

Hougaard, P. (1986). A Class of Multivariate Failure Time Distributions. *Biometrika.* **73**, 671-678.

Hu, T. and Hu, J. (1998). Comparison of order statistics between dependent and independent random variables. *Statistical and Probability Letters.* **37**. 1-6.

Hu, T., Khaledi, B.-E. and Shaked, M. (2003). Multivariate hazard rate orders. *Journal of Multivariate Analysis*, **84**, 173-189.

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability.* **34**. 587-608.

Hurst, H.E. (1951). Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civil Engineers*, **116**, 770–799.

I.A.A. (2004). A global framework for Insurer Solvency Assessment. http://www.actuaries.org/LIBRARY/Papers/

Ingersoll, J. (1987). Theory of financial decision making. Rowman and Littlefield.

Jasiak, J. (1999). Long Memory in Economics. *Journal de la Societe Francaise de Statistique*, **140**, 61-67.

Jensen, D.R. and Mayer, L.S. (1977). Some variational results and their applications in multiple inference. *Annals of Statistics*. **5**, 922-931.

Jirina, M. (1976). On the asymptotic normality of Kendall's rank correlation statistic. *Annals of Statistics*. **4**. 214-215.

Joag-Dev, K. (1984). Measures of dependence. *in* Handbook of Statitics, P.R. Krishnaiah & P.K. Sen Eds.

Jodgeo, K. (1978). On the probability bound of Marshall and Olkin. *The Annals of Statistics*. **6**. 232-234.

Joe, H. (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistical and Probability Letters*, **9**, 75-81.

Joe, H., Smith, R.L. and Weissman, I. (1992). Bivariate Threshold Methods for Extremes. *Journal of the Royal Statistical Society. Series B (Methodological)*. **54**. 171-183.

Joe, H. (1993). Parametric family of of multivariate distributions with given margins. *Journal of the Multivariate Analysis*. **46**. 262-282.

Joe, H. (1997). Multivariate models and dependence concepts. Chapman-Hall.

Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, **3**, 135-146.

Jones, M.C. (1996). The local dependence function. *Biometrika*. **83**. 899-904.

Jones, M.C. (1998). Constant local dependence. *Journal of Multivariate Analysis*. **64**. 148-155.

Jouini, M. and Clemen, R.T. (1996). Copula Models for Aggregating Expert Opinions. *Operations Research*, **44**, 444-457.

Junker, M., Szimayer, A. and Wagner, N. (2006). Nonlinear term structure dependence: Copula functions, empirics, and risk implications. *Journal of Banking & Finance*, **30**, 1171-1199.

Juri, A. and Wüthrich, M.V. (2003). Copula convergence theorems for tail events. *Insurance: Mathematics and Economics* **30**, 411-427.

Juri, A. and Wüthrich, M.V. (2004). Tail dependence from a distributional point of view. *Extremes* **6**, 213-246.

Kaas, R., van Heerwaarden, A.E., and Goovaerts, M.J. (1994). Ordering of actuarial risks. CAIRE, Brussels.

Kaledhi, B.E. and Kochar, S.C. (2001). Dependence properties of multivariate mixture distributions and their applications. *Annals of the Institute of Statistical Mathematics*. **53**. 620-630.

Karl, T.R. & Trenberth, K.E. (2003). Modern Global Climate Change. *Science*, **302**, 1719.

Karl, T.R., & Knight, R.W. (1997=. The 1995 Chicago heat wave: How likely is a recurrence? *Bulletin of the American Meteorological Society*, **78**, 1107-1119.

Karunamuni, R.J. and Albert, T. (2005a). A Locally Adaptive Method of Boundary Correction in Kernel Density Estimation. *Journal of Statistical Planning and Inference*, to appear.

Karunamuni, R.J. and Albert, T. (2005b). A Generalized Reflection Method of Boundary Correction in Kernel Density Estimation. *Canadian Journal of Statistics*, to appear.

Kelker, D. (1970). Distribution theory of spherical distributions and a locationŰscale parameter generalization. *Sankhia*, **32**, 419Ű430.

Kelly, K.S. & Krzysztofowicz, R. (2000). Precipitation uncertainty processor for probabilistic river stage forecasting. *Water resources Research*, **36**, 2643-2653.

Kelly, K.S. & Krzysztofowicz, R. (2001). Hydrologic uncertainty processor for probabilistic river stage forecasting : precipitation dependent model. *Journal of hydrology*, **249**, 46-68.

Kim, J.S., Proschan, F. and Sethuraman, J. (1988). Stochastic comparisons of order statistics, with applications in reliability. *Communications in Statistics - Theory Meth.*. **17**. 2151-2172.

Kimberling, C.H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes Math.*, **10**, 152Ű164.

Kimeldorf, G., and Sampson, A.R. (1989). A framework for positive dependence. *Annals of the Institute of Statistical Mathematics.*. **41**. 31-45.

Klugman, S.A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance : Mathematics and Economics*. **24** 139-148.

Klüppelberg, C. & Mikosh, T. (1996). Self-normalised and randomly centred spectral estimated. *Procedings of the Athens International Conference on Applied Probability and Time Series*, Vol **2**, 259-271, Springer.

KMV-Corporation (1997). Modelling default risk. *Technical document*.

Kochar, S.and Korwar, R. (1996). Stochastic orders for spacings of heterogeneous exponential random variables. *Journal of Multivariate Analysis* **57** 69-83.

Kochar, S.C. and Rojo, J. (1996). Some New Results on Stochastic Comparisons of Spacings from Heterogeneous Exponential Distributions. *Journal of Multivariate Analysis*, **59** (2) 272-281.

Koscielny-Bunde, E., Kantelhardt, J.W., Braun, P., Bunde, A., Havlin, S. (2006). Long-term persistence and multifractality of river runoff records. *Journal of Hydrology*, **322**, 120-137.

Kovats, R.S. & Koppe, C. (2005). Heatwaves past and future impacts on health. *in* Ebi, K., J. Smith and I. Burton (eds). Integration of Public Health with Adaptation to Climate Change: Lessons learned and New Directions. Taylor & Francis Group, Lisse, The Netherlands.

Kruskal, W.H. (1958). Ordinal Measures of Association. *Journal of the American Statistical Association*. **53**. 814-861.

Kulpa, T. (1999). On approximation of copulas. *International Journal of Mathematics and Mathematical Science*. **22**. 259-269.

Lancaster, H.O (1958). The structure of bivariate distributions. *Annals of Mathematical Statistics*, **29**, 719-736.

Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research* **2**, 99-120.

LANZINGER, H. AND STADTMÜLLER, U. (2002). Tauberian Theorems and Limit Distributions for Upper Order Statistics. *Publications de l'Institut de Mathématiques de Belgrade*. **71**, 41-53.

LEDFORD, A.W. AND TAWN, J.A. (1996). Statistics for Near Independence in Multivariate Extreme Values. *Biometrika*. **83** 169-187.

LEDFORD, A.W. AND TAWN, J.A. (1997). Modelling Dependence Within Joint Tail Regions. *Journal of the Royal Statistical Society. Series B (Methodological)*. **59** 475-499.

LEDFORD, A.W. AND TAWN, J.A. (1998). Concomitant tail behaviour for extremes. *Advances in Applied Probability*. **30** 197-215.

LEFÈVRE, CL. AND UTEV, S. (1997). Mixed Poisson approximation in the collective epidemic model. *Stochastic Processes and their Applications* **69**, 217-246.

LEHMANN, E.L. (1955). Ordered Families of Distributions. *Annals of Mathematical Statistics*. **26**. 399-419.

LEHMANN, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*. **37**. 1137-1153.

LEHMANN, E.L. (1975). Nonparametrics. Statistical Methods Based on Ranks. Holden-Day Series in Probability and Statistics.

LEHMANN, E.L. (1986). Testing statistical hypotheses. Wiley Interscience.

LEHMANN, E.L. (1998). Nonparametrics: statistical methods based on ranks. Prentice Hall.

LEHR, M. & LII, K.S. (1998). Maximum Likelihood Estimates of Non-Gaussian ARMA Models. *Working paper*.

LI, H. , SCARSINI, M. AND SHAKED, M. (1996). Linkages: a tool for the construction of multivariate distributions with given nonoverlapping multivariate marginals, *Journal of Multivariate Analysis*, **56**, 20-41.

LI, X., MIKUSINSKI, P., SHERWOOD, H. AND TAYLOR, M.D. (1996). On Approximation of Copulas. in *istributions with Given Marginals and Moment Problems*. Benes and Stepan Eds. 107-1-16.

LI, X., MIKUSINSKI, P. AND TAYLOR, M.D. (1998). Strong approximation of copulas. *Journal of Mathematical Analysis and Applications*. **225**. 608-623.

LI, D.X. (1999). On default correlation: a copula function approach. *RiskMetrics Group, Working Paper*.

LII, K.S. & ROSENBLATT, M. (1996). Maximum Likelihood Estimates of Non-Gaussian nonminimum phase ARMA sequences. *Statistica Sinica*, **6**, 1-22.

LILLO, R.E., NADA, A.K. AND SHAKED, M. (2001). Preservation of some likelihood ratio stochastic orders by order statistics. *Statistics & Probability Letters*. **51** 111-119.

LIN, D. AND YING, Z. (1993). A Simple Nonparametric Estimator of the Bivariate Survival Function under Univariate Censoring. *Biometrika*, **80**, 573-581.

LIU, C. & RUBIN, D.B. (1994). The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence. *Biometrika*, **81**, 633-648.

LLOYD, E.H. (1970). Occurrence interval in the presence of persistence. *Journal of Hydrology*, **10**, 291Ű298.

LO, A.W. (1991). Long term memory in stock market prices. *Econometrica*, **59**, 1276-1313.

Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960-1990 ?. *Journal of International Money and Finance.* **14**. 3-26.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, *Journal of Finance.* **56**. 651-678.

Loretan, M. and English, W.B. (2000). Evaluating correlation breakdowns dug periods of market volatiliy. *Internation Finance Discussion Paper.* **658**.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M. & Wanner, H. (2004). European sea-sonal and annual temperature variability, trends, and extremes since 1500. *Science*, **303**, 1499Ű1503.

Ma, C. (1997). A note on stochastic ordering of order statistics. *Journal of Applied Probability.* **34**. 785-789.

Madsen, R.W. (1993). Generalized binomial distribution. *Communications in statistics. Theory and methods.* **22**. 3065-3086.

Malevergne, Y. and Sornette, D. (2002). Investigating Extreme Dependences: Concepts and Tools. *Working Paper.*

Makarov, G.D. (1981). Estimates for the distribution function of the sum of two random variables with given marginal distributions, *Teor. Veroyatnost. i Primenen.* **26**, 815-817 (in Russian), translated in *Theory Probab. Appl.* **26** 803-806.

Mandelbrot, B.B. (1965). Une classe de processus stochastiques homothétiques à soi: application à la loi climatique de H.E. Hurst. *Comptes Rendus de l'Académie des Sciences.* **260**, 3274-3277

Mandelbrot, B.B. & Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev*, **10**, 422-437.

Mandelbrot, B., & Wallis, J.R. (1968). Noah, Joseph and operational hydrology. *Water Ressources Research*, **4**, 909-918.

Mardia, K.V.(1962). Multivariate Pareto Distributions. *The Annals of Mathematical Statistics.* **33**. 1008-1015.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, **7**, 77-91.

Markowitz, H. (1976). Portfolio selection. Yale University Press.

Marron, J.S. and Ruppert, D. (1994). Transformations to Reduce Boundary Bias in Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological).* **56**, 653-671.

Marshall, A.W. and Olkin, I. (1967). A generalized bivariate exponential distribution. *Journal of Applied Probability.* **4**. 291-302.

Marshall, A.W. and Olkin, I. (1979). Inequalities: theory of majorization and its applications. Academic Press.

Marshall, A.W. and Olkin, I. (1983). Domains of Attraction of Multivariate Extreme Value Distributions. *The Annals of Probability.* **11**. 168-177.

Marshall, A.W. and Olkin, I. (1988). Families of Multivariate Distributions. *Journal of the American Statistical Association.* **84**. 834-841.

Marshall, A.W. and Olkin, I. (1991). Functional equations for multivariate exponential distributions. *Journal of Multivariate Analysis.* **39**. 209-215.

Marshall, A.W. and Shaked, M. (1979). Multivariate Shock Models for Distributions with Increasing Hazard Rate Average. *The Annals of Probability.* **7**. 343-358.

Martins, E.S.P.R., & Clarke, R.T. (1993). Likelihood-based confidence intervals for estimating floods with given return periods. *Journal of Hydrology*, **147**, 61-81.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*. **18**. 1269-1283.

Maurer, W. and Margolin, B.H. (1976). The multivariate inclusion-exclusion formula and order statistics from dependent variates. *Annals of Statistics*. **4**. 1190-1199.

Maulik, K. and Resnick, S. (2003). Characterizations and examples of hidden regular variation. *Technical Reports, Cornell University.*

McCullagh, P., & Nelder J.A. (1988). Generalized Linear Models. Chapman and Hall.

Meehl, G.A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305**, 994-997.

Meerschaert, M. (1988). Regular variation in $\mathbb{R}^k$. *Procedings of the Americian Mathematical Society*. **102**. 341-348.

Meerschaert, M.M. and Scheffler, H.P. (2001). Limit distributions for sums of independent random vectors. Wiley-Interscience.

Merton, M. (1974). On the Pricing of Corporate Debt: the risk structure of interest rates. *Journal of Finance*, **28**, 449-470.

Mikhailov, V.G. (1974). Asymptotic independence of vector components of multivariate extreme value statistics *Theory Prob. Appl.* **19**. 817-821.

Mikosch, T. (2001). Modelling dependence and tails of financial time series. in *Extreme Values in Finance, Telecommunications, and the Environment*, Finkenstaedt, B. and Rootzen, H. Eds. Chapman and Hall.

Mikosch, T. (2005). Copulas: tales and facts. *Preprint.*

Montanari, A. (2003). Long-range dependence in hydrology. *in* Theory and applications of long-range dependence, Doukhan, Oppenheim & Taqqu, eds. Birkhäuser. 461-472.

Moran, P.A.P. (1950). Recent developments in ranking theory. *Journal of the Royal Statistical Society. Series B (Methodological)*. **12**. 153-162.

Morton, I.D. and Bowers, J. (1996). Extreme value analysis in a multivariate offshore environment. *Applied Ocean Research*, **18** 303-317.

Muliere, P. and Scarsini, M. (1987). Characterization of Marshall-Olkin type class of distributions. *Annals of the Institute of Statistical Mathematics*, **39** 429-441.

Müller, A. (1996). Ordering of risks: a comparative study via Stop-Loss transforms. *Insurance Mathematics and Economics*. **17**. 215-222.

Müller, A. (1997). Stop-loss order for portfolios of dependent risks. *Insurance Mathematics and Economics*. **21**. 219-223.

Müller, A. and Stoyan, D. (2001). Comparison methodes for stochastic models and risks. Wiley.

Müller, A. and Scarsini, M. (2001). Stochastic Comparison of Random Vectors with a Common Copula. *Mathematics of Operations Research*. **26**, 723-740.

Müller, A. and Scarsini, M. (2004). Archimedean copulae and positive dependence. *Journal of Multivariate Analysis*, **93**, 434Ű445.

Müller, H. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78**, 521-520.

Nagaraja, H.N. and Baggs, G.E. (1996). Order statistics of bivariate exponentiel random variables. *in* Statistical Theory and Applications, *eds.* Nagaraja, Sen and Morrison (1996), Springer Verlag. 129-141.

Nelsen, R.B., Quesada Molina, J.J., Schweizer, B. and Sempi, C. (1996). Derivability of some operations on distribution functions. in *Distributions with Fixed Marginals and Related Topics*, L. Rüschendorf, B. Schweizer and M.D. Taylor ed. IMS, 233-243.

Nelsen, R.B. (1999). An introduction to copulas. Springer Verlag.

Nelsen, R.B., Quesada Molina, J.J., Rodríguez-Lallena, J.A. and Úbeda-Flores, M.. (2002). Multivariate archimedean quasi-copulas. in *Distributions with Given Marginals and Statistical Modelling*, C.M. Cuadras et al. ed. Kluwer, 179-185.

NERC. (1975). Flood Studies Report Vol 1 - Hydrological Studies. Natural Environment Research Council, London.

Norberg, R. (1989). Actuarial analysis of dependent lives. *Bulletin de l'Association Suisse des Actuaires.* **89**. 243-254.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society. Series B (Methodological).* **44**. 414-422.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* . **84**. 487-493.

Oakes, D. (2001). *Biometrika* centenary: survival analysis. *Biometrika.* **88**. 99-142.

Olsen, R. (1982). Distributional tests for selectivity bias and a more robust likelihood estimator. *International Economic Review.* **23**. 223-240.

Olsen, E.T., Darsow, W.F. and Nguyen, B. (1996). Copulas and Markov operators. in *Proceedings of the conference on distributions with fixed marginals and related topics.* L. Rüschendorf, B. Schweizer and D. Taylor ed. IMS Lecture Notes, 244-259.

Ould Haye, M. & Viano, M.C. (2003). Limit theorems under seasonal long-memory. *in* Theory and applications of long-range dependence, Doukhan, Oppenheim & Taqqu, eds. Birkhäuser. 101-110.

Owen, J. and Rabinovitch, R. (1983). On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice. *Journal of Finance.* **38**. 745-752.

Palma, W. & Chan, N.H. (2005). Efficient estimation of seasonaly long-range dependent process. *Journal of Time Series Analysis*, **26**, 863-892.

Pandey, G., Lovejoy, S., & Schertzer, D. (1998). Multifractal analysis of daily river flows including extremes for basins of five to two million square kilometres, one day to 75 years. *Journal of Hydrology*, **208**, 62-81.

Papadatos, N. & Rychlik, T. (2003). Bounds on expectations of L-statistics from without replacement samples. *Journal of Statistical Planning and Inference.* In Press.

Parkes, C.M., Benjamin, B. and Fitzgeralds, R.G. (1969). Broken heart: a statistical study of increased mortality among widowers. *British Medical Journal.* 740-743.

Patton, A. (2006). Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, **47**, 527-556. *Previously circulated as* Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula. *University of California, San Diego, Discussion Paper* **01-09**.

Pellerey, F. and Shaked, M. (1997). Characterizations of the IFR and DFR aging notions by means of the dispersive order. *Statistics & Probability Letters*, **33** 389-393.

Pelletier, J.D. & Turcotte, D.L. (1997). Long-range persistence in climatological and hydrological time series: analysis, modeling and application to drought hazard assessment. *Journal of Hydrology*, **203**, 198-208.

Pickands, J. (1976). A class of multivariate negative exponential distributions. *Preprint, University of Pennsylvania.*

Pickands, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Association* **62**, 859-878.

Pirard, P., Vandentorren, S., Pascal, M., Laaidi, K., Le Tertre, A., Cassadou, S. & Ledrans, M. (2005). Summary of the mortality impact assessment of the 2003 heat wave in France. *EuroSurveillance*, **10**, 153-156.

Plackett, R.L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika*. **34**. 120-122.

Polasek, W. & Pai, J. (1998). Autoregressive moving average models with t and hyperbolic innovations. *Universitat Basel, Institut fur Statistik und Okonometrie Working Paper*, **0023**.

Pollard, D. (2002). A user's guide to measure theoretic probability. Cambridge University Press.

Pollock, D.S.G. (1998). Time-Series Analysis Signal Processing and Dynamics. Academic Press.

Puente C.E. (1996). A new approach to hydrologic modeling: derived distributions revisited. *Journal of Hydrology*, **187**, 65-80.

Pratt, J.W. (1964) Risk Aversion in the Small and in the Large. *Econometrica*, **32**, 122-136.

Preston, M. and Baratta, P. (1948). An experimental study of auction-value of uncertain outcome. *American Journal of Psychology*, **61**, 183-193.

Reiss, D. and Thomas, M. (2001). Statistical Analysis of extreme values: from insurance, finance, hydrology and other fields. Birkhaüser Verlag.

Renault, O. and Scaillet, O. (2004). On the way to recovery: a nonparametric bias free estimation of recovery rate densities. *Journal of Banking and Finance*, **28**, 2915-2931.

Rényi,A. (1959). On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae* **10**, 441-451.

Resnick, S.I. (1986). Point processes, regular variation and weak convergence. *Advances in Applied Probabilities*, **18**, 66-138.

Resnick, S. I. (1987). *Extreme values, Regular variation, and point processes.* Springer.

Resnick S. and Feigin, P. (1996). Pitfalls of Fitting Autoregressive Models for Heavy-tailed Time Series. *Cornell Reports* **1163**.

Resnick S. (2002a). The Extremal Dependence Measure and Asymptotic Independence. *Cornell Reports*. **1346**.

Resnick S. (2002b). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*. **5**. 3003-336.

Resnick S. (2004). On the Foundations of Multivariate Heavy Tail Analysis. *Journal of Applied Probability*. **41**. 191-Ũ212.

ROMANO, J.P. AND SIEGEL, A.F. (1986). Counterexamples in Probability and Statistics. Wadsworht and Brooks.

ROSENBLATT, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics.* **27** 832-837.

ROSENBLATT, M. (1985). Stationary Sequences and Random Fields. Birkhäuser.

RUPPERT, R. AND CLINE, D.B.H. (1994). Bias Reduction in Kernel Density Estimation by Smoothed Empirical Transformations (in Curve Estimation). *The Annals of Statistics.* **22**, 185-210.

RÜSCHENDORF, L. (1976). Asymptotic Distributions of Multivariate Rank Order Statistics. *The Annals of Statistics.* **4**, 912-923.

RÜSCHENDORF, L. (1982). Random variables with maximum sums, *Advances in Applied Probabability*, **14**, 623-Ű632.

RVAČEVA, E.L. (1962). On the domains of attraction of multidimensional distributions. in *Selected Translations Math. Stat. Prob.* vol **2**, 183-207.

RYAN, M.J., AND VAITHIANATHAN, R. (2003). Adverse selection and insurance contracting: a rank dependent utility analysis. *Contributions to Theoretical Economics.* **3**.

RYCHLIK, T. (1993). Bounds for expectation of L-estimates for dependent samples. *Statistics.* **24**. 1-7.

RYCHLIK, T. (1995). Bounds for order statistics based on dependent variables with given nonidentical distributions. *Journal of Multivariate Analysis.* **23**. 351-358.

RYCHLIK, T. (1998). Bounds for expectations of L-estimates. *in* Handbooks of Statistics, Vol 16, Balakrishnan and Rao Eds 105-145.

RYCHLIK, T. (2001). Mean-variance bounds for order statistics from dependent DFR, IFR, DFRA and IFRA samples. *Journal of Statistical Planning and Inference.* **92** 21-38.

RYCHLIK, T. (2002). Best upper quantile evaluations for NWU distributions. *Statistics and Probability Letters.* **58**. 175-184.

SAMAROV, A. & TAQQU, M. (1988). On the efficiency of the sample mean in long-memory noise. *Journal of Time Series Analysis*, **9**, 191-200.

SANCETTA, A. AND SATCHELL, S.E. (2001). Bernstein approximations to the copula function and portfolio optimization. *DAE Working Paper.* **0105**.

SANCETTA, A. AND SATCHELL, S.E. (2003). Nonparametric Estimation of Multivariate Distributions with Given Marginals: L^2 theory. *DAE Working Paper.* **0320**.

SCAILLET, O. (2004). Density estimation using inverse and reciprocal Gaussian kernels. *Journal of Nonparametric Statistics.* 18, 434-458.

SCARSINI, M. (1984). On measures of concordance. *Stochastica*, **8**, 201-208.

SCARSINI, M. (1988) Multivariate stochastic dominance with fixed dependence structure, *Operations Research Letters*, **7**, 237-240.

SCARSINI, M. AND SPIZZICHINO, F. (1999). Simpson-type paradoxes, dependence and ageing. *Journal of Applied Probability*, **36**, 119-131.

SCARSINI, M. (1999). Incorporating support constraints into nonparametric estimators of densities. *Communications Statistics Ű Theory and Methods*, **14**, 1123Ű1136.

SCHERVISH, M.J. (1995). Theory of statistics. Springer Verlag.

Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, **90**, 139-156.

Schlesinger, H. (1999). Decomposing catastrophic risk. *Insurance: Mathematics and Economics*, **24**, 95-101.

Schmeidler, D. (1986). Integral representation without additivity. *Proceedings of the American Mathematical Society*. **97**. 255-261.

Schmeidler, D. (1989). Subjective Probability and Expected Utility without Additivity. *Econometrica*. **57**. 571-587..

Schönbucher, P. and Schubert, D. (2001). Copula-dependent default risk in intensity models. *Preprint*.

Schuster, E. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics: Theory and Methods*, **14**, 1123-1136.

Schweizer, B. and Wolff, E.F. (1976). Sur une mesure de dépendance pour les variables aléatoires. *Comptes Rendus de l'Académie des Sciences de Paris*. **283**, 659-661.

Schweizer, B. and Wolff, E.F. (1981). On nonparametric measures of dependence for random variables. *Annals of Statistics*. **9**, 879-885.

Schweizer, B. and Sklar, A. (1983). Probabilistic metric spaces. North-Holland, New York.

Scott, D.W. (1992). Multivariate Density Estimation - theory, practice and visualization. Wiley, New York.

Sen, P.K. (1970). A note on order statistics for heterogeneous. *Annals of Applied Probability*. **41**. 2137-2139.

Seneta, E. (2002). Karamata's Characterization Theorem, Feller, and Regular Variation in Probability Theory. *Publications de l'Institut de Mathématiques de Belgrade*. **71**, 79-89.

Shaked, M. (1977). A Concept of Positive Dependence for Exchangeable Random Variables. *The Annals of Statistics*, **5** 505-515.

Shaked, M. and Shantikumar, J.G. (1987). Multivariate hazard rates and stochastic ordering. *Advances in Applied Probability*. **19**. 123-137.

Shaked, M. and Shantikumar, J.G. (1990). Multivariate stochastic orderings and positive dependence in reliability. *Mathematics of Operational Research*. **15**. 545-552.

Shaked, M. and Shantikumar, J.G. (1991). Dynamic multivariate aging notions in reliability theory. *Stochastic Processes and their Applications*. **38**. 85-97.

Shanthikumar, G. and Yao, D.D. (1991) Bivariate characterization of some stochastic order relations, *Advances in Applied Probability*, **23**, 642-659.

Shaked, M. and Shanthikumar, G. (1994). Stochastic Orders and Their Applications. San Diego, Academic Press.

Shaked, M. and Shanthikumar, G. (1995). Hazard rate ordering of k-out-of-n systems, *Statistics & Probability Letters*, **23**, 1-8.

Shaked, M. and Spizzichino, F. (1998). Positive dependence properties of conditionally independent random lifetimes. *Mathematics of Operations Research*, **23**, 944-956.

Shea G.S. (1991). Uncertainty and implied variance bounds in long memory models of interest rate term structure. *Empirical economics*, **16**, 257-312.

Shih, J.H . and Louis, T.A. (1995). Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics*, **55**, 1384-1399.

Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, **11**, 195-210.

Silverman, B.. (1986). Density estimation for statistics and data analysis. Chapman & Hall.

Sklar, A. (1959). Fonctions de ré partition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229-231.

Sklar, A. (1959). Random variables, joint distributions and copulas. *Kybernetica*, **9**, 449-460.

Silvapulle, P., Kim, G. and Silvapulle, M.J. (2004). Robustness of a semiparametric estimator of a copula. *Working Paper.*

Smirnov, N.V. (1952). Limit distributions for the terms of a variational series. *American Mathematical Society Translations.* **67**, 1-64.

Smith, R.L. (1986). Extreme value theory based on the r largest annual events. *Journal of Hydrology.* **86**, 27-43.

Smith, R.L. (1993). Long-range dependence and global warming. *in* Statistics for the Environment, Barnett & Turkman eds., Wiley, 141-161.

Smith, R.L., and Huang, L.S. (1993). Modeling High Threshold Exceedances of Urban Ozone. *Technical Report, National Institute of Statistical Sciences.* **6**.

Smith, R.L. (1994). Multivariate Threshold Methods. *Technical Report, National Institute of Statistical Sciences.* **7**.

Smith, R.L., Tawn, J.A. and Coles, S.G. (1997). Markov chain models for threshold exceedances. *Biometrika.* **84**, 249-268.

Song, J.X., Wassell, J.T. and Kapadia, A. (2004). Relative mortality for correlated lifetime data. *Computational Statistics & Data Analysis.* **45**, 849-864.

Spizzichino, F. (2001). Sujective probability models for lifetimes. Chapman & Hall.

Stadtmüller, U. and Trautner, R. (1981). Tauberian theorems for Laplace transforms in dimension $d{>}1$. *Journal für die reine und angewandte Mathematik*, **323**, 127-138.

Stărică, C. (1999). Multivariate extremes for models with constant correlations. *Journal of Empirical Finance*, **6**, 515-553.

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Probability* **12**, 361-379.

Strassen, V. (1965). The existence of probability measures with given marginals *Annals of Mathematical Statistics* **36** 423-439.

Stute, W. (1982). The Oscillation Behavior of Empirical Processes. *Annals of Probability* **10**, 86-107..

Stute, W. (1984). The Oscillation Behavior of Empirical Processes: The Multivariate Case. *Annals of Statistics* **8**, 1348-1360.

Sungur, E.A. (1999). Truncation invariant dependence structures. - 1999 *Communications in statistics. Theory and methods.* **28**, 2553-2568.

Taqqu, M. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **31**, 287–302.

TAQQU, M. (1986). A bibliographical guide to self-similar processes and long-range dependence. *in* Dependence in Probability and Statistics, E. Eberlein and M.S. Taqqu, eds., 137–162. Birkhäuser.

TAQQU, M., AND TEVEROVSKY, L. (1997) Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *Journal of Time Series Analysis*, **18**, 279-304.

TAWN, J.A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika.* **75**. 397-415.

TAWN, J.A. (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika.* **77**. 245-253.

TCHEN, A.H. (1980). Inequalities for distributions with given marginals. *Annals of the Probabilities.* **8**. 814-827.

TENBUSCH, A. (1994). Two-dimensional Bernstein polynomial density estimators. *Metrika.* **41**, 233-253.

TERRELL, G.R. AND SCOTT, D.W.(1980). On improving convergence rates for nonnegative kernel density estimators. *Annals of Statistics*, **8**, 1160-1163.

TIAGO DE OLIVEIRA, J.(1958). Extremal distributions. *Revista da Faculdade de Ciências de Lisboa (2) A*, **7**, 219-228.

TIAGO DE OLIVEIRA, J. (1963). Structure theory of bivariate extremes, extensions. *Estudos de Matemática, Estatística e Econometria*,, **7**, 165-195.

TIAGO DE OLIVEIRA, J. (1965). Statistical Decision For Bivariate Extremes. *Portugaliae Mathematica*, **24**, 145-154.

TIAGO DE OLIVEIRA, J. (1970). Biextremal Distributions; Statistical Decision. *Trab. Estadística Inv. Oper*, **241**, 107-117.

TIAGO DE OLIVEIRA, J. (1979). Bivariate Extremes: Foundations and Statistics. in *Multivariate Analysis V*, Krishnaiah, ed., 349-366.

TITTERINGTON, D., SMITH, A. & MAKOV, U. (1985). Statistical Analysis of Finite Mixture Distributions. Wiley.

TODOROVIC, P. & ZELENHASIC, E. (1970). A stochastic model for flood analysis. *Water Resource*,**6**, 1641-1648.

TODOROVIC, P. & ROUSSELLE J. (1970). Some problems of flood analysis. *Water Resource*, **7**, 1144-115.

TSUKAHARA, H. (1992). A rank estimator in the two-sample transformation model with randomly censored data. *Annals of the Institute of Statistical Mathematics*, **44**, 313-333.

TUCKER, A. (1992). A reexamination of finite- and infinite-variance distributions as models of daily stock returns. *Journal of Business and Economic Statistics*, **10**, 73-81.

VAN DER LAAN, M.J. (1996). Efficient estimation of the bivariate censoring model and repairing NPMLE. *The Annals of Statistics*, **24**, 596-627.

VAN DER VAART, A.W. AND WELLNER, J.A. (1996). Weak convergence and empirical processes, with application to statistics. Springer Verlag.

VAN DER VAART, A.W.(1998). Asymptotic statistics. Cambridge University Press.

VAUPEL, J.W., MANTON, K.G. AND STALLARD, E. (1979). The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography.* **16**, 439-454.

VITALE, R.A. (1975). A Bernstein polynomial approach to density estimation. in *Statistical inference and related topics*, Puri, M.L. ed. Academic Press. 87-100.

VYNCKE, D. (2003). Comonotonicity: the perfect dependence. *PhD Thesis, Katholieke Universiteit Leuven.*

WAND, M.P., MARRON, J.S. AND RUPPERT, D. (1991). Transformations in Density Estimation: Rejoinder (in Theory and Methods). *Journal of the American Statistical Association.* **86**, 360-361.

WAND, M.P. AND JONES, M.C. (1995). Kernel smoothing. Chapman & Hall.

WANG, S.S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, **17**, 43-54.

WANG, S.S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*,**26**, 71-92.

WANG, S.S., YOUNG, V.R. AND PANJER, H.H. (1997). Axiomatic characterization of insurance prices. *Actuarial Research Clearing House.* **1**, 15-31.

WANG, S. AND WELLS, M.T. (2000). Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association*, **95**, 62-72.

WANG, S., NELSEN, R. AND VALDEZ, E.A. (2005). Distortion of multivariate distributions: adjustment for uncertainty in aggregating risks. *Mimeo.*

WANG, S.S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, **67**, 15-36.

WEI, G. AND HU, T. (1950). Supermodular dependence ordering on a class of multivariate copulas. *Statistics & Probability Letters*, **57**, 375-385 .

WEISSMAN, I. (1975). On Location and Scale Functions of a Class of Limiting Processes with Application to Extreme Value Theory. *The Annals of Probability*, **3**, 178-181.

WILDE, T. (1998). CreditRisk+, A Credit Risk Management Framework. *Credit Suisse First Boston.*

WILKS, S.S. (1947). Order statistics. *in* Collected Papers, Anderson Eds.John Wiley & Sons.

WILLIAMSON, R. E. (1956). Multiply monotone functions and their Laplace transform. *Duke Mathematical Journal*, **23**, 189-207.

WILLIAMSON, R.C. (1989). Probabilistic arithmetic. PhD thesis, Department of Electrical Engineering, University of Queensland.

WÜTHRICH, M. (2004). Bivariate extension of the Pickands-Balkema-de Haan theorem. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **40**, 33-41.

YAARI, M.E. (2004). The dual theory of choice under risk. *Econometrica*, **55**, 95-115.

YAJIMA, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Annals of Statistics*, **16**, 791-807.

YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Annals of Statistics*, **19**, 158-177.

ZHANG, S. AND KARUNAMUNI, R.J. (1998). On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference*, **70**, 301-316.

Zhang, S., Karunamuni, R.J. and Jones, M.C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, **94**, 1231-1241.

Zheng, M. and Klein, J.P. (1995). Estimates of Marginal Survival for Dependent Competing Risks Based on an Assumed Copula. *Biometrika*, **82**, 127-138.

ISBN [90-8649-039-5]

D/ [D/2006/10.705/36]