HAL
open science

Construction et estimation de copules en grande dimension

Gildas Mazo

To cite this version:

Gildas Mazo. Construction et estimation de copules en grande dimension. Autres [stat.ML]. Université de Grenoble, 2014. Français. NNT : 2014GRENM058 . tel-01130963v2

HAL Id: tel-01130963

https://theses.hal.science/tel-01130963v2

Submitted on 26 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques Appliquées**

Arrêté ministériel : 7 août 2006

Présentée par

Gildas Mazo

Thèse dirigée par **Stéphane Girard**
et codirigée par **Florence Forbes**

préparée au sein **du Laboratoire Jean Kuntzmann**
et de **l'École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Construction et estimation de copules en grande dimension

Thèse soutenue publiquement le **17 novembre 2014**,
devant le jury composé de :

M. Fabrizio DURANTE
Assistant Professor, Free University of Bozen-Bolzano (Italie), Rapporteur
M. Johan SEGERS
Professeur, Université Catholique de Louvain (Belgique), Rapporteur
Mme Anne-Catherine FAVRE-PUGIN
Professeur, Université Joseph Fourier, Examinateur
M. Ivan KOJADINOVIC
Professeur, Université de Pau et des Pays de l'Adour, Examinateur
M. Stéphane GIRARD
Chargé de Recherche, Inria Rhône-Alpes, Directeur de thèse
Mme Florence FORBES
Directeur de Recherche, Inria Rhône-Alpes, Co-Directeur de thèse

Remerciements

Tout d'abord, je remercie mes directeurs de thèse, Florence Forbes et Stéphane Girard, pour m'avoir proposé ce sujet de thèse très ouvert et actuel. En particulier, merci à Stéphane pour son suivi ; j'ai également beaucoup apprécié sa clairvoyance et son humour. Je ne pense pas prendre beaucoup de risques en affirmant que nous nous sommes très bien entendu tout au long de ces trois années.

Je remercie Fabrizio Durante et Johan Segers, pour avoir accepté sans hésitation et respectivement s'être proposé de rapporter cette thèse. Je suis très honoré de l'intérêt qu'ils ont porté à mon travail. Merci également à Ivan Kojadinovic et Anne-Catherine Favre pour m'avoir fait l'honneur de faire partie de mon jury. Merci en particulier à Ivan pour les suggestions et remarques détaillées sur mon manuscrit.

Je remercie Benjamin Renard pour avoir répondu à ma demande en me proposant son expertise et en me fournissant les données hydrologiques analysées dans cette thèse. Je remercie également Trung, « mon » étudiant, avec qui j'ai pu collaborer pour implémenter un algorithme d'inférence.

Grâce à ses membres, il y a toujours eu une très bonne ambiance et très bonne humeur dans mon équipe à Inria, l'équipe MISTIS. Je les remercie chaleureusement pour cela. Peut-être que cette atmosphère a été possible grâce à la simplicité de chacun. Un grand merci à eux, mes amis. Évidemment, j'inclus dans le lot mes compères de l'équipe IBIS !

Merci à mes parents, pour tout – en particulier, je n'oublie pas que c'est grâce à ma mère que j'ai commencé une licence en statistique !

Enfin, merci à Quynh, pour sa présence bienveillante.

Résumé

Ces dernières décennies, nous avons assisté à l'émergence du concept de copule en modélisation statistique. Cet essor est justifié par le fait que les copules permettent de faire une analyse séparée des marges et de la structure de dépendance induite par une distribution statistique. Cette séparation facilite l'incorporation de lois non gaussiennes et la prise en compte des dépendances non linéaires entre les variables aléatoires. La finance et l'hydrologie sont deux exemples de domaines où les copules sont très utilisées. Puisqu'il existe beaucoup de familles de copules bivariées, il sera toujours possible à l'utilisateur d'en choisir une qui lui convienne. Malheureusement, on ne peut pas en dire autant dans le cas multivarié. La gamme de ces modèles n'est pas encore assez riche pour pouvoir en choisir un qui satisfasse toutes les propriétés que l'on souhaiterait *a priori*. Cette thèse s'inscrit dans ce contexte. Nous proposons deux classes de copules multivariées avec des propriétés originales, ce qui permet d'élargir la gamme des modèles existants. La première classe proposée s'écrit comme un produit de copules bivariées, où chaque copule bivariée se combine aux autres via un graphe en arbre. Elle permet de prendre en compte les différents degrés de dépendance entre les différentes paires de variables. La seconde classe est un modèle à facteurs, avec une composante singulière, basée sur une famille nonparamétrique de copules bivariées. Elle permet d'obtenir un bon équilibre entre flexibilité et maniabilité. Puisque les copules de la deuxième classe proposée possèdent une composante singulière, les méthodes classiques d'inférence ne permettent pas d'estimer leurs paramètres. Pour cette raison – et c'est aussi une contribution de cette thèse –, nous abordons également l'estimation de copules dans le cas général, et exhibons les propriétés asymptotiques d'un estimateur des moindres carrés pondérés basé sur les coefficients de dépendance sans faire appel à des hypothèses de régularité sur les copules. Les modèles et méthodes proposés sont appliqués sur des données hydrologiques (pluies et débits de rivières).

Abstract

In the last decades, copulas have been more and more used in statistical modeling. Their popularity owes much to the fact that they allow to separate the analysis of the margins from the analysis of the dependence structure induced by the underlying distribution. This renders easier the modeling of non Gaussian distributions, and, moreover, it allows to take into account non linear dependencies between random variables. Finance and hydrology are two examples of scientific fields where the use of copulas is nowadays standard. Since there exists many families of bivariate copulas, it is always possible for the user to choose one that suits his/her needs. Unfortunately, the multivariate case is not that simple. The range of these models is still not rich enough for the user to choose one that satisfies all the desired properties. This thesis addresses this issue. We propose two classes of multivariate copulas with novel properties, resulting in an enlargement of the range of the existing models. The first model writes as a product of bivariate copulas and is underlain by a tree structure where each edge represents a bivariate copula. Hence, we are able to model different pairs with different dependence properties. The second one is a factor model, with a singular component, built on a nonparametric class of bivariate copulas. It exhibits a good balance between tractability and flexibility. Since the copulas belonging to the second proposed class have a singular component, the standard methods of inference do not permit to estimate their parameters. For this reason – and this is a contribution of our thesis as well –, we also deal with the estimation of copulas in general, and establish the asymptotic properties of a least-squares estimator based on dependence coefficients without imposing regularity conditions on the copulas. The models and methods have been applied to hydrological data (flow rates and rain falls).

Table des matières

Résumé **iv**

Abstract **v**

Introduction **1**

I Copules **4**

1 Les copules ou l'étude de la dépendance **5**
1.1 Définition 6
1.2 Mesurer la dépendance 7
1.2.1 Spectre de dépendance 8
1.2.2 Coefficients de dépendance 8
1.3 Deux classes de copules particulières 11
1.3.1 Copules des valeurs extrêmes 11
1.3.2 Copules avec une composante singulière 12

2 Modèles de copules en grande dimension **15**
2.1 Copules archimédiennes 15
2.2 Copules archimédiennes imbriquées 16
2.3 *Vines* 17
2.4 Copules elliptiques 19

3 Inférence **21**
3.1 Estimation 22
3.1.1 La méthode du maximum de vraisemblance 22
3.1.2 La méthode des moments basée sur les coefficients de dépendance 23
3.1.3 Les méthodes non paramétriques 25
3.2 Tests 26

II Deux nouvelles classes de copules et leur estimation **27**

4 Un modèle de copules basé sur des produits de copules bivariées **28**

5 Estimation de copules multivariées par la méthode des moindres carrés pondérés basée sur les coefficients de dépendance **47**

6 Une classe de copules maniable et flexible **73**

Conclusion **102**

Introduction

Le besoin de recourir à des modèles statistiques non gaussiens, ou non normaux [1] a toujours existé en statistique, mais il était considéré pendant longtemps comme moins la règle que l'exception. Récemment, ce besoin s'est fait de plus en plus pressant. Dans plusieurs domaines d'application, comme par exemple l'hydrologie ou la finance [2], on reconnait l'utilité de ces modèles qui sont capables de prendre en compte les dépendances de types non affines, et surtout, les dépendances entre les valeurs extrêmes des facteurs d'intérêts. En effet, il est bien connu que les lois gaussiennes, en particulier, sont incapables de modéliser de telles dépendances [85]. Considérons tout de suite trois exemples.

Retour sur investissement. Le retour sur investissement sur d années d'un placement financier est donné par $1000(1+X_1)\times\cdots\times(1+X_d)$, où X_i est le taux d'intérêt sur l'année i. Supposons, par exemple, que chaque X_i soit distribué uniformément entre 0.05 et 0.15. Si les taux étaient indépendants d'une année à l'autre, nous pourrions calculer la distribution du retour sur les d années ; mais, évidemment, ils ne le sont pas. Il nous faut alors trouver un modèle pour la loi jointe des taux X_i. Cet exemple est tiré de [57].

Gestion de porte-feuille. Lorsque l'on possède un porte-feuille d'actifs financiers, on souhaite savoir comment est distribuée la perte potentielle qui lui est associée. Ainsi, soit P_i^t le prix du i-ème actif de notre porte-feuille à un temps de référence t et soit $X_i = -(\log P_i^{t+1} - \log P_i^t)$ la perte enregistrée pour cet actif à un pas de temps dans le futur. La perte totale associée à notre porte-feuille qui contient d actifs s'écrit alors $X_1 + X_2 + \cdots + X_d$. Pour calculer sa loi, nous avons besoin de la loi jointe de $(X_1, \ldots, X_d)$.

Estimation de niveaux critiques en hydrologie. On dispose de d pluviomètres disposés dans une région d'intérêt. On note X_i de fonction de répartition F_i la quantité de pluie maximale enregistrée sur une année dans le i-ème

1. Bien que ces deux termes soient acceptés par tous comme étant équivalents, nous préférons utiliser le terme « gaussien » dans un contexte de modélisation, car, si le terme « normal » était employé, cela sous entendrait que les autres modèles ne sont pas normaux au sens littéral du terme. Or, et c'est justement le message de cette introduction de thèse, ce n'est plus vrai aujourd'hui. On pourra se consoler en gardant le terme « normal » dans un contexte de statistique mathématique, puisque ce terme sera toujours justifié par le théorème central limite.

2. Surtout la finance : il y a 5 fois plus de combinaisons « *copula AND finance* » que « *copula AND hydrology* » renvoyées par Google Scholar.

pluviomètre. On souhaite évaluer la probabilité d'apparition de l'évènement selon lequel toutes les variables dépassent leur quantile d'ordre 99%, c'est à dire $\{X_1 > F_1^{-1}(0.99), \ldots, X_d > F_d^{-1}(0.99)\}$. Pour cela, nous avons besoin de la loi de $\min(X_1, \ldots, X_d)$, ce qui serait possible si nous avions celle de $(X_1, \ldots, X_d)$. Cet exemple sera repris dans le chapitre 6.

Dans les trois exemples précédents, l'étude de la queue de distribution de la loi considérée est d'une importance capitale. En effet, ce sont les évènements de cette queue de distribution qui impactent le plus fortement des pertes financières ou des inondations subies. Or, les évènements de la queue de distribution sont eux mêmes principalement issus de la co-occurence des valeurs extrêmes de nos variables. C'est pourquoi il est primordial de modéliser correctement les dépendances entre les valeurs extrêmes.

Ainsi, non seulement on souhaite construire des modèles multivariés non gaussiens, mais en plus, il faut souvent le faire sous la contrainte que les lois marginales des facteurs d'intérêts sont données (c'est le cas des exemples ci-dessus). Pour répondre à cette attente, les copules se sont imposées comme un outil incontournable. En résumé – nous détaillerons plus au chapitre 1 –, [3] une copule est un modèle qui permet de refléter fidèlement la dépendance entre les facteurs considérés. Il est donc important pour l'utilisateur de disposer d'une gamme aussi complète que possible de modèles de copules afin de s'assurer que l'un d'entre eux satisfera à ses besoins. Dans le cas bivarié, c'est à dire lorsqu'il n'y a que deux variables à étudier, il y a de nombreuses familles de copules parmi lesquelles il trouvera certainement celle qui lui convient. Malheureusement, on ne peut pas en dire autant dans le cas multivarié. La gamme de ces modèles n'est pas encore assez riche pour pouvoir en choisir un qui satisfasse toutes les propriétés que l'on souhaiterait *a priori*. Bien souvent, l'utilisateur devra accepter de perdre un peu d'une propriété pour en gagner une autre. Ce compromis est notamment vrai lorsque l'on considère la flexibilité et la maniabilité d'un modèle. De l'aveu même de deux des chercheurs les plus reconnus dans ce domaine [47, 69], la construction de copules multivariées est un problème difficile car, comme le souligne Joe [47], « *one cannot just write down a family of functions and expect it to satisfy the necessary condition for multivariate cumulative distribution functions* ». La phrase de Nelsen qui introduit la partie 3.5 de son livre [69] sur la construction de copules multivariées, est aussi restée célèbre : « *First, a word of caution : Constructing [multivariate] copulas is difficult* ».

Cette thèse apporte sa contribution à l'étude des copules à travers deux axes. Premièrement, nous enrichissons la gamme des modèles multivariés de deux classes aux propriétés originales. Les copules de la première classe s'écrivent comme un produit de copules bivariées, où chaque copule bivariée se combine aux autres via un graphe en arbre. Ce modèle permet de prendre en compte les différents degrés de dépendance entre les différentes paires. La seconde classe de copules est un modèle à facteurs, avec une composante singulière, basé sur une classe nonparamétrique de copules bivariées. Elle permet d'obtenir un bon équilibre entre flexibilité et maniabilité. Deuxièmement, nous envisageons l'estimation de copules dans le cas général, c'est à dire pour lesquelles il n'existe pas nécessairement de dérivées partielles (comme par exemple les copules appartenant à la seconde classe proposée), et établissons les propriétés asymptotiques

3. Le lecteur pourra admirer ici une magnifique composition syntaxique : le fameux tiret-virgule.

d'un estimateur des moindres carrés pondérés basé sur les coefficients de dépendance. Chacun de nos modèles et méthodes proposés sont appliqués sur des données hydrologiques (pluies et débits de rivières).

Le plan de la thèse est le suivant. La partie I contient une introduction aux copules dans le chapitre 1, présente une revue de la littérature sur les principaux modèles de copules multivariées au chapitre 2, et aborde les problèmes d'inférence dans le chapitre 3. La partie II présente nos contributions. Chacun des chapitres la composant est constitué d'une brève introduction suivie d'un article soumis pour publication, en anglais. La classe basée sur un produit de copules est introduite dans le chapitre 4 et le modèle à facteur est introduit dans le chapitre 6. Le chapitre 5 présente notre méthode d'estimation. Enfin, une conclusion viendra clore la thèse.

Première partie

Copules

Chapitre 1

Les copules ou l'étude de la dépendance

Les copules permettent d'étudier la dépendance entre plusieurs variables aléatoires, avec l'idée que cette dépendance ne doit pas contenir d'information provenant des lois marginales des variables elles-mêmes. Pour ce faire, on les « uniformise », c'est à dire qu'on se prémuni de « l'effet d'optique » dû au fait que ces variables peuvent avoir des lois marginales très différentes. En particulier, les copules permettent d'imposer une structure de dépendance à des lois marginales (ou des variables aléatoires) données séparément. Par exemple, lorsque nous avions donné comme exemple l'estimation des niveaux critiques associés à un évènement extrême en hydrologie dans l'introduction de cette thèse, nous avions vu que X_i était la quantité de pluie maximale observée sur une année à une certaine station i. Or, nous savons, d'après la théorie des valeurs extrêmes (voir par exemple [11, 13, 76]), que la loi F_i du maximum d'un échantillon, devrait être raisonnablement bien approchée par la loi des valeurs extrêmes généralisée (*Generalized Extreme Value* ou *GEV* en anglais), donnée par

$$GEV(x; \mu_i, \sigma_i, \xi_i) = \exp\left[-\left(1 + \xi_i \frac{x - \mu_i}{\sigma_i}\right)^{-1/\xi_i}\right],$$

où $\sigma_i > 0$, $-\infty < \mu_i, \xi_i < \infty$ et $1 + \xi_i(x - \mu_i)/\sigma_i > 0$. Ainsi, pour chaque station i, la distribution F_i est connue (aux paramètres près). Mais comment modéliser les dépendances entre les différentes stations ? C'est ici un problème typique que l'on peut vouloir résoudre avec les copules [1]. Les autres exemples présentés lors de l'introduction de cette thèse sont aussi des cas d'école pour les copules. Voici un dernier exemple, traitant du risque de crédit et tiré de [61] [2]. Lorsqu'un établissement de crédit prête à plusieurs entreprises, ces dernières remboursent à échéance, sauf si elles font faillite ; elles sont alors en défaut de paiement. Pour

1. Pourtant, les statisticiens spécialistes de la théorie des valeurs extrêmes utilisent très peu les copules. Par exemple, l'ouvrage de S. Coles [11] ne les mentionne pas du tout.

2. Depuis un article du *Wired Magazine* daté du 23 février 2009, cette publication est devenue tristement célèbre : la formule liant la probabilité que plusieurs emprunteurs fassent défaut ensemble avec la copule (gaussienne) a été appelée « *the formula that killed Wall Street* ». Evidemment, c'est moins la formule elle même que l'utilisation qui en a été faite qui était erronée. Toutefois, cela illustre bien les enjeux, pouvant être considérables, de la modélisation.

évaluer le risque de crédit supporté par le prêteur, on commence par évaluer la probabilité que chaque entreprise, prise individuellement, fasse défaut. En fait, on peut modéliser ces probabilités par les outils classiques de l'analyse de survie en statistique. Une fois les modèles de survie choisis, on peut estimer les paramètres de ces modèles de plusieurs façons, voir [61]. Mais, pour prendre en compte la dépendance entre les emprunteurs, il faut pouvoir spécifier une loi jointe étant données les lois marginales.

Dans les mises en situation précédentes, il faut spécifier une loi jointe *étant données les lois marginales*. Les copules aident à faire cela. Elles facilitent la modélisation en la découpant en deux étapes : la modélisation des marges puis celle de la structure de dépendance. Comme nous le verrons dans le chapitre 3, ce découpage se retrouve aussi dans l'inférence, qui s'en trouve aussi facilitée (d'un point de vue pratique ; évidemment, d'une point de vue théorique, on introduit plutôt de nouveaux challenges).

Les copules connaissent un essor remarquable depuis une dizaine d'années, comme en témoigne le tableau 1.1. En décembre 2010, le site internet `ScienceWatch.com` a même élu la discipline « *Copula modeling* » comme « *top topic* » parmi tous les domaines de la catégorie « *Mathematics* » [88].

années	nombre de publications
1973-1983	1
1983-1993	9
1993-2003	68
2003-2013	824

TABLE 1.1 – Nombre de publications dans la base de données *MathSciNet* avec dans le titre « *copula* » en fonction de l'année.

Le reste du chapitre est organisé comme suit. Dans la partie 1.1, nous donnons la définition d'une copule. Dans la partie 1.2, nous montrons comment quantifier la dépendance à l'aide des copules. Enfin, dans la partie 1.3, nous présentons deux classes de copules particulières : les copules des valeurs extrêmes et les copules avec une composante singulière.

1.1 Définition

Soient $X_1, \ldots, X_d$ des variables aléatoires de fonctions de répartition $F_1, \ldots, F_d$, et soit F la fonction de répartition du vecteur $(X_1, \ldots, X_d)$. La copule, souvent notée C, associée à la loi cible F, est la fonction de répartition du vecteur $(F_1(X_1), \ldots, F_d(X_d))$. Elle est donc aussi la fonction qui à $(u_1, \ldots, u_d)$ associe le nombre $F(F_1^{-1}(u_1), \ldots, F_1^{-1}(u_1))$.

Definition 1 (Copule)**.** *Une copule d-variée est une fonction définie sur* $[0,1]^d$ *telle que*

1. $C(u_1, \ldots, u_d) = 0$ *si* $u_i = 0$ *pour au moins un indice* i *dans* $\{1, \ldots, d\}$,
2. *pour chaque pavé* $B = [a_1, b_1] \times \cdots \times [a_d, b_d]$ *inclu dans le cube unité* $[0,1]^d$*, le volume de ce pavé* $\sum sgn(u_1, \ldots, u_d) C(u_1, \ldots, u_d)$ *est positif, où*

la somme est prise sur tous les sommets $(u_1, \ldots, u_d)$ *de* B *et*

$$sgn(u_1, \ldots, u_d) = \begin{cases} 1 & \text{si } u_k = a_k \text{ pour un nombre pair de } k \in \{1, \ldots, d\}, \\ -1 & \text{si } u_k = a_k \text{ pour un nombre impair de } k \in \{1, \ldots, d\}, \end{cases}$$

3. *les marges univariées de* C *sont uniformes, c'est à dire* $C(1, \ldots, u_i, \ldots, 1) = u_i$, $i = 1, \ldots, d$ *(dans le membre de gauche,* u_i *est à la* i*-ème position).*

Il existe une unique copule C associée à F, à condition que les marges F_i soient continues. La réciproque est également vraie. Ce résultat, précisé dans le théorème suivant, appelé *théorème de Sklar* [86] est le résultat fondamental justifiant la modélisation basée sur les copules.

Theorem 1. *Soit* F *une fonction de répartition* d*-variée de marges continues* $F_1, \ldots, F_d$*. Alors il existe une unique copule* C *telle que*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)), \quad (x_1, \ldots, x_d) \in (-\infty, \infty)^d. \tag{1.1}$$

Réciproquement, si C *est une copule et si* $F_1, \ldots, F_d$ *sont des fonctions de répartitions, alors la fonction* F *définie par* (1.1) *est une fonction de répartition de marges* $F_1, \ldots, F_d$*.*

L'équation (1.1) révèle que la donnée de la copule C *et* des marges F_i permet de reconstruire la loi cible F. Ainsi, on interprète la copule C associée à la loi F comme la structure de dépendance « pure » – c'est à dire une fois enlevé l'effet de distorsion des lois marginales – qu'il y a entre les variables d'intérêt. Mathématiquement, cela se traduit par le fait que la copule est invariante par transformation croissante des marges. Si $g_1, \ldots, g_d$ sont des fonctions strictement croissantes, la copule associée à $(X_1, \ldots, X_d)$ est égale à la copule associée à $(g_1(X_1), \ldots, g_d(X_d))$. Si les copules sont absoluement continues (par rapport à la mesure de Lebesgue), le théorème 1 se traduit par la décomposition de la densité f de F en le produit de ses marginales f_i et de la densité c de la copule C, c'est dire que l'on a

$$f(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) f_1(x_1) \ldots f_d(x_d). \tag{1.2}$$

Le nom « copule » vient de ce que la copule « couple » les marges F_i entre elles. Les copules, en plus de permettre une analyse séparée des marges et de la structure de dépendance sous jacente à une loi cible, ont aussi l'avantage de fournir un langage commun aux statisticiens. Deux livres sont devenus incontournables dans ce domaine : le livre de Joe [47] et celui de Nelsen [69]. Cette année, un nouvel ouvrage écrit par Joe vient d'être publié [48]. Dans [25], on pourra trouver un article très pédagogique, accessible et complet sur la modélisation à l'aide des copules. Enfin, les copules sont très faciles à utiliser dans la pratique grâce au package `copula` [41] du langage de programmation statistique `R` [75].

1.2 Mesurer la dépendance

Dans cette section, nous présentons les principaux outils basés sur les copules permettant de quantifier la dépendance entre deux variables aléatoires. Lorsqu'il y a plus de deux variables aléatoires, des extensions sont possibles mais non évidentes et peu utilisées. Nous avons donc choisi de ne pas les introduire, mais nous faisons référence à la littérature.

1.2.1 Spectre de dépendance

Toute copule bivariée C est bornée par les copules associées à la dépendance « parfaite » (ou « complète ») comme suit :

$$W(u_1, u_2) \leq C(u_1, u_2) \leq M(u_1, u_2), \quad (u_1, u_2) \in [0, 1]^2, \tag{1.3}$$

où $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ est la copule de *dépendance négative parfaite* et $M(u_1, u_2) = \min(u_1, u_2)$ est la copule de *dépendance positive parfaite* (la propriété de dépendance parfaite est également appelée *co-monotonicité*). Les bornes dans (1.3) sont appelées *les bornes de Fréchet-Hoeffding* [69] Section 2.2. Pour une généralisation de ces bornes en dimension quelconque, voir par exemple [69]. La dépendance négative complète entre deux variables aléatoires X_1 et X_2 est définie par la relation $X_2 = f(X_1)$ (presque sûrement, ou p.s.) où f est une fonction strictement décroissante. On peut alors facilement montrer que la copule associée à la loi de (X_1, X_2) est donnée par $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$. Le vecteur aléatoire (U_1, U_2) qui a pour loi cette copule vérifie $U_2 = 1 - U_1$ (p.s.). La dépendance positive complète entre deux variables aléatoires X_1 et X_2 est définie par la relation $X_2 = f(X_1)$ où f est une fonction strictement croissante. On peut alors facilement montrer que la copule associée à la loi de (X_1, X_2) est donnée par $M(u_1, u_2) = \min(u_1, u_2)$. Le vecteur aléatoire (U_1, U_2) qui a pour loi cette copule vérifie $U_2 = U_1$ (p.s.). Si les deux variables aléatoires X_1 et X_2 sont indépendantes, leur copule est donnée par $C(u_1, u_2) = u_1 u_2$. On note en général cette copule par le symbole Π, c'est à dire que $\Pi(u_1, u_2) = u_1 u_2$.

Une famille de copules (C_θ), où θ est le paramètre indexant la famille, est dite *complète* (*comprehensive* en anglais) si elle peut atteindre les bornes de Fréchet-Hoeffding en passant par la copule d'indépendance. Par exemple, c'est le cas de la famille de Clayton [10], donnée par

$$C_\theta(u, v) = \left[\max(u^{-\theta} + v^{-\theta} - 1, 0)\right]^{-1/\theta}, \quad \theta \in [-1, \infty). \tag{1.4}$$

Lorsque $\theta = -1$, respectivement 0, on a $C_{-1} = W$, respectivement $C_0 = \Pi$. Lorsque $\theta \to \infty$, on a $C_\infty = M$. Le paramètre est donc une mesure de la dépendance modélisée par la copule. Néanmoins, d'une part, quantifier la dépendance avec les paramètres de différentes familles ne permet pas de les comparer entre elles. D'autre part, quid des copules qui n'appartiennent à aucune famille paramétrique ? Il nous faut donc des outils pour pouvoir comparer les dépendances entre copules. C'est l'objet de la partie 1.2.2, qui traite des coefficients de dépendance.

1.2.2 Coefficients de dépendance

Les coefficients de dépendance permettent de quantifier la dépendance entre deux variables aléatoires, et comparer la quantité de dépendance entre plusieurs couples de variables. Ci-dessous, nous présentons les plus utilisés [3], c'est à dire

3. Le coefficient de corrélation que l'on trouve dans tous les manuels de statistique – celui de Pearson – n'est pas adapté pour mesurer la dépendance de lois non gaussiennes. Le coefficient de Pearson vaut 1 si et seulement si X_2 est une fonction affine de X_1. Si X_2 est une fonction strictement croissante de X_1 autre qu'une fonction affine, la dépendance est complète mais le coefficient de Pearson est plus petit que 1 en valeur absolue. Si la loi de (X_1, X_2) est une loi gaussienne, alors cette fonction strictement croissante *doit* être une fonction affine.

le τ de Kendall et ρ de Spearman, définis respectivement comme

$$\begin{aligned}\tau =& P\left[(X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) > 0\right] - P\left[(X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) < 0\right]\\ \rho =& 3\left\{P\left[(X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(3)}) > 0\right] - P\left[(X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(3)}) < 0\right]\right\}\end{aligned}$$

où $(X_1^{(1)}, X_2^{(1)})$, $(X_1^{(2)}, X_2^{(2)})$ et $(X_1^{(3)}, X_2^{(3)})$ sont trois copies indépendantes et identiquement distribuées de (X_1, X_2). En fait, on peut calculer que

$$\tau = 4\int_{[0,1]^2} CdC - 1, \text{ et } \rho = 12\int_{[0,1]^2} Cd\Pi - 3,$$

ce qui montre que le tau de Kendall et le rho de Spearman ne dépendent que de la copule. Ces deux coefficients de dépendance valent 1 quand la dépendance est positive et parfaite, -1 quand la dépendance est négative et parfaite, et 0 dans le cas de l'indépendance. Ainsi, ces coefficients sont invariants par transformation strictement croissante des variables X_1 et X_2. On pourra consulter [47,69] pour plus de détails. Concernant les extensions multivariées de ces coefficients, on peut les trouver respectivement dans [47, 74] et [79].

Pour quantifier la dépendance entre les très grandes valeurs de X_1 et de X_2, on utilise en général les coefficients de dépendance dits « de queue » inférieurs et supérieurs (*lower tail dependence coefficients* et *upper tail dependence coefficients* en anglais), définis respectivement comme

$$\lambda^{(L)} = \lim_{u\downarrow 0} P\left[F_2(X_2) \leq u | F_1(X_1) \leq u\right], \text{ et } \lambda^{(U)} = \lim_{u\uparrow 1} P\left[F_2(X_2) > u | F_1(X_1) > u\right].$$

Comme le rho de Spearman et le tau de Kendall que nous avons vu précédemment, ces coefficients ne dépendent que de la copule :

$$\lambda^{(L)} = \lim_{u\downarrow 0} \frac{C(u,u)}{u}, \text{ et } \lambda^{(U)} = \lim_{u\uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}. \tag{1.5}$$

La figure 1.1 montre des simulations de 10 000 paires distribuées selon une copule gaussienne, de Gumbel, de Clayton, et de Student (la copule gaussienne et de Student sont des copules elliptiques, et seront vues dans la partie 2.4 ; la copule de Gumbel est une copule archimédienne et sera vue dans la partie 2.1). Les paramètres ont été choisis de sorte que $\rho = 0.5$. Pour la copule de Student, le second paramètre requis (le degré de liberté) a été choisi tel que $\lambda^{(U)} = \lambda^{(L)} \approx 0.4$. Un oeil averti serait capable d'affecter ces familles à chacun des quatre dessins constituant la figure. En effet, pour la copule gaussienne, $\lambda^{(L)} = \lambda^{(U)} = 0$, pouvant s'interpréter comme le fait que les valeurs extrêmes de l'échantillon sont indépendantes. On le voit sur le dessin (a) : l'accumulation des points sur le sommet en haut à droite s'étend sur les côtés adjacents du carré unité, alors que, pour la copule de Student (d), pour laquelle $\lambda^{(L)} = \lambda^{(U)} \approx 0.4$, il n'y a pas de points sur ces bords. Pour la copule de Clayton (c), on retrouve le comportement de la copule de Student sur le coin inférieur gauche, et sur le coin supérieur droit, on retrouve le comportement de la copule gaussienne. Ceci s'explique par le fait que pour le copule de Clayton, $\lambda^{(L)} \approx 0.52$ mais $\lambda^{(U)} = 0$. Pour la copule de Gumbel (b), c'est exactement l'inverse : $\lambda^{(L)} = 0$ mais $\lambda^{(U)} \approx 0.43$.

D'autres propriétés, d'ordre plutôt qualitatif, ont été définies et étudiées pour les copules et plus généralement les distributions statistiques multivariées.

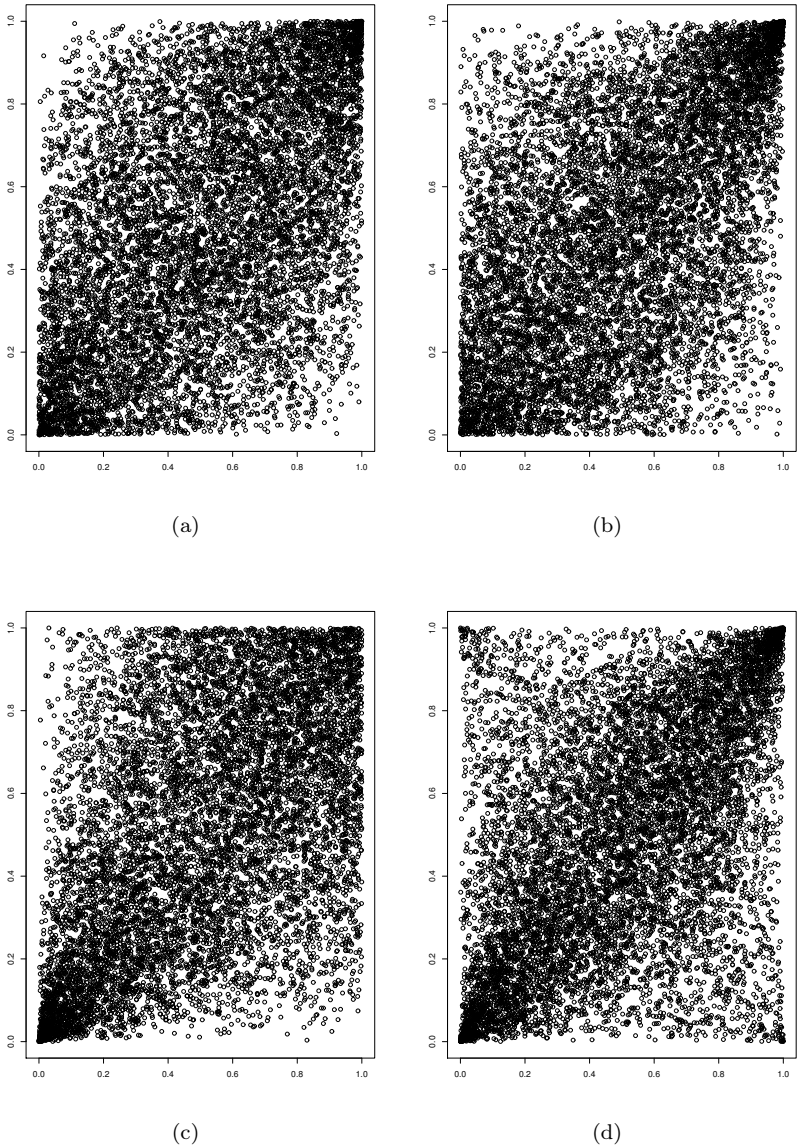

FIGURE 1.1 – Echantillon de 10 000 paires distribuées selon la copule gaussienne (a), la copule de Gumbel (b), la copule de Clayton (c), et la copule de Student (d). Les paramètres ont été choisis tels que $\rho = 0.5$, et $\lambda^{(U)} = \lambda^{(L)} \approx 0.4$ pour la copule de Student.

Ces propriétés incluent par exemple (en anglais) *positive quadrant dependence (PQD), increasing in the concordance ordering* et *stochastic increasing*. Par exemple, la propriété PQD indique que la co-occurence de deux petites valeurs d'une paire de variables distribuées avec une copule PQD arrive plus souvent qu'avec la copule d'indépendance. Nous reportons le lecteur à [69] chapitre 5 ou [47] chapitre 2 pour plus de détails sur ces propriétés.

1.3 Deux classes de copules particulières

Dans cette partie, nous présentons deux classes de copules particulières auxquelles on fera appel à plusieurs reprises au cours de cette thèse. La première classe est la classe des copules des valeurs extrêmes. Ces copules apparaissent lorsqu'on étudie la distribution statistique de maxima d'échantillons, comme nous le verrons dans la partie 1.3.1. Les copules qui constituent la deuxième classe, présentée dans la partie 1.3.2, sont les copules qui ne sont pas absolument continues par rapport à la mesure de Lebesgue (on dit aussi qu'elles possèdent une composante singulière).

1.3.1 Copules des valeurs extrêmes

Les copules extrêmes sont les copules associées aux maxima d'un échantillon, disons de taille n, de vecteurs aléatoires indépendants et identiquement distribués, normalisés convenablement, quand $n \to \infty$. Soit

$$(X_1^{(1)}, \dots, X_d^{(1)}), \dots, (X_1^{(n)}, \dots, X_d^{(n)})$$

un échantillon i.i.d. de vecteurs aléatoires de loi F et de copule C et soit $M_i^{(n)} = \max(X_i^{(1)}, \dots, X_i^{(n)})$ le maximum pris sur la i-ème composante. La copule de $(M_1^{(n)}, \dots, M_d^{(n)})$ est donnée par $(u_1, \dots, u_d) \mapsto C^n(u_1^{1/n}, \dots, u_d^{1/n})$. Si cette copule a une limite quand $n \to \infty$, cette limite sera une copule des valeurs extrêmes $C_\#$. La classe des copules des valeurs extrêmes coïncide avec la classe des copules max-stables, c'est à dire les copules $C_\#$ telles que $C_\#^n(u_1^{1/n}, \dots, u_d^{1/n}) = C_\#(u_1, \dots, u_d)$ pour tout entier $n \geq 1$ et tout $(u_1, \dots, u_d) \in [0,1]^d$. Les copules des valeurs extrêmes correspondent aux copules associées aux lois extrêmes, c'est à dire les lois limites, aux marges non dégénérées, de la suite

$$\left(\frac{M_1^{(n)} - b_1^{(n)}}{a_1^{(n)}}, \dots, \frac{M_d^{(n)} - b_d^{(n)}}{a_d^{(n)}} \right),$$

où $a_i^{(n)}$ et $b_i^{(n)}$ sont des constantes de normalisation bien choisies pour $i = 1, \dots, d$.

Le coefficient de dépendance de queue supérieur d'une copule des valeurs extrêmes bivariée a la forme particulière

$$\lambda^{(U)} = 2 + \log C_\#(e^{-1}, e^{-1}).$$

Ce coefficient est un coefficient de dépendance naturel pour les copules des valeurs extrêmes à cause de la représentation suivante sur la diagonale principale du carré unité :

$$C_\#(u, u) = u^{2-\lambda}, \tag{1.6}$$

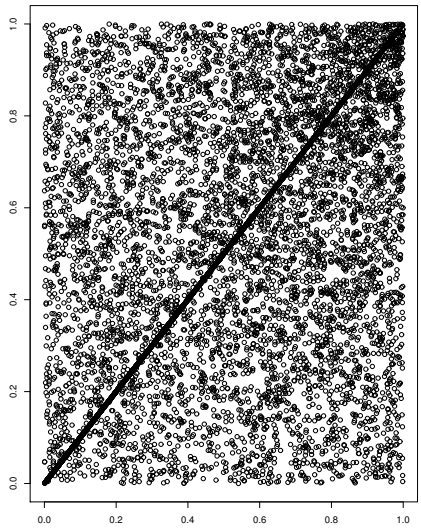

FIGURE 1.2 – Echantillon de 10 000 paires distribuées selon la copule de Cuadras-Augé avec $\theta = 1/2$. La probabilité pour une paire de tomber sur la diagonale du carré unité est de 1/3.

où $\lambda := \lambda^{(U)}$. Si $\lambda = 0$ alors $C_{\#}(u,u) = \Pi(u,u) = u^2$. Si $\lambda = 1$ alors $C_{\#}(u,u) = M(u,u) = \min(u,u) = u$. Les copules Π et M sont les copules d'indépendance et de dépendance positive parfaite, voir la partie 1.2.1. Dans le cas des copules des valeurs extrêmes, cette interpolation entre Π et M permet d'interpréter λ comme un coefficient qui mesure la dépendance en général, et pas seulement dans les queues de distribution. Par exemple, la copule de Gumbel, qui sera vue dans la partie 2.1 (puisque c'est une copule archimédienne), et dont un échantillon a été représenté sur la figure 1.1b, est une copule des valeurs extrêmes avec $\lambda = 2-2^{\theta}$, où $\theta \geq 1$ est son paramètre. Lorsque $\theta = 1$, $\lambda = 0$, indiquant que les variables sont indépendantes, et $\lambda \to 1$ quand $\theta \to \infty$, indiquant la co-monotonicité de celles-ci. Un autre exemple est donné par la copule de Cuadras-Augé, qui, puisqu'elle possède une composante singulière, sera vue dans la partie 1.3.2. Pour cette copule, $\lambda = \theta$, où $\theta \in [0,1]$ est son paramètre. On pourra consulter [11] pour de plus amples détails à propos des statistiques des valeurs extrêmes, et voir, par exemple [33] pour une revue de la littérature sur les copules des valeurs extrêmes.

1.3.2 Copules avec une composante singulière

Les copules avec une composante singulière sont les copules qui ne sont pas absolument continues (par rapport à la mesure de Lebesgue). Elles s'écrivent

$$C(u_1, \ldots, u_d) = A(u_1, \ldots, u_d) + S(u_1, \ldots, u_d)$$

avec

$$A(u_1,\dots,u_d)=\int_{[0,u_1]\times\cdots\times[0,u_d]}\frac{\partial^d C(x_1,\dots,x_d)}{\partial x_1\dots\partial x_d}\mathbf{1}\Big\{\frac{\partial^d C(x_1,\dots,x_d)}{\partial x_1\dots\partial x_d}\text{ existe}\Big\}dx_1\dots dx_d$$

étant la partie absolument continue et $S=C-A$ la composante singulière de la copule.

La loi avec une composante singulière la plus connue est sans doute la loi de *Marshall-Olkin* [64], voir aussi [69] section 3.1.1, dont la copule de survie est donnée par

$$\begin{aligned}\overline{C}_\theta(u_1,\dots,u_d)=&P[F_1(X_1)>u_1,\dots,F_d(X_d)>u_d]\\ =&(1-u_1)^{\theta_1}\dots(1-u_d)^{\theta_d}\min(u_1^{1-\theta_1},\dots,u_d^{1-\theta_d}),\end{aligned}$$

où $\theta=(\theta_1,\dots,\theta_d)\in[0,1]^d$. La copule de Marshall-Olkin est, en principe [4], intéressante pour modéliser des systèmes qui présentent des « chocs ». Plus précisément, soient $Z_1,\dots,Z_d$ et Z_0 des variables aléatoires indépendantes de lois exponentielles qui représentent les instants où les chocs arrivent dans le système, causant des dommages à ses composants. Les variables $Z_1,\dots,Z_d$ représentent des chocs endogènes qui n'affectent que les composants $1,\dots,d$ respectivement, et Z_0 représente un choc exogène qui affecte tout le système $(1,\dots,d)$. Soient $X_1=\min(Z_1,Z_0),\dots,X_d=\min(Z_d,Z_0)$ les instants où les composants $1,\dots,d$ subissent un choc, qui leur est fatal. On peut alors montrer que la copule de survie associée à $(X_1,\dots,X_d)$ est la copule de survie de Marshall-Olkin, où les paramètres sont déterminés par les paramètres des lois exponentielles de $Z_1,\dots,Z_d$ et Z_0. Dans le cas bivarié, la copule de Marshall-Olkin est donnée par

$$C_{(\theta_1,\theta_2)}(u_1,u_2)=\min(u_1^{1-\theta_1}u_2,u_1u_2^{1-\theta_2})=\begin{cases}u_1^{1-\theta_1}u_2 & \text{si } u_1^{\theta_1}\geq u_2^{\theta_2},\\ u_1u_2^{1-\theta_2} & \text{si } u_1^{\theta_1}\leq u_2^{\theta_2},\end{cases}$$

pour $0\leq\theta_1,\theta_2\leq 1$. Pour cette copule, on peut voir facilement que même les dérivées partielles de $C_{(\theta_1,\theta_2)}$ n'existent pas sur la diagonale principale du carré unité. La composante singulière est donnée par

$$S(u_1,u_2)=\int_0^{\min(u_1^{\theta_1},u_2^{\theta_2})}t^{1/\theta_1+1/\theta_2-2}dt.$$

En particulier, on a que $P[U_1^{\theta_1}=U_2^{\theta_2}]=\theta_1\theta_2/(\theta_1+\theta_2-\theta_1\theta_2)$. Dans le cas où $\theta_1=\theta_2\equiv\theta$, la copule de Marshall-Olkin se réduit à la copule de Cuadras-Augé [12], donnée par

$$C_\theta(u_1,u_2)=\min(u_1,u_2)\max(u_1,u_2)^{1-\theta}\tag{1.7}$$

(cette copule peut aussi être vue comme un cas particulier de [3]). Sur la figure 1.2, où un échantillon de taille 10 000 est représenté, on voit bien la composante singulière sur la diagonale principale du carré unité. Les chocs seraient

4. Malgré l'attrait de cette interprétation en termes de chocs, l'intérêt de la copule de Marshall-Olkin reste surtout théorique. En effet, il est difficile de trouver des publications présentant des applications avec des jeux de données réelles lorsque l'on effectue des requêtes par mots clés dans les moteurs de recherche. La seule application que nous ayons trouvée [56] est plutôt décevante : elle consistait en l'analyse de 37 matchs de foot ; d'ailleurs, les auteurs eux mêmes admettent que l'application était présentée uniquement pour « illustrer » leur méthode d'estimation.

les points de la diagonale principale. La référence originale dans lequel figure ce modèle en termes de lois exponentielles est [64].

Une autre classe de copules avec composantes singulières, et qui peuvent aussi être interprétées comme des modèles avec des chocs, est donnée ci-après. La classe des copules de « Durante » [17, 19] consiste en les copules de la forme

$$C(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d) f(\max(u_1, \ldots, u_d)), \tag{1.8}$$

où $f : [0,1] \to [0,1]$, appelé le *générateur* de C, est une fonction dérivable et strictement croissante telle que $f(1) = 1$ et $t \mapsto f(t)/t$ est strictement décroissante. L'interprétation en termes de chocs s'obtient en remarquant que (1.8) est la loi de $(U_1, \ldots, U_d)$, avec $U_i = \max(Z_i, Z_0)$, $i = 1, \ldots, d$; $Z_1, \ldots, Z_d$ sont des variables indépendantes distribuées selon une même loi f, et Z_0 est une variable indépendante de $(Z_1, \ldots, Z_d)$ distribuée selon $t \mapsto t/f(t)$. Dans le cas multivarié, puisqu'il n'y a qu'un seul générateur pour déterminer la structure de dépendance, cette classe n'est pas très utile pour les applications. Dans le cas bivarié, en revanche, la classe de copules de Durante est flexible et maniable (voir le chapitre 6). On peut également obtenir des copules bien connues comme cas particuliers. Ainsi, si $f(t) = t^{1-\theta}$, $\theta \in [0,1]$, on obtient la famille de Cuadras-Augé (1.7), et si $f(t) = (1-\theta)t + \theta$, $\theta \in [0,1]$, on obtient la famille de Fréchet [24], donnée par

$$C_\theta(u_1, u_2) = (1-\theta)\Pi(u_1, u_2) + \theta M(u_1, u_2),$$

et qui est la moyenne arithmétique entre la copule d'indépendance Π et la copule de la dépendance positive parfaite M. La classe de copules bivariées de Durante sera utilisée dans la construction du modèle que nous proposons au chapitre 6. En particulier, dans ce chapitre, nous construisons des copules multivariées, flexibles et maniables, dont les marges bivariées appartiennent à cette classe. On peut donc voir notre travail comme une extension au cas multivarié de la classe des copules bivariées de Durante plus flexible pour les applications que (1.8).

A la vue de la figure 1.2, il ne vient pas à l'esprit que l'on pourrait modéliser un phénomène lisse, présent dans la nature, tel que par exemple les débits de rivières ou la quantité de pluie qui tombe sur plusieurs sites répartis dans l'espace, avec un modèle possédant une composante singulière. C'est pourtant ce qu'on fait les auteurs dans [20, 77]. Leurs résultats suggèrent que certaines caractéristiques d'intérêt de la loi sous jacente ont pû être approchées par un modèle à chocs, alors même qu'il est clair que, par exemple, la probabilité que deux débits dans deux rivières différentes soient exactement égaux est nulle. Cette approche [5], qui s'intéresse moins à modéliser la distribution sous jacente qu'à estimer certaines caractéristiques de celle-ci, comme par exemple le niveau critique associé à une évènement de pluie extrême (ces derniers seront vus aux chapitres 6 et 3), bien que douteuse en dimension petite, devient intéressante quand la dimension augmente. C'est aussi l'approche que nous avons suivie, comme nous le verrons au chapitre 6.

5. Cette stratégie se rapproche de la méthode « boîte noire » utilisée en *machine learning*. Les deux approches, « boîte noire » et « modèle », sont discutées par exemple dans [7], où l'auteur se fait par ailleurs l'avocat de la première.

Chapitre 2

Modèles de copules en grande dimension

Les copules en grande dimension, ou simplement multivariées, sont plus difficiles à construire que les copules bivariées. A cause de cela, le terme *multivarié* se réfère souvent au cas où le nombre de variables d est supérieur à 2 strictement. Le terme *grande dimension*, quand à lui, est très subjectif. La « grande dimension », telle qu'elle est parfois entendue par la communauté des chercheurs dans le domaine des copules, peut commencer à partir de $d = 3$ [1]...

Dans ce chapitre, nous présentons les principaux [2] modèles de copules de la littérature. En grande dimension, il y a principalement trois familles de modèles : les copules archimédiennes (partie 2.1) et leur extensions, les copules archimédiennes imbriquées (2.2), les *Vines* (partie 2.3), et les copules elliptiques (partie 2.4).

2.1 Copules archimédiennes

Une copule archimédienne est une copule qui s'écrit

$$C(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)) \tag{2.1}$$

où ψ est une fonction décroissante et continue de $[0, \infty)$ dans $[0, 1]$, strictement décroissante sur $[0, \inf\{x : \psi(x) = 0\})$, et telle que $\psi(0) = 1$ et $\psi(x) \to 0$ quand $x \to \infty$. En fait, ces conditions sur ψ sont nécessaires mais pas suffisantes. Ces dernières ont été établies dans [67]. Dans le cas où $\inf\{x : \psi(x) = 0\} = \infty$, (2.1) est une copule bien définie si et seulement si ψ est complètement monotone [49], c'est à dire que $(-1)^i \psi^{(i)}(s) \geq 0$ pour tout i et tout $s \geq 0$, où $\psi^{(i)}$ est la i-ème dérivée de ψ. De plus, dans ce cas, on peut montrer que ψ est une transformée de

1. A l'heure des *big data*, on pourrait avoir du mal à cacher sa déception. Cependant, il convient de garder à l'esprit que ce ne sont pas les mêmes questions scientifiques qui sont posées, et que ce ne sont pas non plus les mêmes modèles qui sont utilisés. Par exemple, dans les *genome-wide association studies* en bioinformatique – où on analyse plusieurs milliers de variables –, on s'autorise les modèles linéaires gaussiens pour modéliser le bruit [46]. Dans le domaine des copules, l'aléa n'est pas du bruit, il est considéré comme intrinsèque, et les modèles linéaires gaussiens en seraient une trop mauvaise approximation.

2. cet adjectif comporte inévitablement une part de subjectivité

Laplace d'un vecteur strictement positif. Autrement dit, il existe une fonction de répartition H telle que

$$\psi(s) = \int_0^\infty \exp(-sy)dH(y), \quad s \geq 0.$$

En outre, dans ce cas, il existe de manière unique d fonctions de répartition $G_1, \ldots, G_d$ telles que

$$(2.1) = \int_0^\infty (G_1 \ldots G_d)^\alpha dH(\alpha) = \psi\left(-\sum_{i=1}^d \log G_i\right).$$

Par exemple, la copule de Gumbel [36], donnée par

$$C(u_1, \ldots, u_d) = \exp\left\{-\left[(-\log u_1)^\theta + \cdots + (-\log u_d)^\theta\right]^{1/\theta}\right\}, \tag{2.2}$$

est une copule archimédienne de la forme (2.1) avec $\psi^{-1}(t) = (-\log t)^\theta$ pour $\theta \geq 1$ (notons que c'est aussi une copule des valeurs extrêmes). La copule de Clayton (1.4) est aussi une copule archimédienne avec $\psi^{-1}(t) = (t^{-\theta} - 1)/\theta$, $\theta > 0$. Les copules archimédiennes possèdent l'avantage d'être simples, explicites et interprétables. Ainsi, pour une copule archimédienne bivariée, le tau de Kendall est donné par

$$\tau = 1 + 4\int_0^1 \frac{\psi^{-1}(t)}{(\psi^{-1})'(t)}dt.$$

Par exemple, celui de la copule de Clayton vaut $\theta/(\theta + 2)$. En général, le générateur ψ est déterminé par un ou deux paramètre(s), dont on pourra en trouver une liste dans [69] section 4.

Les copules archimédiennes ont cependant un défaut de poids : leurs quelques paramètres sont supposés rendre compte de toute la richesse de la structure de dépendance entre toutes les variables quelle que soit la dimension considérée. Ces copules sont *échangeables*, c'est à dire que

$$C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots, u_{\pi(d)})$$

pour toute permutation π de $(1, \ldots, d)$. Ceci implique en particulier que toutes les paires de variables ont la même loi statistique.

Les applications des copules archimédiennes en grande dimension couvrent, entre autre, la modélisation et l'évaluation du risque associé à des portefeuilles contenant un grand nombre d'actifs financiers, comme dans l'exemple introductif que nous avons vu dans l'introduction, voir aussi [42] et [65]. Du fait de la propriété d'échangeabilité, de meilleurs résultats sont attendus si ces portefeuilles sont relativement homogènes, mais, comme il est fait remarquer dans [42], les copules archimédiennes ne sont pas utilisées pour ajuster au mieux les données ; on les utilise plutôt en vertu de leur maniabilité, d'une point de vue numérique notamment, et on attend qu'elle résume tout de même la dépendance de manière globale.

2.2 Copules archimédiennes imbriquées

Les copules archimédiennes imbriquées (CAI), ou hiérarchiques, sont une tentative d'assouplir la structure de dépendance des copules archimédiennes

classiques. Elles sont apparues dans [47] Section 4.2, puis ont fait l'objet d'études numériques très poussées [40,68,72] et commencent à être utilisées dans plusieurs applications en finance et économétrie [43, 78] et hydrologie [83].

Une CAI est une copule construite en imbriquant des copules archimédiennes les unes dans les autres. Par exemple, en dimension 3, la copule

$$C(u_1, u_2, u_3) = C_{\psi_0}(u_1, C_{\psi_{23}}(u_2, u_3)) \tag{2.3}$$

est une CAI car la copule C_{ψ_0} prend pour second argument une autre copule $C_{\psi_{23}}$. En notant ψ_0 et ψ_{23} les générateurs des copules archimédiennes C_{ψ_0} et $C_{\psi_{23}}$, (2.3) se réécrit

$$C(u_1, u_2, u_3) = \psi_0\left(\psi_0^{-1}(u_1) + \psi_0^{-1}(\psi_{23}(\psi_{23}^{-1}(u_2) + \psi_{23}^{-1}(u_3)))\right). \tag{2.4}$$

Le même principe s'applique pour construire des copules en plus grande dimension. L'avantage par rapport aux copules archimédiennes classiques réside dans la possibilité de construire des structures de dépendance plus souples. Ainsi, dans (2.3), la loi de (U_2, U_3) diffère de celle de (U_1, U_2). Malheureusement, même si cette rustine apportée aux copules archimédiennes a été l'objet de plusieurs études (voir les références citées plus haut), elle apparait comme un bien maigre réconfort face aux problèmes qu'il reste à résoudre, et, pire, qu'elle engendre. Tout d'abord, le manque de souplesse des copules archimédiennes n'a pas été complétement éliminé. Par exemple, pour reprendre notre CAI (2.3), les paires (U_1, U_2) et (U_1, U_3) ont la même distribution. De plus, les fonctions qui s'écrivent sous la forme (2.4) ne sont pas nécessairement des copules. Les conditions nécessaires sur les générateurs demeurent inconnues. La condition suffisante que l'on trouve dans [47] Section 4.2 ou dans [68] est difficile à vérifier en pratique. Dans le cas particulier où les générateurs sont de la même famille, cette condition est vérifiée si la suite des paramètres des générateurs croit en descendant dans la structure d'imbrication. Par exemple, dans les expressions (2.4) ou (2.3), cela reviendrait à dire que, si on choisissait des générateurs de la famille de Gumbel, il faudrait que les paramètres vérifient $\theta_{\psi_0} \leq \theta_{\psi_{23}}$, ce qui peut être assez restrictif.

2.3 *Vines*

Les modèles *Vines* (qui signifie « vignes, grappes, plantes grimpantes » en français [3]) sont basés sur la décomposition d'une densité f en un produit de densités conditionnelles de copules bivariées multiplié par le produit des densités marginales.

D'après la formule de dépendance conditionnelle, une densité de probabilité f peut se décomposer comme

$$f(x_1, \ldots, x_d) = f_d(x_d) f_{d-1|d}(x_{d-1}|x_d) \ldots f_{1|2\ldots d}(x_1|x_2, \ldots, x_d). \tag{2.5}$$

Chaque terme du membre de droite de (2.5) peut lui même se décomposer en un produit de densités de copules conditionnelles multiplié par un produit de

3. Cette appellation vient du fait que la représentation graphique de ces modèles, qui n'est pas abordé dans cette thèse, ressemblerait à des vignes.

marginales en utilisant la relation (1.2), vue dans la partie 1.1, et rappelée ci-dessous :

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \dots f_d(x_d).$$

Ainsi, f dans (2.5) s'écrit comme un produit de densités de copules conditionnelles multiplié par un produit de densités marginales. Par exemple, en dimension $d = 3$, une décomposition possible est :

$$f_{123}(x_1, x_2, x_3) = f_3(x_3) f_{2|3}(x_2|x_3) f_{1|23}(x_1|x_2, x_3).$$

On réécrit les termes de la décomposition. D'abord,

$$\begin{aligned} f_{2|3}(x_2|x_3) =& \frac{f_{23}(x_2, x_3)}{f_3(x_3)} \\ =& \frac{c_{23}(F_2(x_2), F_3(x_3)) f_2(x_2) f_3(x_3)}{f_3(x_3)} \\ =& c_{23}(F_2(x_2), F_3(x_3)) f_2(x_2). \end{aligned}$$

Ensuite,

$$\begin{aligned} f_{1|23}(x_1|x_2, x_3) =& \frac{f_{12|3}(x_1, x_2|x_3) f_3(x_3)}{f_{23}(x_2, x_3)} \\ =& \frac{c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) f_{1|3}(x_1|x_3) f_{2|3}(x_2|x_3) f_3(x_3)}{f_{23}(x_2, x_3)} \\ =& c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) f_{1|3}(x_1|x_3) \\ =& c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) c_{13}(F_1(x_1), F_3(x_3)) f_1(x_1). \end{aligned}$$

Au final, on a

$$\begin{aligned} f(x_1, x_2, x_3) =& c_{23}(F_2(x_2), F_3(x_3)) c_{13}(F_1(x_1), F_3(x_3)) \\ & c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)) f_1(x_1) f_2(x_2) f_3(x_3). \end{aligned} \tag{2.6}$$

D'après (1.2), le produit de densités de copules conditionnelles dans (2.6) est une décomposition de la densité de la copule c associée à f. Plus la dimension augmente, plus le nombre de décompositions possibles augmente. Les *Vines* ou *Vines* régulières [4, 5] sont un type de décompositions possible, mais encore trop large puisque les cas particuliers appelés *Vines canoniques* (*C-vines* en anglais) et *Vines* « *dessinables* » (*D-vines*) ont été introduites quelques années après [58]. Les décompositions *C-vines* et *D-vines* peuvent être représentées par des modèles graphiques consistant en une suite d'arbres. On pourra consulter les références précédentes pour plus de détails.

L'atout principal des *Vines* est leur grande flexibilité. En effet, passé l'étape du choix de la décomposition de f, aucun modèle n'est encore défini. Quelle que soit la décomposition choisie, on sait qu'il existe des densités de copules qui permettent de retrouver f *exactement*. En pratique, cela laisse le choix à l'utilisateur de quelles paires il va modéliser sans faire de restrictions sur f. Enfin, une fois la décomposition de f choisie, on peut faire une modélisation fine paire par paire et tirer profit de la grande richesse de la gamme de familles de copules bivariées qui existe dans la littérature.

Les inconvénients des *Vines* sont les suivants. D'abord, dans la décomposition de la densité, en pratique, on fait l'hypothèse que les copules conditionnelles ne dépendent pas des valeurs conditionnantes. Par exemple, dans (2.6), la loi de (U_1, U_2) sachant $U_3 = u_3$, $c_{12|3}(\cdot, \cdot|u_3)$, est supposée ne pas dépendre en fait de u_3, c'est à dire que la loi est la même quelles que soient les valeurs prises par $U_3 = u_3$. Cette hypothèse, faite dans la pratique afin de pouvoir choisir pour les copules conditionnelles des modèles paramétriques bivariés abondant dans la littérature, a été discutée dans [2]. De plus, étant donné le très grand nombre de possibilités lors de la modélisation par un modèle *Vines* – possibilités offertes à la fois par le choix de la décomposition et le choix des familles paramétriques à utiliser dans cette décomposition –, il n'est pas encore clair comment choisir le « meilleur » modèle *Vines* et comment tester la robustesse de ce choix. Enfin, les modèles *Vines* ne sont pas des modèles très maniables pour l'utilisateur. Les coûts de calcul nécessaires pour la simulation ou l'estimation sont plus importants que pour d'autres modèles de copules et l'étude des propriétés de dépendance est également moins aisée. On pourra consulter [1] pour un résumé à la fois complet et accessible de la modélisation de données par *Vines* et [59] pour une référence plus exhaustive.

2.4 Copules elliptiques

Une copule elliptique est une copule associée à une loi elliptique. Une loi elliptique est une transformation affine d'une loi sphérique. Un vecteur $Y = (Y_1, \ldots, Y_d)$ est distribué selon une loi sphérique si Y a la même loi que QY pour toute matrice orthogonale Q, c'est à dire pour toute matrice telle que $Q^T Q = QQ^T = I_d$, où I_d est la matrice identité de taille d. Autrement dit, une loi shérique est une loi invariante par rotation. La densité f_Y d'une loi sphérique s'écrit $f_Y(t) = g(\|t\|^2)$, $t \in \mathbb{R}^d$, où g est une fonction univariée appelée le *générateur de densité* de la loi sphérique que l'on note $S_d(g)$. Un vecteur $Y \sim S_d(g)$ a la représentation

$$Y = RS \tag{2.7}$$

où S est un vecteur aléatoire distribué uniformément sur la sphère unitaire $\{s \in \mathbb{R}^d : s^T s = 1\}$ et $R \geq 0$ est une variable aléatoire indépendante de S. Un vecteur $X = (X_1, \ldots, X_d)$ est distribué selon une loi elliptique $E_d(\mu, \Sigma, g)$ s'il s'écrit $X = \mu + \Sigma^{1/2} Y$ où $Y \sim S_d(g)$ et Σ est une matrice définie positive telle que $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. La densité f_X de X s'écrit

$$f_X(t) = |\Sigma|^{-1/2} g\left((t-\mu)^T \Sigma^{-1} (t-\mu)\right), \quad t \in \mathbb{R}^d$$

et par conséquent est constante sur les ellipsoïdes de la forme $\{x : (x-\mu)^T \Sigma^{-1} (x-\mu) = c\}$ pour une certaine constante c. La matrice de variance-covariance de X, lorsqu'elle existe, est donnée par $E(R^2)\Sigma/d$ où E est symbolise l'espérance mathématique et R est définie dans (2.7).

Comme nous l'avons dit plus haut, une copule elliptique en dimension d est la copule associée à une loi elliptique $E_d(\mu, \Sigma, g)$. Puisqu'une copule est invariante par standardisation des lois marginales, $E_d(\mu, \Sigma, g)$ et $E_d(0, P, g)$ ont la même copule, où P est la matrice de corrélation obtenue à partir de la matrice Σ. Deux propriétés remarquables des copules elliptiques sont les suivantes. D'abord, on

peut montrer [63] que le tau de Kendall d'une loi elliptique est donné par

$$\tau = \frac{2}{\pi} \arcsin(\rho_{ij}), \tag{2.8}$$

où ρ_{ij} est l'élément de la i-ème ligne et j-ème colonne de P. Ensuite, une loi elliptique $E_d(\mu, \Sigma, g)$ est une loi symmétrique par rapport à son rayon μ, c'est à dire que $X - \mu$ est distribué comme $\mu - X$. Cette propriété implique que les coefficients de dépendance de queues inférieurs et supérieurs sont égaux

$$\lambda^{(L)} = \lambda^{(U)}. \tag{2.9}$$

Deux exemples de copules très connues sont les copules de Student et gaussiennes. La copule de Student avec degré de liberté ν est une copule elliptique pour laquelle

$$g_\nu(x) = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d}} \left(1 + \frac{x}{\nu}\right)^{-(\nu+d)/2}, \quad \nu > 2, x \geq 0$$

et la copule gaussienne est une copule pour laquelle

$$g(x) = (2\pi)^{-d/2} \exp(-x/2), \quad x \geq 0.$$

Le coefficient de dépendance de queue de la copule de Student bivariée issue de la loi $E_2(0, \rho, g_\nu)$ est donné par

$$\lambda^{(U)} = \lambda^{(L)} = 2t_{\nu+1}\left(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}\right),$$

où t_ν est la fonction de répartition de la loi univariée de Student standard. Le coefficient de dépendance de queue de la copule gaussienne vaut 0 (quand le coefficient de corrélation est strictement inférieur à 1) : elle n'a pas de dépendance de queue.

L'avantage des copules elliptiques est que l'on peut moduler l'équilibre entre la flexibilité et la maniabilité dans la modélisation. On peut par exemple réduire le nombre de paramètres en imposant une structure particulière à la matrice de variance-covariance Σ, voir par exemple [51]. De plus, la relation bijective entre les paramètres de la matrice P et les taus de Kendall (2.8) permet d'estimer ces paramètres par la méthode d'inversion du tau de Kendall, voir par exemple [16] et la partie 3.1.2 de cette thèse. Les copules elliptiques ont l'inconvénient que les coefficients de dépendance supérieurs et inférieurs sont égaux (2.9). Elles ne sont donc pas des modèles réalistes pour modéliser des données présentant des dépendances de queue seulement pour les grandes valeurs ou les petites valeurs. En outre, l'ajustement de copules elliptiques à des données résulte parfois – voire souvent, lorsque le nombre de variables est grand par rapport à la taille de l'échantillon – en une matrice de covariance mal conditionnée ou non inversible. Or, dans les applications, il est nécessaire de calculer l'inverse afin de pouvoir évaluer la densité [60].

Notre présentation des loi sphériques et elliptiques s'est appuyée sur les documents [23] et [66] Section 3.3 que l'on pourra consulter pour plus de détails. Pour un exposé complet et accessible sur la copule de Student et ses extensions, comme par exemple une copule des valeurs extrêmes reliée à la copule de Student, on pourra consulter [16]. Cet article traite aussi de l'inférence. La copule de Student est très utilisée en finance et gestion du risque, voir par exemple [66] et ses références.

Chapitre 3

Inférence

Supposons que l'on dispose d'un échantillon d'une copule appartenant à une famille paramétrique (C_θ), et que l'on souhaite estimer le paramètre θ, possiblement multivarié, de notre copule. Notons notre échantillon par

$$(X_1^{(1)}, \ldots, X_d^{(1)}), \ldots, (X_1^{(n)}, \ldots, X_d^{(n)}), \tag{3.1}$$

et remarquons que si les fonctions de répartition $F_1, \ldots, F_d$ de $X_1, \ldots, X_d$ étaient connues, l'échantillon

$$(F_1(X_1^{(1)}), \ldots, F_d(X_d^{(1)})), \ldots, (F_1(X_1^{(n)}), \ldots, F_d(X_d^{(n)})),$$

serait un échantillon de la copule C elle même, et alors on pourrait se ramener aux méthodes d'inférence classiques de la statistique. Mais, puisque les marges sont en fait inconnues, nous ne disposons pas d'un échantillon de notre copule. Pour surmonter cette difficulté, principalement deux approches peuvent être adoptées. Dans l'approche *paramétrique*, on suppose que les lois marginales appartiennent elles aussi à une famille indexée par un paramètre. Pour estimer les marges, il suffit alors d'estimer leur paramètre. Dans l'approche *semi-paramétrique*, on ne fait pas l'hypothèse que les lois marginales appartiennent à une quelconque famille. On estime les marges non paramétriquement, par exemple avec la version de l'estimateur empirique donnée par

$$\widehat{F}_i(x) = \frac{1}{n+1} \sum_{k=1}^{n} \mathbf{1}(X_i^{(k)} \leq x). \tag{3.2}$$

Quelle que soit la façon dont on modélise et estime les lois marginales, le paramètre θ de la copule doit ensuite être estimé. Il y principalement deux stratégies. La première est basée sur la maximisation d'une certaine fonction de vraisemblance, et la deuxième est une méthode des moments basée sur les coefficients de dépendance. Si l'on souhaite se passer de l'hypothèse selon laquelle la copule appartient à une famille paramétrique, il faut estimer la copule avec des méthodes non-paramétriques. Puisque ces méthodes ont peu de chances de succès en grande dimension, nous ne les aborderons pas en détail, mais donnons quelques références à la littérature.

La suite de ce chapitre est organisée comme suit. La partie 3.1 traite de l'estimation de copules, en présentant les méthodes basées sur la vraisemblance

(partie 3.1.1) et sur les coefficients de dépendance (partie 3.1.2). Pour les méthodes non paramétriques, les références à la littérature sont données dans la partie 3.1.3. La partie 3.2 traite brièvement des tests d'adéquation.

3.1 Estimation

3.1.1 La méthode du maximum de vraisemblance

La densité associée à notre échantillon (3.1) a été donnée dans (1.2) ; nous la rappelons ci-dessous :

$$f(x_1, \ldots, x_d; \theta) = c(F_1(x_1), \ldots, F_d(x_d); \theta) \prod_{i=1}^{d} f_i(x_i), \tag{3.3}$$

où c est la densité de la copule d'intérêt, F_i sont les fonctions de répartition marginales et f_i les densités. Ci-dessous, les méthodes présentées, paramétriques et semi-paramétriques, cherchent toutes les deux à maximiser une approximation de la vraisemblance basée sur (3.3). La différence entre les deux méthodes tient au fait que dans l'approche paramétrique, nous supposons que les marges $F_1, \ldots, F_d$ appartiennent à une certaine famille paramétrique, ce qui n'est pas le cas de méthode semi-paramétrique. Evidemment, l'intérêt de l'approche paramétrique réside dans le fait que, si le modèle ajusté pour les lois marginales est raisonnable, cette approche permet une réduction de la variabilité et une augmentation de la maniabilité du modèle. En revanche, si le modèle est mal ajusté, les résultats peuvent donner lieu à des interprétations fausses [21]. C'est pour cette raison que l'approche semi-paramétrique est intéressante.

L'approche paramétrique suppose que chaque marge F_i appartient à une famille de lois indexée par un paramètre α_i. Ainsi, la densité (3.3) se réécrit comme

$$f(x_1, \ldots, x_d; \alpha_1, \ldots, \alpha_d; \theta) = c(F_1(x_1; \alpha_1), \ldots, F_d(x_d; \alpha_d); \theta) \prod_{i=1}^{d} f_i(x_i; \alpha_i). \tag{3.4}$$

Pour estimer le vecteur des paramètres $(\alpha_1, \ldots, \alpha_d, \theta)$, on peut vouloir maximiser la vraisemblance *stricto sensu*

$$L(\alpha_1, \ldots, \alpha_d, \theta) := \prod_{k=1}^{n} f(X_1^{(k)}, \ldots, X_d^{(k)}; \alpha_1, \ldots, \alpha_d, \theta).$$

Cependant, cette vraisemblance peut être compliquée, voire impossible à calculer, ou alors l'optimisation numérique peut être trop lente ou trop complexe. Dans ces situations, on fera appel à une méthode en deux étapes, qui tire profit de la représentation (3.4). Avec cette méthode, appelée la méthode IFM (*Inference Functions for Margins*, [47], Section 10), on procède en deux étapes.

1) Le paramètre α_i est estimé par $\tilde{\alpha}_i$ en maximisant la vraisemblance marginale

$$L_i(\alpha_i) = \prod_{k=1}^{n} f_i(X_i^{(k)}; \alpha_i).$$

2) Une fois les paramètres marginaux estimés dans l'étape précédente, le vecteur de paramètres θ de la copule est estimé par $\tilde{\theta}$ en maximisant la partie de la vraisemblance qui dépend de θ, c'est à dire que l'on maximise

$$L(\theta) = \prod_{k=1}^{n} c(F_1(X_1^{(k)}; \tilde{\alpha}_1), \ldots, F_d(X_d^{(k)}; \tilde{\alpha}_d); \theta).$$

L'estimateur IFM est consistant et asymptotiquement normal sous certaines conditions de régularité, voir par exemple [82]. L'approche IFM est intéressante d'un point de vue numérique car l'optimisation par ordinateur a plus de chances de succès que la maximisation de la vraisemblance *stricto sensu*. Au vu de ce qui précède, on peut se demander quelle est l'efficacité relative de l'estimateur IFM par rapport à l'estimateur du maximum de vraisemblance. Dans [47] section 10, il est suggéré de comparer (numériquement) les matrices asymptotiques de variance-covariances des deux estimateurs, ou bien de comparer les deux estimateurs au moyen de simulations numériques. Dans les comparaisons pour quelques modèles faites dans [47], l'auteur rapporte une efficacité relative proche de 1, où l'efficacité relative a été mesurée comme le rapport de l'erreur quadratique moyenne de l'estimateur IFM avec celle de l'estimateur du maximum de vraisemblance.

L'approche semi-paramétrique ne fait pas l'hypothèse que les marges $F_1, \ldots, F_d$ appartiennent à une quelconque famille paramétrique. On les estime directement par l'estimateur non paramétrique donné dans (3.2), rappelé ci-dessous,

$$\widehat{F}_i(x) = \frac{1}{n+1} \sum_{k=1}^{n} \mathbf{1}(X_i^{(k)} \leq x).$$

Pour estimer θ, on remplace les marges par leur estimation dans (3.3), et on maximise la partie de la vraisemblance faisant intervenir θ, c'est à dire,

$$L(\theta) = \prod_{k=1}^{n} c(\widehat{F}_1(X_1^{(k)}), \ldots, \widehat{F}_d(X_d^{(k)}); \theta).$$

L'estimateur qui en résulte est consistant et asymptotiquement normal sous des conditions de régularité peu restrictives, voir [26]. Toutefois, malgré ces propriétés de convergence, il n'est pas, en général, efficace [32], sauf dans le cas de la copule gaussienne bivariée [50]. La construction d'estimateurs asymptotiquement efficaces est un sujet de recherche très récent, se focalisant pour l'instant sur les modèles gaussiens : l'obtention d'une borne inférieure pour la matrice de variance-covariance asymptotique, et la preuve que cette borne pouvait être atteinte par un estimateur semi-paramétrique a été réalisé dans [44]. Dans [81], les auteurs ont construit explicitement un estimateur atteignant la borne inférieure.

3.1.2 La méthode des moments basée sur les coefficients de dépendance

Dans la littérature, l'estimation basée sur les moments est souvent entendue comme un nom générique se référant en fait à la méthode basée sur l'inversion

du rho de Spearman ou du tau de Kendall (qui peuvent en effet se voir de la sorte). Ces méthodes tirent profit de la relation, plus ou moins explicite, qu'il peut y avoir entre le rho de Spearman et le tau de Kendall et le paramètre de la copule θ. Nous avons vu dans la partie 1.2.2 la définition de ces coefficients de dépendance. La version empirique de ces coefficients pour la paire (X_i, X_j) est respectivement donnée par

$$\hat{\rho}_{i,j} = \frac{\sum_{k=1}^{n} \left(\widehat{U}_i^{(k)} - \overline{\widehat{U}}_i\right)\left(\widehat{U}_j^{(k)} - \overline{\widehat{U}}_j\right)}{\left[\sum_{k=1}^{n} \left(\widehat{U}_i^{(k)} - \overline{\widehat{U}}_i\right)^2 \sum_{k=1}^{n} \left(\widehat{U}_j^{(k)} - \overline{\widehat{U}}_j\right)^2\right]^{1/2}}, \text{ et} \tag{3.5}$$

$$\hat{\tau}_{i,j} = \binom{n}{2}^{-1} \sum_{k<l} \operatorname{sign}\left((X_i^{(k)} - X_i^{(l)})(X_j^{(k)} - X_j^{(l)})\right), \tag{3.6}$$

où $\widehat{U}_i^{(k)} = \widehat{F}_i(X_i^{(k)})$, $\overline{\widehat{U}_i} = \sum_{k=1}^{n} \widehat{U}_i^{(k)}/n$, $i = 1, \ldots, d$ et $\operatorname{sign}(x) = 1$ si $x > 0$, -1 si $x < 0$ et 0 si $x = 0$. Depuis [39], on sait que ces deux estimateurs sont consistants et asymptotiquement non biaisés et normaux. Dans le cas bivarié ($d = 2$) et lorsqu'il n'y a qu'un seul paramètre réel à estimer, la méthode par inversion du tau de Kendall s'applique à faire correspondre l'estimation sous le modèle avec son estimation empirique. En d'autres termes, l'estimateur $\hat{\theta}$ vérifie

$$\tau(\hat{\theta}) = \hat{\tau}_{1,2},$$

et donc, si $\theta \mapsto \tau(\theta)$ est inversible

$$\hat{\theta} = \tau^{-1}(\hat{\tau}_{1,2}). \tag{3.7}$$

Les propriétés asymptotiques de (3.7) s'obtiennent immédiatement d'après les propriétés asymptotiques de $\hat{\tau}_{1,2}$ lui même et la méthode « delta » (*delta-method* en anglais, voir par exemple [84]). Si au lieu du tau de Kendall, on souhaite utiliser le rho de Spearman, voire d'autres coefficients de dépendance, la méthode fonctionne de la même manière. Notons enfin que la méthode par inversion du tau de Kendall ou du rho de Spearman est semi-paramétrique, puisque dans les expressions (3.6) et (3.5), les marges $F_1, \ldots, F_d$ sont implicitement estimées par (3.2). On pourra consulter [25] pour une introduction accessible de ces méthodes, et [30] pour plus de détails dans le cas du tau de Kendall et des copules archimédiennes. Des généralisations de la méthode des moments basée sur les coefficients de dépendance ont été proposées dans la littérature pour pouvoir estimer des copules plus complexes.

Copules elliptiques. La méthode par inversion du tau de Kendall est très utilisée pour estimer les paramètres de la matrice de corrélation P des copules elliptiques (vues dans la partie 2.4), car, pour chaque marge bivariée de ces copules, il y a une correspondance un à un entre l'élément de la i-ème ligne et de la j-ème colonne de P et le tau de Kendall (2.8). Voir par exemple [66] chapitre 5.5, ou [16] dans le contexte des copules de Student. Dans le cas d'un modèle parcimonieux, c'est à dire lorsque l'on impose une structure à la matrice de variance-covariance, cette correspondance est brisée. Cependant, on peut tout de même estimer le vecteur des paramètres en minimisant la fonction (3.8), comme expliqué ci-dessous.

Le cas multivarié, mais avec un seul paramètre réel. Au delà du cas bivarié ($d > 2$), mais lorsqu'il n'y a qu'un seul paramètre à estimer – c'est le cas par exemple des copules archimédiennes, voir la partie 2.1 – une extension a été étudiée dans [27]. Comme il y a cette fois plusieurs paires, les auteurs étudient l'estimateur qui vérifie

$$\tau(\hat{\theta}) = \frac{1}{d(d-1)/2} \sum_{i<j} \hat{\tau}_{i,j},$$

où les $\hat{\tau}_{i,j}$ sont les taus de Kendall empiriques des paires (X_i, X_j). Toujours en se basant sur [39], on peut établir les propriétés asymptotiques de $\hat{\theta}$.

Le cas général. L'estimateur revêt la forme

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(\hat{r} - r(\theta)\right)^T \widehat{W} \left(\hat{r} - r(\theta)\right), \tag{3.8}$$

où $\widehat{W}$ est une matrice de poids et $r = (r_{1,2}, \ldots, r_{d-1,d})$, $\hat{r} = (\hat{r}_{1,2}, \ldots, \hat{r}_{d-1,d})$. Pour construire l'estimateur, $r_{i,j}$ doit être remplacé par un coefficient de dépendance entre X_i et X_j et $\hat{r}_{i,j}$ par sa contrepartie empirique. Par exemple, le cas des copules elliptiques avec $r_{i,j} = \mathrm{Cor}(X_i, X_j)$ a été considéré dans [51]. Le cas plus général où $r_{i,j}$ peut être n'importe quel coefficient de dépendance (sous des conditions de convergence de sa contrepartie empirique) a été considéré dans [71]. Dans cet article, les auteurs partent du principe que les coefficients de dépendance $r_{i,j}$ ne peuvent pas être calculés, même numériquement. Leur point de vue est motivé par le fait que les copules qu'ils considèrent [70, 71] sont définies implicitement. Pour résoudre ce problème, ils proposent d'approcher $r_{i,j}$ par simulation. Nous renvoyons à [71] pour plus de détails. L'estimateur a été prouvé consistant et asymptotiquement normal sous des conditions de régularité naturelles.

Dans le cas général, pour établir les propriétés asymptotiques de (3.8), les auteurs de [71] ont besoin de l'existence (et de la continuité) des dérivées partielles des copules sous jacentes. Donc, si l'on souhaite estimer les paramètres de copules pour lesquelles ces dérivées n'existent pas, comme par exemple les copules singulières vues dans la partie 1.3.2 et le chapitre 6, il n'y a aucun argument théorique pour utiliser cette méthode. Dans ce contexte, l'objet du chapitre 5 est de lever cette hypothèse. Ainsi, nous pouvons estimer les paramètres des copules proposées dans le chapitre 6.

3.1.3 Les méthodes non paramétriques

La plupart des méthodes non-paramétriques sont basées sur la copule empirique [14] définie par

$$\widehat{C}(u_1, \ldots, u_d) = \widehat{F}(\widehat{F}_1^{-1}(u_1), \ldots, \widehat{F}_d^{-1}(u_d)),$$

où $\widehat{F}$ est un estimateur de la fonction de répartition – par exemple la fonction de répartition empirique, ou bien un estimateur construit à l'aide d'un noyau –, et $\widehat{F}_i^{-1}$ est un estimateur non paramétrique de la fonction quantile généralisée, par exemple, $\widehat{F}_i^{-1}(x) = \inf\{y : \widehat{F}_i(y) \geq x\}$. La consistance et normalité asymptotique de la copule empirique ont été établies dans [22] sous des conditions de

régularité sur les dérivées partielles. Voir aussi [80] pour la possibilité de relaxer certaines de ces conditions.

Dans le cas des copules des valeurs extrêmes, on peut envisager d'autres méthodes qui se basent sur une représentation de celles-ci en terme d'une fonction de dépendance univariée qui doit vérifier certaines propriétés, voir par exemple [33]. Dans le cas où les lois marginales F_i sont supposées connues, on pourra consulter [9, 15, 38] dans le cas bivarié, et [73] dans le cas multivarié. Dans le cas où elles sont supposées inconnues, voir [8, 31] dans le cas bivarié et [34, 35] dans le cas multivarié.

Récemment, les méthodes non paramétriques ont été utilisées pour estimer des copules *Vines* [37]. Puisque les copules *Vines* sont construites à partir de copules bivariées, ces méthodes ont de meilleures chances de succès dans ce cas.

Enfin, la copule de Durante vue dans la partie 1.3.2 a aussi été le terrain de jeux de méthodes non paramétriques, voir [18].

3.2 Tests

Etant choisie une famille paramétrique pour la copule d'intérêt, il est critique de se demander si ce choix était le bon. Pour répondre à cette question, on fait donc le test H_0 : « la copule appartient à la famille paramétrique choisie » contre H_1 : « la copule n'appartient pas à la famille ». D'après [6, 29], les tests les plus puissants sont basés sur le processus

$$\sqrt{n}(\widehat{C} - C_{\hat{\theta}}),$$

où $\widehat{C}$ est la copule empirique et $C_{\hat{\theta}}$ est l'estimation paramétrique obtenue sous H_0. En particulier, la statistique de test de Cramér-von Mises

$$\int_{[0,1]^d} n(\widehat{C} - C_{\hat{\theta}}) d\widehat{C}$$

donne les meilleurs résultats. Pour obtenir des p valeurs, on peut recourir au *bootstrap* [28] mais au prix d'un coût de calcul élevé. Une alternative, appelée approche *multiplier*, proposée dans [53,55] et implémentée dans le paquet `copula` [54] du logiciel `R` , permet de réduire ce coût. Enfin, des tests existent pour évaluer si une copule est une copule des valeurs extrêmes, voir par exemple [52].

Deuxième partie

Deux nouvelles classes de copules et leur estimation

Chapitre 4

Un modèle de copules basé sur des produits de copules bivariées

Un produit de copules n'est pas, en général, une copule. Un contre exemple évident serait de prendre $\tilde{C}_1(u,v) = \tilde{C}_2(u,v) = uv$ la copule d'indépendance et de remarquer que

$$C(u,v) = \tilde{C}_1(u,v)\tilde{C}_1(u,v) = u^2v^2 \tag{4.1}$$

n'est pas une copule, car ses marges ne sont pas uniformes. Toutefois, en modifiant les arguments des copules amenées à composer le produit, et en leur imposant certaines contraintes, il est possible de réussir à ce que ce dernier vérifie toutes les conditions pour être une copule bien définie. Dans l'exemple précédent (4.1), il suffit d'élever les arguments de $\tilde{C}_1$ et $\tilde{C}_2$ à la puissance $1/2$, c'est à dire $C(u,v) = \tilde{C}_1(u^{1/2}, v^{1/2})\tilde{C}_1(u^{1/2}, v^{1/2}) = uv$ pour que C soit bien une copule. De manière générale, considérons le produit

$$C(u_1, \ldots, u_d) = \prod_{e=1}^{K} \tilde{C}_e \left(g_{e1}(u_1), \ldots, g_{ed}(u_d)\right) \tag{4.2}$$

où les g_{ei} sont des fonctions de $[0,1]$ dans $[0,1]$ et les $\tilde{C}_e$ sont des copules arbitraires. Dans [62], l'auteur donne des conditions suffisantes sur les g_{ei} pour que (4.2) soit une copule bien définie. Dans la suite, nous parlerons donc de produit de copules en sous entendant le produit (4.2), et pas le produit *stricto sensu*. Entre autre, g_{ei} doit satisfaire la contrainte

$$\prod_{e=1}^{K} g_{ei}(v) = v, \quad v \in [0,1]. \tag{4.3}$$

Par exemple, si on prend la forme paramétrique $g_{ei}(v) = v^{\theta_{ei}}$, la contrainte devient

$$\sum_{e=1}^{K} \theta_{ei} = 1.$$

Au total, il y a d contraintes de la sorte à satisfaire, une pour chaque indice i.

La classe de copules (4.2) n'a pour l'instant un intérêt que théorique. En effet, en pratique, les problèmes apparaissent, comme le choix des g_{ei} et la satisfaction des contraintes (4.3). En outre, comment construire un modèle parcimonieux à partir de (4.2) ? Même si de telles copules pouvaient être construites, comment aborder l'inférence ? En effet, la méthode du maximum de vraisemblance nécessite de calculer la densité de la copule étudiée. Puisque la copule qui nous occupe est un produit, il est clair que son calcul sera très complexe. Même si, du fait des difficultés énoncées plus haut, peu de modèles ont été construits en pratique, on citera tout de même [20], où les auteurs utilisent (4.2) pour construire des copules des valeurs extrêmes. Cependant, dans ces travaux, les contraintes sur les paramètres n'ont pas pu être levées dans le cas où l'on souhaite un paramètre par paire de variables.

Dans ce contexte, notre contribution s'articule autour de deux axes. D'une part, nous proposons une classe de copules construite à partir de produits de copules bivariées et dont les contraintes (4.3) sont automatiquement satisfaites. D'autre part, nous faisons le lien avec un algorithme *message-passing* récent [45] qui permet de calculer la densité associée à un un produit de fonctions de répartition, et donc, a fortiori, d'un produit de copules. Nous montrons dans le Théorème 1 de notre article que cette nouvelle classe est obtenue avec seulement deux hypothèses naturelles. Notre copule s'écrit

$$C(u_1, \ldots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{ij}(u_i^{1/n_i}, u_j^{1/n_j}), \tag{4.4}$$

où les $\tilde{C}_{ij}$ sont des copules complètement arbitraires, E est l'ensemble des arrêtes $\{ij\}$ d'un graphe arbitraire $(\{1, \ldots, d\}, E)$ et n_k est le nombre de voisins du sommet k dans ce graphe. Par exemple, avec $E = \{\{12\}, \{24\}, \{23\}, \{35\}\}$ comme sur la Figure 1 de l'article, notre copule s'écrit

$$C(u_1, u_2, u_3, u_4, u_5) = \tilde{C}_{12}(u_1, u_2^{1/3})\tilde{C}_{24}(u_2^{1/3}, u_4)\tilde{C}_{23}(u_2^{1/3}, u_3^{1/2})\tilde{C}_{35}(u_3^{1/2}, u_5).$$

La marginale bivariée de (4.4) associée aux variables indexées par k et l s'écrit

$$C_{kl}(u_k, u_l) = \begin{cases} u_k^{(n_k-1)/n_k} u_l^{(n_l-1)/n_l} \tilde{C}_{kl}(u_k^{1/n_k}, u_l^{1/n_l}) & \text{si } \{kl\} \in E, \\ u_k u_l & \text{sinon.} \end{cases}$$

A la vue de la formule du dessus, il apparait immédiatement que deux variables ne sont pas connectées dans le graphe associé au modèle sont marginalement indépendantes. Dans l'article, nous montrons par ailleurs que les coefficients de dépendance des marginales bivariées sont bornés, et les bornes sont d'autant plus sévères que le graphe est connecté. L'utilisateur d'un tel modèle doit alors faire face à un compromis. D'un côté, il souhaite connecter les variables afin qu'elles ne soient pas marginalement indépendantes, et de l'autre, plus ces variables sont connectées, plus les dépendances modélisées sont faibles. L'inférence de la copule (4.4) peut se faire par maximum de vraisemblance, bien que son expression s'écrive sous la forme d'un produit, grâce à l'utilisation d'un algorithme de *message-passing* [45] qui tire profit de la structure de graphe associée à la copule. Nous donnons l'intuition de cet algorithme récent dans l'appendice de notre article, et nous l'avons implémenté dans notre cas. Ce paquet `R` est disponible librement sur le serveur du CRAN [87]. L'article présenté ci-dessous a été soumis pour publication, et est disponible à l'adresse `http://hal.archives-ouvertes.fr/hal-00910775`.

A class of multivariate copulas based on products of bivariate copulas

Gildas Mazo (gildas.mazo@free.fr), Stéphane Girard and Florence Forbes

Inria and Laboratoire Jean Kuntzmann, Grenoble, France

Abstract

Copulas are a useful tool to model multivariate distributions. While there exist various families of bivariate copulas, much fewer has been done when the dimension is higher. In this paper we propose a class of multivariate copulas based on products of transformed bivariate copulas. No constraints on the parameters refrain the applicability of the proposed class. Furthermore the analytical forms of the copulas within this class allow to naturally associate a graphical structure which helps to visualize the dependencies and to compute the likelihood efficiently even in high dimension.

Keywords: maximum-likelihood inference, graphical models, message-passing algorithm, multivariate, copula.

1 Introduction

The modelling of random multivariate events is a central problem in various scientific domains and the construction of multivariate distributions able to properly model the variables at play is challenging. A useful tool to deal with this problem is the concept of copulas. Let $(X_1, \ldots, X_d)$ be a random vector with distribution function F. Let F_i be the (continuous) marginal distribution function of X_i, $i = 1, \ldots, d$. By Sklar's Theorem [17], there exists a unique function C such that

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{1}$$

This function C is called the copula of F and is the d-dimensional distribution function of the random vector $(F_1(X_1), \ldots, F_d(X_d))$. For a general account on copulas, see, e.g. [16]. Copulas are interesting since they permit to impose a dependence structure on pre-determined marginal distributions.

While there exist many copulas in the bivariate case, it is less clear how to construct copulas in higher dimension. In the presence of non-Gaussianity and/or tail dependence, various constructions have been adopted, such as, for instance, Archimedean copulas [9], Vines [1] or elliptical copulas [5]. Because Archimedean copulas possess only a few parameters, they lack flexibility in high-dimension. Vines, on the opposite, achieve greater flexibility but at the price of increased complexity in the modeling process. The use of elliptical copulas

goes together with assuming a similar dependence pattern among all pairs of variables. This may be undesirable in applications. Moreover, they have in general as many as $O(d^2)$ parameters and it is difficult to carry out maximum likelihood inference [3].

Another approach [14] aims at constructing a multivariate copula as a product of transformed bivariate copulas. This approach possesses several advantages. A probabilistic interpretation is available and thus the generation of random vectors is straightforward. The resulting copula is explicit, leading to explicit bounds on dependence coefficients of the bivariate marginals. The class of copulas which can be constructed from this approach is large and can cover a wide range of dependencies. Finally the analysis of extreme values can be performed by constructing extreme-value copulas.

However, although many copulas with different features can be built, the use of this approach for practical applications remains challenging. Indeed, two pitfalls render inference difficult: first, they are constraints on the parameters, and second, the product form complicates the computation of the density – hence, of the potential likelihood – even numerically.

The main contribution of this paper is to revisit the product of transformed copulas in order to propose a new multivariate copula model of practical interest. First, there are no constraints on the parameters anymore. Moreover, a graphical structure associated to the copulas within this class permits to visualize the dependencies and to efficiently compute the likelihood, even in high dimension.

The rest of this paper is organized as follows. Section 2 reviews the product of transformed copulas and important properties such as random generation and the ability to construct extreme-value models. Section 3 presents the new copula model and enlightens the link with the product of transformed copulas. Section 4 discusses the dependence properties of bivariate marginals of the proposed class by providing bounds on some of the most popular dependence coefficients such as Spearman's rho, Kendall's tau, and tail dependence coefficients. In Section 5, we apply the proposed copula model to a simulated and a real dataset. The appendix gathers the proofs of this paper.

2 Product of transformed copulas

It is easily seen that a product of copulas is not a copula in general. Nonetheless the next theorem due to Liebscher [14] shows that, up to marginal transformations, a product of copulas can lead to a well defined copula.

Theorem 1. *Assume $\tilde{C}_1, \ldots, \tilde{C}_K : [0,1]^d \to [0,1]$ are copulas. Let $g_{ei} : [0,1] \to [0,1]$ for $e = 1, \ldots, K$, $i = 1, \ldots, d$ be functions with the property that each of them is strictly increasing or is identically equal to 1. Suppose that $\prod_{e=1}^{K} g_{ei}(v) = v$ for $v \in [0,1], i = 1, \ldots, d$, and $\lim_{v \to 0} g_{ei}(v) = 0$ for $e = 1, \ldots, K$, $i = 1, \ldots, d$. Then*

$$C(u_1, \ldots, u_d) = \prod_{e=1}^{K} \tilde{C}_e \left(g_{e1}(u_1), \ldots, g_{ed}(u_d) \right) \tag{2}$$

is also a copula.

The probabilistic interpretation of (2) is as follows. Let

$$(U_1^{(1)}, \dots, U_d^{(1)}), \dots, (U_1^{(K)}, \dots, U_d^{(K)})$$

be K independent random vectors having distribution function $\tilde{C}_1, \dots, \tilde{C}_K$ respectively. Let g_{ei}, $e = 1, \dots, K$, $i = 1, \dots, d$ be as in Theorem 1 and define $g_{ei}^{-1}(v) := 0$ for $v \leq g_{ei}(0)$ and $J_i = \{e \in \{1, \dots, K\} : g_{ei} \neq 1\}$. Then C is the joint distribution function of the random vector

$$\left(\max_{e \in J_1} g_{e1}^{-1}(U_1^{(e)}), \dots, \max_{e \in J_d} g_{ed}^{-1}(U_d^{(e)}) \right). \tag{3}$$

If there exists a random generation procedure for $\tilde{C}_e$, $e = 1, \dots, K$ then thanks to (3) a random generation procedure for C can be derived as well.

The statistical analysis of extreme values should theoretically be carried out with the help of extreme-value copulas [8]. Recall that a copula $C_\#$ is an extreme-value copula if there exists a copula C such that

$$C_\#(u_1, \dots, u_d) = \lim_{n \uparrow \infty} C^n(u_1^{1/n}, \dots, u_d^{1/n}), \tag{4}$$

for every $(u_1, \dots, u_d) \in [0,1]^d$. A copula $C_\#$ is said to be max-stable if for every integer $n \geq 1$ and every $(u_1, \dots, u_d) \in [0,1]^d$

$$C_\#^n(u_1^{1/n}, \dots, u_d^{1/n}) = C_\#(u_1, \dots, u_d).$$

Extreme-value copulas correspond exactly to max-stable copulas [8]. Theorem 1 can be used to construct extreme-value copulas as shown in the next proposition due to [4].

Proposition 1. *In (2), let $g_{ei}(v) = v^{\theta_{ei}}$, $v \in [0,1]$ with $\theta_{ei} \in [0,1]$ and $\sum_{e=1}^K \theta_{ei} = 1$ for $i = 1, \dots, d$. If $\tilde{C}_e$, $e = 1, \dots, K$ is max-stable then so is C.*

Out of the context of extreme values, applications of Theorem 1 can be found, for instance, in the analysis of directional dependence [13] ($K = d = 2$), finance [2] ($K = d = 2$) and hydrology [4] ($K = 2$, $d = 3$).

We are not aware of applications of Theorem 1 in practice when $K > 2$ or $d > 3$. As pointed out in the introduction, the product form (2) renders the density $\frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d}$, hence the likelihood, complicated to compute even numerically. Furthermore, the constraints $\prod_{e=1}^K g_{ei}(v) = v$, $v \in [0,1]$, $i = 1, \dots, d$ in Theorem 1 are not easy to deal with in practice.

The next section aims at overcoming these drawbacks.

3 Product of transformed copulas revisited

The product over $e \in \{1, \dots, K\}$ in (2) can be taken over $e \in E$, where E is an arbitrary finite set, yielding

$$C(u_1, \dots, u_d) = \prod_{e \in E} \tilde{C}_e \left(g_{e1}(u_1), \dots, g_{ed}(u_d) \right). \tag{5}$$

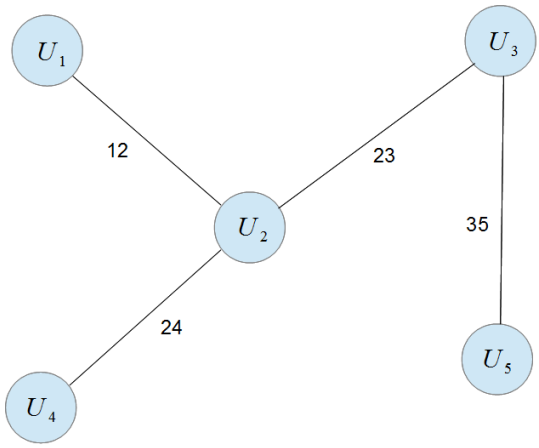

Figure 1: Graphical representation of the set $E = \{\{12\}, \{24\}, \{23\}, \{35\}\}$. $N(1) = \{\{12\}\}$, $N(2) = \{\{12\}, \{23\}, \{24\}\}$, $N(3) = \{\{23\}, \{35\}\}$, $N(4) = \{\{24\}\}$ and $N(5) = \{\{35\}\}$.

In particular, an element $e \in E$ can represent a pair of the variables at play. More precisely, let $U_1, \ldots, U_d$ be d standard uniform random variables. Denote by $\{ij\}$ the index of the pair (U_i, U_j) and let $E \subset \{\{ij\} : i, j = 1, \ldots, d,\ j > i\}$ be a subset of the set of the pair indices. The cardinal of E, denoted by $|E|$, is less or equal to $d(d-1)/2$. The pair index $e \in E$ is said to contain the variable index i if $e = \{ik\}$ for $k \neq i$. Let us introduce $N(i) = \{e \in E : e \text{ contains } i\}$. $N(i)$ is called the set of neighbors of i and has cardinal $|N(i)| = n_i$. It is natural to associate a graph to the set E as follows: an element $e = \{ij\} \in E$ is an edge linking U_i and U_j in the graph whose nodes are the variables $U_1, \ldots, U_d$. The example $E = \{\{12\}, \{24\}, \{23\}, \{35\}\}$ is illustrated in Figure 1. For $u = (u_1, \ldots, u_d) \in [0,1]^d$, consider the functional

$$C(u_1, \ldots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{ij}(u_i^{1/n_i}, u_j^{1/n_j}), \tag{6}$$

where the $\tilde{C}_{ij}$'s are bivariate copulas. Keeping in mind the graphical representation, C in (6) is a product over the edges. For instance, when $E = \{\{12\}, \{24\}, \{23\}, \{35\}\}$ as in Figure 1, (6) writes

$$C(u_1, u_2, u_3, u_4, u_5) = \tilde{C}_{12}(u_1, u_2^{1/3})\tilde{C}_{24}(u_2^{1/3}, u_4)\tilde{C}_{23}(u_2^{1/3}, u_3^{1/2})\tilde{C}_{35}(u_3^{1/2}, u_5).$$

In the following, (6) is referred to as the Product of Bivariate Copulas (PBC) copula, or PBC model. The next theorem establishes that (6) is a copula and makes the link with Theorem 1.

Theorem 2. *If in* (5)*:*

(i) for $e = \{ij\} \in E$, $\tilde{C}_e$ takes exactly two arguments non identically equal to one, namely, g_{ei} and g_{ej}, and

(ii) for $i = 1, \ldots, d$ and $e \in N(i)$, g_{ei} does not depend on e;

then the only copula which can be constructed from (5) *is the PBC model* (6)*, where* $\tilde{C}_{ij}$ *is defined by*

$$\tilde{C}_{ij}(u,v) = \tilde{C}_{\{ij\}}(1,\ldots,1,u,1,\ldots,1,v,1,\ldots,1), \quad (u,v) \in [0,1]^2,$$

and where in $(1,\ldots,1,u,1,\ldots,1,v,1,\ldots,1)$*,* u *and* v *are at the* i*-th and* j*-th positions respectively.*

Condition (i) in Theorem 2 simply means that only bivariate copulas are allowed in the construction. The simplification (ii) achieves two goals: first to reduce the number of parameters (an important feature in high-dimension), and second to intrinsically satisfy the constraints $\prod_{e\in E} g_{ei}(v) = v$, $v \in [0,1]$, $i = 1,\ldots,d$ in the assumptions of Theorem 1. If assumption (ii) in Theorem 2 was not made, one could take $g_{ei}(v) = v^{\theta_{ei}}, e \in E$, $i = 1,\ldots,d$, $\theta_{ei} \in [0,1]$ with the constraints

$$\sum_{e\in N(i)} \theta_{ei} = \sum_{k:\{ki\}\in E} \theta_{ki,i} = 1, \quad i = 1,\ldots,d. \tag{7}$$

These constraints would be difficult to handle in practice, and, furthermore, the number of parameters would increase quadratically with the dimension. Indeed, one would have $(|E|-1)d$ parameters θ_{ei} plus an additional number $|E|$ of parameters for each copula $\tilde{C}_e$. If the graph associated to E is a tree, for instance, then $|E| = d-1$, yielding $O(d^2)$ parameters. As a comparison, in the PBC model (6), there are no constraints and only $O(d)$ parameters in total.

From (1), the PBC copula (6) is associated to a distribution function F with continuous marginals F_i, $i = 1,\ldots,d$, such that

$$F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots,F_d(x_d)), \quad (x_1,\ldots,x_d) \in \mathbb{R}^d. \tag{8}$$

By substituting (6) into (8), it is easy to see that F writes

$$F(x_1,\ldots,x_d) = \prod_{\{ij\}\in E} F_{ij}(x_i,x_j), \quad (x_1,\ldots,x_d) \in \mathbb{R}^d, \tag{9}$$

where F_{ij}, $\{ij\} \in E$, is a bivariate distribution function such that the first (respectively the second) marginal $F_{ij,1}$ (respectively $F_{ij,2}$) only depends on i (respectively j). It is interesting to note that the converse is also true as stated in the following proposition.

Proposition 2. *The distribution function corresponding to the PBC copula* (6) *writes as* F *in* (9)*. Conversely, the copula corresponding to the distribution function* F *in* (9) *writes as the PBC copula* (6)*.*

4 Dependence properties and max-stability

Let C be the PBC copula (6). First the dependence properties of a pair (U_k, U_l) whose copula is the bivariate copula $C_{kl}(u_k,u_l) = C(1,\ldots,1,u_k,1,\ldots,1,u_l,1,\ldots,1)$ are studied. The conditions under which the PBC model (6) is an extreme-value copula are given afterwards.

Proposition 3. *The bivariate marginal* C_{kl} *is given by*

$$C_{kl}(u_k,u_l) = \begin{cases} u_k^{(n_k-1)/n_k} u_l^{(n_l-1)/n_l} \tilde{C}_{kl}(u_k^{1/n_k}, u_l^{1/n_l}) & \text{if } \{kl\} \in E, \\ u_k u_l & \text{otherwise.} \end{cases} \tag{10}$$

Example 1. *If in* (10) $\tilde{C}_{kl}$ *is a Marshall-Olkin copula (see for instance [16], p.53) with parameters* $0 \leq \alpha, \beta \leq 1$ *(denoted by MO*(α, β)*), that is,*

$$\tilde{C}_{kl}(u_k, u_l) = \min(u_k^{1-\alpha} u_l, u_l^{1-\beta} u_k),$$

then C_{kl} *is MO*$(\alpha/n_k, \beta/n_l)$*. If* $\alpha = \beta$ *then* $\tilde{C}_{kl}$ *is a Cuadras-Augé copula and* C_{kl} *is MO*$(\alpha/n_k, \alpha/n_l)$*. If* $\alpha = \beta = 0$ *then both* $\tilde{C}_{kl}$ *and* C_{kl} *are the independence copula. If* $\alpha = \beta = 1$ *then* $\tilde{C}_{kl}$ *is the Fréchet upper bound copula and* C_{kl} *is MO*$(1/n_k, 1/n_l)$*.*

Remark 1. *If in* (10) *one puts* $\kappa = 1/n_k$ *and* $\lambda = 1/n_l$*, then the copulas take the form* $C_{kl}(u_k, u_l) = u_k^{1-\kappa} u_l^{1-\lambda} \tilde{C}_{kl}(u_k^{\kappa}, u_l^{\lambda})$*. This class of copulas, sometimes referred to as Khoudraji copulas, was proposed in [6] Proposition 2.*

Let (U, V) be a random vector with copula C. The dependence between U and V is positive if, roughly speaking, U and V tend to be large or small together. Below are recalled a few definitions of statistical concepts about positive dependence. The copula C has the TP2 (totally positive of order 2) property if and only if

$$C(u_1, u_2)C(v_1, v_2) \geq C(u_1, v_2)C(v_1, u_2), \text{ for all } u_1 < v_1 \text{ and } u_2 < v_2. \quad (11)$$

Also, C is said to be PQD (positive quadrant dependent) if $C(u, v) \geq uv$ for all $(u, v) \in [0, 1]^2$. The random variable V is said to be LTD (left tail decreasing) in U if for all $v \in [0, 1]$, the function $u \mapsto P(V \leq v | U \leq u)$ is decreasing in u. The dependence between U and V can be quantified through dependence measures such as the Kendall's tau or Spearman's rho respectively given by

$$\tau = 4 \int_{[0,1]^2} C(u, v)\, dC(u, v) - 1, \quad (12)$$

$$\rho = 12 \int_{[0,1]^2} C(u, v)\, du\, dv - 3. \quad (13)$$

The dependence in the upper and lower tails can be respectively measured with

$$\lambda^{(U)} = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \in [0, 1], \quad \lambda^{(L)} = \lim_{u \downarrow 0} \frac{C(u, u)}{u} \in [0, 1].$$

See [16] and [12] for further details about these concepts. Let us denote by τ_{kl}, ρ_{kl}, $\lambda_{kl}^{(U)}$ and $\lambda_{kl}^{(L)}$ the Kendall's tau, Spearman's rho, upper tail dependence coefficient and lower tail dependence coefficient of the copula C_{kl} in (10) respectively. As shown in Proposition 3, C_{kl} is a bivariate marginal of the PBC copula (6) and one may apply the results of [14] to obtain the following.

Proposition 4. *If in* (10) $\tilde{C}_{kl}$ *is TP2, LTD or PQD then* C_{kl} *is also TP2, LTD or PQD respectively.*

Explicit bounds in terms of the number of neighbors for the dependence coefficients of PBC bivariate marginals are given in the next proposition. The behavior of (10) when the number of neighbors tends to infinity is also studied.

Proposition 5. *We have* $\lambda_{kl}^{(L)} = 0$ *and* $\lambda_{kl}^{(U)} \leq \min(1/n_k, 1/n_l)$. *The lower and upper bounds for* ρ_{kl} *and* τ_{kl} *are respectively given by*

$$a_\rho(n_k, n_l) \leq \rho_{kl} \leq b_\rho(n_k, n_l),$$
$$a_\tau(n_k, n_l) \leq \tau_{kl} \leq b_\tau(n_k, n_l),$$

with

$$\begin{aligned}
a_\rho(n_k, n_l) &= \frac{6\beta(2n_k - 1, 2n_l - 1)n_k n_l}{(2n_k + 2n_l - 1)(n_k + n_l - 1)} - \frac{3}{(2n_k - 1)(2n_l - 1)}, \\
b_\rho(n_k, n_l) &= \frac{3}{2n_k + 2n_l - 1}, \\
a_\tau(n_k, n_l) &= \frac{\beta(2n_l - 1, 2n_k - 1)}{n_k + n_l - 1} - \frac{2}{(2n_k - 1)(2n_l - 1)}, \\
b_\tau(n_k, n_l) &= \frac{1}{n_k + n_l - 1},
\end{aligned}$$

where β *denotes the* β*-function,* $\beta(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$. *Furthermore, as* $\max(n_k, n_l) \to \infty$, *we have* $C_{kl}(u, v) \to uv$ *for all* $(u, v) \in [0, 1]^2$.

The above results show that we are facing a tradeoff: on the one hand, the larger the cardinal of E (or the more connected the graph associated to E), the less the pairs in E are able to model strong dependencies. On the other hand, the smaller the cardinal of E, the more there are independent pairs (since there are less pairs in E). To illustrate Proposition 5, numerical values of the bounds are computed in Table 1 for different numbers of neighbors (n_k, n_l).

coefficient	ρ_{kl}	τ_{kl}	λ_{kl}
(n_k, n_l)			
$(1, 2)$	$[-0.60, 0.60]$	$[-0.50, 0.50]$	$[0.00, 0.50]$
$(2, 2)$	$[-0.30, 0.43]$	$[-0.21, 0.33]$	$[0.00, 0.50]$
$(1, 3)$	$[-0.43, 0.43]$	$[-0.33, 0.33]$	$[0.00, 0.33]$
$(2, 3)$	$[-0.19, 0.33]$	$[-0.13, 0.25]$	$[0.00, 0.33]$
$(3, 3)$	$[-0.12, 0.27]$	$[-0.08, 0.20]$	$[0.00, 0.33]$

Table 1: Lower and upper bounds [lower, upper] for Spearman'rho ρ_{kl}, Kendall's tau τ_{kl} and upper tail dependence coefficient λ_{kl} depending on the number of neighbors (n_k, n_l).

Finally, it is easy to construct extreme-value copulas belonging to the PBC class (6). Indeed, the following result follows from Proposition 1.

Proposition 6. *If in the PBC copula* (6), $\tilde{C}_{ij}$ *is an extreme-value copula for* $\{ij\} \in E$, *then* C *is also an extreme-value copula.*

All copulas C_{kl} in Example 1 are max-stable since Marshall-Olkin copulas are max-stable. Thus the associated PBC is an extreme-value copula. If $\tilde{C}_{kl}$ in (10) is a (max-stable) Gumbel copula, that is,

$$\tilde{C}_{kl}(u_k, u_l) = \exp - \left[(-\log u_k)^\theta + (-\log u_l)^\theta\right]^{1/\theta}, \quad \theta \geq 1, \tag{14}$$

then C_{kl} is also max-stable, hence, the PBC is an extreme-value copula.

5 Numerical applications to simulated and real datasets

In this section, PBC copulas models are applied to simulated and real datasets. The methods used to simulate and infer the copulas are presented in Section 5.1. The considered families for the bivariate copulas $\tilde{C}_{ij}$ in (6) are the following: the Ali-Mikhail-Haq (AMH), Farlie-Gumbel-Morgenstern (FGM), Frank, Gumbel, and Joe families. See [16] or [12] for details about these families. The corresponding PBC copula models (6) are therefore referred to as PBC AMH, PBC FGM, PBC Frank, PBC Gumbel and PBC Joe respectively. In Section 5.2, the two inference procedures presented in Section 5.1 are compared. Section 5.3 applies PBC models to an hydrological dataset.

5.1 Computational aspects

In this section, we assume that the copulas $\tilde{C}_{ij}$ of the PBC model (6) depend on parameters θ_{ij}'s and that we are given a sample of i.i.d data vectors from (6).

Data simulation from a PBC copula is straightforward thanks to the probabilistic interpretation given in (3). The generation procedure is given below.

- For all $\{ij\} \in E$, generate $(U_i^{(ij)}, U_j^{(ij)}) \sim \tilde{C}_{ij}$.
- For all $i = 1, \ldots, d$, compute $U_i = \max_{k \in \{1,\ldots,d\}:\{ki\} \in E} \left\{ \left(U_i^{(ki)} \right)^{n_i} \right\}$.

The resulting vector $(U_1, \ldots, U_d)$ has distribution (6).

The inference of PBC copulas can be performed by maximum-likelihood based methods. As it is well known, the estimators resulting from these methods have the advantage to be consistent and asymptotically unbiased under mild conditions. Properly scaled, their asymptotic distribution is Gaussian and confidence intervals or tests can be derived.

The first considered approach is the pairwise maximum-likelihood method [15]. This approach consists in maximizing the sum of the likelihoods corresponding to all the pairs of variables. In our case, it simplifies to maximizing $|E|$ univariate functions independently. However, unlike the full joint maximum likelihood estimator, the pairwise maximum-likelihood estimator is not guaranteed to be efficient.

The second considered approach is the standard full joint maximum-likelihood method. Indeed, it is possible to compute the full joint likelihood of a PBC copula when the graph associated to it is a tree thanks to a message-passing algorithm [11]. A brief explanation of how this algorithm works is given in Appendix B. The reader is referred to [11] for the complete algorithm and [10] for a detailed explanation. We implemented this algorithm in the `R` package `PBC` [18].

5.2 A simulation experiment to compare pairwise likelihood and full joint likelihood approaches

We generated 100 datasets of dimension $d = 9$ and size $n = 500$ according to a PBC copula whose tree graph is given in Figure 2. The amount of time required

to maximize the full joint likelihood for one dataset replication was 36, 21, 18, 21, and 21 seconds for PBC AMH, PBC FGM, PBC Frank, PBC Gumbel, and PBC Joe respectively with a 8 GiB memory and 3.20 GHz processor computer. The $d-1=8$ coordinates θ_i of the parameter vectors were chosen to be regularly spaced within the intervals $[-0.9, 0.9], [-0.9, 0.9], [-9, 11], [2, 20]$ and $[1, 20]$ respectively.

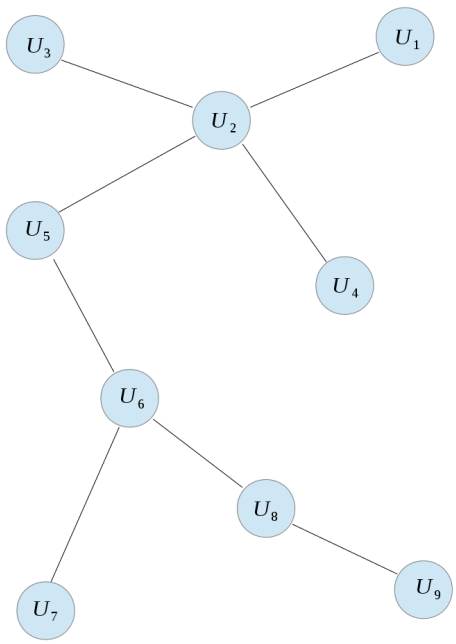

Figure 2: Tree graph associated to the simulated PBC copulas.

The following criteria were calculated in order to assess the results of the experiment. The variance ratio (VR) is defined as

$$VR = \frac{\sum_{e=1}^{d-1} \widehat{\mathrm{Var}}\left(\hat{\theta}_e^{FULL}\right)}{\sum_{e=1}^{d-1} \widehat{\mathrm{Var}}\left(\hat{\theta}_e^{PW}\right)},$$

where $\hat{\theta}_e^{FULL}$, $\hat{\theta}_e^{PW}$ is the coordinate estimated by maximization of the full joint likelihood, pairwise likelihood, respectively, and where $\widehat{\mathrm{Var}}$ is the empirical variance on the replications. For each dataset replication, the mean absolute error on Spearman's rho ρ (MAE_ρ) and the Kendall's tau τ (MAE_τ) is defined as

$$\mathrm{MAE}_\rho = \frac{1}{d-1}\sum_{e=1}^{d-1} |\rho(\theta_e) - \rho(\hat{\theta}_e^{FULL})|, \quad \mathrm{MAE}_\tau = \frac{1}{d-1}\sum_{e=1}^{d-1} |\tau(\theta_e) - \tau(\hat{\theta}_e^{FULL})|.$$

The MAEs were averaged over the 100 replications to get a single value per model.

Copula	VR	MAE_ρ	MAE_τ
PBC AMH	0.96	0.03	0.02
PBC FGM	0.98	0.03	0.02
PBC Frank	0.79	0.02	0.01
PBC Gumbel	0.68	0.00	0.00
PBC Joe	0.71	0.00	0.00

Table 2: Variance ratio (VR) and mean absolute errors (MAEs) for each of the tested PBC models. The MAEs were averaged over the dataset replications.

The results are reported in Table 2. It appears that for PBC AMH and PBC FGM the precision was not improved by maximizing the full joint likelihood relative to the pairwise approach: the VR for those models are close to 1. For the Frank, Gumbel and Joe families, however, the variance decreases by at least 20% in average. These families, in contrast to the AMH and FGM families, are comprehensive, meaning that they include the lower and upper bounds for copulas. The MAEs are quite low for all the models. This indicates that the maximization of the full joint likelihood with the message-passing algorithm of Appendix B performs well.

5.3 Application to an hydrological dataset

In this section, PBC copula models are applied to an hydrological dataset consisting of $d = 3$ stations and $n = 36$ observations of flow rate annual maxima. The sites are located on three french rivers at the following places: La Celle-en-Morvan on the river la Selle (S), Rigny-sur-Arroux on l'Arroux (A), and Isclades-et-Rieutord on la Loire (L). These rivers are embedded in the sense that Selle flows into Arroux which flows into Loire. Thus, the graph is naturally set up as

$$\text{S} - \text{A} - \text{L}.$$

The same models as in Section 5.2 were tested, that is, PBC AMH, PBC FGM, PBC Frank, PBC Gumbel, and PBC Joe. The Gumbel copula was also considered here as a benchmark. This family is standard in hydrology for fitting trivariate distributions [19]. The estimation of the parameters was performed by maximization of the full joint likelihood, as explained in Section 5.1. In order to assess the fit of the models, the empirical Spearman's rho and Kendall's tau coefficient estimates were compared to their counterpart under the models. Since the number of parameters is the same for all models, the likelihood values for the different models were also compared. The results are reported in Table 3.

One can observe that PBC AMH and PBC FGM perform very poorly compared to the other models. This was expected since the AMH and FGM families are not comprehensive, roughly meaning that they do not allow much dependence (see, e.g., [16]). The standard Gumbel copula performs poorly too, with one of the smallest log-likelihood values. One can also see that, since it has a single parameter, the dependence coefficients between the different pairs are equal

	$\rho_{\mathrm{S,A}}$	$\rho_{\mathrm{A,L}}$	$\rho_{\mathrm{S,L}}$	$\tau_{\mathrm{S,A}}$	$\tau_{\mathrm{A,L}}$	$\tau_{\mathrm{S,L}}$	log-likelihood
empirical data	0.70	0.30	0.13	0.5	0.21	0.08	
PBC AMH	0.25	0.25	0	0.17	0.17	0	7.05
PBC FGM	0.20	0.20	0	0.13	0.13	0	5.46
PBC Frank	0.44	0.30	0	0.30	0.21	0	9.09
PBC Gumbel	0.43	0.31	0	0.30	0.21	0	9.20
PBC Joe	0.41	0.27	0	0.29	0.18	0	8.05
Gumbel	0.39	0.39	0.39	0.27	0.27	0.27	6.16

Table 3: Optimized log-likelihood and pairwise dependence coefficients for the empirical data and the tested PBC copulas. The symbol ρ and τ stand for Spearman's rho and Kendall's tau respectively. For instance, $\rho_{\mathrm{S,A}}$ is Spearman's rho coefficient between the variables S and A.

to each other. The PBC copulas with comprehensive families present a much better fit. The dependence coefficient with the smallest value, that of the pair (A,L), is very well approximated by the PBC Frank, PBC Gumbel, and PBC Joe. In particular, PBC Frank and PBC Gumbel both provide, for instance, a Kendall's tau of 0.21, which is the same as the empirical value. Also, these copulas possess the highest log-likelihood values, 9.20 and 9.09, a step above the third highest, 8.05. The dependence coefficient of the pair (S,A), which presents more dependence (0.7 for Spearman's rho and 0.5 for the Kendall's tau) is underestimated. Although the theoretical upper bound for the Kendall's tau is 0.5 (see Table 1), the closest copulas are PBC Frank and PBC Gumbel with a Kendall's tau of 0.3 for both. Given that the third pair (S,L) presents low values for Spearman's rho (0.13) and the Kendall's tau (0.08), its distribution might by approximated by the independence copula, as PBC models do. The test for independence proposed in [7], implemented in the `R` package `copula`, gave a p-value of 0.41. The Gumbel copula, instead, seems to overestimate the dependence in the third pair (S,L).

6 Discussion

In this paper, we have constructed a class of multivariate copulas, called PBC copulas, based on bivariate copulas. Therefore, this novel class benefits from the many bivariate families existing in the literature. No constraints on the parameters refrain the applicability of the PBC class and a natural graph structure helps to visualize the dependencies between the variables. Full joint multivariate inference can be performed, and shown to perform well, with the message-passing algorithm presented in the appendix. However, PBC copula models still suffer from weaknesses. First, the more there are edges in the graph, the more the bounds on the dependence coefficients are restrictive. Second, it was shown numerically that dependence coefficients of high magnitude were prone to be underestimated. In view of these remarks, it may be advisable to keep the number of neighbors in the graph associated to PBC models as low as possible, and to be careful with highly dependent data.

Acknowledgment. The authors thank "Banque HYDRO du Ministère de l'Écologie, du Développement durable et de l'Énergie" for providing the data and Benjamin Renard for fruitful dicussions about statistical issues in hydrological science.

Appendix

A Proofs

Proof of Theorem 2

From Theorem 1, it is straightforward to see that (6) is a copula. Let us now prove that (6) is the only copula arising from (5). Condition (i) implies that if $e \notin N(i)$ then $g_{ei} = 1$, $i = 1, \ldots, d$. Hence, the constraint over the functions reduces to $\prod_{e \in N(i)} g_{ei}(v) = v$, $v \in [0,1]$. In view of condition (ii), one has $g_{ei} = g_i$ for $e \in N(i)$, hence $(g_i(v))^{n_i} = v$. Therefore

$$g_{ei}(v) = \begin{cases} v^{1/n_i} & \text{if } e \in N(i), \\ 1 & \text{otherwise.} \end{cases}$$

To conclude it suffices to rewrite the product in (5) as

$$\prod_{e \in E} \tilde{C}_e(1, \ldots, 1, u_i^{1/n_i}, 1, \ldots, 1, u_j^{1/n_j}, 1, \ldots, 1) = \prod_{\{ij\} \in E} \tilde{C}_{ij}(u_i^{1/n_i}, u_j^{1/n_j})$$

which corresponds to (6).

Proof of Proposition 2

Let us first prove that (9) is the distribution function of (6). By (1) we have

$$\begin{aligned} F(x_1, \ldots, x_d) =& C(F_1(x_1), \ldots, F_d(x_d)) \\ =& \prod_{\{ij\} \in E} \tilde{C}_{ij}(F_i(x_i)^{1/n_i}, F_j(x_j)^{1/n_j}) \\ =:& \prod_{\{ij\} \in E} \Phi_{ij}(x_i, x_j). \end{aligned}$$

The first margin of Φ_{ij} is given by $\Phi_{ij,1}(x) = \Phi_{ij}(x, \infty) = F_i(x_i)^{1/n_i}$ which depends only on i. The same holds for the second margin $\Phi_{ij,2}$.

Let us prove that (6) is the copula associated to (9). Let $\Phi_{ij,k}$, $k = 1, 2$ be the k-th univariate marginal of $\Phi_{ij}, \{ij\} \in E$. The copula associated to F is given by

$$C_F(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) = \prod_{\{ij\} \in E} \Phi_{ij}\left(F_i^{-1}(u_i), F_j^{-1}(u_j)\right).$$

Let $\tilde{C}_{ij}$ be the copula associated to Φ_{ij}. We have

$$\Phi_{ij}(x_i, x_j) = \tilde{C}_{ij}\left(\Phi_{ij,1}(x_i), \Phi_{ij,2}(x_j)\right)$$

so that $\Phi_{ij}\left(F_i^{-1}(u_i), F_j^{-1}(u_j)\right) = \tilde{C}_{ij}\left(\Phi_{ij,1} \circ F_i^{-1}(u_i), \Phi_{ij,2} \circ F_j^{-1}(u_j)\right)$, and

$$C_F(u_1, \ldots, u_d) = \prod_{\{ij\} \in E} \tilde{C}_{ij}\left(\Phi_{ij,1} \circ F_i^{-1}(u_i), \Phi_{ij,2} \circ F_j^{-1}(u_j)\right). \tag{15}$$

Moreover, since $\tilde{C}_F$ is a copula, it follows

$$\begin{aligned}
u_k =& C_F(1,\dots,1,u_k,1,\dots,1)\\
=& \prod_{j>k:\{kj\}\in E} \tilde{C}_{kj}\left(\Phi_{kj,1}\circ F_k^{-1}(u_k),1\right) \prod_{j<k:\{jk\}\in E} \tilde{C}_{jk}\left(1,\Phi_{jk,2}\circ F_k^{-1}(u_k)\right)\\
=& \prod_{j:\{kj\}\in E} \Phi_{kj,1}\circ F_k^{-1}(u_k).
\end{aligned}$$

Now by assumption $\Phi_{kj,1}=\Phi_{jk,2}=\Phi_k$ only depends on k and therefore $u_k^{1/n_k}=\Phi_k\circ F_k^{-1}(u_k)$ which implies $\Phi_k(z)=F_k(z)^{1/n_k}$, $z\in\mathbb{R}$. By pluging Φ_k into (15) the result follows.

Proof of Proposition 3

If $\{kl\}\in E$, then

$$\begin{aligned}
C_{kl}(u_k,u_l) =& C(1,\dots,1,u_k,1,\dots,1,u_l,1,\dots,1)\\
=& \left(\prod_{e\in N(k)\setminus\{kl\}} \tilde{C}_e(u_k^{1/n_k},1)\right)\left(\prod_{e\in N(l)\setminus\{kl\}} \tilde{C}_e(u_l^{1/n_l},1)\right)\times\\
& \tilde{C}_{kl}(u_k^{1/n_k},u_l^{1/n_l})\\
=& u_k^{(n_k-1)/n_k}u_l^{(n_l-1)/n_l}\tilde{C}_{kl}(u_k^{1/n_k},u_l^{1/n_l}).
\end{aligned}$$

Otherwise,

$$\begin{aligned}
C_{kl}(u_k,u_l) &= \left(\prod_{e\in N(k)} \tilde{C}_e(u_k^{1/n_k},1)\right)\left(\prod_{e\in N(l)} \tilde{C}_e(u_l^{1/n_l},1)\right)\\
&= u_k^{n_k/n_k}u_l^{n_l/n_l}\\
&= u_k u_l.
\end{aligned}$$

Proof of Proposition 5

The Fréchet-Hoeffding bounds for copulas (see e.g. [16], p. 11) applied to $\tilde{C}_{kl}$ in (10) yield

$$W_{kl}(u_k,u_l)\le C_{kl}(u_k,u_l)\le M_{kl}(u_k,u_l), \tag{16}$$

where

$$\begin{aligned}
W_{kl}(u_k,u_l) &= u_k^{1-1/n_k}u_l^{1-1/n_l}\max(u_k^{1/n_k}+u_l^{1/n_l}-1,0),\\
M_{kl}(u_k,u_l) &= u_k^{1-1/n_k}u_l^{1-1/n_l}\min(u_k^{1/n_k},u_l^{1/n_l}).
\end{aligned}$$

We have $M_{kl}(u,u)/u\to 0$ as $u\downarrow 0$. It is easily seen that $W_{kl}(u,u)/u\to 0$ as $u\downarrow 0$ which implies $C_{kl}(u,u)/u\to 0$. It is straightforward to see that $(1-2u+M_{kl}(u,u))/(1-u)\to 1/\max(n_k,n_l)$ as $u\uparrow 1$. To compute the lower and upper bounds for ρ_{kl} and τ_{kl}, it suffices to substitute W_{kl} and M_{kl} into (13) and (12). Lengthy but elementary computations lead to the results. Finally, letting n_k or n_l going to infinity in (16) yields that C_{kl} tends to independence.

B Algorithm to compute the full joint likelihood of PBC copulas

Denote the parameter vector by $\boldsymbol{\theta} = (\theta_{ij})_{\{ij\}\in E}$. Recall that the graph is assumed to be a tree, that is, there is no cycles in the graph (then $|E| = d-1$). Let $V = \{1, \dots, d\}$ and $u = (u_1, \dots, u_d)$ a vector in $[0,1]^d$. For a subset $A \subset V$, the notation $\partial_{u_A} C(u;\boldsymbol{\theta})$ stands for the derivative of C with respect to all the variables in A. For instance the density (hence the likelihood) writes

$$\frac{\partial^d C(u;\boldsymbol{\theta})}{\partial_{u_1} \dots \partial_{u_d}} = \partial_{u_V} C(u;\boldsymbol{\theta}) = c(u;\boldsymbol{\theta}), \tag{17}$$

and the gradient with respect to the parameter vector,

$$\left(\frac{\partial c(u;\boldsymbol{\theta})}{\theta_{ij}}\right)_{\{ij\in E\}}.$$

To keep the notation simple, the dependence on the parameter vector $\boldsymbol{\theta}$ is dropped in the remaining of the section. The purpose here is not to give the algorithm, but rather to provide an intuitive idea of it.

Let us write

$$C(u_1, \dots, u_d) = \prod_{\{ij\}\in E} \tilde{C}_{ij}(u_i^{1/n_i}, u_j^{1/n_j}) =: \prod_{\{ij\}\in E} \Phi_{ij}(u_i, u_j).$$

and let an arbitrary variable index i (the root) be given. Let τ_s^i denote the subtree rooted at the variable indexed by i and containing the edge indexed by e (see Figure 3). The idea is to note that, since the graph is a tree, the copula C can be decomposed over the subtrees rooted at i:

$$C(u) = \prod_{e\in E} \Phi_e(u) =: \prod_{e\in N(i)} T_{\tau_e^i}(u), \quad u = (u_1, \dots, u_d),$$

where $T_{\tau_e^i}(u)$ corresponds to the product of all edges located in the subtree τ_e^i. Since the $T_{\tau_e^i}(u)$'s do not share any variables (except the root), the derivative and the product operations commute, more precisely,

$$\begin{aligned}
\partial_{u_V} C(u) &= \partial_{u_i, u_{V\setminus i}} \left[\prod_{e\in N(i)} T_{\tau_e^i}(u)\right] \\
&= \partial_{u_i} \left[\prod_{e\in N(i)} \partial_{u_{\tau_e^i \setminus i}} T_{\tau_e^i}(u)\right] \\
&= \partial_{u_i} \left[\prod_{e\in N(i)} \mu_{e\to i}(u)\right].
\end{aligned} \tag{18}$$

The quantity $\mu_{e\to i}(u) := \partial_{u_{\tau_e^i \setminus i}} T_{\tau_e^i}(u)$ is called a message from the edge indexed by e to the variable indexed by i. Now consider $T_{\tau_e^i}(u)$ and let j be the neighbor variable index of e. One can go deeper in the tree, that is, we have

$$T_{\tau_e^i}(u) = \Phi_e(u_i, u_j) T_{\tau_j^e}(u)$$

where τ_j^e is the subtree rooted at the edge indexed by e and containing the variable indexed by j (see Figure 3). Hence,

$$\partial_{u_{\tau_e^i \setminus i}} T_{\tau_e^i}(u) = \partial_{u_j} \left[\phi_e(u_i, u_j) \partial_{u_{\tau_j^e \setminus j}} T_{\tau_j^e}(u) \right] = \partial_{u_j} \left[\phi_e(u_i, u_j) \mu_{j \to e}(u) \right].$$

A second type of message has been defined: $\mu_{j \to e}(u) := \partial_{u_{\tau_j^e \setminus j}} T_{\tau_j^e}(u)$ is called a message from the variable index j to the edge index e. Again,

$$T_{\tau_j^e}(u) = \prod_{e' \in N(j) \setminus e} T_{\tau_{e'}^j}(u),$$

hence,

$$\partial_{u_{\tau_j^e \setminus j}} T_{\tau_j^e}(u) = \prod_{e' \in N(j) \setminus e} \partial_{u_{\tau_{e'}^j \setminus j}} T_{\tau_{e'}^j}(u) = \prod_{e' \in N(j) \setminus e} \mu_{e' \to j}(u),$$

where the message $\mu_{e' \to j}(u)$ has been already defined in (18). To summarize, the calculation of $\mu_{e \to i}(u)$ requires the calculation of $\mu_{j \to e}(u)$, which, in turn, requires the calculation of $\mu_{e' \to j}(u)$, where $e = \{ij\}$ and e' is an edge index attached to j. The algorithm presented above allows to compute recursively all the messages from the leaves to the root. Once all the messages have been computed, the density is given by the derivative with respect to the root of the product of all the messages (18).

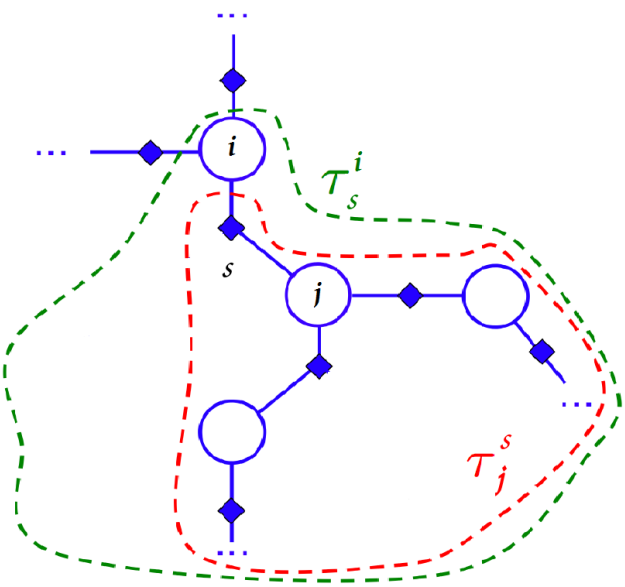

Figure 3: Examples of subtrees. This figure is partly drawn from [10].

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

[2] L.B.G Andersen and V.V. Piterbarg. *Interest Rate Modeling.* Atlantic Financial Press, 2010.

[3] S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005.

[4] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.

[5] G. Frahm, M. Junker, and A. Szimayer. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286, 2003.

[6] C. Genest, K. Ghoudi, and L-P. Rivest. Understanding relationships using copulas. *North American Actuarial Journal*, 2(3):143–149, 1998.

[7] C. Genest and B. Rémillard. Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369, 2004.

[8] G. Gudendorf and J. Segers. Extreme-value copulas. In *Copula Theory and Its Applications*, page 127–145. Springer, 2010.

[9] M. Hofert, M. Mächler, and A. J. McNeil. Archimedean copulas in high dimensions: Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1):25–63, 2012.

[10] J. C. Huang. *Cumulative distribution networks: Inference, estimation and applications of graphical models for cumulative distribution functions.* PhD thesis, University of Toronto, 2009.

[11] J.C. Huang and N. Jojic. Maximum-likelihood learning of cumulative distribution functions on graphs. *Journal of Machine Learning Research W&CP Series*, 9:342–349, 2010.

[12] H. Joe. *Multivariate models and dependence concepts.* Chapman & Hall/CRC, 2001.

[13] D. Kim and J.M. Kim. Analysis of directional dependence using asymmetric copula-based regression models. *Journal of Statistical Computation and Simulation*, 84(9):1990–2010, 2014.

[14] E. Liebscher. Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis*, 99(10):2234–2250, 2008.

[15] B. G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–39, 1988.

[16] R.B. Nelsen. *An introduction to copulas.* Springer, 2006.

[17] M. Sklar. *Fonctions de répartition à n dimensions et leurs marges.* Publications de l'Institut de Statistique de l'Université de Paris, 8:229-231, 1959.

[18] T. Van Pham and G. Mazo. *PBC: product of bivariate copulas.* http://cran.r-project.org, 2014. R package version 1.2.

[19] L. Zhang and V. P. Singh. Gumbel–Hougaard copula for trivariate rainfall frequency analysis. *Journal of Hydrologic Engineering*, 12(4):409–419, 2007.

Chapitre 5

Estimation de copules multivariées par la méthode des moindres carrés pondérés basée sur les coefficients de dépendance

Supposons que l'on dispose d'un échantillon d'une fonction de répartition F d-variée ($d > 2$) dont la copule appartient à une famille indexée par un vecteur de paramètres $\boldsymbol{\theta}$ et que l'on souhaite estimer ces paramètres. Supposons aussi que cette copule n'admette pas de dérivées partielles sur son ensemble de définition et que par conséquent, comme nous l'avions fait remarquer dans la partie 3.1.2, il n'est pas possible de s'assurer des propriétés asymptotiques des méthodes bien connues de la littérature. Certaines de ces copules ont déjà été présentées dans la partie 1.3.2. D'autres ont été utilisées dans la pratique avec succès [20, 77]. Enfin, et surtout, c'est le cas de la classe de copules construite dans le chapitre 6.

Dans ce chapitre, pour estimer le paramètre $\boldsymbol{\theta}$, nous reconsidérons l'estimateur des moindres carrés pondérés basé sur les coefficients de dépendances (3.8) de la partie 3.1.2 et établissons ses propriétés asymptotiques sans supposer que les copules admettent des dérivées partielles. Rappelons que l'estimateur est obtenu en minimisant la fonction

$$l(\boldsymbol{\theta}) = \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta})\right)^T \widehat{\boldsymbol{W}} \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta})\right), \tag{5.1}$$

où $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) = (\mathcal{D}_{1,2}(\boldsymbol{\theta}), \ldots, \mathcal{D}_{d-1,d}(\boldsymbol{\theta}))$, $\widehat{\boldsymbol{\mathcal{D}}} = (\widehat{\mathcal{D}}_{1,2}, \ldots, \widehat{\mathcal{D}}_{d-1,d})$ et $\widehat{\boldsymbol{W}}$ est une matrice de poids ; $\mathcal{D}_{i,j}(\boldsymbol{\theta})$ est le coefficient de dépendance entre la i-ème et la j-ème variable d'intérêt sous le modèle, et $\widehat{\mathcal{D}}_{i,j}$ est une estimation empirique de ce coefficient. Par exemple, pour $\mathcal{D}_{i,j}(\boldsymbol{\theta})$, on peut prendre le coefficient rho de Spearman (1.5), et pour $\widehat{\mathcal{D}}_{i,j}$, on prendrait bien sûr sa version empirique (3.5). Les résultats sont obtenus sous des conditions naturelles d'identifiabilité des modèles. Nous entendons par là que l'ensemble des coefficients de dépendance

des paires détermine la copule à l'intérieur d'une famille donnée. En particulier, l'application

$$\boldsymbol{\theta} \mapsto (\mathcal{D}_{1,2}(\boldsymbol{\theta}), \ldots, \mathcal{D}_{d-1,d}(\boldsymbol{\theta}))$$

devra être bijective, continue, et son inverse continue également. Les résultats théoriques sont illustrés sur des simulations et des données hydrologiques, où nous estimons des copules possèdant une composante singulière. L'article présenté ci-dessous a été soumis pour publication, et est disponible à l'adresse `http://hal.archives-ouvertes.fr/hal-00979151`.

Weighted least-squares inference for multivariate copulas based on dependence coefficients

Gildas Mazo, Stéphane Girard and Florence Forbes

Inria Grenoble Rhône-Alpes & Laboratoire Jean Kuntzmann, France

Abstract

In this paper, we address the issue of estimating the parameters of general multivariate copulas, that is, copulas whose partial derivatives may not exist. To this aim, we consider a weighted least-squares estimator based on dependence coefficients, and establish its consistency and asymptotic normality. The estimator's performance on finite samples is illustrated on simulations and a real dataset.

Keywords: partial derivatives, singular component, weighted least-squares, method-of-moments, dependence coefficients, parametric inference, multivariate copulas.

1 Introduction

The concept of copulas is useful to model multivariate distributions. Given a multivariate random vector of interest, copulas allow to separate the analysis of the margins from the dependence structure. Standard books covering this subject include [9, 24, 29]. See also [13] for an introduction to this topic.

Some copulas possess a singular component, meaning that they are not absolutely continuous (with respect to the Lebesgue measure). For instance, let us consider the copula given below, introduced in [8]:

$$C(u_1, u_2, u_3, u_4) = \prod_{i=1}^{4} u_i^{1-\sum_{j\neq i}\theta_{ij}} \prod_{i<j} \min(u_i, u_j)^{\theta_{ij}}, \tag{1}$$

$$\sum_{j\neq i} \theta_{ij} \leq 1,\ i = 1, \ldots, 4,\ \theta_{ij} = \theta_{ji} \in [0, 1].$$

One can see that, on the diagonal of the unit hypercube, the partial derivatives do not exist. Yet, most inference methods for multivariate copulas make the assumption that these derivatives exist, and even are continuous. This is the case, for example, for the minimum-distance estimator [35], the simulated method of moments [31], and, of course, likelihood-based methods (Section 10.1 [24], [14]). When this assumption is not made, some methods can still be applied but only in specific situations. For example, when there are only two dimensions, one can rely on the inversion of the Kendall's tau, see [18] and [13]. When there is an arbitrary number of dimensions but only one parameter to estimate, an

extension of this method can also be found in [15]. Also, if the copulas of interest are elliptical copulas, one can use the analysis of covariance structures [25]. The issue that the partial derivatives need to exist and be continuous on the unit hypercube in order to properly apply most inference methods was also raised in [3, 34]. In these papers, the authors show how to weaken the classical differentiability assumptions in the empirical copula process theory.

In order to estimate the parameters of general multivariate copulas, we consider a weighted least-squares (WLS) estimator based on dependence coefficients. The consistency and asymptotic normality of the estimator are derived without assuming that the copulas of interest have partial derivatives at all. This method is therefore broadly applicable and allows to estimate the parameters of any kind of copulas, provided that one can calculate their dependence coefficients.

In Section 2 of this paper, the consistency and asymptotic normality of the WLS estimator are established. The theoretical results are illustrated on simulated and real datasets in Section 3. The proofs are postponed to the Appendix.

2 Asymptotic properties of the WLS estimator based on dependence coefficients

We derive the consistency and asymptotic normality of a generic WLS estimator in Section 2.1 and give three examples based on the Spearman's rho, the Kendall's tau, and the extremal dependence coefficients in Section 2.2.

Let $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ with $\mathbf{X}^{(k)} = (X_1^{(k)}, \ldots, X_d^{(k)})$, $k = 1, \ldots, n$, be independent and identically distributed copies of a vector $\mathbf{X} = (X_1, \ldots, X_d)$ with distribution F and copula C. The marginal distributions $F_1, \ldots, F_d$ are assumed to be continuous. The copula C is assumed to belong to the family $(C_{\boldsymbol{\theta}})$ for $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$. The true parameter vector is denoted by $\boldsymbol{\theta}_0$, that is, $C = C_{\boldsymbol{\theta}_0}$. Let $p = d(d-1)/2$ be the number of variable pairs (X_i, X_j), for $i = 1, \ldots, d-1$, $j = 2, \ldots, d$, $i < j$. Let us define the vector map

$$\begin{aligned} \boldsymbol{\mathcal{D}} : \Theta &\to \boldsymbol{\mathcal{D}}(\Theta) \subset \mathbb{R}^p \\ \boldsymbol{\theta} &\mapsto (\mathcal{D}_{1,2}(\boldsymbol{\theta}), \ldots, \mathcal{D}_{d-1,d}(\boldsymbol{\theta})) \,, \end{aligned} \tag{2}$$

where $\mathcal{D}_{i,j}(\cdot)$ can represent, but is not limited to, a well chosen dependence coefficient between the variables X_i and X_j (see Section 2.2 for examples). The space $\boldsymbol{\mathcal{D}}(\Theta)$ stands for the image of Θ by the multivariate map $\boldsymbol{\mathcal{D}}$. The coordinates of $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta})$ are the $\mathcal{D}_{i,j}(\boldsymbol{\theta})$ sorted in the lexicographic order. When the map $\boldsymbol{\mathcal{D}}$ is differentiable, its Jacobian matrix at $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$ is denoted by

$$\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial \mathcal{D}_{1,2}(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \mathcal{D}_{1,2}(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial \mathcal{D}_{1,2}(\boldsymbol{\theta})}{\partial \theta_q} \\ \vdots & & & \vdots \\ \frac{\partial \mathcal{D}_{d-1,d}(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \mathcal{D}_{d-1,d}(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial \mathcal{D}_{d-1,d}(\boldsymbol{\theta})}{\partial \theta_q} \end{pmatrix}.$$

Besides, let $\widehat{\boldsymbol{\mathcal{D}}} = (\widehat{\mathcal{D}}_{1,2}, \ldots, \widehat{\mathcal{D}}_{d-1,d})$ be an empirical (nonparametric) estimator of $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$. To simplify the notations, we shall write $\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}_0) = \dot{\boldsymbol{\mathcal{D}}}$, $\mathcal{D}_{i,j}(\boldsymbol{\theta}_0) = \mathcal{D}_{i,j}$

and $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$. Vectors are assumed to be column vectors and ^{T} denotes the transpose operator.

The WLS estimator of $\boldsymbol{\theta}_0$ studied in this paper is defined as

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta})\right)^T \widehat{\mathbf{W}} \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta})\right), \tag{3}$$

where $\widehat{\mathbf{W}} = \widehat{\mathbf{W}}_n$ is a sequence $(n = 1, 2, \dots)$ of symmetric and positive definite matrices with full rank. Let us denote $\hat{\ell}(\boldsymbol{\theta})$ the loss function to be minimized in (3). In general, the minimizer $\hat{\boldsymbol{\theta}}$ of $\hat{\ell}(\cdot)$ may not exist, or may not be unique. However, it will be seen in Section 2.1 that the existence and uniqueness of $\hat{\boldsymbol{\theta}}$ hold with probability tending to one as the sample size increases. Since $\widehat{\mathbf{W}}$ is positive definite, the loss function $\hat{\ell}$ is such that $\hat{\ell}(\boldsymbol{\theta}) \geq 0$ for all $\boldsymbol{\theta} \in \Theta$ and vanishes at $\hat{\boldsymbol{\theta}}$ if and only if $\hat{\boldsymbol{\theta}} \in \boldsymbol{\mathcal{D}}^{-1}(\{\widehat{\boldsymbol{\mathcal{D}}}\})$, where $\boldsymbol{\mathcal{D}}^{-1}(\{\widehat{\boldsymbol{\mathcal{D}}}\})$ denotes the set of all $\boldsymbol{\theta}$ in Θ such that $\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) = \widehat{\boldsymbol{\mathcal{D}}}$. In this case, the WLS estimator does not depend on the weights and $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\mathcal{D}}}$. Moreover, if the multivariate map $\boldsymbol{\mathcal{D}}$ is one-to-one, then the WLS estimator takes the form $\hat{\boldsymbol{\theta}} = \boldsymbol{\mathcal{D}}^{-1}(\widehat{\boldsymbol{\mathcal{D}}})$.

2.1 Asymptotic properties of the generic WLS estimator

The assumptions needed to derive the asymptotic properties of the WLS estimator are given below. The symbol $\|\cdot\|$ denotes the Euclidean norm.

Assumptions. *(A1) The true parameter vector $\boldsymbol{\theta}_0$ lies in the interior of Θ. Moreover, there exists $\varepsilon_0 > 0$ such that the set $\{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \varepsilon_0\}$ is closed (and thus compact) in $\mathbb{R}^q$.*

(A2) As $n \to \infty$, the sequence of weight matrices $\widehat{\mathbf{W}}$ converges in probability to a symmetric and positive definite matrix $\mathbf{W}$ with full rank.

(A3) The map $\boldsymbol{\mathcal{D}}$ defined in (2) is a twice continuously differentiable homeomorphism such that $\dot{\boldsymbol{\mathcal{D}}}$ is of full rank.

(A4) As $n \to \infty$, the empirical estimator $\widehat{\boldsymbol{\mathcal{D}}}$ is such that

$$\widehat{\boldsymbol{\mathcal{D}}} \xrightarrow{P} \boldsymbol{\mathcal{D}}, \text{ and, } \sqrt{n}\left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}\right) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is some symmetric, positive definite $p \times p$ matrix noted as follows

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{1,2;1,2} & \Sigma_{1,2;1,3} & \cdots & \Sigma_{1,2;d-1,d} \\ \Sigma_{1,3;1,2} & \Sigma_{1,3;1,3} & \cdots & \Sigma_{1,3;d-1,d} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_{d-1,d;1,2} & \Sigma_{d-1,d;1,3} & \cdots & \Sigma_{d-1,d;d-1,d} \end{pmatrix}. \tag{4}$$

Assumption (A1), which is rather standard, see, e.g. [10], is not too restrictive for most copula models. Indeed, a parameter lying in the parameter space boundaries often means that the copula of interest is in fact the independence copula or one of the Fréchet-Hoeffding bounds, that is, a copula where the dependence is "perfect", see for instance [29] chapter 2. This is not an issue

because one does not encounter perfect dependence in practice. As for independence, one might carry out a statistical test as in [16], and, based on the results, decide whether independence holds or not. If not, then one can safely assume that the parameters lie in the interior of the parameter space.

A sequence of weight matrices verifying Assumption (A2) can always be constructed. A trivial example is $\widehat{\mathbf{W}} = \mathbf{I}_p$, where $\mathbf{I}_p$ is the identity matrix of size p. The construction of optimal weights is addressed in Proposition 1 below.

The estimation of the copula parameter vector is performed by matching the theoretical and empirical dependence coefficients. Hence, a successful match should ensure that the resulting parameter vector estimate is close to the true value. This identifiability condition, also made in [10] in order to estimate extreme-value copulas with a singular component, is the essence of Assumption (A3). Assumption (A3) is hard to check in practice, except maybe for particular simple cases when $p = q$, as in Section 3.2. Nonetheless, one can always test the estimation procedure on simulated data, by, for instance, comparing the expected asymptotic distribution to its estimated counterpart; see Section 3.1 for an example of how to proceed.

The last assumption, (A4), naturally states that one should have convergence of the dependence coefficient empirical estimator to ensure convergence of the WLS estimator.

Theorem 1. *Assume that (A1)–(A4) hold. Then, as $n \to \infty$ and with probability tending to one, the WLS estimator defined by* (3) *exists and is unique. Moreover, it is consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \boldsymbol{\Xi}\right), \text{ where} \tag{5}$$
$$\boldsymbol{\Xi} = \left(\dot{\boldsymbol{\mathcal{D}}}^T \mathbf{W} \dot{\boldsymbol{\mathcal{D}}}\right)^{-1} \dot{\boldsymbol{\mathcal{D}}}^T \mathbf{W} \boldsymbol{\Sigma} \mathbf{W} \dot{\boldsymbol{\mathcal{D}}} \left(\dot{\boldsymbol{\mathcal{D}}}^T \mathbf{W} \dot{\boldsymbol{\mathcal{D}}}\right)^{-1}.$$

As usual, the results of Theorem 1 allow to derive the asymptotic distribution of quadratic forms in $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}})$. These asymptotics serve to build confidence regions and statistical tests for the parameters and the dependence coefficients. Let χ^2_q denote the Chi square distribution with q degrees of freedom. Let us write $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ to emphasize that in general these matrices depend on $\boldsymbol{\theta}$. The continuity of matrices with respect to the parameter vector $\boldsymbol{\theta}$ is meant elementwise. Corollary 1, given below, may serve to build confidence regions around $\hat{\boldsymbol{\theta}}$ or $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}})$.

Corollary 1. *Suppose that the assumptions of Theorem 1 hold.*

(i) If $\boldsymbol{\Xi}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta}$ in Θ and $\boldsymbol{\Sigma}(\cdot)$ is continuous at $\boldsymbol{\theta}_0$, then, as $n \to \infty$,

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \chi^2_q.$$

(ii) Define $\widehat{\boldsymbol{\Sigma}}$ such that $\widehat{\boldsymbol{\Sigma}}$ is invertible and converges to $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ in probability as $n \to \infty$. Then as $n \to \infty$,

$$n(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0))^T \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)) \xrightarrow{d} \chi^2_p.$$

For a particular value $\boldsymbol{\theta}_1^\star \in \mathbb{R}^r$, $r \leq q-1$, the test $H_0 : \boldsymbol{\theta}_{01} = \boldsymbol{\theta}_1^\star$ against $H_1 : \boldsymbol{\theta}_{01} \neq \boldsymbol{\theta}_1^\star$, where $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{01}, \boldsymbol{\theta}_{02}) \in \mathbb{R}^r \times \mathbb{R}^{q-r}$, may be carried out using the asymptotic approximation suggested by Corollary 2, given next. In general, write $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^r \times \mathbb{R}^{q-r}$ for $\boldsymbol{\theta} \in \Theta$, and, likewise, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$. Let $\boldsymbol{\Xi}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the $r \times r$ asymptotic covariance matrix corresponding to $\boldsymbol{\theta}_1$, that is, the upper left part of $\boldsymbol{\Xi}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

Corollary 2. *Under the assumptions of Corollary 1 (i), as $n \to \infty$,*

$$n(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^\star)^T \boldsymbol{\Xi}_1(\boldsymbol{\theta}_1^\star, \hat{\boldsymbol{\theta}}_2)^{-1}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^\star) \xrightarrow{d} \chi_r^2.$$

The test H_0: "the chosen parametric model is the true model of the underlying copula" against H_1 "the chosen parametric model is false" may be carried out by using the asymptotic approximation suggested by Corollary 3 below, adapted from [22].

Corollary 3. *Suppose that the assumptions of Theorem 1 and Corollary 1 (ii) hold. For $\boldsymbol{\theta} \in \Theta$, define*

$$\mathbf{A}(\boldsymbol{\theta}) := \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \right)^{-1} \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T,$$

$\widehat{\mathbf{A}} = \mathbf{A}(\hat{\boldsymbol{\theta}})$, and denote by k the rank of $\mathbf{I}_p - \mathbf{A}(\boldsymbol{\theta}_0)$. If $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta}$ in Θ and $\boldsymbol{\Sigma}(\cdot)$ is continuous at $\boldsymbol{\theta}_0$, then as $n \to \infty$,

$$n \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right)^T (\mathbf{I}_p - \widehat{\mathbf{A}})[(\mathbf{I}_p - \widehat{\mathbf{A}})\widehat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \widehat{\mathbf{A}}) + \widehat{\mathbf{A}}]^{-1}(\mathbf{I}_p - \widehat{\mathbf{A}}) \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) \to \chi_k^2.$$

The asymptotic covariance matrix $\boldsymbol{\Xi}$ in (5) depends on the weight matrix $\mathbf{W}$. The optimal weight matrix $\mathbf{W}^\star$, in the sense that it allows to minimize the asymptotic covariance matrix $\boldsymbol{\Xi}$, is given in Proposition 1 below (due to [22]). The above mentioned ordering of covariance matrices is to be understood in the following sense. The notation $\mathbf{A} \geq 0$ means that the matrix $\mathbf{A}$ is nonnegative definite. For two nonnegative definite matrices $\mathbf{A}$ and $\mathbf{B}$, define $\mathbf{A}$ to be less or equal than $\mathbf{B}$ if $\mathbf{B} - \mathbf{A} \geq 0$. It is easily checked that $\mathbf{A} \leq \mathbf{B}$ implies $\mathrm{tr}(\mathbf{A}) \leq \mathrm{tr}(\mathbf{B})$, where $\mathrm{tr}(\cdot)$ stands for the trace operator of matrices. Thus, the distribution with the smallest covariance matrix is the one for which the sum of the variances is minimum. In view of (6), an optimal estimator, that is, an estimator that leads to the smallest asymptotic covariance matrix, can be constructed by letting the sequence of weight matrices converge to $\boldsymbol{\Sigma}^{-1}$.

Proposition 1. *Suppose that $\boldsymbol{\Sigma}$ defined in (A4) is invertible. Then the asymptotic covariance matrix $\boldsymbol{\Xi}$ is minimum for $\mathbf{W}^\star$ such that*

$$\mathbf{W}^\star \dot{\boldsymbol{\mathcal{D}}} \propto \boldsymbol{\Sigma}^{-1} \dot{\boldsymbol{\mathcal{D}}}, \tag{6}$$

where the symbol $\propto$ denotes proportionality.

An estimate of the optimal weight matrix $\boldsymbol{\Sigma}^{-1}$ can be based on empirical data or constructed as follows. Define the *zero-step estimator* $\hat{\boldsymbol{\theta}}^0$ to be the WLS estimator (3) with $\widehat{\boldsymbol{W}} = \mathbf{I}_p$. Define the *one-step estimator* $\hat{\boldsymbol{\theta}}^1$ to be the WLS estimator with $\widehat{\boldsymbol{W}} = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\theta}}^0)$, where $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}^0)$ is an estimate of $\boldsymbol{\Sigma}$ based on the zero-step estimator. For instance, one may simulate data according

to $C = C(\hat{\boldsymbol{\theta}}^0)$ and use them to construct $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}^0)$. This one-step estimator is then an optimal estimator. The performances of the zero-step and the optimal estimators will be compared in Section 3.1.

When there are as many pairs as parameters, the WLS estimator does not depend on the weights, as stated in the next proposition.

Proposition 2. *Suppose that the assumptions of Theorem 1 hold. If $p = q$ then, as $n \to \infty$ and with probability tending to one,*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mathcal{D}}^{-1}\left(\widehat{\boldsymbol{\mathcal{D}}}\right), \tag{7}$$

and

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \left(\dot{\boldsymbol{\mathcal{D}}}^T\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\right). \tag{8}$$

2.2 Examples of three dependence coefficients verifying Assumption (A4)

Three examples of a dependence coefficient for which the pair of vectors $(\boldsymbol{\mathcal{D}}, \widehat{\boldsymbol{\mathcal{D}}})$ satisfies Assumption (A4) are provided. These coefficients are the Spearman's rho, the Kendall's tau, and the extremal dependence coefficient. They are widely used in practice, and that is why we illustrate our methodology on them. But others can be used, as long as (A4) holds. See [24,29] for more about these coefficients and [23] for their asymptotic properties. Recall that F_i is the distribution of X_i and let

$$\widehat{F}_i(x) = \frac{1}{n+1}\sum_{k=1}^{n}\mathbf{1}(X_i^{(k)} \le x), \quad x \in \mathbb{R}.$$

Put $U_i = F_i(X_i)$ and $\widehat{U}_i^{(k)} = \widehat{F}_i(X_i^{(k)})$. Recall that $F_{i,j}$ is the distribution function of (X_i, X_j) and that $C_{i,j}$ denotes its copula.

Example 1 (Spearman's rho)**.** *The Spearman's rho dependence coefficient of the pair (X_i, X_j) is given by*

$$\mathcal{D}_{i,j} = 12\int_{[0,1]^2} C_{i,j}(u,v)\,du\,dv - 3. \tag{9}$$

Its empirical counterpart is defined as

$$\widehat{\mathcal{D}}_{i,j} = \frac{\sum_{k=1}^{n}\left(\widehat{U}_i^{(k)} - \overline{\widehat{U}}_i\right)\left(\widehat{U}_j^{(k)} - \overline{\widehat{U}}_j\right)}{\left[\sum_{k=1}^{n}\left(\widehat{U}_i^{(k)} - \overline{\widehat{U}}_i\right)^2\sum_{k=1}^{n}\left(\widehat{U}_j^{(k)} - \overline{\widehat{U}}_j\right)^2\right]^{1/2}},$$

where $\overline{\widehat{U}_i} = \sum_{k=1}^{n}\widehat{U}_i^{(k)}/n$. This empirical estimator is a U-statistic [23], whose theory is well known, see, e.g. [36] Chapter 12. In particular, it has been established in [23] Section 9 that Assumption (A4) holds with

$$\begin{aligned}\Sigma_{i,j;k,l} = 9\int_{[0,1]^2} &[(1-2u_i)(1-2u_j) - 2u_i - 2u_j + 4I_i(u_j) + 4I_j(u_i) - \rho_{ij}] \\ &\times [(1-2u_k)(1-2u_l) - 2u_k - 2u_l + 4I_k(u_l) + 4I_l(u_k) - \rho_{kl}]\,dC(u_1,\ldots,u_d),\end{aligned}$$

where $I_j(u) = \int_0^1 C_{ij}(u,v)dv$ *and* $I_i(v) = \int_0^1 C_{ij}(u,v)du$*. Let us note that this result holds without any condition on the underlying copula.*

Example 2 (Kendall's tau)**.** *The Kendall's tau dependence coefficient of the pair* (X_i, X_j) *is given by*

$$\mathcal{D}_{i,j} = 4\int_{[0,1]^2} C_{i,j}(u,v)\, dC_{i,j}(u,v) - 1. \tag{10}$$

Its empirical counterpart is defined as

$$\widehat{\mathcal{D}}_{i,j} = \binom{n}{2}^{-1} \sum_{k<l} sign\left((X_i^{(k)} - X_i^{(l)})(X_j^{(k)} - X_j^{(l)})\right), \tag{11}$$

where $sign(x) = 1$ *if* $x > 0$*,* -1 *if* $x < 0$ *and* 0 *if* $x = 0$*. This empirical estimator is also a U-statistic and the same discussion as in Example 1 also yields that Assumption (A4) holds with*

$$\begin{aligned}\Sigma_{i,j;k,l} = 4\int_{[0,1]^2} &[4C_{i,j}(u_i,u_j) + 1 - \mathcal{D}_{i,j} - 2u_i - 2u_j] \\ &\times [4C_{k,l}(u_k,u_l) + 1 - \mathcal{D}_{k,l} - 2u_k - 2u_l]\, dC(u_1,\ldots,u_d).\end{aligned} \tag{12}$$

Again, let us note that this result holds without any condition on the underlying copula.

The third example deals with extreme-value copulas, which are theoretically well grounded for performing a statistical analysis of extreme values, such as maxima of samples. Recall that a copula $C_\#$ is an extreme-value copula if there exists a copula $\tilde{C}$ such that

$$C_\#(u_1,\ldots,u_d) = \lim_{m\uparrow\infty} \tilde{C}^m(u_1^{1/m},\ldots,u_d^{1/m}), \quad (u_1,\ldots,u_d) \in [0,1]^d,$$

see, e.g. [20]. The class of extreme-value copulas corresponds exactly to the class of max-stable copulas, that is, the copulas $C_\#$ such that

$$C_\#^m(u_1^{1/m},\ldots,u_d^{1/m}) = C_\#(u_1,\ldots,u_d), \quad m \geq 1,\ (u_1,\ldots,u_d) \in [0,1]^d.$$

The extremal dependence coefficient is implicitely defined by the following representation of bivariate extreme-value copulas on the diagonal of the unit square:

$$C_\#(u,u) = u^{2-\lambda}, \quad \lambda \in [0,1]. \tag{13}$$

If $\lambda = 0$ then $C_\#(u,u) = \Pi(u,u) = u^2$, where Π stands for the independence copula. If $\lambda = 1$ then $C_\#(u,u) = M(u,u) = \min(u,u) = u$, where M stands for the Fréchet-Hoeffding upper bound for copulas, that is, the case of perfect dependence. In the case of extreme-value copulas, the extremal dependence coefficient corresponds to the well known upper tail dependence coefficient

$$\lambda = \lim_{u\uparrow 1} \frac{1 - 2u + C_\#(u,u)}{1-u},$$

which measures the dependence in the tails. Nonetheless, for extreme-value copulas, the interpolation between Π and M on the diagonal of the unit square (13)

makes the extremal dependence coefficient a natural coefficient of general dependence, and not just a coefficient that measures dependence in the tails. For further information about extreme-value statistics, see, e.g. [5]. An account about extreme-value copulas can be found in [20].

Estimators of the extremal dependence coefficient for which the asymptotic properties are derived under unknown margins can be found in [2, 19]. However, in order to obtain the results, the existence of partial derivatives for the underlying copulas was assumed. Hence, these estimators cannot be used since we aim at estimating the parameters of copulas for which these derivatives may not exist.

If the marginal distributions are assumed to be known, however, various estimators of the extremal dependence coefficient and their asymptotic properties can be found in the literature [4,7,11,21,32]. A review can be found in [20]. Our choice of the estimator presented in Example 3 below, that of [11], is arbitrary. One can choose an other estimator in the literature and adapt the results.

Example 3 (Extremal dependence coefficient)**.** *Assume that the copula of interest C is an extreme-value copula and let $\mathcal{D}_{i,j}$ be the extremal dependence coefficient of the pair (X_i, X_j), implicitely defined in* (13)*, and given by*

$$\mathcal{D}_{i,j} = 2 + \log C_{i,j}(e^{-1}, e^{-1}). \tag{14}$$

Its empirical counterpart, as defined in [11], is given by

$$\widehat{\mathcal{D}}_{i,j} = 3 - \frac{1}{1 - \sum_{k=1}^{n} \max(U_i^{(k)}, U_j^{(k)})/n}.$$

By adapting [11] to the multivariate case, Assumption (A4) holds with

$$\Sigma_{i,j;k,l} = (3 - \mathcal{D}_{i,j})^2 (3 - \mathcal{D}_{k,l})^2 \; Cov\left(\max(U_i, U_j), \max(U_k, U_l)\right). \tag{15}$$

In practice, the margins are usually unknown . However, assuming that F is an extreme-value distribution, the margins should be Generalized Extreme-Value (GEV) distributions, see [5]. Therefore, to estimate the parameters of the GEV, one can apply a method known as "inference functions for margins" (IFM, see [24] Section 10) whose asymptotic properties are well established.

3 Illustrations on simulated and real datasets

In order to assess the WLS estimator's performance on finite samples, numerical experiments are undertaken in Section 3.1 and a real dataset application is presented in Section 3.2. In both the experiments and the application, we aim at estimating the parameters of multivariate copulas possessing a singular component.

3.1 Estimating the parameters of multivariate copulas possessing a singular component

By substituting the Fréchet copulas [12]

$$C_{0k}(u_0, u_k) = \theta_k \min(u_0, u_k) + (1 - \theta_k) u_0 u_k, \quad \theta_k \in [0, 1]$$

into the one-factor copula [27]

$$C(u_1, \ldots, u_d) = \int_0^1 \prod_{k=1}^{d} \frac{\partial C_{0k}(u_0, u_k)}{\partial u_0} \, du_0,$$

one obtains a copula C with a singular component and whose bivariate margins are given by the following Fréchet copulas

$$C_{ij}(u_i, u_j) = \theta_i \theta_j \min(u_i, u_j) + (1 - \theta_i \theta_j) u_i u_j, \quad \theta_i, \theta_j \in [0, 1]. \tag{16}$$

The Spearman's rho and Kendall's tau coefficients of (16) are respectively equal to $\theta_i \theta_j$ and $\theta_i \theta_j (\theta_i \theta_j + 2)/3$. The extreme-value copula $C_\#$ associated to C can be derived by calculating the limit

$$C_\#(u_1, \ldots, u_d) = \lim_{n \uparrow \infty} C^n(u_1^{1/n}, \ldots, u_d^{1/n}).$$

It appears that the bivariate margins of $C_\#$ are Cuadras-Augé copulas [6]

$$C_{\#,ij}(u_i, u_j) = \min(u_i, u_j) \max(u_i, u_j)^{1-\theta_i \theta_j}, \quad \theta_i, \theta_j \in [0, 1]$$

with extremal dependence coefficient given by $\theta_i \theta_j$. Similarly to C, $C_\#$ possesses a singular component.

		$d = 4$		$d = 10$	
		zero-step	one-step	zero-step	one-step
$n = 50$	(S1)	0.11	0.11	0.10	0.12
	(S2)	0.10	0.10	0.09	0.10
	(S3)	0.18	0.18	0.17	0.20
$n = 200$	(S1)	0.06	0.06	0.05	0.05
	(S2)	0.05	0.05	0.04	0.04
	(S3)	0.10	0.10	0.09	0.09
$n = 500$	(S1)	0.04	0.04	0.03	0.03
	(S2)	0.03	0.03	0.03	0.03
	(S3)	0.06	0.06	0.06	0.05

Table 1: MAEs for the three studied situations with respect to the data sample size n and dimension d. The results are averaged over 200 replications.

The two copulas C and $C_\#$ are considered in the following numerical experiment. For each combination (d, n) with $d = 4, 10$ and $n = 50, 200, 500$, we generated 200 datasets according to these copulas. The true parameter vector coordinates θ_{0k}, $k = 1, \ldots, d$, were chosen to be regularly spaced between 0.3 and 0.9. Three situations were studied:

(S1) the parameters of C are estimated with the Spearman's rho (see Example 1),

(S2) the parameters of C are estimated with the Kendall's tau (see Example 2), and

(S3) the parameters of $C_\#$ are estimated with the extremal dependence coefficient (see Example 3).

In situation (S1), Assumption (A3) is verified, see [28]. For each situation (Si) above, the zero-step and one-step WLS estimators of Section 2.1 were tested (recall that the one-step estimator is optimal, see Proposition 1). For each dataset and each situation (Si), the mean absolute error, defined as

$$\mathrm{MAE} = \frac{1}{d}\sum_{k=1}^{d} |\hat{\theta}_k - \theta_{0k}|$$

was computed and averaged over the 200 replications. These criteria are reported in Table 1. It appears from this Table that there is almost no difference between the zero-step and one-step estimators. This lack of weighting effect was also mentioned in [31] Section 3. This suggests that the zero-step estimator is already near optimal. The comparison of the rows (S1) and (S2) shows that the choice between the Spearman's rho and Kendall's tau in the WLS estimator has very little impact on its performance. Estimating the parameters of an extreme-value copula with the extremal dependence coefficient, however, appears to be less accurate–see the (S3) row of the table. Finally, the comparison of the two columns $d = 4$ and $d = 50$ shows that the dimension of the inference problem does not seem to affect the estimator's performance. This property makes it very attractive to deal with high-dimensional applications. To complete the study of the estimator's abilities, its asymptotic distribution derived in Theorem 1 is tested. Since this distribution is multivariate, we checked the Chi-square approximation of Corollary 1 instead. The values $n(\hat{\boldsymbol{\theta}}^{(r)} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi}(\hat{\boldsymbol{\theta}}^{(r)})^{-1}(\hat{\boldsymbol{\theta}}^{(r)} - \boldsymbol{\theta}_0)$, $r = 1, \ldots, 200$, should be approximately χ^2_d distributed (here the dimension is equal to the number of parameters), where $\hat{\boldsymbol{\theta}}^{(r)}$ denotes the parameter vector estimated on the r-th dataset replication. This approximation, shown in Figure 1, seems rather satisfactory.

3.2 Measuring uncertainty for multivariate return periods in hydrology

In hydrology, the severity and frequency of extreme events must be quantified. Such potentially dangerous events are underlain by the behavior of a random vector $(X_1, \ldots, X_d)$ distributed according to a certain distribution F with continuous margins $F_1, \ldots, F_d$ and copula C. Suppose that C is determined by a parameter vector $\boldsymbol{\theta}$ in Θ. For a certain potentially dangerous event, define the *return period* T and the *critical level* p through the relationship

$$T = \frac{1}{1 - K_{\boldsymbol{\theta}}(p)}, \tag{17}$$

where $K_{\boldsymbol{\theta}}(t) = P(C(F_1(X_1), \ldots, F_d(X_d)) \leq t)$, $t \in [0, 1]$, is called the Kendall's distribution function associated to C, see [30]. The return period can be interpreted as the average time elapsing between two dangerous events. For instance, $T = 30$ years means that the event happens once every 30 years in average. The critical level can be viewed as a measure of how dangerous the underlying event is. The following question naturally arises: given a certain return period, what is the critical level of the underlying event? To answer this question, it suffices to invert (17) to get p as a function of T:

$$p_T(\boldsymbol{\theta}) = K_{\boldsymbol{\theta}}^{-1}(1 - 1/T).$$

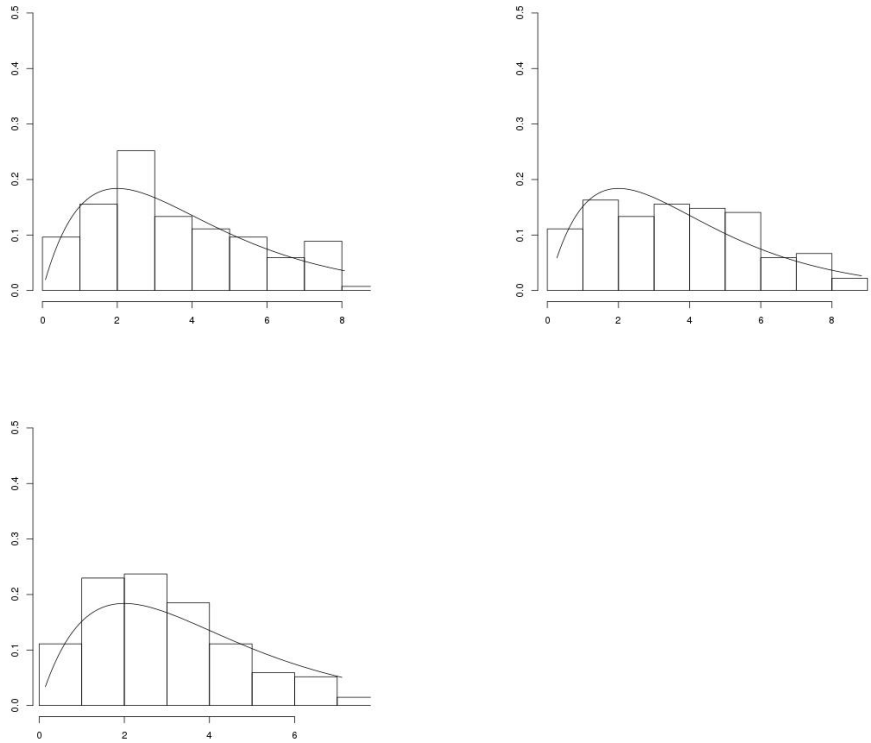

Figure 1: Histograms of $n(\hat{\boldsymbol{\theta}}^{(r)} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi} (\hat{\boldsymbol{\theta}}^{(r)})^{-1} (\hat{\boldsymbol{\theta}}^{(r)} - \boldsymbol{\theta}_0)$, $r = 1, \ldots, 200$ together with the density of a χ^2_d distribution. The considered experiment parameters were $n = 500$ and $d = 4$. Upper left: (S2). Upper right: (S1). Bottom: (S3).

Let $\boldsymbol{\theta}_0$ denote the true parameter vector and let $p_T = p_T(\boldsymbol{\theta}_0)$. The estimation of p_T, or, in other words, the estimation of $\boldsymbol{\theta}_0$, was performed in [8] for all the pairs of $d = 3$ sites in Italy (Airole, Merelli and Poggi). The parametric model proposed for C was the extreme-value copula

$$C(u_1, \ldots, u_d) = \left(\prod_{i=1}^{d} u_i^{1-\theta_i} \right) \min_{i=1,\ldots,d} (u_i^{\theta_i}), \quad \theta_i \in [0,1],\, i = 1, \ldots, d. \tag{18}$$

As it can be seen from (18), this copula has a singular component. The authors chose to base the inference on the Kendall's tau (see Example 2). For $\boldsymbol{\theta}$ in $[0,1]^d$, the Kendall's tau coefficients are given by

$$\tau_{i,j}(\boldsymbol{\theta}) = \frac{\theta_i \theta_j}{\theta_i + \theta_j - \theta_i \theta_j}, \quad i < j. \tag{19}$$

By inverting (19), one obtains

$$\hat{\theta}_i = \frac{1}{2} \left(1 + \frac{1}{\hat{\tau}_{i,j}} + \frac{1}{\hat{\tau}_{i,k}} - \frac{1}{\hat{\tau}_{j,k}} \right), \tag{20}$$

where i, j, k denote the indexes of the three sites and $\hat{\tau}_{i,j}$ is given by (11). Observe that this is the solution of the equation (7), and, under the light of Proposition 2 (since $p = q = d = 3$), we see that this estimator has the smallest asymptotic variance within the class (3). However, in [8], the asymptotic behavior of $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ was not derived. This is done next, and we shall see that it allows to quantify the uncertainties around the critical levels.

The asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is established by applying Theorem 1. It suffices to verify that assumption (A3) holds, which is easily checked from (19). Hence, as $n \to \infty$

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Xi}), \tag{21}$$

where $\boldsymbol{\Xi}$ is given by (8) and (12). Now, the derivation of the asymptotic behavior of the critical levels is straightforward. From (21), we get by the delta-method that, as $n \to \infty$

$$\sqrt{n}\left(p_T(\hat{\boldsymbol{\theta}}) - p_T\right) \xrightarrow{d} N(0, s_T^2), \tag{22}$$

with $s_T^2 = \dot{\boldsymbol{p}}_T \boldsymbol{\Xi} \dot{\boldsymbol{p}}_T^T$, and where $\dot{\boldsymbol{p}}_T$ is the Jacobian of $p_T(\cdot)$ at the true parameter value. It follows that confidence intervals can be computed from the finite-sample approximation of (22), provided that the sample size is large enough. In [8], the critical levels in terms of return periods were reported for the three pair of sites (Airole, Merelli), (Airole, Poggi) and (Merelli, Poggi). We added to their figure 95% confidence intervals for the critical levels (Figure 2).

The test based on Corollary 3 has no power to detect a wrong model in this situation. Indeed, since $\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\mathcal{D}}}$, the test statistic is always zero. Other tests can be performed to achieve such a task, see the original paper [8].

When studying extreme events, it is common to have only a limited amount of data. For instance, in [8], only $n = 34$ (multivariate) observations were available. With such a small sample size, the approximation of the distribution of $\sqrt{n}(p_T(\hat{\boldsymbol{\theta}}) - p_T)$ to a normal distribution may be questionable. To assess the goodness of this approximation for small and moderate sample sizes, we carried out the following numerical experiment. $N = 500$ datasets of size $n \in \{34, 100\}$ were generated according to (18) with $\boldsymbol{\theta}_0 = (0.6, 0.7, 0.2)$. For the m-th dataset ($m = 1, \ldots, N$), the parameter vector estimate $\hat{\boldsymbol{\theta}}^{(m)}$ was computed. Let $s_T(\hat{\boldsymbol{\theta}}^{(m)})$ be the asymptotic standard deviation in (22) at $\hat{\boldsymbol{\theta}}^{(m)}$ where $s_T(\boldsymbol{\theta})$ is regarded as a function of $\boldsymbol{\theta}$. The critical levels $p_T(\hat{\boldsymbol{\theta}}^{(m)})$ together with the 95% confidence bands $p_T(\hat{\boldsymbol{\theta}}^{(m)}) \pm 1.96 s_T(\hat{\boldsymbol{\theta}}^{(m)})/\sqrt{n}$ were computed for $T = 10, 20, 30$. Some of the $\hat{\boldsymbol{\theta}}^{(m)}$ did not lie in their theoretical bounds $[0, 1]$, which led to numerical difficulties for computing $s_T(\hat{\boldsymbol{\theta}}^{(m)})$. Therefore, these were dropped from the experiment. The results reported Table 2 show that the finite sample approximation is rather good for $n = 100$. Even for $n = 34$, this approximation appears to be good for the pair Airole-Merelli. Despite these encouraging results for moderate and small samples, we finish by stressing that the number of missing outputs (recall that this happens when $\hat{\boldsymbol{\theta}}^{(m)}$ do not belong to $[0, 1]$) were quite high: 354 and 298 over the 500 dataset replications for $n = 34$ and $n = 100$ respectively. Consequently, it would be of interest to improve the estimator (20) to reduce this vexing effect.

One can observe from Figure 2 that the curves for the pairs (Airole, Poggi) and (Merelli, Poggi) are similar comparing to that of the pair (Airole, Merelli). Hence to illustrate the use of Corollary 2, we performed the test $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. The change of parameters $\mu_1 := \theta_1 - \theta_2$, $\mu_2 := \theta_1 + \theta_2$ and $\mu_3 := \theta_3$ was applied to the copula model (18). By Corollary 2, the test statistics $n\hat{\mu}_1^2/\Xi_1(0, \hat{\mu}_2, \hat{\mu}_3)$ converges in distribution to a χ_1^2 variable. We obtained a p-value of 95%, indicating that there is no statistical arguments against the null hypothesis. This high p-value also suggests that this test has little power for $n = 34$ data. The p-value for testing $\theta_2 = \theta_3$ and $\theta_1 = \theta_3$ were 83% and 84% respectively. The search of powerful tests for copulas is still an active area of research [1, 17, 26].

	pair	(Airole, Merelli)			(Airole, Poggi)			(Merelli, Poggi)		
n	T	10	20	30	10	20	30	10	20	30
34		0.95	0.95	0.93	0.89	0.84	0.82	0.90	0.87	0.82
100		0.95	0.94	0.94	0.96	0.94	0.93	0.96	0.94	0.93

Table 2: Proportion of inclusions within the 95% confidence intervals for the true value p_T.

4 Discussion

In this paper, we considered a weighted least-squares (WLS) estimator in order to estimate the parameters of general multivariate copulas, that is, copulas for which the partial derivatives may not exist. We established its asymptotic properties and studied its performance on finite samples. In particular, the numerical experiments revealed that the weights may have little impact on the accuracy. Moreover, and this is interesting for practical purposes, the accuracy of the WLS estimator does not seem to depend on the dimension of the statistical problems being addressed. In our work, we provided three dependence coefficients which can be used to form the WLS estimator: the Spearman's rho, the Kendall's tau, and the extremal dependence coefficient. We chose popular dependence coefficients, but others can be used. Even combinations of them may be considered, as long as the formed vector $\widehat{\boldsymbol{\mathcal{D}}}$ verifies Assumption (A4). In the hydrological application of Section 3.2, this may help to make the system of equations (20) more robust numerically.

Acknowledgment. The authors thank Fabrizio Durante and Gianfausto Salvadori for sharing the dataset used in Section 3.2. They also thank two anonymous referees for their careful reading and helpful comments.

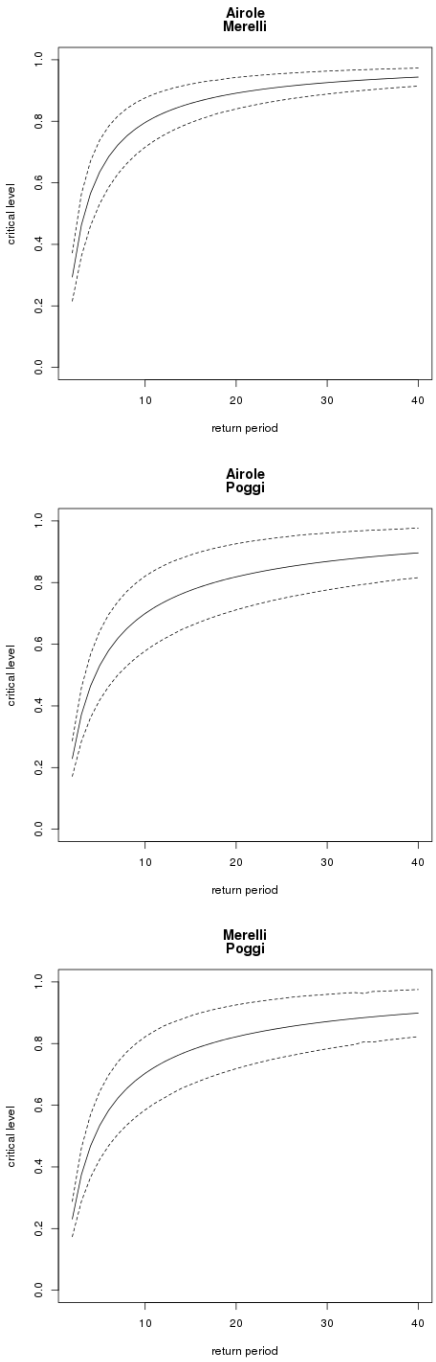

Figure 2: Critical levels $p_T(\hat{\boldsymbol{\theta}})$ for $T = 2, \ldots, 40$ together with 95% confidence intervals.

Appendix: proofs

In order to prove Theorem 1, we first establish two lemmas. These lemmas, as well as their proofs, are adapted from [10]. It will appear that the proof of the theorem is a straightforward application of these lemmas.

Let Θ and ε_0 be defined as in assumption (A1). Define the vector map

$$\begin{aligned} \boldsymbol{\varphi} : \Theta \subset \mathbb{R}^q &\to \boldsymbol{\varphi}(\Theta) \subset \mathbb{R}^p \\ \boldsymbol{\varphi}(\boldsymbol{\theta}) &\mapsto (\varphi_1(\boldsymbol{\theta}), \ldots, \varphi_p(\boldsymbol{\theta}))^T, \end{aligned} \tag{23}$$

and assume that $\boldsymbol{\varphi}$ is twice continuously differentiable. Denote by $\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta})$ the $p \times q$ Jacobian matrix of $\boldsymbol{\varphi}$ at $\boldsymbol{\theta}$ and define $\dot{\boldsymbol{\varphi}} := \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)$. Let

$$\mathbf{Y}_n = (Y_{n,1}, \ldots, Y_{n,p})^T$$

be a random vector in $\mathbb{R}^p$ depending on an integer n and assume that $\mathbf{Y}_n \xrightarrow{P} \boldsymbol{\varphi}(\boldsymbol{\theta}_0)$ as $n \to \infty$. Let $\widehat{\mathbf{W}} = \widehat{\mathbf{W}}_n$ be a $p \times p$ symmetric and positive definite matrix with full rank and suppose that $\widehat{\mathbf{W}}$ converges in probability to a symmetric and positive definite matrix $\mathbf{W}$ with full rank as $n \to \infty$. Then the Cholesky decomposition entails that $\widehat{\mathbf{W}} = \widehat{\mathbf{V}}^T \widehat{\mathbf{V}}$ for some $p \times p$ matrix $\widehat{\mathbf{V}}$. Denote by $\hat{\Theta}_n$ the set of all minimizers of the loss function

$$\begin{aligned} \ell_n(\boldsymbol{\theta}) &= (\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}))^T \widehat{\mathbf{W}} (\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta})) \\ &= \left\| \widehat{\mathbf{V}} (\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta})) \right\|^2, \quad \boldsymbol{\theta} \in \Theta, \end{aligned} \tag{24}$$

where $\| \cdot \|$ stands for the Euclidean norm. Observe that this set may contain several or no elements. Let $\mathbf{H}_n(\boldsymbol{\theta})$ be the Hessian matrix of ℓ_n at $\boldsymbol{\theta}$, that is, the matrix whose (k, l) element is given by

$$H_{n,kl}(\boldsymbol{\theta}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l}.$$

Let $\mathbf{Q}(\boldsymbol{\theta})$ be the $q \times q$ matrix whose (k, l) element writes

$$Q_{kl}(\boldsymbol{\theta}) = \left(\frac{\partial^2 \varphi_1(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l}, \ldots, \frac{\partial^2 \varphi_p(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l} \right) \mathbf{W}^T \left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \boldsymbol{\varphi}(\boldsymbol{\theta}_0) \right),$$

and $\mathbf{H}(\boldsymbol{\theta})$ be the $q \times q$ matrix defined by

$$\mathbf{H}(\boldsymbol{\theta}) = 2 \left(\mathbf{Q}(\boldsymbol{\theta}) + \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta})^T \mathbf{W}^T \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}) \right).$$

Finally write $\overline{B}_\varepsilon(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \Theta : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq \varepsilon\}$ the closed ball around $\boldsymbol{\theta}_0$ with radius $\varepsilon > 0$ and assume that there exists $\varepsilon_0 > 0$ such that $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$ is closed. Then $\overline{B}_\varepsilon(\boldsymbol{\theta}_0)$ is compact for all $0 < \varepsilon \leq \varepsilon_0$.

Lemma 1. *(i) The elementwise convergence $\mathbf{H}_n(\boldsymbol{\theta}) \xrightarrow{P} \mathbf{H}(\boldsymbol{\theta})$ holds uniformly for all $\boldsymbol{\theta}$ in $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$.*

(ii) If $\dot{\boldsymbol{\varphi}}$ is of full rank then, with probability tending to 1, $\mathbf{H}_n(\boldsymbol{\theta})$ is positive definite for all $\boldsymbol{\theta}$ in some closed neighborhood of $\boldsymbol{\theta}_0$.

Proof. (i) It is easily seen that $\mathbf{H}_n(\boldsymbol{\theta}) = 2\left(\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta})^T \widehat{\mathbf{W}}^T \dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}) + \mathbf{Q}_n(\boldsymbol{\theta})\right)$ where $\mathbf{Q}_n(\boldsymbol{\theta})$ is a $q \times q$ matrix such that its (k,l) element is given by

$$Q_{n,kl}(\boldsymbol{\theta}) = \left(\frac{\partial^2 \varphi_1(\boldsymbol{\theta})}{\partial\theta_k \partial\theta_l}, \ldots, \frac{\partial^2 \varphi_p(\boldsymbol{\theta})}{\partial\theta_k \partial\theta_l}\right) \widehat{\mathbf{W}}^T \left(\boldsymbol{\varphi}(\boldsymbol{\theta}) - \mathbf{Y}_n\right).$$

Let $\widehat{W}_{ji}$ denote the element of $\widehat{\mathbf{W}}$ in the j-th row and i-th column. For all $\boldsymbol{\theta}$ in $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$,

$$\begin{aligned}
\frac{1}{2}|H_{n,kl}(\boldsymbol{\theta}) - H_{kl}(\boldsymbol{\theta})| \leq & \sum_{i,j=1}^{p} \left|\frac{\partial\varphi_i(\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial\varphi_j(\boldsymbol{\theta})}{\partial\theta_l}\right| \left|\widehat{W}_{ij} - W_{ij}\right| \\
& + \sum_{i,j=1}^{p} \left|\frac{\partial^2\varphi_i(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}\right| \left|\widehat{W}_{ij} - W_{ij}\right| \left|\varphi_j(\boldsymbol{\theta}) - Y_{n,j}\right| \\
& + \sum_{i,j=1}^{p} \left|\frac{\partial^2\varphi_i(\boldsymbol{\theta})}{\partial\theta_k\partial\theta_l}\right| \left|W_{ij}\right| \left|\varphi_j(\boldsymbol{\theta}_0) - Y_{n,j}\right|.
\end{aligned}$$

Since φ_j and its first and second order derivatives are continuous on the closed and thus compact set $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$, they are uniformly bounded by some constant on this set. Therefore, as $n \to \infty$,

$$\sup_{\boldsymbol{\theta}\in\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)} |H_{n,kl}(\boldsymbol{\theta}) - H_{kl}(\boldsymbol{\theta})| \xrightarrow{P} 0,$$

which follows from the weak consistency of $\mathbf{Y}_n$ and $\widehat{\mathbf{W}}$.

(ii) The matrix $H(\boldsymbol{\theta}_0)$ is positive definite. Hence its eigenvalues are positive. Since an eigenvalue is a continuous function of the parameter vector [33] (Theorem 3.13), for each eigenvalue of H, there exists a neighborhood such that if $\boldsymbol{\theta}$ belongs to this neighborhood, then this eigenvalue at $\boldsymbol{\theta}$ is positive. Hence, since there is only a finite number of eigenvalues, it follows that there exists a neighborhood $\overline{B}_\delta(\boldsymbol{\theta}_0)$ $(\delta > 0)$ on which $H(\boldsymbol{\theta})$ is positive definite.

Now define

$$A_{ij} = \left\{\sup_{\boldsymbol{\theta}\in\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)} |H_{n,ij}(\boldsymbol{\theta}) - H_{ij}(\boldsymbol{\theta})| \leq \inf_{x\in\mathbb{R}^q, x\neq 0,\, \boldsymbol{\theta}\in\overline{B}_\delta(\boldsymbol{\theta}_0)} \frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2\sum_{i,j=1}^q |x_i x_j|}\right\}$$

and put $A = \bigcap_{i,j} A_{ij}$. On the event A, for all $\mathbf{x} \neq \mathbf{0}$ and for all $\boldsymbol{\theta}$ in $\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$, we have

$$\begin{aligned}
\left|\mathbf{x}^T\left(\mathbf{H}(\boldsymbol{\theta}) - \mathbf{H}_n(\boldsymbol{\theta})\right)\mathbf{x}\right| &\leq \sum_{i,j=1}^{q} |x_i x_j| \inf_{x\in\mathbb{R}^q, x\neq 0,\, \boldsymbol{\theta}\in\overline{B}_\delta(\boldsymbol{\theta}_0)} \frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2\sum_{i,j=1}^q |x_i x_j|} \\
&\leq \inf_{\boldsymbol{\theta}\in\overline{B}_\delta(\boldsymbol{\theta}_0)} \frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2}.
\end{aligned}$$

If, moreover, $\boldsymbol{\theta} \in \overline{B}_\delta(\boldsymbol{\theta}_0)$, then

$$\mathbf{x}^T\mathbf{H}_n(\boldsymbol{\theta})\mathbf{x} \geq \frac{\mathbf{x}^T\mathbf{H}(\boldsymbol{\theta})\mathbf{x}}{2} > 0$$

because $\mathbf{H}(\boldsymbol{\theta})$ is positive definite on $\overline{B}_\delta(\boldsymbol{\theta}_0)$. Hence on A and for all $\boldsymbol{\theta}$ in $\overline{B}_\delta(\boldsymbol{\theta}_0)\bigcap\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)$, the matrix $\mathbf{H}_n(\boldsymbol{\theta})$ is positive definite. By (i), $P(A)\to 1$ as $n\to\infty$, which concludes the proof.

Lemma 2. *(i) If $\boldsymbol{\varphi}$ in (23) is an homeomorphism, then for all ε such that $0<\varepsilon\le\varepsilon_0$, as $n\to\infty$,*

$$P\left[\hat{\Theta}_n\neq\emptyset \text{ and } \hat{\Theta}_n\subset\overline{B}_\varepsilon(\boldsymbol{\theta}_0)\right]\to 1.$$

(ii) If, moreover, $\dot{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)$ is of full rank then as $n\to\infty$,

$$P\left[\text{ card } \hat{\Theta}=1\right]\to 1,$$

where card denotes the cardinal of a set. Define $\hat{\boldsymbol{\theta}}$ to be the unique element of $\widehat{\Theta}$ if card $\hat{\Theta}=1$, and any arbitrary point otherwise. Then $\hat{\boldsymbol{\theta}}\xrightarrow{P}\boldsymbol{\theta}_0$ as $n\to\infty$.

(iii) If in addition to the assumptions of (i) and (ii)

$$\sqrt{n}(\mathbf{Y}_n-\boldsymbol{\varphi}(\boldsymbol{\theta}_0))\xrightarrow{d}N_p(\mathbf{0},\boldsymbol{\Sigma})$$

then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\xrightarrow{d}N_q\left(\mathbf{0},\left(\dot{\boldsymbol{\varphi}}^T\mathbf{W}\dot{\boldsymbol{\varphi}}\right)^{-1}\dot{\boldsymbol{\varphi}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\dot{\boldsymbol{\varphi}}\left(\dot{\boldsymbol{\varphi}}^T\mathbf{W}\dot{\boldsymbol{\varphi}}\right)^{-1}\right)$$

Proof. (i) Let $0<\varepsilon\le\varepsilon_0$. Since $\boldsymbol{\varphi}$ is a homeomorphism and $\widehat{\mathbf{W}}$ has full rank, $\widehat{\mathbf{V}}\boldsymbol{\varphi}$ is also an homeomorphism. Hence there exists $\delta>0$ such that $\boldsymbol{\theta}\in\Theta$ and $\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta})-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|\le\delta$ imply $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\varepsilon$. Thus for every $\boldsymbol{\theta}\in\Theta$ with $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\varepsilon$ we have $\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta})-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|>\delta$. On the event $A_n=\{\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta}_0)-\mathbf{Y}_n\right)\|\le\delta/2\}$ and for $\boldsymbol{\theta}$ outside $\overline{B}_\varepsilon(\boldsymbol{\theta}_0)$, the inequality

$$\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta})-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|\le\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta})-\mathbf{Y}_n\right)\|+\|\widehat{\mathbf{V}}\left(\mathbf{Y}_n-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|$$

implies

$$\begin{aligned}\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta})-\mathbf{Y}_n\right)\|&\ge\|\widehat{\mathbf{V}}\left(\boldsymbol{\varphi}(\boldsymbol{\theta})-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|-\|\widehat{\mathbf{V}}\left(\mathbf{Y}_n-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|\\&>\delta-\delta/2\\&=\delta/2\\&\ge\|\widehat{\mathbf{V}}\left(\mathbf{Y}_n-\boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right)\|.\end{aligned}$$

Therefore

$$\min_{\boldsymbol{\theta}\in\overline{B}_\varepsilon(\boldsymbol{\theta}_0)}\|\widehat{\mathbf{V}}\left(\mathbf{Y}_n-\boldsymbol{\varphi}(\boldsymbol{\theta})\right)\|\le\inf_{\boldsymbol{\theta}\notin\overline{B}_\varepsilon(\boldsymbol{\theta}_0)}\|\widehat{\mathbf{V}}\left(\mathbf{Y}_n-\boldsymbol{\varphi}(\boldsymbol{\theta})\right)\|,$$

where in the left hand side the minimum is attained because $\overline{B}_\varepsilon(\boldsymbol{\theta}_0)$ is compact. By consistency of $\mathbf{Y}_n$ and $\widehat{\mathbf{W}}$, we have $P(A_n)\to 1$. It follows that the event $\left\{\hat{\Theta}_n\neq\emptyset \text{ and } \hat{\Theta}_n\subset\overline{B}_\varepsilon(\boldsymbol{\theta}_0)\right\}$ has probability tending to 1.

(ii) Without loss of generality denote by $\overline{B}_\eta(\boldsymbol{\theta}_0)$, $\eta < \varepsilon_0$, the closed neighborhood of Lemma 1 (ii). Assume that the event

$$\left\{\hat{\Theta} \neq \emptyset,\ \hat{\Theta} \subset \overline{B}_\eta(\boldsymbol{\theta}_0) \text{ and } \mathbf{H}_n(\boldsymbol{\theta}) \text{ is positive definite for all } \boldsymbol{\theta} \text{ in } \overline{B}_\eta(\boldsymbol{\theta}_0)\right\} \quad (25)$$

happens. Let $\boldsymbol{\theta} \in \overline{B}_\eta(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}^\star$ be a vector in $\hat{\Theta}$. A Taylor expansion of ℓ_n in (24) at $\boldsymbol{\theta}^\star$ gives

$$\ell_n(\boldsymbol{\theta}) = \ell_n(\boldsymbol{\theta}^\star) + (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \nabla \ell_n(\boldsymbol{\theta}^\star) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\theta}^\star),$$

where $\tilde{\boldsymbol{\theta}} = t\boldsymbol{\theta} + (1-t)\boldsymbol{\theta}^\star$, $t \in (0,1)$ and $\nabla \ell_n$ denotes the gradient of ℓ_n. In view of Lemma 2 (i), $\boldsymbol{\theta}^\star$ is in some open neighborhood of $\boldsymbol{\theta}_0$ and thus $\nabla \ell_n(\boldsymbol{\theta}^\star) = 0$. The fact that $\tilde{\boldsymbol{\theta}} \in \overline{B}_\eta(\boldsymbol{\theta}_0)$ entails that $\mathbf{H}_n(\tilde{\boldsymbol{\theta}})$ is positive definite. Therefore, we have shown that $\ell_n(\boldsymbol{\theta}) > \ell_n(\boldsymbol{\theta}^\star)$ for all $\boldsymbol{\theta}$ in $\overline{B}_\eta(\boldsymbol{\theta}_0)$. This implies that the cardinal of $\hat{\Theta}$ is 1 when (25) holds. By Lemma 1 (ii) and Lemma 2 (i), the event (25) has probability tending to 1, hence, $P[\text{ card } \hat{\Theta} = 1] \to 1$. Now let $\hat{\boldsymbol{\theta}}$ be as in Lemma 2 (ii) and let $\varepsilon > 0$. Without loss of generality, assume that $\varepsilon \leq \varepsilon_0$. Then

$$\lim_{n\to\infty} P\left[\hat{\boldsymbol{\theta}} \in \overline{B}_\varepsilon(\boldsymbol{\theta}_0)\right] = \lim_{n\to\infty} P\left[\hat{\boldsymbol{\theta}} \in \overline{B}_\varepsilon(\boldsymbol{\theta}_0) \text{ and } \text{ card } \hat{\Theta} = 1\right] = 1,$$

the last equality holding because of Lemma 2 (i). Thus the consistency of $\hat{\boldsymbol{\theta}}$ is proved.

(iii) A Taylor expansion for the gradient $\nabla \ell_n$ of ℓ_n in equation (24) around $\boldsymbol{\theta}_0$ entails

$$\nabla \ell_n(\hat{\boldsymbol{\theta}}) = \nabla \ell_n(\boldsymbol{\theta}_0) + \mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\tilde{\boldsymbol{\theta}} = t\hat{\boldsymbol{\theta}} + (1-t)\boldsymbol{\theta}_0$, $t \in (0,1)$. By the same arguments as in the proof of Lemma 2 (ii), $\nabla \ell_n(\hat{\boldsymbol{\theta}}) = 0$, hence,

$$\begin{aligned}
\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \sqrt{n}\left(\nabla \ell_n(\hat{\boldsymbol{\theta}}) - \nabla \ell_n(\boldsymbol{\theta}_0)\right) \\
&= -\sqrt{n}\nabla \ell_n(\boldsymbol{\theta}_0) \\
&= 2\dot{\boldsymbol{\varphi}}^T \widehat{\mathbf{W}} \sqrt{n}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right).
\end{aligned}$$

For $\mathbf{x}$ in $\mathbb{R}^q$, we have

$$\begin{aligned}
P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x}\right] =& P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{ card } \hat{\Theta} = 1\right] \\
&+ P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{ card } \hat{\Theta} \neq 1\right].
\end{aligned} \quad (26)$$

Since the second term in the sum in the right hand side of (26) tends to 0, we have that

$$\begin{aligned}
&\lim_{n\to\infty} P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x} \text{ and } \text{ card } \hat{\Theta} = 1\right] \\
&= \lim_{n\to\infty} P\left[\sqrt{n}\mathbf{H}_n(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \mathbf{x}\right] \\
&= \lim_{n\to\infty} P\left[2\dot{\boldsymbol{\varphi}}^T \widehat{\mathbf{W}} \sqrt{n}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right) \leq \mathbf{x}\right].
\end{aligned}$$

By the assumptions of Lemma 2 (iii) and by consistency of $\widehat{\mathbf{W}}$, we have

$$2\dot{\boldsymbol{\varphi}}^T\widehat{\mathbf{W}}\sqrt{n}\left(\mathbf{Y}_n - \boldsymbol{\varphi}(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N_q\left(0, 4\dot{\boldsymbol{\varphi}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T\dot{\boldsymbol{\varphi}}\right).$$

If $\mathbf{H}_n(\tilde{\boldsymbol{\theta}})$ converges in probability to $\mathbf{H}(\boldsymbol{\theta}_0) = 2\dot{\boldsymbol{\varphi}}\mathbf{W}\dot{\boldsymbol{\varphi}}$, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_q\left(\mathbf{0}, \left(\dot{\boldsymbol{\varphi}}^T\mathbf{W}^T\dot{\boldsymbol{\varphi}}\right)^{-1}\dot{\boldsymbol{\varphi}}^T\mathbf{W}\Sigma\mathbf{W}^T\dot{\boldsymbol{\varphi}}\left[\left(\dot{\boldsymbol{\varphi}}^T\mathbf{W}^T\dot{\boldsymbol{\varphi}}\right)^{-1}\right]^T\right).$$

Therefore, to conclude the proof, it suffices to prove that $\mathbf{H}_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{P} \mathbf{H}(\boldsymbol{\theta}_0)$.

Let $\varepsilon > 0$. Assume that

$$\sup_{\boldsymbol{\theta}\in\overline{B}_{\varepsilon_0}(\boldsymbol{\theta}_0)} |H_{n,ij}(\boldsymbol{\theta}) - H_{ij}(\boldsymbol{\theta})| < \frac{\varepsilon}{2}.$$

The map $\boldsymbol{\theta} \mapsto H_{n,ij}(\boldsymbol{\theta})$ is continuous, hence, there exists $\delta > 0$ such that $|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| < \delta$ implies $|H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{n,ij}(\boldsymbol{\theta}_0)| < \varepsilon/2$. Assume that $\hat{\boldsymbol{\theta}} \in \overline{B}_\delta(\boldsymbol{\theta}_0)$ and suppose without loss of generality that $\delta \leq \varepsilon_0$. Then it holds that

$$\begin{aligned}|H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{ij}(\boldsymbol{\theta}_0)| &\leq |H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{n,ij}(\boldsymbol{\theta}_0)| + |H_{n,ij}(\boldsymbol{\theta}_0) - H_{ij}(\boldsymbol{\theta}_0)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

By Lemma 1 (i) and Lemma 2 (i) we have shown that for all $\varepsilon > 0$, the event $\left\{|H_{n,ij}(\tilde{\boldsymbol{\theta}}) - H_{ij}(\boldsymbol{\theta}_0)| \leq \varepsilon\right\}$ has probability tending to 1. Hence the proof is finished.

Proof of Theorem 1

The proof of Theorem 1 is a direct application of Lemma 2 with $\boldsymbol{\varphi} = \boldsymbol{\mathcal{D}}$ and $\mathbf{Y}_n = \hat{\boldsymbol{\mathcal{D}}}$.

Proof of Corollary 1

(i) The limiting covariance matrix of $\hat{\boldsymbol{\theta}}$, viewed as a function of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\Xi}(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\mathbf{W}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})\left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})\right)^{-1}.$$

By assumption, $\dot{\boldsymbol{\mathcal{D}}}(\cdot)$ and $\boldsymbol{\Sigma}(\cdot)$ are continuous at $\boldsymbol{\theta}_0$, hence so is $\boldsymbol{\Xi}(\cdot)$. Therefore, since $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}_0$, we also have that $\boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})$ converges in probability to $\boldsymbol{\Xi}(\boldsymbol{\theta}_0)$. Moreover, since $\boldsymbol{\Xi}(\boldsymbol{\theta})$ is invertible and nonnegative definite for all $\boldsymbol{\theta}$ in Θ, we have $\boldsymbol{\Xi}(\boldsymbol{\theta}) = \boldsymbol{\Xi}^{1/2}(\boldsymbol{\theta})\boldsymbol{\Xi}^{1/2}(\boldsymbol{\theta})$ where $\boldsymbol{\Xi}^{1/2}(\boldsymbol{\theta})$ is also invertible. Therefore, by Theorem 1, as $n \to \infty$,

$$\sqrt{n}\boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_q),$$

leading to the desired result.

(ii) By Assumption (A4),

$$\sqrt{n}\left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0))$$

as $n \to \infty$. The arguments in the proof of (i) can be easily adapted to prove (ii).

Proof of Corollary 2

The proof of Corollary 2 is similar to that of Corollary 1 (i).

Proof of Corollary 3

Note $\boldsymbol{\mathcal{D}}_0 := \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$ and write

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} = \boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 + \boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}}. \tag{27}$$

A Taylor expansion yields

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 = \widetilde{\dot{\boldsymbol{\mathcal{D}}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{28}$$

where $\widetilde{\dot{\boldsymbol{\mathcal{D}}}} := \dot{\boldsymbol{\mathcal{D}}}(\tilde{\boldsymbol{\theta}})$ with $\tilde{\boldsymbol{\theta}}$ being a vector between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Substitute (28) into (27) to get

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} = \widetilde{\dot{\boldsymbol{\mathcal{D}}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}}. \tag{29}$$

From (28), we have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}})^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T (\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0). \tag{30}$$

Substitute (30) into (29) to obtain

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} = \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \left(\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \right)^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 \right) + \left(\boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}} \right).$$

Since

$$\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\mathcal{D}}_0 = \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) + \left(\widehat{\boldsymbol{\mathcal{D}}} - \boldsymbol{\mathcal{D}}_0 \right),$$

we have

$$\left(\mathbf{I}_p - \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \left(\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \right)^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \right) \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) = \left(\mathbf{I}_p - \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \left(\widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \widetilde{\dot{\boldsymbol{\mathcal{D}}}} \right)^{-1} \widetilde{\dot{\boldsymbol{\mathcal{D}}}}^T \right) \left(\boldsymbol{\mathcal{D}}_0 - \widehat{\boldsymbol{\mathcal{D}}} \right).$$

Take $\boldsymbol{\theta} \in \Theta$ and define $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) := \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \left(\dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta}) \right)^{-1} \dot{\boldsymbol{\mathcal{D}}}(\boldsymbol{\theta})^T$. Likewise, write $\widetilde{\mathbf{A}} := \mathbf{A}(\widetilde{\boldsymbol{\theta}})$. By Assumption (A4) and because $\boldsymbol{\mathcal{D}}$ is continuously differentiable, as $n \to \infty$,

$$(\mathbf{I}_p - \widetilde{\mathbf{A}})\sqrt{n} \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) \xrightarrow{d} N\left(\mathbf{0}, (\mathbf{I}_p - \mathbf{A}_0)\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{A}_0)\right) \tag{31}$$

where $\mathbf{A}_0 := \dot{\boldsymbol{\mathcal{D}}}_0 \left(\dot{\boldsymbol{\mathcal{D}}}_0^T \dot{\boldsymbol{\mathcal{D}}}_0 \right)^{-1} \dot{\boldsymbol{\mathcal{D}}}_0^T$ and $\dot{\boldsymbol{\mathcal{D}}}_0 := \boldsymbol{\mathcal{D}}(\boldsymbol{\theta}_0)$. Now write $\mathbf{I}_p - \mathbf{A}_0 = \mathbf{Q}\boldsymbol{\Delta}\mathbf{Q}^T$, where $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_p$, and $\boldsymbol{\Delta} = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$ with the number of ones being equal to k. Pre-multiply the left member of (31) by

$$\mathbf{Q}^T \left[(\mathbf{I}_p - \mathbf{A}_0)\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{A}_0) + \mathbf{A}_0 \right]^{-1/2} = \left[\boldsymbol{\Delta}\mathbf{Q}^T\boldsymbol{\Sigma}\mathbf{Q}\boldsymbol{\Delta} + \mathbf{I}_p - \boldsymbol{\Delta} \right]^{-1/2} \mathbf{Q}^T,$$

Note that the matrix between the brackets in the right-hand side is bloc-diagonal. It then can be verified that the limit normal distribution in the right member will have covariance matrix $\boldsymbol{\Delta}$, entailing

$$n \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right)^T (\mathbf{I}_p - \widetilde{\mathbf{A}})[(\mathbf{I}_p - \mathbf{A}_0)\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{A}_0) + \mathbf{A}_0]^{-1}(\mathbf{I}_p - \widetilde{\mathbf{A}}) \left(\boldsymbol{\mathcal{D}}(\hat{\boldsymbol{\theta}}) - \widehat{\boldsymbol{\mathcal{D}}} \right) \to \chi^2_k.$$

Put $\widehat{\mathbf{A}} := \mathbf{A}(\hat{\boldsymbol{\theta}})$. Since $\widetilde{\mathbf{A}} \to \mathbf{A}_0$ in probability, we can replace $\widetilde{\mathbf{A}}$ and $\mathbf{A}_0$ by $\widehat{\mathbf{A}}$ to get the desired result.

Proof of Proposition 1

(This proof is adapted from [22] but is given here for sake of completeness.) Without loss of generality, assume that $\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}} = \alpha\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mathcal{D}}}$ for some scalar α. Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{W})$ and note $\hat{\boldsymbol{\theta}}(\mathbf{W}^\star)$ the estimator for which $\mathbf{W} = \mathbf{W}^\star$. Denote by $\boldsymbol{\Xi}(\mathbf{W})$ and $\boldsymbol{\Xi}(\mathbf{W}^\star)$ the associated limiting covariance matrices of Theorem 1. We have

$$
\begin{aligned}
&\boldsymbol{\Xi}(\mathbf{W}) - \boldsymbol{\Xi}(\mathbf{W}^\star) \\
&= \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1} - \alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1} \\
&= \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}} - \dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1} \\
&= \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}^{1/2}\left(\mathbf{I}_p - \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2}\right)\boldsymbol{\Sigma}^{1/2}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}} \\
&\quad \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1},
\end{aligned}
$$

where $\boldsymbol{\Sigma}^{1/2}$ is the symmetric and invertible matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$. Write $\mathbf{A} = \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2}$. Note that $\mathbf{A}$ is idempotent, that is, $\mathbf{A}^2 = \mathbf{A}$. Indeed,

$$
\begin{aligned}
\mathbf{A}^2 =& \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2} \\
=& \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2} \\
=& \boldsymbol{\Sigma}^{-1/2}\dot{\boldsymbol{\mathcal{D}}}\alpha\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}^\star\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\boldsymbol{\Sigma}^{-1/2} \\
=& \mathbf{A}.
\end{aligned}
$$

Hence $\mathbf{I}_p - \mathbf{A}$ is idempotent as well and therefore

$$
\boldsymbol{\Xi}(\mathbf{W}) - \boldsymbol{\Xi}(\mathbf{W}^\star) = \left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\boldsymbol{\Sigma}^{1/2}(\mathbf{I}_p - \mathbf{A})(\mathbf{I}_p - \mathbf{A})\boldsymbol{\Sigma}^{1/2}\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\left(\dot{\boldsymbol{\mathcal{D}}}^T\mathbf{W}\dot{\boldsymbol{\mathcal{D}}}\right)^{-1}
$$

which is easily seen to be nonnegative definite.

Proof of Proposition 2

The gradient of the loss function (3) is equal to $\mathbf{0}$ if and only if

$$
\dot{\boldsymbol{\mathcal{D}}}^T\widehat{\mathbf{W}}\left(\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\mathcal{D}}}\right) = 0.
$$

But since $\dot{\boldsymbol{\mathcal{D}}}$ is of full rank and $p = q$, the kernel of $\dot{\boldsymbol{\mathcal{D}}}^T$ is null, hence

$$
\widehat{\mathbf{W}}\left(\boldsymbol{\mathcal{D}}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\mathcal{D}}}\right) = 0.
$$

The fact that $\widehat{\mathbf{W}}$ is of full rank concludes the proof.

References

[1] D. Berg. Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701, 2009.

[2] A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics*, 39(4):1963–2006, 2011.

[3] A. Bücher, J. Segers, and S. Volgushev. When uniform weak convergence fails: Empirical processes for dependence functions and residuals via epi- and hypographs. *The Annals of Statistics*, 42(4):1598–1634, 2014.

[4] P. Capéraà, A.L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 1997.

[5] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[6] C. M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353, 1981.

[7] P. Deheuvels. On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5):429–439, 1991.

[8] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.

[9] F. Durante and C. Sempi. Copula theory: An introduction. In *Copula Theory and Its Applications*, pages 3–31. Springer, 2010.

[10] J. Einmahl, A. Krajina, and J. Segers. An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics*, 40(3):1764–1793, 2012.

[11] M. Ferreira. Nonparametric estimation of the tail-dependence coefficient. *REVSTAT–Statistical Journal*, 11(1):1–16, 2013.

[12] M. Fréchet. Remarques au sujet de la note précédente. *CR Acad. Sci. Paris Sér. I Math*, 246:2719–2720, 1958.

[13] C. Genest and A. C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

[14] C. Genest, K. Ghoudi, and L. P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.

[15] C. Genest, J. Nešlehová, and N. Ben Ghorbal. Estimators based on Kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2):157–177, 2011.

[16] C. Genest and B. Rémillard. Test of independence and randomness based on the empirical copula process. *Test*, 13(2):335–369, 2004.

[17] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, 44(2):199–213, 2009.

[18] C. Genest and L. P. Rivest. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American statistical Association*, 88(423):1034–1043, 1993.

[19] C. Genest and J. Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022, 2009.

[20] G. Gudendorf and J. Segers. Extreme-value copulas. In *Copula Theory and Its Applications*, pages 127–145. Springer, 2010.

[21] P. Hall and N. Tajvidi. Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(5):835–844, 2000.

[22] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[23] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

[24] H. Joe. *Multivariate models and dependence concepts.* Chapman & Hall/CRC, Boca Raton, FL, 2001.

[25] C. Klüppelberg and G. Kuhn. Copula structure analysis. *Journal of the Royal Statistical Society: Series B*, 71(3):737–753, 2009.

[26] I. Kojadinovic and J. Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21(1):17–30, 2011.

[27] P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101, 2013.

[28] G. Mazo, S. Girard, and F. Forbes. A flexible and tractable class of one-factor copulas. *Statistics and Computing*, to appear, 2015.

[29] R. B. Nelsen. *An introduction to copulas.* Springer, 2006.

[30] R. B. Nelsen, J. J. Quesada-Molina, J.A. Rodríguez-Lallena, and M. Úbeda-Flores. Kendall distribution functions. *Statistics & Probability Letters*, 65(3):263–268, 2003.

[31] D. H. Oh and A. J. Patton. Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502):689–700, 2013.

[32] J. Pickands. Multivariate extreme value distributions. *Proceedings of the 43rd Session of the International Statistical Institute*, 2:859–878, 1981.

[33] J. R. Schott. *Matrix analysis for statistics.* Wiley, 2005.

[34] J. Segers. Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782, 2012.

[35] H. Tsukahara. Semiparametric estimation in copula models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 33(3):357–375, 2005.

[36] A.W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Chapitre 6

Une classe de copules maniable et flexible

La flexibilité et la maniabilité de copules multivariées sont deux propriétés désirées mais plutôt antagonistes. Une copule donnée est souvent soit maniable, soit flexible, mais rarement les deux à la fois. Prenons les copules archimédiennes, vues dans la partie 2.1 du chapitre 2 : elles sont maniables (on peut par exemple calculer leur tau de Kendall facilement) mais pas du tout flexibles (elles n'ont qu'un ou deux paramètres). A l'opposé, on pourrait prendre les *Vines* (vues dans la partie 2.3), qui sont extrêmement flexibles, mais au prix d'une modélisation complexe. Ce sont les copules elliptiques qui réalisent le meilleur compromis entre flexibilité et maniabilité. Les copules elliptiques ont un nombre de paramètres de l'ordre de d^2 (rappelons que d est le nombre de variables), mais on peut rendre ce nombre beaucoup plus petit en imposant une structure à la matrice de corrélation. Le principal reproche fait aux copules elliptiques réside dans le fait que leur queues de distribution supérieures et inférieures sont symétriques. Ces copules ont été vues dans la partie 2.4 chapitre 2.

Dans ce chapitre, nous construisons une classe (appelée classe FDG) de copules multivariées tout à fait intéressante en ceci qu'elle allie à la fois flexibilité et maniabilité. Sa construction est très simple : elle ne nécessite rien d'autre que de se donner $U_0, U_1, \ldots, U_d$ des variables aléatoires distribuées selon une loi uniforme standard, et telles que (i) les U_i sont indépendantes sachant la variable U_0, et (ii) le couple $(U_0, U_i) \sim C_{f_i} \equiv C_{0i}$ est distribué selon une copule de la classe de Durante (vue dans la partie 1.3.2) pour chaque indice i. En utilisant la propriété d'indépendance conditionnelle, la loi jointe C de $(U_1, \ldots, U_d)$ s'écrit très facilement comme

$$C(u_1, \ldots, u_d) = \int_0^1 C_{1|0}(u_1|u_0) \ldots C_{d|0}(u_d|u_0)\, du_0, \tag{6.1}$$

où $C_{i|0}(u_i|u_0)$ est la loi conditionnelle de U_i sachant $U_0 = u_0$. En tirant profit de la forme de la copule de Durante, rappelée ci-dessous,

$$C_{0i}(u_0, u_i) = \min(u_0, u_i) f_i(\max(u_0, u_i)),$$

où f_i est le générateur de la copule, nous pouvons même calculer l'intégrale (6.1) et obtenir une expression explicite de la copule. En particulier, les marges appartiennent toujours à la classe de Durante. Egalement, et c'est très intéressant,

on peut obtenir toutes les combinaisons de dépendance de queue inférieures et supérieures. Par exemple, les marges bivariées peuvent être dépendantes dans la queue supérieure, mais indépendantes dans la queue inférieure, ou l'inverse, ou dépendantes dans les deux queues. La classe FDG présente aussi un intérêt pour réaliser une analyse statistique de valeurs extrêmes, comme par exemple étudier la distribution statistique des maxima d'un échantillon, car il est possible de calculer explicitement les copules des valeurs extrêmes associées aux copules FDG. Dans l'article présenté ci-dessous, une illustration sur un jeu de données réelles tiré de l'hydrologie est fournie : des niveaux critiques associés à des évènements hydrométriques extrêmes sont estimés. L'article présenté ci-dessous a été soumis pour publication, et est disponible à l'adresse `http://hal.archives-ouvertes.fr/hal-00979147`.

A flexible and tractable class of one-factor copulas

Gildas Mazo, Stéphane Girard and Florence Forbes

MISTIS, Inria - Laboratoire Jean Kuntzmann, France

Abstract

Copulas are a useful tool to model multivariate distributions. While there exist various families of bivariate copulas, the construction of flexible and yet tractable copulas suitable for high-dimensional applications is much more challenging. This is even more true if one is concerned with the analysis of extreme values. In this paper, we construct a class of one-factor copulas and a family of extreme-value copulas well suited for high-dimensional applications and exhibiting a good balance between tractability and flexibility. The inference for these copulas is performed by using a least-squares estimator based on dependence coefficients. The modeling capabilities of the copulas are illustrated on simulated and real datasets.

Keywords: tractable, flexible, extreme-value copula, factor copula, multivariate, high-dimension, copula.

1 Introduction

The modelling of random multivariate events (i.e., of dimension strictly greater than 2) is a central problem in various scientific domains and the construction of multivariate distributions able to properly model the variables at play is challenging. The challenge is even more diffcult if the data provide evidence of tail dependencies or non Gaussian behaviors. To address this problem, the concept of copulas is a useful tool as it permits to impose a dependence structure on pre-determined marginal distributions. Standard books covering this subject include [19,27]. See also [16] for an introduction to this topic. The most common copula models used in high dimensional applications are discussed below.

The popular Archimedean copulas are tractable and allow to model a different behavior in the lower and upper tails. For instance, the Gumbel copula is upper, but no lower, tail dependent; the opposite holds for the Clayton copula. Nevertheless, the dependence structure of Archimedean copulas is severely restricted because they are exchangeable, implying that all the pairs of variables have the same distribution. More details about these copulas can be found in the above mentionned books.

Nested Archimedean copulas are a class of hierarchical copulas generalizing the class of Archimedean copulas. They allow to introduce asymmetry in the dependence structure but only between groups of variables. This hierarchical structure is not desirable when no prior knowledge of the random phenomenon under consideration is available. Furthermore, constraints on the parameters

restrict the tractability of these copulas. These copulas first appeared in [19] Section 4.2.

The class of elliptical copulas arises from the class of elliptical distributions. These copulas are interesting in many respect but they are tail symmetric, meaning that the lower tail dependence coefficient is equal to the upper tail dependence coefficient (these coefficients are defined in Section 2.2). This may not be the case in applications. See, e.g., [26] Section 5 or [14] for an introduction to these copulas.

Pair copula constructions and Vines are flexible copula models based on the decomposition of the density as a product of conditional bivariate copulas. However, these models are difficult to handle. Furthermore, the conditional bivariate copulas are typically assumed not to depend on the conditioning variables. This so called *simplifying assumption* can be misleading, as remarked in [2]. Pair-copula constructions first appeared in [19] Section 4.5. See also [4, 5, 23] for theoretical developments and [1] for a practical introduction to modelling with Vines.

As shown above, most copula models are either tractable or flexible, but rarely both. In this paper, we propose a tractable and yet flexible class of one-factor copulas well suited for high-dimensional applications. This class is nonparametric, and, therefore, encompasses many distributions with different features. Unlike elliptical copulas, the members of this class allow for tail asymmetry. Furthermore, we have derived the associated extreme-value copulas, and, therefore, the analysis of extreme values can be carried out with the presented models. Finally, we show how to perform theoretically well-grounded, and practically fast and accurate, inference of these copulas, thanks to the ability of calculating explicitly the dependence coefficients.

The remaining of this paper is as follows. Section 2 presents the proposed class of one-factor copulas, Section 3 deals with inference, and, in Section 4, the proposed copulas are applied to simulated and real datasets. The proofs are postponed to the Appendix.

2 A tractable and flexible class of one-factor copulas

The class of copulas proposed in this paper, referred to as the FDG class (see Section 2.2 for an explanation of this acronym), can be embedded in the framework of one-factor models. We therefore introduce the later in Section 2.1. The construction and properties of FDG copulas are given in Section 2.2. Parametric examples are proposed in Section 2.3. The extreme-value copulas associated to the FDG class are derived in Section 2.4.

2.1 One-factor copulas

By definition, the coordinates of a random vector distributed according to a one-factor copula [22] are independent given a latent factor. More precisely, let $U_0, U_1, \ldots, U_d$ be standard uniform random variables such that the coordinates of $(U_1, \ldots, U_d)$ are conditionally independent given U_0. The variable U_0 plays the role of a latent, or unobserved, factor. Let us write C_{0i} the distribution of (U_0, U_i) and $C_{i|0}(\cdot|u_0)$ the conditional distribution of U_i given $U_0 = u_0$ for

$i = 1, \ldots, d$. It is easy to see that the distribution of $(U_1, \ldots, U_d)$, called a one-factor copula, is given by

$$C(u_1, \ldots, u_d) = \int_0^1 C_{1|0}(u_1|u_0) \ldots C_{d|0}(u_d|u_0)\, du_0. \tag{1}$$

The copulas C_{0i} are called the *linking copulas* because they link the factor U_0 to the variables of interest U_i. The one-factor model has many advantages to address high dimensional problems. We recall and briefly discuss them below.

Nonexchangeability. The one-factor model is nonexchangeable. Recall that a copula C is said to be *exchangeable* if $C(u_1, \ldots, u_d) = C(u_{\pi(1)}, \ldots, u_{\pi(d)})$ for any permutation π of $(1, \ldots, d)$. This means in particular that all the bivariate marginal distributions are equal to each other. For example, Archimedean copulas are exchangeable copulas. Needless to say, this assumption may be too strong in practice.

Parsimony. The one-factor model is parsimonious. Indeed, only d linking copulas are involved in the construction of the one-factor model, and since they are typically governed by one parameter, the number of parameters in total is no more than d, which increases only linearly with the dimension. Parsimony is more and more desirable as the dimension increases.

Random generation. The conditional independence property of the one factor model allows to easily generate data $(U_1, \ldots, U_d)$ from this copula.

1 Generate $U_0, V_1, \ldots, V_d$ independent standard uniform random variables.

2 For $i = 1, \ldots, d$, put $U_i = C_{i|0}^{-1}(V_i|U_0)$ where $V \mapsto C_{i|0}^{-1}(V|U_0)$ denotes the inverse of $V \mapsto C_{i|0}(V|U_0)$.

Dependence properties of the one-factor model have been studied in [22]. The authors investigated how positive dependence properties of the linking copulas extend to the bivariate margins

$$C_{ij}(u_i, u_j) := C(1, \ldots, 1, u_i, 1, \ldots, 1, u_j, 1, \ldots, 1).$$

These properties included positive quadrant dependence, increasing in the concordance ordering, stochastic increasing, and tail dependence. For details about these dependence concepts, see [19] Section 2. The copulas proposed in this paper, presented in Section 2.2 and Section 2.4, possess simple expressions and therefore the properties mentionned above can be made more precise.

2.2 Construction and properties of FDG copulas

The class of FDG copulas is constructed by choosing appropriate linking copulas for the one-factor copula model (1). The class of linking copulas which served to build the FDG copulas is referred to as the *Durante* class [9] of bivariate copulas, which can also be viewed as part of the framework of [3]. The Durante class consists of the copulas C of the form

$$C(u, v) = \min(u, v) f(\max(u, v)), \tag{2}$$

where $f : [0,1] \to [0,1]$, called the *generator* of C, is a differentiable and increasing function such that $f(1) = 1$ and $t \mapsto f(t)/t$ is decreasing. Hence the choice of the acronym, FDG, which stands for "one-Factor copula with Durante Generators". The advantages of taking Durante linking copulas are twofold: the integral (1) can be calculated and the resulting multivariate copula is non-parametric.

Theorem 1. *Let C be defined by* (1) *and assume that C_{0i} belongs to the Durante class* (2) *with given generator f_i. Then*

$$C(u_1,\dots,u_d) = u_{(1)} \left[\left(\prod_{j=2}^{d} u_{(j)} \right) \int_{u_{(d)}}^{1} \prod_{j=1}^{d} f_j'(x)dx + f_{(1)}(u_{(2)}) \left(\prod_{j=2}^{d} f_{(j)}(u_{(j)}) \right) \right. \tag{3}$$
$$\left. + \sum_{k=3}^{d} \left(\prod_{j=2}^{k-1} u_{(j)} \right) \left(\prod_{j=k}^{d} f_{(j)}(u_{(j)}) \right) \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} f_{(j)}'(x)dx \right],$$

where $u_{(i)} := u_{\sigma(i)}$, $f_{(i)} := f_{\sigma(i)}$ and σ is the permutation of $(1,\dots,d)$ *such that $u_{\sigma(1)} \le \cdots \le u_{\sigma(d)}$.*

The particularity of the copula expression (3) is that it depends on the generators through their reordering underlain by the permutation σ. For instance, with $d = 3$ and $u_1 < u_3 < u_2$ we have $u_{(1)} = u_1$, $u_{(2)} = u_3$, $u_{(3)} = u_2$, $\sigma = \{1,3,2\}$ and $f_{(1)} = f_{\sigma(1)} = f_1$, $f_{(2)} = f_{\sigma(2)} = f_3$, $f_{(3)} = f_{\sigma(3)} = f_2$. This feature gives its flexibility to the model. Observe also that $C(u_1,\dots,u_d)$ writes as $u_{(1)}$ multiplied by a functional of $u_{(2)},\dots,u_{(d)}$, form that is similar to (2). Although the expression of a FDG copula has the merit to be explicit, it is rather cumbersome. Hence, we shall continue its analysis through the prism of its bivariate margins.

Proposition 1. *Let C_{ij} be a bivariate margin of the FDG copula* (3)*. Then C_{ij} belongs to the Durante class* (2) *with generator*

$$f_{ij}(t) = f_i(t)f_j(t) + t\int_t^1 f_i'(x)f_j'(x)dx.$$

In other words,

$$C_{ij}(u_i,u_j) = C_{f_{ij}}(u_i,u_j) = \min(u_i,u_j)f_{ij}(\max(u_i,u_j)).$$

In view of Proposition 1, the FDG copula can be regarded as a multivariate generalization of the Durante class of bivariate copulas. In fact, such a generalization was already proposed in the literature [11]:

$$C_f(u_1,\dots,u_d) = u_{(1)} \prod_{i=2}^{d} f(u_{(i)}),$$

where f is a generator in the usual sense of the Durante class of bivariate copulas. Nonetheless, since there is only one generator to determine the whole copula in arbitrary dimension, this generalization lacks flexibility to be used in

applications. This issue is overcome by the FDG copula. To illustrate this further, its pairwise dependence coefficients are given next. (Note that, since the bivariate margins of the FDG copula belong to the Durante class of bivariate copulas, a more detailed account of their properties can be found in the original paper [9]). Recall that Spearman's rho ρ, Kendall's tau τ, the lower $\lambda^{(L)}$ and upper $\lambda^{(U)}$ tail dependence coefficients of a general bivariate copula C are respectively given by

$$\rho = 12 \int_{[0,1]^2} C(u,v)dudv - 3,\ \tau = 4 \int_{[0,1]^2} C(u,v)dC(u,v) - 1, \tag{4}$$
$$\lambda^{(L)} = \lim_{u \downarrow 0} \frac{C(u,u)}{u},\ \text{and}\ \lambda^{(U)} = \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}.$$

In the case where C belongs to the Durante class with generator f, these coefficients are respectively given by

$$\rho = 12 \int_0^1 x^2 f(x)dx - 3,\ \tau = 4 \int_0^1 x f(x)^2 dx - 1,$$
$$\lambda^{(L)} = f(0),\ \text{and}\ \lambda^{(U)} = 1 - f'(1). \tag{5}$$

Hence, to get the dependence coefficients of the FDG bivariate margins, it is enough to apply the above formulas and Proposition 1. The obtained coefficient expressions are given in Proposition 2 below.

Proposition 2. *The Spearman's rho, Kendall's tau, the lower and upper tail dependence coefficients of the FDG bivariate margins C_{ij} are respectively given by*

$$\rho_{ij} = 12 \int_0^1 x^2 f_i(x) f_j(x) dx + 3 \int_0^1 x^4 f_i'(x) f_j'(x) dx - 3,$$
$$\tau_{ij} = 4 \int_0^1 x \left(f_i(x) f_j(x) + x \int_x^1 f_i'(t) f_j'(t) dt \right)^2 dx - 1,$$
$$\lambda_{ij}^{(L)} = \lambda_i^{(L)} \lambda_j^{(L)}\ \textit{and}$$
$$\lambda_{ij}^{(U)} = \lambda_i^{(U)} \lambda_j^{(U)},$$

where $\lambda_i^{(L)} := f_i(0)$, $\lambda_i^{(U)} := 1 - f_i'(1)$, $i = 1, \ldots, d$ are the lower and upper tail dependence coefficients of the bivariate linking copulas respectively.

2.3 Examples of parametric families

Four examples of families indexed by a real parameter for the generators $f_1, \ldots, f_d$ are given below.

Example 1 (Cuadras-Augé generators)**.** *In* (3)*, let*

$$f_i(t) = t^{1-\theta_i}, \quad \theta_i \in [0,1]. \tag{6}$$

A copula belonging to the Durante class with generator (6) *gives rise to the well known Cuadras-Augé copula with parameter θ_i [7]. By Proposition 1, the*

generator for the bivariate margin C_{ij} of the FDG copula is given by

$$f_{ij}(t) = \begin{cases} t^{2-\theta_i-\theta_j}\left(1 - \frac{(1-\theta_i)(1-\theta_j)}{1-\theta_i-\theta_j}\right) + t\frac{(1-\theta_i)(1-\theta_j)}{1-\theta_i-\theta_j} & \text{if } \theta_i+\theta_j \neq 1 \\ t(1-(1-\theta)\theta\log t) & \text{if } \theta = \theta_j = 1-\theta_i. \end{cases}$$

The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by

$$\rho_{ij} = \frac{3\theta_i\theta_j}{5-\theta_i-\theta_j},\ \lambda_{ij}^{(L)} = 0 \text{ and } \lambda_{ij}^{(U)} = \theta_i\theta_j.$$

The Kendall's tau is given by

$$\tau_{ij} = \begin{cases} \frac{\theta_i\theta_j(\theta_i\theta_j+6-2(\theta_i+\theta_j))}{(\theta_i+\theta_j)^2-8(\theta_i+\theta_j)+15} & \text{if } \theta_i+\theta_j \neq 1 \\ \frac{\theta(\theta-1)(\theta^2-\theta-4)}{8} & \text{if } \theta = \theta_i = 1-\theta_j. \end{cases}$$

Example 2 (Fréchet generators)**.** *In* (3)*, let*

$$f_i(t) = (1-\theta_i)t + \theta_i, \quad \theta_i \in [0,1]. \tag{7}$$

A copula belonging to the Durante class with generator (7) *gives rise to the well known Fréchet copula with parameter θ_i [15]. By Proposition 1, the generator for the bivariate margin C_{ij} of the FDG copula is given by*

$$f_{ij}(t) = (1-\theta_i\theta_j)\,t + \theta_i\theta_j.$$

By noting that f_{ij} is of the form (7) *with parameter $\theta_i\theta_j$, one can see that the bivariate margins of the FDG copula based on Fréchet generators are still Fréchet copulas. The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by*

$$\rho_{ij} = \lambda_{ij}^{(L)} = \lambda_{ij}^{(U)} = \theta_i\theta_j,$$

Kendall's tau is given by

$$\tau_{ij} = \frac{\theta_i\theta_j(\theta_i\theta_j+2)}{3}.$$

Example 3 (Durante-sinus generators)**.** *In* (3)*, let*

$$f_i(t) = \frac{\sin(\theta_i t)}{\sin(\theta_i)}, \quad \theta_i \in (0, \pi/2]. \tag{8}$$

This generator was proposed in [9]. By Proposition 1, the generator for the bivariate margin C_{ij} of the FDG copula is given by

$$f_{ij}(t) = \frac{\sin(\theta_i t)\sin(\theta_j t)}{\sin(\theta_i)\sin(\theta_j)} + \frac{t\theta_i\theta_j}{2(\theta_j^2-\theta_i^2)\sin(\theta_i)\sin(\theta_j)} \times$$

$$\Big\{(\theta_i+\theta_j)\left[\sin\left((\theta_i-\theta_j)t\right) + \sin\left(\theta_j-\theta_i\right)\right]$$

$$+ (\theta_j-\theta_i)\left[\sin(\theta_i+\theta_j) - \sin\left((\theta_i+\theta_j)t\right)\right]\Big\} \text{ if } \theta_i \neq \theta_j, \text{ and}$$

$$f_{ij}(t) = \frac{4\sin(t\theta)^2 + t\theta\left(2(1-t)\theta + \sin(2\theta) - \sin(2t\theta)\right)}{4\sin(\theta)^2} \text{ if } \theta_i = \theta_j = \theta.$$

The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by

$$\rho_{ij} = 12(\sin\theta_i \sin\theta_j)^{-1} \int_0^1 x^2 \sin(\theta_i x)\sin(\theta_j x) + \frac{1}{4}\theta_i\theta_j x^4 \cos(\theta_i x)\cos(\theta_j x)dx - 3, \tag{9}$$

$$\lambda_{ij}^{(L)} = 0 \text{ and } \lambda_{ij}^{(U)} = \left(1 - \frac{\theta_i}{\tan(\theta_i)}\right)\left(1 - \frac{\theta_j}{\tan(\theta_j)}\right). \tag{10}$$

Example 4 (Durante-exponential generators)**.** *In* (3)*, let*

$$f_i(t) = \exp\left(\frac{t^{\theta_i} - 1}{\theta_i}\right), \quad \theta_i > 0 \tag{11}$$

This generator was proposed in [9]. By Proposition 1, the generator for the bivariate margin C_{ij} of the FDG copula is given by

$$f_{ij}(t) = \exp\left(\frac{t^{\theta_i} - 1}{\theta_i} + \frac{t^{\theta_j} - 1}{\theta_j}\right) + t\int_t^1 \exp\left(\frac{x^{\theta_i} - 1}{\theta_i} + \frac{x^{\theta_j} - 1}{\theta_j}\right) x^{\theta_i+\theta_j-2}dx.$$

The Spearman's rho, the lower and upper tail dependence coefficients are respectively given by

$$\rho_{ij} = 12\int_0^1 \exp\left(\frac{x^{\theta_i} - 1}{\theta_i} + \frac{x^{\theta_j} - 1}{\theta_j}\right)\left(x^2 + \frac{1}{4}x^{2+\theta_i+\theta_j}\right)dx - 3,$$

$$\lambda_{ij}^{(L)} = \exp\left(-\frac{1}{\theta_i} - \frac{1}{\theta_j}\right), \text{ and } \lambda_{ij}^{(U)} = 0.$$

Remark 1. *The calculation of the integral in* (9) *with $\theta_i = \theta_j = \pi/2$ shows that for the FDG copula with Durante-sinus generators, Spearman's rho is such that*

$$0 \le \rho_{ij} \le \frac{3\pi^4 - 100\pi^2 + 840}{40\pi^2} \simeq 0.37.$$

The Spearman's rho values for all the other models in the examples above spread the entire interval $[0, 1]$.

The four examples above allow to get all possible types of tail dependencies, as shown in Table 1. The Cuadras-Augé and Durante-sinus families allow for upper but no lower tail dependence, the Durante-exponential family allows for lower but no upper tail dependence, and the Fréchet family allows for both. In the Fréchet case, furthermore, the lower and upper tail dependence coefficients are equal: this is called tail symmetry, a property of elliptical copulas.

2.4 Extreme-value attractors associated to FDG copulas

Extreme-value copulas are theoretically well-grounded copulas to perform a statistical analysis of extreme values such as maxima of random samples. Recall that a copula $C_\#$ is an extreme-value copula if there exists a copula $\tilde{C}$ such that

$$C_\#(u_1, \ldots, u_d) = \lim_{n\uparrow\infty} \widetilde{C}^n(u_1^{1/n}, \ldots, u_d^{1/n}), \quad (u_1, \ldots, u_d) \in [0, 1]^d, \tag{12}$$

family of generators	$\lambda_{ij}^{(L)}$	$\lambda_{ij}^{(U)}$
Cuadras-Augé	0	$\theta_i\theta_j$
Fréchet	$\theta_i\theta_j$	$\theta_i\theta_j$
Durante-sinus	0	$(1-\frac{\theta_i}{\tan\theta_i})(1-\frac{\theta_j}{\tan\theta_j})$
Durante-exponential	$\exp(-\frac{1}{\theta_i}-\frac{1}{\theta_j})$	0

Table 1: Lower $\lambda_{ij}^{(L)}$ and upper $\lambda_{ij}^{(U)}$ tail dependence coefficients for the four families presented in Section 2.3.

see, e.g. [17]. The extreme-value copula $C_\#$ is called the *attractor* of $\widetilde{C}$ and $\widetilde{C}$ is said to belong to the *domain of attraction* of $C_\#$. The class of extreme-value copulas corresponds exactly to the class of max-stable copulas, that is, the copulas $C_\#$ such that

$$C_\#^n(u_1^{1/n},\ldots,u_d^{1/n}) = C_\#(u_1,\ldots,u_d), \quad n\geq 1,\ (u_1,\ldots,u_d)\in[0,1]^d.$$

The upper tail dependence coefficient of a (bivariate) extreme-value copula $C_\#$ has the particular form

$$\lambda^{(U)} = 2+\log C_\#(e^{-1},e^{-1}). \tag{13}$$

This coefficient is a natural dependence coefficient for extreme-value copulas because of the following representation on the diagonal of the unit square:

$$C_\#(u,u) = u^{2-\lambda}, \tag{14}$$

where $\lambda := \lambda^{(U)}$. If $\lambda = 0$ then $C_\#(u,u) = \Pi(u,u) = u^2$, where Π stands for the independence copula. If $\lambda = 1$ then $C_\#(u,u) = M(u,u) = \min(u,u) = u$, where M stands for the Fréchet-Hoeffding upper bound for copulas, that is, the case of perfect dependence. In the case of extreme-value copulas, this interpolation between Π and M allows to interpret λ as a coefficient that measures general dependence, not only dependence in the tails. In order to emphasize this interpretation, λ will be referred to as the *extremal dependence coefficient* of an extreme-value copula. See [6] for more about extreme-value statistics, and, see, e.g. [17] for an account about extreme-value copulas.

In the case of FDG copulas, the limit (12) can be calculated. This leads to a new family of extreme-value copulas, referred to as the family EV-FDG. The bivariate margins $C_{\#,ij}$ of this new family are Cuadras-Augé copulas. These results are precised in Theorem 2 and Proposition 3, given next.

Theorem 2. *Assume that the generators f_i of the FDG copula are twice continuously differentiable on $[0,1]$. Then, the attractor $C_\#$ of the FDG copula exists and is given by*

$$C_\#(u_1,\ldots,u_d) = \prod_{i=1}^d u_{(i)}^{\chi_i}, \tag{15}$$

where

$$\chi_i = \left(\prod_{j=1}^{i-1}(1-\lambda_{(j)})\right)\lambda_{(i)} + 1 - \lambda_{(i)},$$

with the convention that $\prod_{j=1}^{0}(1-\lambda_{(j)}) = 1$ *and where* $\lambda_i = 1 - f_i'(1)$. *As in* (3), $u_{(i)} = u_{\sigma(i)}$ *and* $f_{(i)}'(1) = f_{\sigma(i)}'(1)$ *where* σ *is the permutation of* $(1, \ldots, d)$ *such that* $u_{(1)} \leq \cdots \leq u_{(d)}$.

Proposition 3. *Let* $C_{\#,ij}$ *be a bivariate margin of an EV-FDG copula* (15). *Then* $C_{\#,ij}$ *is a Cuadras-Augé copula with parameter (and therefore extremal dependence coefficient)* $\lambda_i\lambda_j$. *In other words,*

$$C_{\#,ij}(u_i, u_j) = \min(u_i, u_j)\max(u_i, u_j)^{1-\lambda_i\lambda_j}. \tag{16}$$

Remark 2. *In view of Table 1, the FDG copulas with Cuadras-Augé and Fréchet generators both lead to the same EV-FDG copula.*

Multivariate generalizations of the bivariate Cuadras-Augé copula were already proposed in the literature [12,24], but they are less flexible than EV-FDG. Thus, let

$$A(u_1, \ldots, u_d) = u_{(1)} \prod_{i=2}^{d} u_{(i)}^{a_i},$$

where $(a_1 = 1, a_2, a_3, \ldots, a_d)$ is a d-monotone sequence of real numbers, that is, a sequence which satisfies $\nabla^{j-1}a_k \geq 0$, $k = 1, \ldots, d$, $j = 1, \ldots, d-k+1$ where $\nabla^j a_k = \sum_{i=0}^{j}(-1)^i\binom{j}{i}a_{k+i}$, $j, k \geq 1$ and $\nabla^0 a_k = a_k$. This exchangeable copula was proposed in [24]. In particular, the bivariate margins are Cuadras-Augé copulas

$$A_{ij}(u_i, u_j) = \min(u_i, u_j)\max(u_i, u_j)^{a_2}$$

with the same parameter $1 - a_2$. This means that all of them exhibit the same statistical behavior. For instance, all the upper tail dependence coefficients are equal and are given by $1 - a_2$. This is far too restrictive for most applications. Now let

$$B(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i^{1-\sum_{j=1, j\neq i}^{d}\lambda_{ij}} \prod_{i<j} \min(u_i, u_j)^{\lambda_{ij}},$$

where $\lambda_{ij} \in [0, 1], \lambda_{ij} = \lambda_{ji}$ and

$$\sum_{j=1,\ldots,d;\, j\neq i} \lambda_{ij} \leq 1, \quad i = 1, \ldots, d. \tag{17}$$

This copula was proposed in [12]. The bivariate margins B_{ij} are Cuadras-Augé copulas

$$B_{ij}(u_i, u_j) = \min(u_i, u_j)\max(u_i, u_j)^{1-\lambda_{ij}}$$

with parameters λ_{ij}. Unlike the copula A, the tail dependence coefficients can take distinct values from each other. Unfortunately, the constraints (17) are quite restrictive, as it was already stressed by the original authors in [12]. To summarize, the class EV-FDG achieves greater flexibility than its competitors. In particular, one can obtain different bivariate marginal distributions with *no* conditions on the parameters.

3 Parametric inference

Let $(X_1, \ldots, X_d)$ be a random vector following a distribution F with continuous margins $F_1, \ldots, F_d$. Suppose that its copula, C, is a FDG copula defined by (3). Denote by

$$(X_1^{(k)}, \ldots, X_d^{(k)}),\ k = 1, \ldots, n,$$

independent and identically distributed observations obtained from F. Suppose that all the generators f_i of the FDG copula belong to the same parametric family $\{f_\theta,\ \theta \in \Theta \subset \mathbb{R}\}$, that is, there exists $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0d}) \in \Theta^d$ such that $f_{\theta_{0i}} = f_i$. The generators f_i are regarded as functions defined over the product space $[0,1] \times \Theta$ and we write $f_i(t) = f(t, \theta_i)$ for all t in $[0,1]$. The nonparametric inference problem has turned into a parametric one where the parameter vector $\boldsymbol{\theta}_0 \in \Theta^d$ has to be estimated.

In order to estimate the parameters of the FDG and EV-FDG copulas, we consider a least-squares estimator based on dependence coefficients. Its construction is given below. Choose a type of dependence coefficient (Spearman's rho, Kendall's tau, tail dependence coefficient, etc) and denote by $r(\theta_i, \theta_j)$ the chosen dependence coefficient between the variables X_i and X_j. Suppose that the map r is continuous and symmetric in its arguments. Let $p = d(d-1)/2$ be the number of variable pairs (X_i, X_j), $i < j$. Denote by $\mathbf{r}$ be the p-variate map defined on Θ^d such that $\mathbf{r}(\theta_1, \ldots, \theta_d) = (r(\theta_1, \theta_2), \ldots, r(\theta_{d-1}, \theta_d))$. The least-squares estimator based on dependence coefficients is defined as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta^d}{\arg\min} \left\| \hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta}) \right\|^2, \tag{18}$$

where the quantity $\hat{\mathbf{r}} = (\hat{r}_{1,2}, \ldots, \hat{r}_{d-1,d})$ is an empirical estimator of $\mathbf{r}(\boldsymbol{\theta_0})$. The empirical coefficient $\hat{\mathbf{r}}$ has to be chosen such that, as $n \to \infty$,

$$\hat{\mathbf{r}} \xrightarrow{P} \mathbf{r}(\boldsymbol{\theta_0}) \text{ and } \sqrt{n}(\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta_0})) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{19}$$

for some symmetric and positive definite matrix $\boldsymbol{\Sigma}$. For example, the convergences (19) hold for Spearman's rho and Kendall's tau dependence coefficients, see [18].

Example 5 (Spearman's rho)**.** *Let $r(\theta_i, \theta_j)$ be Spearman's rho* (4) *of (X_i, X_j). Let $\hat{U}_i^{(k)} = \sum_{l=1}^n \mathbf{1}(X_i^{(l)} \le X_i^{(k)})/(n+1)$ and put*

$$\hat{r}_{i,j} = \frac{\sum_{k=1}^n \left(\hat{U}_i^{(k)} - \overline{\hat{U}}_i\right)\left(\hat{U}_j^{(k)} - \overline{\hat{U}}_j\right)}{\left[\sum_{k=1}^n \left(\hat{U}_i^{(k)} - \overline{\hat{U}}_i\right)^2 \sum_{k=1}^n \left(\hat{U}_j^{(k)} - \overline{\hat{U}}_j\right)^2\right]^{1/2}},$$

where $\overline{\hat{U}}_i = \sum_{k=1}^n U_i^{(k)}/n$. Then (19) *holds.*

Example 6 (Kendall's tau)**.** *Let $r(\theta_i, \theta_j)$ be Kendall's tau* (4) *of (X_i, X_j) and put*

$$\hat{r}_{i,j} = \binom{n}{2}^{-1} \sum_{k<l} sign\left((X_i^{(k)} - X_i^{(l)})(X_j^{(k)} - X_j^{(l)})\right),$$

where $sign(x) = 1$ if $x > 0$, -1 if $x < 0$ and 0 if $x = 0$. Then (19) *holds.*

The relationship (14) suggests that the extremal dependence coefficient can be used to estimate the parameters of an extreme-value copula. If the margins F_i are known, (19) holds with various empirical estimators of the extremal dependence coefficient, see [25]. In the following example, an estimator proposed in [13] is used.

Example 7 (extremal dependence coefficient)**.** *Assume that C is an extreme-value copulas and put $U_i^{(k)} = F_i(X_i^{(k)})$. Let $r(\theta_i, \theta_j)$ be the extremal dependence coefficient* (4) *of (X_i, X_j). Put*

$$\hat{r}_{i,j} = 3 - \frac{1}{1 - \sum_{k=1}^{n} \max(U_i^{(k)}, U_j^{(k)})/n}.$$

Then (19) *holds.*

The least-squares estimator is unique with probability tending to one, consistent, and asymptotically normal under mild assumptions that are stated in the following proposition, due to [25].

Proposition 4. *Suppose that* (19) *and the following assumptions hold.*

(A1) The map $\mathbf{r}$ is a twice continuously differentiable homeomorphism from Θ^d to its image $\mathbf{r}(\Theta^d)$.

(A2) The Jacobian matrix of the map $\mathbf{r}$ at $\boldsymbol{\theta_0}$, that is,

$$\mathbf{J} := \begin{pmatrix} \left.\frac{\partial r(\boldsymbol{\theta})}{\partial \theta_1}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} & \left.\frac{\partial r(\boldsymbol{\theta})}{\partial \theta_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} & \cdots & \left.\frac{\partial r(\boldsymbol{\theta})}{\partial \theta_d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \\ \vdots & & & \vdots \\ \left.\frac{\partial r(\boldsymbol{\theta})}{\partial \theta_1}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} & \left.\frac{\partial r(\boldsymbol{\theta})}{\partial \theta_2}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} & \cdots & \left.\frac{\partial r(\boldsymbol{\theta})}{\partial \theta_d}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \end{pmatrix},$$

is of full rank.

Then, as $n \to \infty$, the estimator $\hat{\boldsymbol{\theta}}$ defined in (18) *is unique with probability tending to one, consistent for $\boldsymbol{\theta_0}$, and asymptotically normal*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Xi}), \quad \text{where } \boldsymbol{\Xi} = \left(\mathbf{J}^T\mathbf{J}\right)^{-1} \mathbf{J}^T \boldsymbol{\Sigma} \mathbf{J} \left(\mathbf{J}^T\mathbf{J}\right)^{-1}.$$

Remark 3. *From the asymptotic normality of $\hat{\boldsymbol{\theta}}$, standard arguments in mathematical statistics yield, as $n \to \infty$,*

$$n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi}(\hat{\boldsymbol{\theta}})^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \chi_d^2,$$

where χ_d^2 stands for a chi-square distribution with d degrees of freedom. This result will be useful in Section 4.1.1 to assess the accuracy of the inference.

The map $\mathbf{r}$ of one-factor copula models possesses the property that all of its components involve the same bivariate function r. This allows to establish a lemma that gives sufficient conditions for the assumptions of Proposition 4 to hold.

Lemma 1. *(i) Define the univariate function $r_{\theta_j}(\theta_i) := r(\theta_i, \theta_j)$ and assume that it is a twice continuously differentiable homeomorphism. Let $r_{1,2}, \ldots, r_{d-1,d}$ be p elements of $\mathbf{r}(\Theta^d)$. Define $s_{i,j}(\theta) := r_\theta^{-1}(r_{i,j})$ for $\theta \in \Theta$. Then, the function $s_{1,3} \circ s_{1,2} \circ s_{2,3}$ has at least one fix point, that is, the equation*

$$s_{1,3} \circ s_{1,2} \circ s_{2,3}(\theta) = \theta, \quad \theta \in \Theta \tag{20}$$

has at least one solution.

(ii) If, moreover, the function $s_{1,3} \circ s_{1,2} \circ s_{2,3}$ has exactly one fix point, that is, equation (20) has exactly one solution, then the assumptions (A1) and (A2) of Proposition 4 hold.

The fact that the extremal dependence coefficients of the EV-FDG bivariate margins write $\lambda_{i,j}(\boldsymbol{\theta}) = \lambda(\theta_i)\lambda(\theta_j)$, where

$$\lambda(\theta) = 1 - \frac{\partial f(t,\theta)}{\partial t}\Big|_{t=1}, \quad \theta \in \Theta,$$

allows to apply Lemma 1 and therefore to satisfy the assumptions of Proposition 4.

Corollary 1. *Assume that $(X_1, \ldots, X_d)$ has EV-FDG (15) as copula and consider one of the following cases:*

- *The generators are Cuadras-Augé (6) and $\Theta = (0, 1)$.*
- *The generators are Fréchet (7) and $\Theta = (0, 1)$.*
- *The generators are Durante-sinus (8) and $\Theta = (0, \pi/2)$.*

Then the assumptions (A1) and (A2) of Proposition 4 hold with $r(\theta_i, \theta_j)$ and $\hat{\mathbf{r}}$ being as in Example 7.

Remark 4. *For the FDG copula with Fréchet generators presented in Example 2, Spearman's rho dependence coefficient is equal to the extremal dependence coefficient, hence, one can also apply Lemma 1 to this copula. Therefore, also for this model, the assumptions (A1) and (A2) of Proposition 4 hold with $r(\theta_i, \theta_j)$ and $\hat{\mathbf{r}}$ being as in Example 5.*

Except for the FDG copula with Fréchet generators (see Remark 4), the results of Corollary 1 could not be extended to general FDG copulas because the analytical forms of Spearman's rho and Kendall's tau dependence coefficients are not as simple as the forms of the extremal dependence coefficients. Nonetheless, we shall provide empirical evidence by means of numerical experiments in Section 4.1 that the assumptions of Proposition 4 are likely to hold.

4 Applications to simulated and real datasets

The modeling of data with (EV-)FDG copulas is illustrated through numerical experiments in Section 4.1 and a real dataset application in Section 4.2. In the numerical experiments, we first provide empirical evidence to support the assumptions (A1) and (A2) of Proposition 4. We then illustrate, by fitting data of dimension $d = 50$, that the proposed FDG copulas are well suited for

high-dimensional applications. In the real dataset application, critical levels of potentially dangerous hydrological events are estimated. Throughout this section, the four copulas of Example 1–4, are respectively referred to as FDG-CA, FDG-F, FDG-sinus, and FDG-exponential.

The minimization of the loss function (18) was carried out with standard gradient descent algorithms whose implementations can be found in the function `optim` from the R software [28]. In principle, several runs with different starting points should be tested to ensure that the global minimizer is reached. However, we found that a single run was enough to find what appeared to be the global optimum. Thus, the loss functions one encounters when dealing with FDG copulas seem to be easy to minimize in practice.

4.1 Numerical experiments

4.1.1 Empirical evidence in favor of the assumptions of Proposition 4

The first step of this numerical experiment consists of generating 200 datasets of dimension $d = 4$ and size $n = 500$ for each of the copulas FDG-CA, FDG-F, FDG-sinus, and FDG-exponential. The true parameter vectors were chosen so that the coordinates regularly span the parameter space. Hence, they were respectively set to

$$(0.6, 0.7, 0.8, 0.9),\ (0.3, 0.5, 0.7, 0.9),\ (1, 1.2, 1.37, 1.55) \text{ and } (3, 8.7, 14.3, 20).$$

The second step consists of estimating the parameters of the models. To this end, the loss function (18) was used with Spearman's rho dependence coefficient as in Example 5. The amount of time required with a 8 GiB memory and 3.20 GHz processor computer to carry out the simulation of one dataset of size $n = 500$ and dimension $d = 4$ and perform the corresponding inference is given in Table 2 (columns $d = 4$). The computational costs for performing the inference of FDG-exponential and FDG-sinus are larger because their dependence coefficient expressions, given in Example 3 and 4, involve integrals which have to be computed numerically. The third step consists of defining and computing

	Time (ms)			
	Simulation		Inference	
Copula	$d = 4$	$d = 50$	$d = 4$	$d = 50$
FDG-CA	1.3	16	8	920
FDG-F	0.2	1	2	390
FDG-sinus	0.7	7	62	9700
FDG-exponential	2.2	25	228	82700

Table 2: Time required, in milliseconds, to simulate a dataset of size $n = 500$ and dimension d, and to perform the corresponding inference.

error criteria in order to assess the inference method accuracy. For each dataset and each model, the mean absolute error and the relative mean absolute error, respectively defined as

$$\mathrm{MAE}_r = \frac{1}{p}\sum_{i<j} |r_{i,j} - r(\hat{\theta}_i, \hat{\theta}_j)| \text{ and } \mathrm{RMAE} = \frac{1}{d}\sum_{i=1}^{d} \frac{|\hat{\theta}_i - \theta_{0i}|}{\theta_{0i}},$$

were computed and averaged over the replications. These criteria are reported in Table 3 (columns $d = 4$).

	MAE_r		RMAE	
Copula	$d = 4$	$d = 50$	$d = 4$	$d = 50$
FDG-CA	0.03	0.03	0.08	0.06
FDG-F	0.03	0.03	0.08	0.06
FDG-sinus	0.03	0.02	0.08	0.05
FDG-exponential	0.03	0.03	0.08	0.17

Table 3: (Relative) mean absolute errors averaged over the 200 dataset replications for the four FDG copula models.

All the models giving the same small errors, it seems that the method proposed in Section 3 to estimate the parameters of FDG copulas works well in practice. In order to check the validity of the estimator's asymptotic distribution, and since this distribution is multivariate – hence, not straightforward to study –, we rely on the following argument. If the assumptions of Proposition 4 hold, then, by Remark 3, the values $n(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi} (\hat{\boldsymbol{\theta}}^{(k)})^{-1} (\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)$, $k = 1, \ldots, 200$, should be approximately χ^2_d distributed, where $\hat{\boldsymbol{\theta}}^{(k)}$ denotes the parameter vector estimated on the k-th dataset replication. This approximation, shown in Figure 1, appears to be satisfactory. Therefore, the assumptions of Proposition 4 are more likely. Recall that these assumptions lead to the consistency and asymptotic normality of the parameter vector estimator for FDG copulas.

4.1.2 A high-dimensional illustration

The experiment carried out in Section 4.1.1 is repeated but the dimension is increased to $d = 50$. The coordinates of the parameter vector were chosen to be regularly spaced within $[0.3, 0.9], [0.3, 0.9], [1, 1.55]$ and $[3, 20]$ for FDG-CA, FDG-F, FDG-sinus, and FDG-exponential respectively. The amount of time required to simulate a dataset of size $n = 500, d = 50$ and to perform the corresponding inference is given in Table 2 (columns $d = 50$). Simulating all the models even in high dimension is instantaneous because of the conditional independence property seen in (1). Less than 2 minutes are necessary to fit all the models. In particular, simulating or fitting FDG-F is instantaneous. The MAEs/RMAEs computed in this high-dimensional context are reported in Table 3 (columns $d = 50$). The results are quite satisfying for FDG-CA, FDG-F and FDG-sinus: although the sample size n did not increase, the error values are similar to those found in Section 4.1.1. The inference for these models seems not to be sensitive to the dimension. FDG-exponential has a larger RMAE, but it is still below 17%, and its MAE_r is as good as the other models. To summarize, FDG copulas seem to scale up well.

4.2 Application to a hydrological dataset

4.2.1 Data and context

The dataset consists of $n = 32$ observations $(X_1^{(k)}, \ldots, X_d^{(k)})$, $k = 1, \ldots, n$, of annual maxima river flow rates located at $d = 9$ sites across south-east France

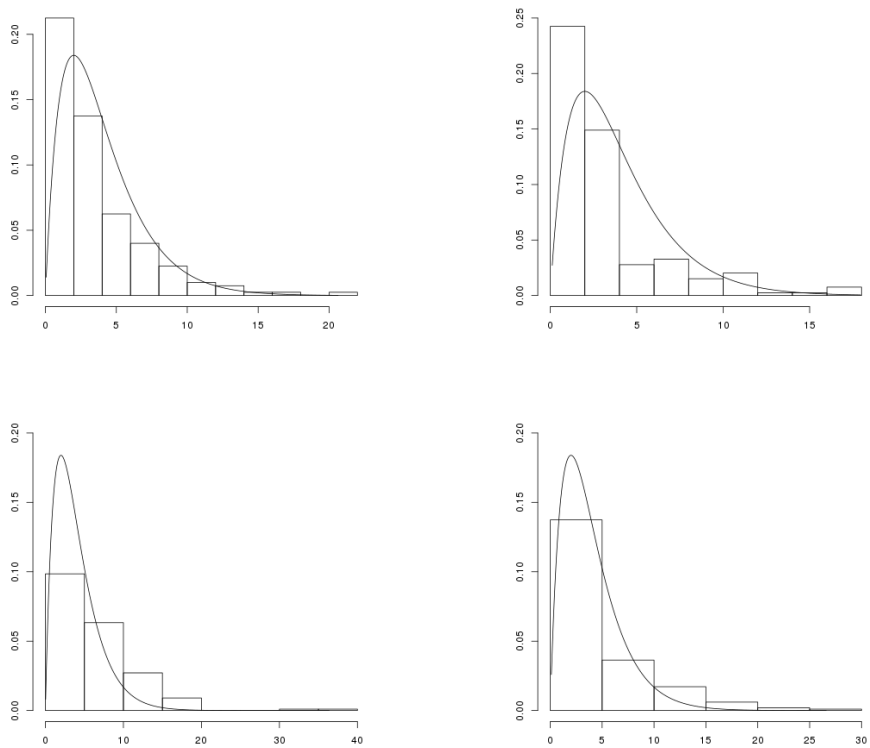

Figure 1: Histograms of $n(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)^T \boldsymbol{\Xi}(\hat{\boldsymbol{\theta}}^{(k)})^{-1}(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}_0)$, $k = 1, \ldots, 200$, together with the density of a χ^2_d distribution. From left to right, top to bottom: FDG-CA, FDG-F, FDG-sinus, and FDG-exponential.

between 1969 and 2007 (some records are missing). Let us denote by F the distribution with continuous margins $F_1, \ldots, F_d$ of the random vector $(X_1, \ldots, X_d)$ whose realizations provide the observed dataset. The location of the sites are shown in Figure 2. The number of variable pairs is $p = 36$. Due to the heterogeneous dispersion of the sites, the span of positive dependence is almost maximum; for instance, Spearman's rho dependence coefficients range from about 0 to 0.9.

In hydrology, it is of interest to get information about the statistical distribution of a potentially dangerous event, such as $\{F_1(X_1) > q, \ldots, F_d(X_d) > q\}$, or, equivalently, $\{\min(F_1(X_1), \ldots, F_d(X_d)) > q\}$, where q is the *critical level* associated to that event. The *return period* T is defined as

$$T = \frac{1}{1 - M(q)}, \quad \text{where } M(q) = P(\min(F_1(X_1), \ldots, F_d(X_d)) \leq q). \tag{21}$$

For instance, a return period of $T = 30$ years and a critical level of $q = 0.7$ means that each X_i exceeds its quantile of order 70% once every 30 years in average. A common question in the study of extreme events is the following. Given a return period T, how dangerous is the corresponding event? In other words,

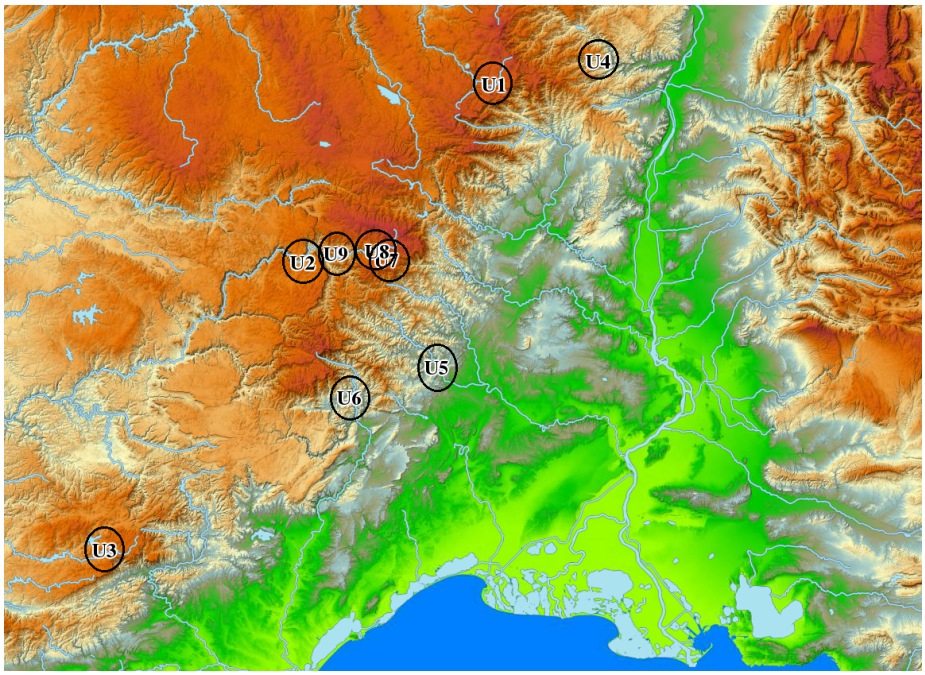

Figure 2: Location of the 9 sites for the flow rate dataset. The sea in dark blue at the bottom (south) is the Mediterranean sea. The rivers are shown in light blue. The river flowing from north to south in the green area is the Rhône. Green indicates low altitude, and orange high altitude. The map of this figure was drawn with Géoportail `www.geoportail.gouv.fr`.

what is the associated critical level q? The answer is obtained by inverting (21)

$$q = M^{-1}\left(1 - \frac{1}{T}\right). \tag{22}$$

Thus, the answer q is the quantile of order $1 - 1/T$ of the distribution M. This quantile can be estimated empirically from the data and parametrically by fitting a model to the data.

Potentially dangerous events happen with the co-occurrence of extremely high flow rates at several locations. Thus, it is clear that the models to describe this dataset should be upper tail dependent. Hence, good candidates are the copulas of Example 1–3, referred to as FDG-CA, FDG-F, and FDG-sinus respectively, and all the extreme-value copulas. However, as it was shown in Remark 1, Spearman's rho of FDG-sinus cannot take values greater than 0.37. Hence, this copula is removed from the candidate models. So the considered models are FDG-CA, FDG-F, and their extreme-value attractor EV-FDG-CAF given in (15) (recall that FDG-CA and FDG-F lead to the same extreme-value copula). Two other popular copula models, the Gumbel and Student copulas, are also fitted to the data. The Gumbel copula is famous among hydrologists [30] and the Student copula is well known in risk management [26]. They serve as a benchmark for our models. A factor structure is assumed for the Student copula, that is, its (i, j)-th element $(i \neq j)$ of its correlation matrix writes $\theta_i\theta_j$,

where $\theta_1, \ldots, \theta_d$ belong to $[-1, 1]$. Recall that a Gumbel copula is an extreme-value copula. More details about the Gumbel and the Student copula can be found respectively in, e.g., [19, 27] and [8].

4.2.2 Method

A practically convenient approach dicted the estimation of the copula parameters. For each copula model, the dependence coefficients with the simplest mathematical forms were chosen to build the loss function (18). In other words, the parameters of FDG-F and FDG-CA were estimated with Spearman's rho as in Example 5. The parameters of EV-FDG-CAF were estimated with the extremal dependence coefficient as in Example 7. The parameters of the Gumbel and Student copulas were estimated with Kendall's tau as in Example 6. Finally, the degree of freedom of the Student copula was estimated by maximizing its likelihood but with all the parameters of the correlation matrix held fixed. This approach improves the speed, tractability, and chances of success of the minimization procedure.

To be valid, the asymptotic properties of the estimator based on the extremal dependence coefficients require the knowledge of the marginal distributions F_i. Assuming that the data come from an extreme-value distribution, the marginal distributions theoretically pertain to the family of the Generalized Extreme Value (GEV) distributions. The good results of the GEV fit to the data margins are depicted by the quantile-quantile plot in Figure 3. In addition to transform the data to have standard uniform margins, these fitted GEV distributions served to calculate the critical levels of (22) as well. This primary step, that of fitting a parametric model to the margins, is standard in extreme-value statistics; see, for instance, [6] Chapter 8 and [29]. Details for the GEV distribution are to be found, e.g., in [6] Chapter 3.

4.2.3 Results

The fit of the tested copulas was assessed by comparing the pairwise dependence coefficients and the critical levels.

Pairwise dependence coefficients. The mean absolute error (MAE), defined as

$$MAE_r = \frac{1}{p} \sum_{i<j} \left| \hat{r}_{i,j} - r(\hat{\theta}_i, \hat{\theta}_j) \right|$$

was computed for Spearman's rho (MAE_ρ), Kendall's tau (MAE_τ) and the extremal dependence coefficient (MAE_λ). They are reported in Table 4. MAE_λ was computed only for EV-FDG-CAF and the Gumbel copula because, with this dependence coefficient, (19) holds only for extreme-value copulas, see Example 7. The Gumbel copula has the largest errors (more than 0.17) and does not seem to fit the data well. This was expected, because this model has only one parameter to account for a $d = 9$ dimensional phenomenon. All the remaining errors (but MAE_ρ for EV-FDG-CAF) are smaller and of the same magnitude. Thus, according to these criteria, the Gumbel copula is not appropriate.

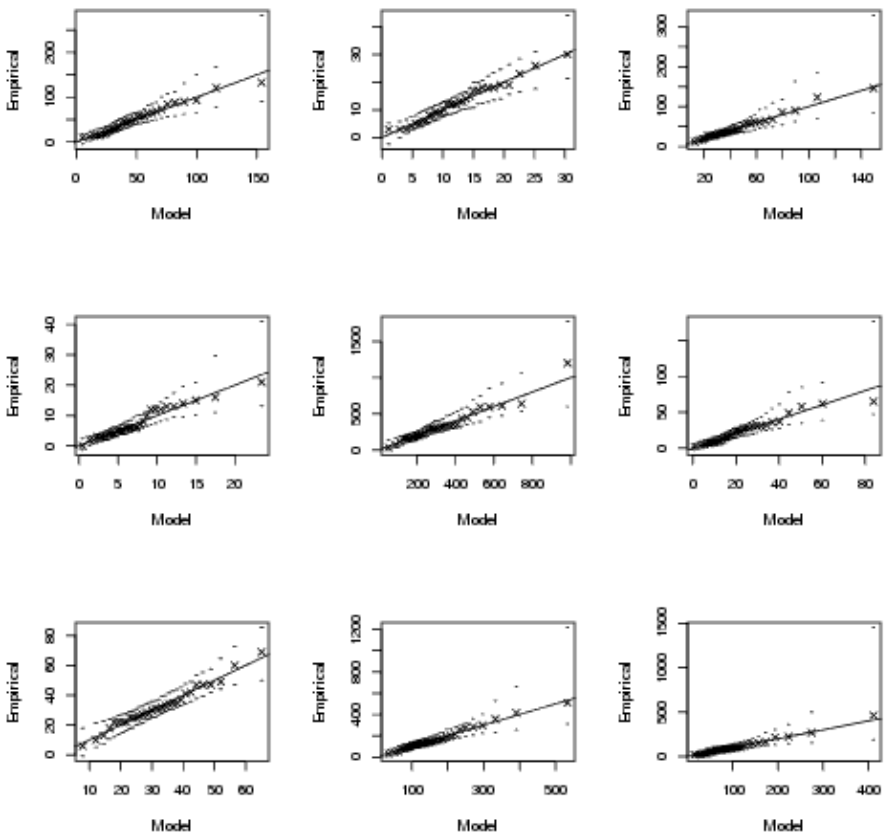

Figure 3: Quantile-Quantile plots with confidence intervals for the GEV fit of the data margins.

	FDG-F	FDG-CA	EV-FDG-CAF	Gumbel	Student
MAE_ρ	0.12	0.12	0.17	0.22	0.12
MAE_τ	0.12	0.11	0.12	0.17	0.10
MAE_λ			0.11	0.45	

Table 4: Mean absolute error for Spearman's rho (MAE_ρ), Kendall's tau (MAE_τ) and extremal (MAE_λ) dependence coefficient of the models.

Critical levels. The critical levels obtained from the empirical data and the models were calculated by making use of (22). In statistical terms, this amounts to compare the quantiles of the distribution M under the empirical data and under the different models. The results are presented in Figure 4, where the independence copula $C(u_1, \ldots, u_d) = u_1 \ldots u_d$ was added to emphasize the need for a joint model on such a dataset. The Gumbel model is confirmed to perform poorly. FDG-F, FDG-CA and EV-FDG-CAF seem to fit the data quite well. In particular, FDG-F and FDG-CA are as close as the Student copula to the

empirical curve.

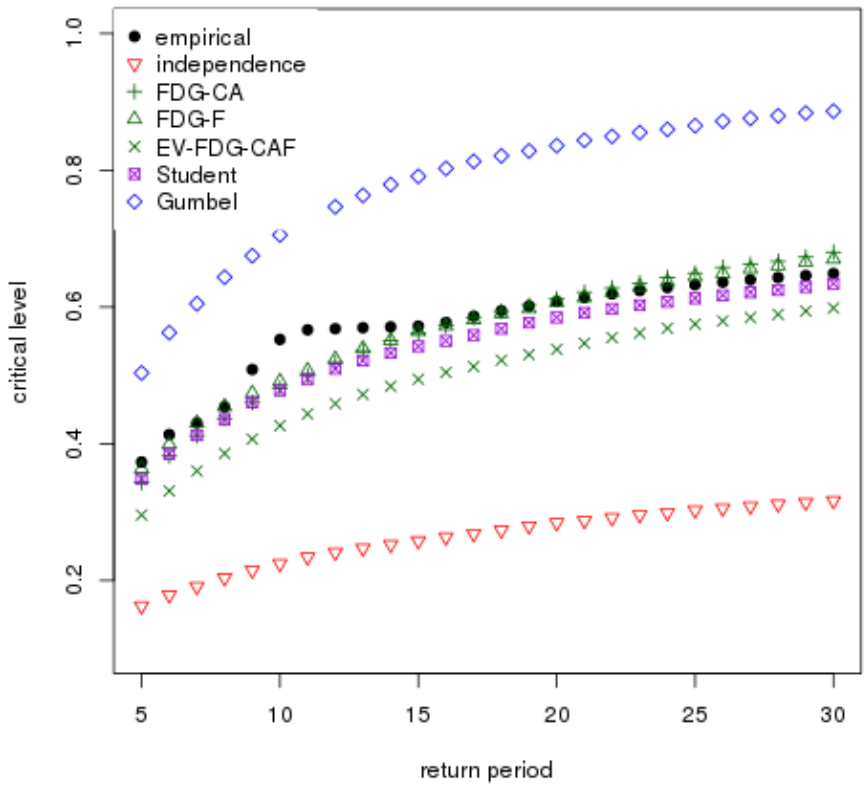

Figure 4: Critical level q as a function of the return period T. "empirical" stands for the empirical critical levels, and "independence" for the independence copula $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$.

With such a small sample size $n = 32$, one must be extremely careful when looking at empirical data, because one is likely to observe a large deviation from the true underlying statistical distribution. In view of this remark, one should select a statistical model based not only on empirical data, but also on the model properties. The class of FDG copulas is very interesting in this respect. Indeed, the practitioner has with this class three models that fit well the data and with different features: FDG-F is upper, lower, and symmetric tail dependent, FDG-CA is upper tail dependent but no lower tail dependent, and EV-FDG-CAF is an extreme-value copula. The user is then free to choose the model that most suits his expert knowledge about the underlying phenomenon at play. The test of extreme-value dependence [20] gave a p-value of 0.21, which means that one does not reject extreme-value copulas at the 5% level. Of course, as before, one must be extremely careful when looking at the p-value because of the small data sample size. Regarding the possibility of testing for a given parametric family,

one can see [21].

5 Discussion

In this article, we have constructed a new class of copulas by combining one-factor copulas, that is, a conditional independent property, together with a class of bivariate copulas called the Durante class of bivariate copulas. This combination led to many advantageous properties. The copulas within the proposed class, referred to as FDG copulas, are tractable, flexible, and cover all types of tail dependencies. The theoretically well grounded least-squares inference estimator is particularly well suited for FDG copulas because their dependence coefficients are easy to compute, if not in closed form. This allows to perform fast and reliable inference in the parametric case. We have demonstrated, furthermore, that FDG copulas work well in practice and are able to model both high-dimensional and real datasets. Finally, we have derived the extreme-value copulas (EV-FDG) associated to FDG copulas, yielding a new extreme-value copula, which can be viewed as a generalisation of the well known Cuadras-Augé copula. This copula benefits from almost all the many advantageous properties of FDG copulas, and therefore opens the door for statistical analyses of extreme data in high dimension.

One may argue that a model with a singular component, as a FDG copula, is not natural nor realistic to model hydrological data. While this may be true in the bivariate case, this argument becomes weaker when the dimension increases. Indeed, in high-dimensional applications, the focus is less on the distribution itself than on a feature of interest of the data, such as, for instance, the critical levels defined in (22). If alleged "unrealistic" models are able to better estimate these features than "realistic" models – compare the fit of the Gumbel copula to the fit of FDG copulas in Section 4.2 – then one should consider using them.

This work raises several research questions. First, how to estimate the generators nonparametrically? The generator of a bivariate Durante copula was estimated nonparametrically in [10], but the matter is more complicated in our case because this bivariate relationship occurs between the variable of interest and the unobserved latent factor. Second, one may add more factors when building an FDG copula. Nonetheless, the model might not be as tractable as it is and therefore it may be less appealing in practice. Finally, FDG copulas possess the conditional independence property, but the extreme-value EV-FDG copulas were not shown to do so. If this property held, this would be of great interest for the simulation of datasets from this model.

Acknowledgment. The authors thank "Banque HYDRO du Ministère de l'Écologie, du Développement durable et de l'Énergie" for providing the data and Benjamin Renard for fruitful discussions about statistical issues in hydrological science.

A Appendix

Proof of Theorem 1

Let $C_{j|0}(\cdot|u_0)$ be the conditional distribution of U_j given $U_0 = u_0$. The U_j's are conditionally independent given U_0, hence,

$$\begin{aligned}
C(u) &= \int_0^1 \prod_{j=1}^d C_{j|0}(u_j|u_0)du_0 \qquad (23)\\
&= \int_0^1 \prod_{j=1}^d \frac{\partial C_{0j}(u_0,u_j)}{\partial u_0} du_0\\
&= \int_0^1 \prod_{j=1}^d \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0} du_0\\
&= \int_0^{u_{(1)}} \prod_{j=1}^d \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0} du_0\\
&\quad + \sum_{k=2}^d \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0} \prod_{j=k}^d \frac{\partial C_{0(j)}(u_0,u_{(j)})}{\partial u_0}\\
&\quad + \int_{u_{(d)}}^1 \prod_{j=1}^d \frac{\partial C_{0(j)}(u_0,u_(j))}{\partial u_0} du_0.
\end{aligned}$$

Since

$$\frac{\partial C_{0j}(u_0,u_j)}{\partial u_0} = \begin{cases} f_j(u_j) & \text{if } u_0 < u_j \\ u_j f_j'(u_0) & \text{if } u_0 > u_j, \end{cases}$$

(23) yields

$$\begin{aligned}
C(u) =& u_{(1)} \prod_{j=1}^d f_j(u_j) + \sum_{k=2}^d \prod_{j=k}^d f_{(j)}(u_{(j)}) \int_{u_{(k-1)}}^{u_{(k)}} \prod_{j=1}^{k-1} u_{(j)} f'_{(j)}(u_0)du_0\\
&+ \prod_{j=1}^d u_j \int_{u_{(d)}}^1 f_j'(u_0)du_0.
\end{aligned}$$

Putting $u_{(1)}$ in factor and noting that $\int_{u_{(1)}}^{u_{(2)}} f'_{(1)}(x)dx = f_{(1)}(u_{(2)}) - f_{(1)}(u_{(1)})$ finishes the proof.

Proof of Proposition 1

It suffices to set all u_k equal to one but u_i and u_j in the formula (3).

Proof of Proposition 2

It suffices to apply the formulas (5) with f_{ij} given in Proposition 1. To compute Spearman's rho, note that

$$\int_0^1 x^2 f_{ij}(x)dx = \int_0^1 x^2 f_i(x)f_j(x)dx + \int_0^1 x^3 \int_x^1 f_i'(z)f_j'(z)dzdx.$$

An integration by parts yields $\int_0^1 x^3 \int_x^1 f_i'(z) f_j'(z) dz dx = (1/4) \int_0^1 x^4 f_i'(x) f_j'(x) dx$ and the result follows.

Proof of Theorem 2

Fix $(u_1, \ldots, u_d) \in [0,1]^d$ and let $n \geq 1$ be an integer. Put

$$\begin{aligned}
\alpha_n :=& u_{(1)}^{1/n} \prod_{j=1}^{d} \frac{f_j(u_j^{1/n})}{u_j^{1/n}}, \\
\beta_n :=& \int_{u_{(d)}^{1/n}}^{1} \prod_{j=1}^{d} f_j'(u_0) du_0, \\
\gamma_n :=& \prod_{j=k}^{d} \frac{f_{(j)}(u_{(j)}^{1/n})}{u_{(j)}^{1/n}}, \\
\delta_{n,k} :=& \int_{u_{(k-1)}^{1/n}}^{u_{(k)}^{1/n}} \prod_{j=1}^{k-1} f_{(j)}'(u_0) du_0,
\end{aligned}$$

and define

$$A_n := \alpha_n + \beta_n + \sum_{k=2}^{d} \gamma_{n,k} \delta_{n,k}.$$

We are going to derive asymptotic equivalent sequences for $\alpha_n, \beta_n \gamma_n$ and δ_n. Let $\sim$ denote the equivalent symbol at infinity (i.e., $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$). By using the well known formulas $e^x \sim 1 + x$ (when $x \to 0$), $\log x \sim x - 1$ (when $x \to 1$) and $f_j(x) \sim 1 + (x-1) f_j'(1)$ (when $x \to 1$) we get

$$\alpha_n \sim \left(1 + \frac{1}{n} \log u_{(1)}\right) \left(1 - \frac{1}{n} \log(u_1 \ldots u_d)\right) \left(1 + \frac{1}{n} \sum_{j=1}^{d} \log u_{(j)} f_{(j)}'(1)\right) \text{ and}$$

$$\gamma_{n,k} \sim \left(1 - \frac{1}{n} \sum_{j=k}^{d} \log u_{(j)}\right) \left(1 + \frac{1}{n} \sum_{j=k}^{d} \log u_{(j)} f_{(j)}'(1)\right).$$

For β_n the equivalence is obtained as follows. Let $F(x)$ be a primitive of $\prod_{j=1}^{d} f_j'(x)$. It follows that $\beta_n = F(1) - F(u_{(d)}^{1/n})$. A Taylor expansion yields

$$F(u_{(d)}^{1/n}) = F(1) + (u_{(d)}^{1/n} - 1) F'(1) + \frac{(u_{(d)}^{1/n} - 1)^2}{2} F''(x_n)$$

where x_n is between $u_{(d)}^{1/n}$ and 1. Since F'' is assumed to be continuous on $[0,1]$, it is uniformly bounded on this set and therefore $(u_{(d)}^{1/n} - 1)^2 F''(x_n)/2 = o(1/n)$ where $o(1/n)$ is a quantity such that $n o(1/n) \to 0$ as $n \to \infty$. Hence, since $u_{(d)}^{1/n} = \exp(\log(u_{(d)})/n) \sim 1 + \log(u_{(d)})/n$, we have as $n \to \infty$

$$F(1) - F(u_{(d)}^{1/n}) \sim -\frac{1}{n} \log(u_{(d)}) F'(1).$$

The same arguments apply to get

$$\beta_n \sim -\frac{1}{n}\log u_{(d)}\prod_{j=1}^{d} f'_j(1)$$

$$\delta_{n,k} \sim \frac{1}{n}\log\left(\frac{u_{(k)}}{u_{(k-1)}}\right)\prod_{j=1}^{k-1} f'_{(j)}(1).$$

The quantity A_n is a polynomial with respect to n^{-1} of order at most three. In (24), the coefficients of order 0, 2, and 3 vanish at infinity. Only remain the terms of order 1, hence,

$$\begin{aligned}\lim_{n\uparrow\infty} n(A_n-1) = &\log u_{(1)} - \log(u_1\dots u_d) + \sum_{j=1}^{d}\log u_{(j)} f'_{(j)}(1) \\ &- \log u_{(d)}\prod_{j=1}^{d} f'_{(j)}(1) + \sum_{k=2}^{d}\prod_{j=1}^{k-1} f'_{(j)}(1)\log\left(\frac{u_{(k)}}{u_{(k-1)}}\right).\end{aligned} \tag{24}$$

From Abel's identity, that is, $\sum_{i=1}^{d-1} a_i(b_{i+1}-b_i) = \sum_{i=1}^{d-1} b_i(a_{i-1}-a_i)+a_{d-1}b_d-a_1b_1$ for two sequences (a_i) and (b_i) of real numbers, we can write

$$\lim_{n\uparrow\infty} n(A_n-1) = \sum_{k=1}^{d}\left(\underbrace{\left(\prod_{j=1}^{k-1} f'_{(j)}(1)\right)(1-f'_{(k)}(1)) + f'_{(k)}(1) - 1}_{=:\chi_k}\right)\log u_{(k)},$$

with the convention that $\prod_{j=1}^{0} f'_{(j)}(1) = 1$. From (3) it follows that

$$\begin{aligned}C^n(u_1^{1/n},\dots,u_d^{1/n}) =& u_1\dots u_d\exp\left[n\log A_n\right] \\ =& u_1\dots u_d\exp\left[n(A_n-1)(1+o(1))\right] \\ \to& \prod_{i=k}^{d} u_{(k)}^{\chi_k}\end{aligned}$$

as $n\to\infty$.

Proof of Lemma 1

(i) Since the p-uple $(r_{1,2},\dots,r_{d-1,d})$ belongs to the image space $\mathbf{r}(\Theta^d)$, the system

$$\begin{aligned}r(\theta_1,\theta_2) =& r_{1,2} \\ &\vdots \\ r(\theta_{d-1},\theta_d) =& r_{d-1,d}\end{aligned}$$

has at least one solution. In particular, there exists $(\theta_1,\theta_2,\theta_3)$ in Θ^3 such that

$$\begin{aligned}r(\theta_1,\theta_2) =& r_{1,2} \\ r(\theta_1,\theta_3) =& r_{1,3} \\ r(\theta_2,\theta_3) =& r_{2,3}.\end{aligned} \tag{25}$$

The system (25) rewrites

$$\begin{aligned} r_{\theta_2}(\theta_1) =& r_{1,2} \\ r_{\theta_3}(\theta_1) =& r_{1,3} \\ r_{\theta_3}(\theta_2) =& r_{2,3}, \end{aligned}$$

or equivalently,

$$\begin{aligned} \theta_1 =& s_{1,2}(\theta_2) \\ \theta_1 =& s_{1,3}(\theta_3) \\ \theta_2 =& s_{2,3}(\theta_3). \end{aligned}$$

This yields

$$s_{1,3}(\theta_3) = s_{1,2} \circ s_{2,3}(\theta_3). \tag{26}$$

Note that $s_{1,3}$ is involutive at θ_3, that is, $s_{1,3} \circ s_{1,3}(\theta_3) = \theta_3$. Indeed, $r(\theta_1, \theta_3) = r_{\theta_3}(\theta_1) = r_{1,3}$ is equivalent to $\theta_1 = r_{\theta_3}^{-1}(r_{1,3}) = s_{1,3}(\theta_3)$. This implies $r(s_{1,3}(\theta_3), \theta_3) = r_{1,3}$, and, composing by $r^{-1}_{s_{1,3}(\theta_3)}$ in both sides, we get $r^{-1}_{s_{1,3}(\theta_3)}(r_{1,3}) = s_{1,3}(s_{1,3}(\theta_3)) = \theta_3$. Therefore, one can compose both sides of (26) by $s_{1,3}$ to get

$$\theta_3 = s_{1,3} \circ s_{1,2} \circ s_{2,3}(\theta_3).$$

Hence (20) has at least one solution.

(ii) If (20) admits exactly one solution θ_3, then θ_2 and θ_1 are also unique. Furthermore, for all $j \geq 3$,

$$\theta_{j+1} = s_{j,j+1}(\theta_j)$$

which concludes the proof that assumption (A1) holds. It is now shown that assumption (A2) holds as well. Define $\partial_1 r$, respectively, $\partial_2 r$, the derivative of r with respect to the first, respectively, second, variable of r. Hence for all θ_i and θ_j in Θ, the quantities $\partial_1 r(\theta_i, \theta_j)$ and $r_{\theta_j}(\theta_i)$ only differ in the notation. The first step in the proof is to consider the case $d = 3$. The Jacobian matrix of $\mathbf{r}$ at $\boldsymbol{\theta_0}$ is given by

$$\mathbf{J} = \begin{pmatrix} \partial_1 r(\theta_{01}, \theta_{02}) & \partial_2 r(\theta_{01}, \theta_{02}) & 0 \\ \partial_1 r(\theta_{01}, \theta_{03}) & 0 & \partial_2 r(\theta_{01}, \theta_{03}) \\ 0 & \partial_1 r(\theta_{02}, \theta_{03}) & \partial_2 r(\theta_{02}, \theta_{03}) \end{pmatrix}.$$

To show that $\mathbf{J}$ has full rank, we show that its determinant

$$\partial_1 r(\theta_{01}, \theta_{02}) \partial_2 r(\theta_{01}, \theta_{03}) \partial_1 r(\theta_{02}, \theta_{03}) + \partial_2 r(\theta_{02}, \theta_{03}) \partial_1 r(\theta_{01}, \theta_{03}) \partial_2 r(\theta_{01}, \theta_{02})$$

is nonzero. Indeed, note that for all θ in Θ, the map $r_\theta : \Theta \to r_\theta(\Theta)$ is a twice continuously differentiable homeomorphism. Furthermore, by assumption, the true parameter vector $\boldsymbol{\theta_0}$ lies in the interior of Θ that is open. Finally, by symmetry, for all $i < j$,

$$\partial r_1(\theta_{0i}, \theta_{0j}) > 0 \quad (\text{respectively } \partial r_1(\theta_{0i}, \theta_{0j}) < 0)$$

is equivalent to

$$\partial r_2(\theta_{0j}, \theta_{0i}) > 0 \quad (\text{respectively } \partial r_2(\theta_{0j}, \theta_{0i}) < 0).$$

For the general case, we proceed by mathematical induction. When the dimension is d, write $\mathbf{J}(\boldsymbol{\theta}) = \mathbf{J}^{(d)}(\boldsymbol{\theta})$ to emphasize the dependence on the dimension. Notice that it was already shown above that $\mathbf{J}^{(3)}(\boldsymbol{\theta})$ has full rank. Now suppose that the kernel of $\mathbf{J}^{(d-1)}(\boldsymbol{\theta})$ is null when the dimension is $d-1$. Let $\mathbf{A} = \mathbf{J}^{(d)}(\boldsymbol{\theta})$. Each row of $\mathbf{A}$ writes

$$(0, \ldots, 0, \partial_1 r(\theta_i, \theta_j), 0 \ldots, 0, \partial_2 r(\theta_i, \theta_j), 0, \ldots, 0)$$

where $\partial_1 r(\theta_i, \theta_j)$ is at the i-th position and $\partial_2 r(\theta_i, \theta_j)$ at the j-th position. There are $d-1$ rows of A which depend on θ_d and $p-d+1$ which do not (recall $p = d(d-1)/2$ is the number of pairs). Since the kernel of a matrix is invariant by permutation, we can without loss of generality put all the rows which do not depend on θ_d on the top. More precisely, decompose $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

such that $\mathbf{A}_{11}$ is a $(p-d+1) \times (d-1)$ matrix containing all the rows which do not depend on θ_d and $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ are $(p-d+1) \times 1$ and $(d-1) \times 1$ matrices respectively. Note that $\mathbf{A}_{12}$ is the null vector of size $p-d+1 \times 1$. Let $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x} = (\mathbf{x}_1^T, x_2)^T$ where $\mathbf{x}_1 \in \mathbb{R}^{d-1}$, $x_2 \in \mathbb{R}$. It follows that $\mathbf{A}\mathbf{x} = \mathbf{0}$ is equivalent to

$$\begin{cases} \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}x_2 = \mathbf{0} \\ \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}x_2 = 0. \end{cases}$$

But $\mathbf{A}_{12} = \mathbf{0}$ and since $\mathbf{A}_{11} = \mathbf{J}^{(d-1)}(\boldsymbol{\theta})$ whose kernel is null, $\mathbf{x}_1 = \mathbf{0}$. Then $\mathbf{A}_{22}x_2 = 0$ and the assumptions imply $x_2 = 0$, which concludes the proof.

Proof of Corollary 1

To prove Corollary 1, it suffices to apply Lemma 1. Since $r(\theta_i, \theta_j)$ denotes the extremal dependence coefficient of the $\mathcal{E}$ copula bivariate marginal $C_{\#,ij}$ defined in (16), we have

$$\begin{aligned} r(\theta_i, \theta_j) &= \lambda(\theta_i)\lambda(\theta_j), \text{ where} \\ \lambda(\theta) &:= 1 - \frac{\partial f(t,\theta)}{\partial t}\Big|_{t=1}. \end{aligned} \tag{27}$$

In the Cuadras-Augé and the Fréchet cases, (27) is given by $\lambda(\theta) = \theta$, and in the sinus case, $\lambda(\theta) = 1 - \theta/\tan(\theta)$. In all these situations, it is easy to see that the map $r_{\theta_j}(\cdot)$ is a twice continuously differentiable homeomorphism. Therefore, Lemma 1 (i) applies.

To apply the second part of Lemma 1, note that equation (20) translates into

$$\lambda(\theta)^2 = \frac{r_{1,3}r_{2,3}}{r_{1,2}}.$$

Since it has a unique solution, Lemma 1 (ii) applies, and the result is proved.

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

[2] E. F. Acar, C. Genest, and J. NešLehová. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74–90, 2012.

[3] C. Amblard and S. Girard. A new extension of bivariate FGM copulas. *Metrika*, 70(1):1–17, 2009.

[4] T. Bedford and R.M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4):245–268, 2001.

[5] T. Bedford and R.M. Cooke. Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

[6] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[7] C.M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4):339–353, 1981.

[8] S. Demarta and A. J. McNeil. The t copula and related copulas. *International statistical review*, 73(1):111–129, 2005.

[9] F. Durante. A new class of symmetric bivariate copulas. *Nonparametric Statistics*, 18(7-8):499–510, 2006.

[10] F. Durante and O. Okhrin. Estimation procedures for exchangeable marshall copulas with hydrological application. *Stochastic Environmental Research and Risk Assessment*, published online, 2014.

[11] F. Durante, J.J. Quesada-Molina, and M. Úbeda Flores. On a family of multivariate copulas for aggregation processes. *Information Sciences*, 177(24):5715–5724, 2007.

[12] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.

[13] M. Ferreira. Nonparametric estimation of the tail-dependence coefficient. *REVSTAT–Statistical Journal*, 11(1):1–16, 2013.

[14] G. Frahm, M. Junker, and A. Szimayer. Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286, 2003.

[15] M. Fréchet. Remarques au sujet de la note précédente. *CR Acad. Sci. Paris Sér. I Math*, 246:2719–2720, 1958.

[16] C. Genest and A. C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

[17] G. Gudendorf and J. Segers. Extreme-value copulas. In P. Jaworski, F. Durante, W.K. Härdle, and T. Rychlik, editors, *Copula Theory and Its Applications*, page 127–145. Springer, 2010.

[18] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

[19] H. Joe. *Multivariate models and dependence concepts.* Chapman & Hall/CRC, Boca Raton, FL, 2001.

[20] I. Kojadinovic, J. Segers, and J. Yan. Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics*, 39(4):703–720, 2011.

[21] I. Kojadinovic and J. Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21(1):17–30, 2011.

[22] P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101, 2013.

[23] D. Kurowicka and R.M. Cooke. Distribution-free continuous Bayesian belief nets. In *Proceedings of Mathematical methods in Reliability Conference, Santa Fe, New Mexico, USA*, 2004.

[24] J.F. Mai and M. Scherer. Lévy-frailty copulas. *Journal of Multivariate Analysis*, 100(7):1567–1585, 2009.

[25] G. Mazo, S. Girard, and F. Forbes. Weighted least-squares inference based on dependence coefficients for multivariate copulas. *http://hal.archives-ouvertes.fr/hal-00979151*, 2014.

[26] A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques, and tools.* Princeton university press, 2010.

[27] R.B. Nelsen. *An introduction to copulas.* Springer, New York, 2006.

[28] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013.

[29] B. Renard and M. Lang. Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Advances in Water Resources*, 30(4):897–912, 2007.

[30] L. Zhang and V. P. Singh. Gumbel–Hougaard copula for trivariate rainfall frequency analysis. *Journal of Hydrologic Engineering*, 12(4):409–419, 2007.

Conclusion

Dans cette thèse, nous avons proposé deux nouvelles classes de copules. La première est construite à partir d'une classe prenant la forme d'un produit généralisé de copules. Etant reconnu que cette classe doit être simplifiée pour pouvoir être utile en pratique, nous avons montré que, si les fonctions g_{ei} de Liebscher ne dépendent en fait pas de l'indice e, et si l'on souhaite tirer profit de la large gamme de familles de copules bivariées disponibles dans la littérature, alors la seule classe possible que l'on peut construire est la classe des copules PBC du chapitre 4. De plus, nous avons montré le lien avec un algorithme *message-passing*, ce qui permet de faire l'estimation des copules PBC par maximum de vraisemblance. Toutefois, nous avons montré également que les copules PBC ne sont pas capables de modéliser des variables très dépendantes, et, par conséquent, sont d'une utilité limitée pour les applications.

La deuxième classe de copules que nous avons proposées (les copules FDG) est, à notre avis, beaucoup plus intéressante pour les applications. En effet, cette classe est très maniable, parcimonieuse – puisque le nombre de paramètres n'augmente que proportionnellement par rapport à la dimension –, et les expressions des coefficients de dépendance sont explicites, ce qui permet d'étudier et de quantifier la dépendance avec aisance. Nous avons montré que la classe FDG permet d'obtenir des dépendances dans les queues inférieures mais pas supérieures, ou l'inverse, ou les deux. Les variables d'intérêts sont conditionnellement indépendantes par rapport à un facteur latent, ce qui permet de simuler très facilement selon ces copules. Ce facteur latent peut être ignoré, ou peut être vu comme un attrait supplémentaire du modèle en cela qu'il facilite son interprétation. Enfin, les copules FDG présentent une composante singulière. Elles peuvent donc être utilisées dans le cadre de modèles à chocs, comme vu dans la partie 1.3.2. Mais nous avons montré que les copules FDG peuvent aussi être utilisées avec succès dans des applications de natures différentes, c'est à dire, même quand il n'y a pas de chocs. En particulier, nous avons estimé le niveau critique associé aux évènements de débits extrêmes de la région des Cévennes (sud de la France). Nos résultats sont robustes à l'intérieur de la classe FDG, ajustent bien les données, et sont en cohérence avec les résultats obtenus avec une copule de Student. Enfin, nous avons montré que l'on pouvait calculer les copules extrêmes associées aux copules FDG (les copules EV-FDG), enrichissant ainsi la gamme des copules extrêmes, d'autant que les copules extrêmes issues des copules FDG réalisent elles aussi un bon compromis entre flexibilité et maniabilité. Elles ont d'ailleurs été testées et utilisées en pratique au même titre que les autres copules.

Dans le cadre de cette thèse, nous avons également apporté notre contribution à l'estimation de copules. Nous avons souligné que les méthodes ac-

tuelles d'estimation n'avaient pas de fondements théoriques si les copules considérées n'ont pas de dérivées partielles, ce qui est le cas des copules FDG. Nous avons montré que, sous des conditions d'identifiabilité naturelles, l'estimateur des moindres carrés pondérés basé sur les coefficients de dépendance existe et est unique avec une probabilité s'approchant de 1 au fur et à mesure que la taille de l'échantillon augmente, et qu'il était asymptotiquement consistant et normal, et ce, même si les copules considérées n'admettent pas de dérivées partielles.

Les perspectives à nos différents travaux ont été discutées dans les chapitres les concernant. Pour les copules PBC, il avait été suggéré que ces dernières étaient peut-être d'un intérêt plus théorique que pratique. Concernant les copules FDG, avait été posé la question de l'estimation non paramétrique des générateurs. Nous avions aussi fait remarquer que plusieurs facteurs cachés pourraient être envisagés lors la construction de ces copules, mais que cela risquerait de les rendre moins maniables. La propriété d'indépendance conditionnelle, établie (par construction) pour les copules FDG, n'a pas été démontrée pour les copules EV-FDG. Or si cette propriété était vraie, elle serait utile, entre autre, pour simuler des données selon ces copules. Enfin, concernant l'estimateur des moindres carrés basé sur les coefficients de dépendance, l'application hydrologique a révélé que ce dernier pouvait être instable numériquement lorsque le nombre de paramètres à estimer est égal aux nombre de paires de variables. Dans ce cas, on pourrait envisager une combinaison de coefficients de différents types (rho de Spearman, tau de Kendall...) afin de rendre le système d'équations à résoudre plus robuste numériquement.

Nos travaux s'inscrivent dans un contexte plus global, dont nous faisons une synthèse ci-dessous. La recherche de modèles multivariés avec autant de qualités que possible était un sujet important il y a une dizaine d'années. Nous avons le sentiment, que, de nos jours, il en existe une gamme assez diversifiée. Ces toutes dernières années, en particulier, ont vu apparaitre une classe de modèles très flexibles, les *Vines*. Bien sûr, il reste encore beaucoup de recherche à effectuer, notamment pour rendre ces modèles plus parcimonieux, et mieux comprendre leur sensibilité par rapport au choix de la décomposition de la densité, ou des familles paramétriques bivariées à incorporer. En outre, le risque est grand, avec ces modèles, de sur-ajuster (*overfit* en anglais) les données. En ce sens, les modèles à facteur, et en particulier à un facteur (comme celui que nous avons proposé), sont très intéressants. Ils peuvent d'ailleurs être vus comme des modèles *Vines* « tronqués ». Nous aimerions terminer cette thèse par le questionnement « méta-statistique » suivant : que voulons-nous faire avec nos modèles ? Le but est-il vraiment d'ajuster les données le mieux possible, comme il en ressort l'impression à la lecture de certaines publications ? A notre avis, en grande dimension, la distribution que nous tentons de modéliser nous importe moins qu'une caractéristique de cette dernière. Autrement dit, c'est moins la loi de $(X_1, \ldots, X_d)$ que celle de $\varphi(X_1, \ldots, X_d)$, avec $\varphi : \mathbb{R}^d \mapsto \mathbb{R}$, qui nous intéresse. Par exemple, dans le cas des applications en hydrologie que nous avons traitées à plusieurs reprises dans cette thèse, nous avons $\varphi(X_1, \ldots, X_d) = \min(X_1, \ldots, X_d)$. Il y a trois possibilités d'estimer la loi de $\varphi(X_1, \ldots, X_d)$:

1. en utilisant un historique des données de $\varphi(X_1, \ldots, X_d)$ lui même,
2. en modélisant la loi de $\varphi(X_1, \ldots, X_d)$ elle même, ou,
3. en modélisant la loi de $(X_1, \ldots, X_d)$.

C'est seulement dans la dernière situation que nous avons besoin d'un modèle

multivarié, et si c'était cas, ce modèle devrait être choisi aussi en fonction de ses capacités relativement à la fonctionnelle φ, et pas seulement relativement à ses capacités à « ajuster les données ». Même un modèle *a priori* peu flexible peux donner des résultats satisfaisants. La reproductibilité des résultats, en particulier, doit être vérifiée avec attention.

Bibliographie

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance : Mathematics and Economics*, 44(2) :182–198, 2009.

[2] E. F. Acar, C. Genest, and J. NešLehová. Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110 :74–90, 2012.

[3] C. Amblard and S. Girard. A new extension of bivariate FGM copulas. *Metrika*, 70(1) :1–17, 2009.

[4] T. Bedford and R. M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1-4) :245–268, 2001.

[5] T. Bedford and R. M. Cooke. Vines : A new graphical model for dependent random variables. *Annals of Statistics*, 30(4) :1031–1068, 2002.

[6] D. Berg. Copula goodness-of-fit testing : an overview and power comparison. *The European Journal of Finance*, 15(7-8) :675–701, 2009.

[7] L. Breiman. Statistical modeling : The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3) :199–231, 2001.

[8] A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics*, 39(4) :1963–2006, 2011.

[9] P. Capéraà, A. L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3) :567–577, 1997.

[10] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1) :141–151, 1978.

[11] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.

[12] C. M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4) :339–353, 1981.

[13] L. de Haan and A. Ferreira. *Extreme value theory : an introduction*. Springer, 2007.

[14] P. Deheuvels. A nonparametric test for independence. *Publications de l'Institut de Statistique de l'Université de Paris*, 26 :29–50, 1981.

[15] P. Deheuvels. On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions. *Statistics & Probability Letters*, 12(5) :429–439, 1991.

[16] S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1) :111–129, 2005.

[17] F. Durante. A new class of symmetric bivariate copulas. *Nonparametric Statistics*, 18(7-8) :499–510, 2006.

[18] F. Durante and O. Okhrin. Estimation procedures for exchangeable Marshall copulas with hydrological application. *Stochastic Environmental Research and Risk Assessment*, published online, 2014.

[19] F. Durante, J. J. Quesada-Molina, and M. Úbeda Flores. On a family of multivariate copulas for aggregation processes. *Information Sciences*, 177(24) :5715–5724, December 2007.

[20] F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2) :143–161, 2010.

[21] D. Fantazzini. The effects of misspecified marginals and copulas on computing the value at risk : A monte carlo study. *Computational Statistics & Data Analysis*, 53(6) :2168–2188, 2009.

[22] J. D. Fermanian, D. Radulovic, and M. Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10(5) :847–860, 2004.

[23] G. Frahm, M. Junker, and A. Szimayer. Elliptical copulas : applicability and limitations. *Statistics & Probability Letters*, 63(3) :275–286, 2003.

[24] M. Fréchet. Remarques au sujet de la note précédente. *Compte Rendu de l'Académie des Sciences de Paris, Série I, Mathématique*, 246 :2719–2720, 1958.

[25] C. Genest and A. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12(4) :347–368, 2007.

[26] C. Genest, K. Ghoudi, and L. P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3) :543–552, 1995.

[27] C. Genest, J. Nešlehová, and N. Ben Ghorbal. Estimators based on Kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2) :157–177, 2011.

[28] C. Genest and B. Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré : Probabilités et Statistiques*, 44(6) :1096–1127, 2008.

[29] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas : A review and a power study. *Insurance : Mathematics and economics*, 44(2) :199–213, 2009.

[30] C. Genest and L. P. Rivest. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American statistical Association*, 88(423) :1034–1043, 1993.

[31] C. Genest and J. Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B) :2990–3022, 2009.

[32] C. Genest and B. J. M. Werker. Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In C. M. Cuadras and J. A. R. Lallena, editors, *Proceedings of the Conference on Distributions With Given Marginals and Statistical Modelling*, pages 103–112, Kluwer Academic Publishers, Dordrecht, 2002.

[33] G. Gudendorf and J. Segers. Extreme-value copulas. In P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, editors, *Copula Theory and Its Applications*, page 127–145. Springer, 2010.

[34] G. Gudendorf and J. Segers. Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis*, 102(1) :37–47, 2011.

[35] G. Gudendorf and J. Segers. Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, 142(12) :3073–3085, 2012.

[36] E. J. Gumbel. Distributions des valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9 :171–173, 1960.

[37] I. H. Haff and J. Segers. Nonparametric estimation of pair-copula constructions with the empirical pair-copula. *arXiv preprint :1201.5133*, 2012.

[38] P. Hall and N. Tajvidi. Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(5) :835–844, 2000.

[39] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3) :293–325, 1948.

[40] M. Hofert. Construction and sampling of nested Archimedean copulas. In P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, editors, *Copula Theory and Its Applications*, pages 147–160. Springer, 2010.

[41] M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula : Multivariate Dependence with Copulas*, 2014. R package version 0.999-8.

[42] M. Hofert, M. Mächler, and A. J. McNeil. Archimedean copulas in high dimensions : Estimators and numerical challenges motivated by financial applications. *Journal de la Société Française de Statistique*, 154(1) :25–63, 2012.

[43] M. Hofert and M. Scherer. CDO pricing with nested Archimedean copulas. *Quantitative Finance*, 11(5) :775–787, 2011.

[44] P. D. Hoff, X. Niu, and J. A. Wellner. Information bounds for Gaussian copulas. *Bernoulli*, 20(2) :604–622, 2014.

[45] J. C. Huang, N. Jojic, and C. Meek. Exact inference and learning for cumulative distribution functions on loopy graphs. *Neural Information Processing Systems*, 2010.

[46] Li J., Das K., Fu G., Li R., and Wu R. The Bayesian lasso for genome-wide association studies. *Bioinformatics*, 27(4) :516–523, 2010.

[47] H. Joe. *Multivariate models and dependence concepts.* Chapman & Hall/CRC, 2001.

[48] H. Joe. *Dependence Modeling with Copulas.* CRC Press, 2014.

[49] C. H. Kimberling. A probabilistic interpretation of complete monotonicity. *Aequationes Mathematicae*, 10(2) :152–164, 1974.

[50] C. A. J. Klaassen and J. A. Wellner. Efficient estimation in the bivariate normal copula model : normal margins are least favourable. *Bernoulli*, 3 :55–77, 1997.

[51] C. Klüppelberg and G. Kuhn. Copula structure analysis. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71(3) :737–753, 2009.

[52] I. Kojadinovic, J. Segers, and J. Yan. Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics*, 39(4) :703–720, 2011.

[53] I. Kojadinovic and J. Yan. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21(1) :17–30, 2011.

[54] I. Kojadinovic and J. Yan. copula : Multivariate dependence with copulas. *R package*, 2014.

[55] I. Kojadinovic, J. Yan, and M. Holmes. Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica*, 21(2) :841–871, 2011.

[56] D. Kundu and A. K. Dey. Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm. *Computational Statistics & Data Analysis*, 53(4) :956–965, 2009.

[57] D. Kurowicka. Introduction : Dependence Modeling. In D. Kurowicka and H. Joe, editors, *Dependence Modeling, Vine Copula Handbook*. World Scientific, 2011.

[58] D. Kurowicka and R.M. Cooke. Distribution-free continuous Bayesian belief nets. In *Proceedings of Mathematical methods in Reliability Conference, Santa Fe, New Mexico, USA*, 2004.

[59] D. Kurowicka and H. Joe, editors. *Vine Copula Handbook*. World Scientific, 2011.

[60] O. Ledoit and Wolf M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5) :603–621, 2003.

[61] D. X. Li. On default correlation : A copula function approach. *The Journal of Fixed Income*, 9(4) :43–54, 2000.

[62] E. Liebscher. Construction of asymmetric multivariate copulas. *Journal of Multivariate Analysis*, 99(10) :2234–2250, 2008.

[63] F. Lindskog, A. McNeil, and U. Schmock. *Kendall's tau for elliptical distributions*. Springer, 2003.

[64] A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317) :30–44, 1967.

[65] A. J. Mc Neil. *Quantitative risk management : concepts, techniques and tools*. Princeton series in finance. Princeton University Press, Princeton, N.J, 2005.

[66] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management : concepts, techniques, and tools*. Princeton university press, 2010.

[67] A. J. McNeil and J. Nešlehová. Multivariate Archimedean copulas, d-monotone functions and l1-norm symmetric distributions. *The Annals of Statistics*, 37(5B) :3059–3097, 2009.

[68] A.J. McNeil. Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6) :567–581, 2008.

[69] R. B. Nelsen. *An introduction to copulas.* Springer, 2006.

[70] D. H. Oh and A. J. Patton. Modelling dependence in high dimensions with factor copulas. *Manuscript, Duke University*, 2012.

[71] D. H. Oh and A. J. Patton. Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association*, 108(502) :689–700, 2013.

[72] O. Okhrin, Y. Okhrin, and W. Schmid. On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, 173(2) :189–204, 2013.

[73] J. Pickands. Multivariate extreme value distributions. In *Proceedings 43rd Session International Statistical Institute*, volume 2, pages 859–878, 1981.

[74] J. F. Quessy, M. Said, and A. C. Favre. Multivariate Kendall's tau for change point detection in copulas. *The Canadian journal of statistics*, 41 :65–82, 2013.

[75] R Core Team. *R : A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013.

[76] S. I. Resnick. *Extreme values, regular variation, and point processes.* Springer, 2007.

[77] G. Salvadori and C. De Michele. Estimating strategies for multiparameter multivariate extreme value copulas. *Hydrology and Earth System Sciences*, 15(1) :141–150, 2011.

[78] C. Savu and M. Trede. Hierarchies of Archimedean copulas. *Quantitative Finance*, 10(3) :295–304, 2010.

[79] F. Schmid and R. Schmidt. Multivariate extensions of Spearman's rho and related statistics. *Statistics & Probability Letters*, 77(4) :407–416, 2007.

[80] J. Segers. Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3) :764–782, 2012.

[81] J. Segers, R. V. D. Akker, and B. J. M. Werker. Semiparametric Gaussian copula models : Geometry and rank-based efficient estimation. *arXiv preprint :1306.6658*, 2013.

[82] R. J. Serfling. *Approximation theorems of mathematical statistics.* Wiley, 1980.

[83] F. Serinaldi and S. Grimaldi. Fully nested 3-copula : procedure and application on hydrological data. *Journal of Hydrologic Engineering*, 12(4) :420–430, 2007.

[84] J. Shao. *Mathematical Statistics.* Springer, 1999.

[85] M. Sibuya. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, 11 :195–210, 1960.

[86] A. Sklar. Fonction de répartition dont les marges sont données. *Publications de l'Institut de Statistique de l'Université de Paris*, 8 :229–231, 1959.

[87] T. Van Pham and G. Mazo. *PBC : product of bivariate copulas.* http ://cran.r-project.org, 2014. R package version 1.2.

[88] Science Watch. `http://archive.sciencewatch.com/dr/tt/2010/10dectt-MATH/`.