UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

PAR
DIAKHOU NDIAYE

RÉGRESSION : APPROCHES PAR COPULES

MARS 2022

AVANT-PROPOS

Le monde des mathématiques m'a très tôt passionné. Ensuite, au fil des années, j'ai fait des découvertes fascinantes dans ce domaine riche et complexe, notamment en statistique. Ces dernières années, ce sont les sciences statistiques qui ont particulièrement attiré mon attention par leur nombre exorbitant d'applications possibles. L'usage des copules, que j'ai découvert grâce à ce projet de recherche, en font partie. Ce mémoire se veut une contribution dans cette branche très intéressante de la statistique.

Je tiens à remercier sincèrement mon directeur de recherche Jean-François Quessy pour m'avoir proposé ce projet passionnant et pour avoir su m'encourager tout au long de ma maîtrise. La réussite de ce projet a été assurée par sa grande disponibilité, ses connaissances, son expérience, ses judicieux conseils ainsi que la confiance qu'il m'a témoignée. J'aimerais aussi exprimer ma profonde gratitude envers Jean-François pour sa contribution intellectuelle au projet et pour m'avoir initié à Matlab.

Je veux également remercier MM. Mhamed Mesfioui et François Meunier, professeurs au Département de mathématiques et d'informatique de l'UQTR, pour avoir gracieusement accepté de lire et de commenter ce travail de recherche. Leur remarques judicieuses ont permis d'améliorer la version finale de mon travail.

Je dédie cet ouvrage à ma maman Diarra NDAO qui m'a soutenue et encouragée durant ces années d'études. À mon très cher époux M. Babacar Ibn Aliou BA qui a partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail tout en étant à l'autre bout du monde. A ma famille, mes proches et à ceux qui me donnent de l'amour et de la vivacité.

Le financement durant mes études a été fourni par un octroi individuel à M. Jean-François Quessy dans le cadre du programme des Subventions à la Découverte du CRSNG ainsi que par l'Institut des Sciences Mathématiques du Québec.

RÉSUMÉ

Ce mémoire de maîtrise développe une approche de régression non paramétrique permettant de modéliser des liens non monotones entre une variable dépendante et un ensemble de variables explicatives. D'une part, le modèle général qui est proposé utilise une expression de l'espérance conditionnelle qui met en lumière le rôle de la copule pour expliquer le lien entre les variables. D'autre part, la copule utilisée est choisie dans la famille des structures de dépendance Khi-deux, ce qui permet la modélisation de liens non monotones. Des illustrations sur des jeux de données simulées montrent que l'approche fonctionne bien, fournissant ainsi une approche alternative aux méthodes de régression non linéaire paramétriques et non paramétriques.

Table des matières

Table des figures

CHAPITRE 1

INTRODUCTION ET OBJECTIFS

1.1 Idée générale de la régression

La construction de la régression repose d'une part sur une modélisation des variables statistiques par des variables aléatoires (réelles ou non), d'autre part sur un recueil de données croisées, c'est-à-dire que pour un même échantillon de population, on dispose d'observations des différentes variables mesurées avec une imprécision éventuelle. La régression consiste alors à formuler un indicateur sur les valeurs de la variable expliquée dépendant uniquement des valeurs des variables explicatives. Cet indicateur pourra ensuite être utilisé sur une population pour laquelle on ne connait que les valeurs des variables explicatives, afin d'estimer les valeurs de la variable expliquée.

Soit donc une variable aléatoire dépendante Y, ainsi qu'un ensemble de variables explicatives $X_1, \ldots, X_d$. Généralement, la variable dépendante Y est reliée aux variables explicatives par une certaine fonction $f : \mathbb{R}^d \to \mathbb{R}$ telle que

$$Y = f(X_1, \ldots, X_d) + \varepsilon,$$

où ε représente un terme d'erreur aléatoire. Généralement, on suppose que ε est une variable aléatoire réelle de moyenne nulle. Suivant par exemple Saporta [25], un modèle

courant consiste à considérer l'espérance conditionnelle

$$m(x_1, \ldots, x_d) = \mathrm{E}\left(Y | X_1 = x_1, \ldots, X_d = x_d\right). \tag{1.1}$$

À noter que l'on pourrait également régresser la variable Y selon des indicateurs autres que la moyenne, par exemple la médiane conditionnelle ou tout autre percentile de la distribution conditionnelle. On pourrait également considérer les mode et variance conditionnelles. À ce sujet, voir Manski [20] pour plus de détails.

Dans le cas d'une fonction de régression définie par l'espérance conditionnelle, on a

$$\mathrm{E}\left\{m(X_1, \ldots, X_d)\right\} = \mathrm{E}\left\{\mathrm{E}\left(Y | X_1 = x_1, \ldots, X_d = x_d\right)\right\} = \mathrm{E}(Y).$$

À partir de l'identité

$$\mathrm{var}(Y) = \mathrm{var}\left\{\mathrm{E}(Y | X_1, \ldots, X_d)\right\} + \mathrm{E}\left\{\mathrm{var}(Y | X_1, \ldots, X_d)\right\},$$

que l'on retrouve dans les ouvrages classiques de statistique mathématique comme Casella & Berger [5] et Shao [26], on déduit également

$$\mathrm{var}\left\{m(X_1, \ldots, X_d)\right\} = \mathrm{var}(Y) - \mathrm{E}\left\{\mathrm{var}(Y | X_1, \ldots, X_d)\right\}.$$

On a alors que

$$\eta = \frac{\mathrm{var}\left\{m(X_1, \ldots, X_d)\right\}}{\mathrm{var}(Y)} = 1 - \frac{\mathrm{E}\left\{\mathrm{var}(Y | X_1, \ldots, X_d)\right\}}{\mathrm{var}(Y)}.$$

On voit donc que le quotient des variances conditionnelle et non-conditionnelle est inférieur ou égal à 1. Ce quotient sera d'autant plus près de 1 que la variance de Y, conditionnellement à $X_1, \ldots, X_d$, sera en moyenne très faible. L'indice η est donc un bon indicateur de la qualité de la régression. Inversement, lorsque η est proche de 0, cela signifie que $\mathrm{var}(Y | X_1, \ldots, X_d) \approx \mathrm{var}(Y)$, c'est-à-dire que les variables explicatives apportent peu d'information sur Y.

À noter que dans le cas particulier d'une seule variable explicative telle que $m(x) = \beta_0 + \beta_1 x$, on a $\eta = \rho^2$, où ρ est la corrélation entre X_1 et Y, c'est-à-dire

$$\rho = \frac{\mathrm{cov}(Y, X_1)}{\sqrt{\mathrm{var}(X_1)}\sqrt{\mathrm{var}(Y)}}\,.$$

Remarque 1.1. *Les développements précédents s'appliquent à une variable dépendante Y qui prend ses valeurs dans $\mathbb{R}$. Si Y est qualitative, alors la régression s'apparente à un problème de classification au sens où l'on cherche à déterminer une modalité à partir des valeurs des autres variables. L'exemple le plus courant concerne le cas d'une variable Y de type Bernoulli, donc à valeur dans $\{0, 1\}$; dans cette situation, on utilise généralement une fonction de régression de type logistique ou logit.*

1.2 Avantages de l'utilisation des copules

La régression basée sur les copules se profile comme une bonne alternative à la régression linéaire pour remédier à certaines de ses limites. Au final, cette approche permet une plus grande flexibilité que la régression linéaire dans le choix de la distribution d'une variable conditionnellement à des variables explicatives.

1.3 Organisation du mémoire

Dans ce travail, nous adoptons l'approche de Noh *et coll.* [22] en modélisant la dépendance entre Y et $\mathbf{X}$ par l'intermédiaire d'une copule. L'idée principale derrière cette approche est d'écrire la fonction de régression en termes de copule et de distributions marginales, une fois la copule et les distributions marginales estimées, la méthode du plug-in est utilisée pour construire un nouvel estimateur. Le chapitre 2 est essentielle-

ment constitué des rappels sur la régression linéaire, non linéaire et non paramétrique. Au Chapitre 3, quelques rappels sur la théorie des copules sont donnés, ce qui met la table pour la description des modèles de régression par copule ; les liens avec les modèles pseudo-linéaires, des exemples avec des copules populaires, ainsi que l'estimation non paramétrique de ces courbes sont également traités. Le Chapitre 4 discute de certaines failles de la régression par copules, ce qui motive à proposer des modèles construits à partir de copules qui permettent d'induire une régression non-monotone ; des études par simulations montrent l'efficacité de l'approche proposée et répond aux critiques émises par Dette *et coll.* [7] concernant la régression par copules. Il s'agit de la principale contribution de ce mémoire.

CHAPITRE 2

RÉGRESSION : QUELQUES RAPPELS

2.1 La régression linéaire multiple

2.1.1 Modèle

Le modèle de régression le plus connu est le modèle de régression linéaire multiple. Dans ce cas, on suppose que la fonction m définie à l'Équation (1.1) est

$$m(x_1, \ldots, x_d) = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d.$$

Dans le cas particulier d'une seule variable explicative, on retrouve le modèle linéaire simple, c'est-à-dire $Y = \beta_0 + \beta_1 X_1 + \varepsilon$, où ϵ est le terme d'erreur. En définissant les vecteurs $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)$ et $\mathbf{X} = (1, X_1, \ldots, X_d)$, on peut écrire de façon élégante et compacte que

$$Y = \mathbf{X}\boldsymbol{\beta}^\top + \varepsilon.$$

Supposons maintenant que l'on observe n fois le vecteur aléatoire $(X_1, \ldots, X_d, Y)$ via $(X_{11}, \ldots, X_{1d}, Y_1), \ldots, (X_{n1}, \ldots, X_{nd}, Y_n)$. On suppose que ces observations proviennent du modèle de régression multiple dans le cas standard où les termes d'erreur

sont supposés indépendants et identiquement distribués. On pose alors

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix}, \quad \mathbb{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1d} \\ 1 & X_{21} & \cdots & X_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nd} \end{pmatrix} \quad \text{et} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix},$$

ce qui permet d'écrire

$$\mathbb{Y} = \mathbb{X}\boldsymbol{\beta}^\top + \boldsymbol{\varepsilon}.$$

Dans le cas du modèle linéaire simple, on observe donc des paires $(X_{11}, Y_1), \ldots, (X_{n1}, Y_n)$. On peut alors construire un nuage de points, tel que l'on peut l'observer avec les points bleus sur la Figure 2.1.

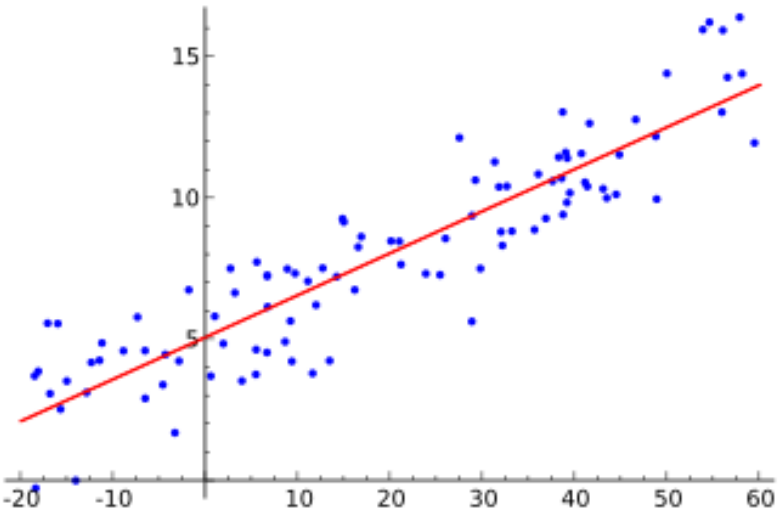

FIGURE 2.1 – Droite de régression tracée à travers un nuage de paires d'observations

2.1.2 Estimation des paramètres

Une étape cruciale pour l'explication de données avec un modèle de régression paramétrique est l'estimation de ses paramètres. En général, l'estimation des paramètres d'un modèle statistique se fait à l'aide de la méthode du maximum de vraisemblance,

des moindres carrés, des moments ou encore des techniques bayésiennes. Dans le cas régression linéaire, on utilise la méthode des moindres carrés. L'idée consiste à minimiser une distance entre le modèle supposé et les observations. Mathématiquement, on cherche à minimiser, en terme du vecteur des paramètres $\boldsymbol{\beta}$, la fonction

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(Y_i - \mathbf{X}_i \boldsymbol{\beta}^\top\right)^2.$$

La solution à ce problème mathématique fournit l'estimateur des moindres carrés ordinaires de $\boldsymbol{\beta}$. D'après le Théorème de Gauss–Markov, cet estimateur est le meilleur estimateur linéaire sans biais du vecteur des coefficients $\boldsymbol{\beta}$; voir par exemple Wasserman & Wasserman [30]. Explicitement, tel qu'on le retrouve par exemple dans l'ouvrage de Cameron *et coll.* [4],

$$\widehat{\boldsymbol{\beta}} = \left(\mathbb{X}^\top \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{Y},$$

en autant bien sûr que la matrice $\mathbb{X}^\top \mathbb{X}$ soit inversible. Sous l'hypothèse de normalité des termes d'erreur, cet estimateur des moindres carrés correspond à l'estimateur du maximum de vraisemblance. De façon explicite, on a l'expression équivalente

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{d+1} \sum_{j=1}^{d+1} \mathbf{X}_j \mathbf{X}_j^\top\right)^{-1} \left(\frac{1}{d+1} \sum_{j=1}^{d+1} \mathbf{X}_j \mathbb{Y}\right),$$

où $\mathbf{X}_j$ est la j-ème colonne de la matrice $\mathbb{X}$. Lorsque $d = 1$, c'est-à-dire dans le cas de la régression linéaire simple, on a

$$\widehat{\beta}_1 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \Bigg/ \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \text{ et } \widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}.$$

On peut voir à la Figure 2.1 la droite en rouge qui correspond à la droite de régression, c'est-à-dire la droite dont la pente est $\widehat{\beta}_1$ et l'ordonnée à l'origine $\widehat{\beta}_0$.

Remarque 2.1. *Lorsque les termes d'erreur ne sont pas tous de même variance et/ou qu'ils sont corrélés, on utilise plutôt la méthode des moindres carrés généralisés ou*

des moindres carrés quasi-généralisés. Pour les décrire, soit Σ la matrice de variance–covariance du vecteur des perturbations $\boldsymbol{\varepsilon}$ (dans le cas standard, Σ est la matrice identité). Alors en se référant par exemple à Cameron et coll. [4], on peut montrer que l'estimateur des moindres carrés généralisés est donné par la formule

$$\widehat{\boldsymbol{\beta}}_{\Sigma} = \left(\mathbb{X}^{\top}\Sigma^{-1}\mathbb{X}\right)^{-1}\mathbb{X}^{\top}\Sigma^{-1}\mathbb{Y}.$$

Cet estimateur suppose la connaissance de la matrice de variance–covariance des termes d'erreur, ce qui n'est généralement par le cas. On doit l'estimer avec la matrice de variance–covariance empirique S. On obtient alors l'estimateur des moindres carrés quasi-généralisés, à savoir $\widehat{\boldsymbol{\beta}}_{S} = \left(\mathbb{X}^{\top}S\mathbb{X}\right)^{-1}\mathbb{X}^{\top}S\mathbb{Y}$.

2.1.3 Qualité de l'ajustement

On définit la valeur prédite ou ajustée par $\widehat{Y} = X\widehat{\beta}$. Le résidu est alors défini comme la différence entre la valeur observée et la valeur prédite, à savoir $\widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$. On définit aussi la somme des carrés des résidus par

$$SSE = \widehat{\varepsilon}^{\top}\widehat{\varepsilon} = \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2.$$

Pour évaluer la qualité de la prévision, on peut utiliser différents critères. Dans un premier temps, on définit la variation expliquée par la régression ainsi que la variation totale respectivement par

$$SSR = \sum_{i=1}^{n}\left(\widehat{Y}_i - \bar{Y}\right)^2 \quad \text{et} \quad SST = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2.$$

Le coefficient de détermination est alors défini par le quotient

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}.$$

Il s'agit, au fond, de la mesure de la proportion de variance dans les valeurs de Y qui est expliquée par les variables explicatives à travers le modèle de régression linéaire. Il est alors clair que R^2 prend ses valeurs dans $[0, 1]$: une valeur de R^2 près de 0 indique un modèle avec un faible pouvoir prédictif, alors qu'un valeur près de 1 indique un fort pouvoir prédictif.

2.1.4 Les limites de la régression linéaire

Beaucoup d'autres méthodes de régression ont été développées pour contrer certaines des limites du modèle linéaire. Si l'hypothèse d'exogénéité des variables explicatives n'est pas vérifiée, l'estimateur des moindres carrés conduit à une estimation biaisée des paramètres du modèle. Dans ce cas, on peut avoir recours à la méthode des variables instrumentales. On appelle «variable instrumentale» une variable qui a un effet sur les variables explicatives suspectées d'endogénéité, mais n'est pas corrélée avec le terme d'erreur. Le vecteur des variables instrumentales $\mathbf{Z}$ est un bon ensemble d'instruments si et seulement s'il est exogène au sens où $\mathrm{E}(\epsilon|\mathbf{Z}) = 0$ et la matrice $\mathrm{E}(\mathbf{Z}\,\mathbf{X}^\top)$ est inversible.

Si les erreurs sur X et sur Y sont de même ordre de grandeur, alors il est plus pertinent d'effectuer une « régression orthogonale » ou « régression géométrique » : pour chaque point expérimental i, l'erreur d_i considérée est la distance du point à la droite modèle, c'est-à-dire la distance prise perpendiculairement à la droite : d'où le terme orthogonal.

Dans le cas où on a deux niveaux d'observations, par exemple la région et les individus, on ne peut plus utiliser la régression linéaire mais plutôt le modèle linéaire hiérarchique, ou modèle linéaire multiniveau, dans lequel on va permettre aux coefficients de varier ; voir par exemple Gelman & Hill [11]. Par exemple, le modèle suivant

est un modèle linéaire hiérarchique :

$$y_{j,i} = \beta_{0,j} + \beta_{1,j} x_{1,j,i} + \ldots + \beta_{K,j} x_{K,j,i} + \epsilon_{j,i} \, .$$

Dans le cas où le nombre de variables explicatives est élevé (c'est-à-dire légèrement inférieur ou même supérieur au nombre d'observations), il peut être intéressant de sélectionner les variables ou de contraindre les coefficients. Robert Tibshirani [28] a développé la méthode du lasso, une méthode de contraction des coefficients.

2.2 Régressions non linéaires

2.2.1 Quelques généralités

Une régression non linéaire consiste à ajuster un modèle de la forme $Y = f_{\mathbf{a}}(\mathbf{X})$, où la fonction $f_{\mathbf{a}} : \mathbb{R}^d \to \mathbb{R}$ est en général non linéaire (sinon on retombe sur le modèle linéaire multiple ...) et $\mathbf{a} = (a_1, \ldots, a_p) \in \mathbb{R}^p$ est un vecteur de paramètres. Deux exemples de modèles non linéaires sont montrés à la Figure 2.2 et à la Figure 2.3.

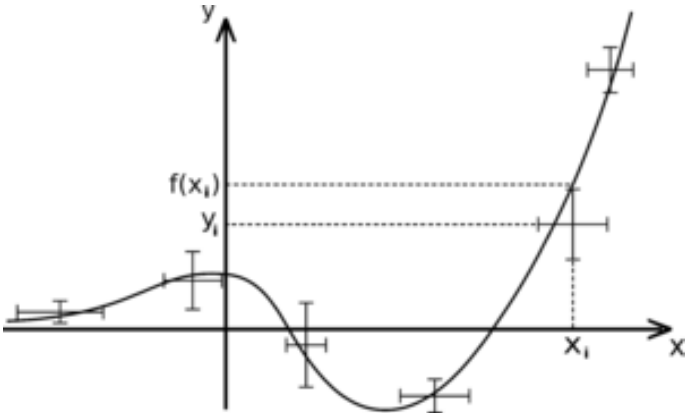

FIGURE 2.2 – Exemple d'une régression non linéaire avec barres d'incertitudes

Pour ajuster un tel modèle à un ensemble d'observations $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$, il

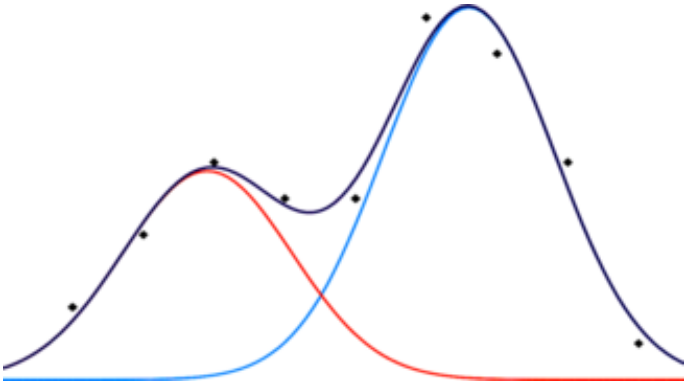

FIGURE 2.3 – Exemple d'une régression non linéaire : décomposition en deux gaussiennes avec six paramètres

s'agit d'obtenir un estimateur de $\mathbf{a}$. Une méthode consiste à minimiser la function

$$S(\mathbf{a}) = \sum_{i=1}^{n} \|Y_i - f_{\mathbf{a}}(\mathbf{X}_i)\|,$$

où $\|\cdot\|$ est une certaine norme sur $\mathbb{R}^2$. Si la norme utilisée est la norme euclidienne usuelle, on parle alors de méthode des moindres carrés. L'estimateur $\widehat{\mathbf{a}}$ est défini comme la solution au système de p équations

$$\frac{\partial S}{\partial a_j} = 0, \quad \text{où } j \in \{1, \ldots, p\},$$

en autant bien entendu que ces dérivées existent. En général, il n'est pas possible de résoudre un tel système de manière analytique.

2.2.2 Algorithme de résolution de Gauss–Newton

L'algorithme de Gauss–Newton, du à Carl Friedrich Gauss, est une méthode de résolution des problèmes de moindres carrés non linéaires. Elle peut être vue comme une modification de la méthode de Newton dans le cas multidimensionnel afin de trouver le minimum d'une fonction à plusieurs variables. Cependant, cet algorithme est spécifique à la minimisation d'une somme de fonctions au carré et présente le

grand avantage de ne pas nécessiter les dérivées secondes, parfois complexes à calculer. Pour le décrire, soient des fonctions $r_1, \ldots, r_n$ à $p \leq n$ variables. L'algorithme de Gauss–Newton consiste à déterminer le minimum de la somme de carrés

$$S(\mathbf{a}) = \sum_{i=1}^{n} r_i^2(\mathbf{a}).$$

En supposant une valeur initiale $\mathbf{a}_0$, la méthode procède par itérations en posant $\mathbf{a}_{s+1} = \mathbf{a}_s + \Delta\mathbf{a}$, où l'incrément $\Delta\mathbf{a}$ vérifie les équations normales $J_\mathbf{r}^\top J_\mathbf{r}\, \Delta\mathbf{a} = -J_\mathbf{r}^\top \mathbf{r}$, où $\mathbf{r} = (r_1, \ldots, r_m)$ et $J_\mathbf{r} \in \mathbb{R}^{n \times p}$ est la matrice jacobienne des dérivées de $\mathbf{r}$ par rapport à $\mathbf{a}$ évalué en $\mathbf{a}_s$. Dans les problèmes d'ajustement de données, où le but est de trouver les paramètres $\mathbf{a}$ d'un modèle $Y = f_\mathbf{a}(\mathbf{X})$, les fonctions $r_1, \ldots, r_n$ correspondent aux résidus, à savoir que $r_i(\mathbf{a}) = Y_i - f_\mathbf{a}(\mathbf{X}_i)$ pour $i \in \{1, \ldots, n\}$. On a alors, dans ce cas,

$$\mathbf{a}_{s+1} = \mathbf{a}_s - \left(J_\mathbf{r}^\top J\right)^{-1} J_\mathbf{r}^\top \mathbf{r}.$$

2.2.3 Algorithme de résolution de Levenberg–Marquardt

L'algorithme de Levenberg–Marquardt, initialement développé par Levenberg [19] et publié plus tard par Marquardt [21], permet d'obtenir une solution numérique au problème de minimisation d'une fonction non linéaire à plusieurs variables. Il repose sur les méthodes sous-jacentes à l'algorithme de Gauss–Newton, ainsi que sur l'algorithme du gradient. Plus stable que celui de Gauss–Newton, l'algorithme de Levenberg–Marquardt est capable d'atteindre la solution même si la valeur initiale est très éloignée du minimum recherché.

Remarque 2.2. *L'algorithme du gradient désigne un algorithme d'optimisation différentiable ; voir par exemple Avriel [2] pour plus de détails. Il est par conséquent destiné à minimiser une fonction réelle différentiable définie sur un espace euclidien (par exemple, $\mathbb{R}^d$, l'espace des d-uplets de nombres réels, muni d'un produit scalaire) ou, plus généralement, sur un espace hilbertien. L'algorithme est itératif et procède*

donc par améliorations successives.

2.3 Régressions non paramétriques

2.3.1 Idée générale

La régression non paramétrique est une forme d'analyse de la régression dans lequel le prédicteur, ou fonction d'estimation, ne prend pas de forme prédéterminée, mais est construit selon les informations provenant des données. On suppose ici un modèle de la forme $Y = f(\mathbf{X})$, sauf que contrairement à la régression non linéaire, la fonction f est totalement indéterminée. La régression non paramétrique exige par le fait même des tailles d'échantillons élevées car les données doivent fournir de l'information non pas sur la valeur de paramètres, mais sur la structure complète de la fonction f ; pour de plus amples détails, voir Ahamada & Flachaire [1].

2.3.2 Modèle de régression additif

Le modèle additif consiste à supposer que $f(\mathbf{x}) = a_0 + f_1(x_1) + \cdots + f_p(x_p)$, ce qui permet de simplifier la problème à l'estimation de p fonctions à une seule variable. En général, on suppose que $f_1, \ldots, f_p$ sont dérivables. Il existe plusieurs variantes à ce modèle. Par exemple, on peut supposer que certaines fonctions sont linéaires, c'est-à-dire que $f_j = a_j x_j$ pour certains $j \in \{1, \ldots p\}$. Une autre possibilité consiste à introduire des termes d'interaction faisant intervenir des fonctions à deux variables de la forme $f_{jj'}(x_j, x_{j'})$ pour certains $j, j' \in \{1, \ldots, p\}$.

2.3.3 Régression locale

La régression locale consiste à faire de la régression par parties : on partitionne d'abord l'espace des variables explicatives et on effectue une régression *locale* sur chaque élément de la partition. Bien que la régression au sein même d'une zone puisse être paramétrique, la méthode est considérée comme non paramétrique. On fait ainsi fréquemment de la régression locale polynomiale ou de la régression locale par spline.

2.3.4 Estimation par noyau

La méthode de l'estimation par noyau consiste à considérer une fonction $\mathcal{K} : \mathbb{R}^p \to \mathbb{R}$ qui est symétrique et semi-définie positive de telle sorte que

$$f(\mathbf{x}) = \sum_k \mathbf{a}_k \, \mathcal{K}\left(\mathbf{x} - \mathbf{x}_k^\star\right),$$

où $\mathbf{x}_1^\star, \ldots, \mathbf{x}_p^\star$ sont des points de l'espace des variables explicatives. Contrairement à la régression locale, la fonction $\mathcal{K}$ est définie sur tout l'espace des variables explicatives. Des choix typiques pour $\mathcal{K}$ sont les formes linéaires, polynomiales et gaussiennes.

2.4 Conclusion du Chapitre 2

Ce chapitre a rappelé quelques notions standards concernant la régression d'une variable Y sur un ensemble de variables explicatives $X_1, \ldots, X_d$. Parmi celles-ci, on retrouve bien entendu la régression linéaire, mais aussi des méthodes plus générales et plus flexibles, comme les régressions non linéaires et non-paramétriques. Le chapitre suivant propose une approche encore plus générale en construisant des courbes de régression à l'aide de modèles de copules.

CHAPITRE 3

RÉGRESSION À BASE DE COPULES

3.1 Mise en situation

Soit un vecteur aléatoire $\mathbf{X} = (X_1, \ldots, X_d)$ constitué de variables explicatives, de même qu'une variable aléatoire Y qui agit à titre de variable dépendante. L'objectif de ce chapitre est de construire un modèle pour la régression de Y sur $\mathbf{X}$, à savoir

$$m(\mathbf{x}) = \mathrm{E}\left(Y|\mathbf{X} = \mathbf{x}\right).$$

Si on suppose que la loi conjointe de $(Y, \mathbf{X})$ est normale, alors $m(\mathbf{x})$ s'écrit sous la forme du modèle de régression multiple à d variables explicatives. On peut toutefois opérer de façon plus générale en supposant que la loi de $(Y, \mathbf{X})$ est $f_{Y,\mathbf{X}}$, de sorte que

$$m(\mathbf{x}) = \int_{\mathbb{R}} y\, f_{Y|\mathbf{X}}(y|\mathbf{x})\, \mathrm{d}y = \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}} y\, f_{Y,\mathbf{X}}(y, \mathbf{x})\, \mathrm{d}y \tag{3.1}$$

On verra comment la notion de copule d'une loi de probabilité multidimensionnelle peut être exploitée afin de construire des modèles extrêmement flexibles pour la fonction de régression $m(\mathbf{x})$. Avant de procéder formellement, la prochaine section rappel les principales définitions et propriétés de base relatives aux copules.

3.2 Copules : quelques rappels utiles

Dans les domaines des probabilités et de la statistique, les copules fournissent des outils statistiques qui permettent de caractériser la dépendance entre plusieurs variables aléatoires. Le terme « copule »a été introduit par le mathématicien américain Abe Sklar dans son désormais célèbre article Sklar [27] paru en 1959 et intitulé « Fonctions de répartition à n dimensions et leurs marges ». À la base, « copule »provient du latin « copulo »qui signifie « lier ensemble, attacher ».

3.2.1 Définition mathématique d'une copule

Soit $\mathbf{u} = (u_1, \ldots, u_d)$, un vecteur dans le carré unitaire d-dimensionnel $[0,1]^d$. Une fonction $C : [0,1]^d \to [0,1]$, où $d \geq 2$, sera appelée une copule si elle satisfait les trois propriétés suivantes :

- ($\mathcal{C}_1$) $C(\mathbf{u}) = 0$ si $\min(u_1, \ldots, u_d) = 0$, c'est-à-dire si au moins une des composantes de $\mathbf{u}$ est nulle ;
- ($\mathcal{C}_2$) Si toutes les composantes de $\mathbf{u}$ sont 1, sauf la j-ème composante (qui vaut alors u_j), on a $C(\mathbf{u}) = u_j$;
- ($\mathcal{C}_3$) La fonction C possède la propriété de d-croissanse, à savoir que sa mesure sur n'importe quel hyper-rectangle de $[0,1]^d$ est non-négative.

Dans le cas $d = 2$, la propriété $\mathcal{C}_1$ implique que $C(0,u) = C(u,0) = 0$, alors que $\mathcal{C}_2$ entraine $C(1,u) = C(u,1) = u$ pour tout $u \in [0,1]$. Enfin, la propriété $\mathcal{C}_3$ assure que pour n'importe quels $u_1 \leq u_1' \in [0,1]$ et $u_2 \leq u_2' \in [0,1]$,

$$C(u_1', u_2') - C(u_1', u_2) - C(u_1, u_2') + C(u_1, u_2) \geq 0.$$

3.2.2 Relation entre une fonction de répartition et une copule

Le théorème suivant est la base de l'application des copules pour modéliser les distributions de vecteurs aléatoires en statistique.

Théorème 3.1 (Théorème de Sklar, 1959)**.** *Si F est une fonction de répartition de dimension $d \geq 2$ dont les lois marginales sont $F_1, \ldots, F_d$, alors il existe une copule $C : [0,1]^d \to [0,1]$ telle que*

$$F(x_1, \ldots, x_d) = C\left\{F_1(x_1), \ldots, F_d(x_d)\right\}. \tag{3.2}$$

Si $F_1, \ldots, F_d$ sont continues, alors C dans la représentation (3.2) *est unique.*

Supposons que $(X_1, \ldots, X_d)$ suit une loi F de marges continues $F_1, \ldots, F_d$. Alors l'unique copule de F correspond à la fonction de répartition du vecteur aléatoire

$$(U_1, \ldots, U_d) = (F_1(X_1), \ldots, F_d(X_d)).$$

En effet, si F_j^{-1} est l'inverse de la fonction de répartition F_j, alors

$$\begin{aligned} \mathbb{P}\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) &= \mathbb{P}\left\{F_1(X_1) \leq u_1, \ldots, F_d(X_d) \leq u_d\right\} \\ &= \mathbb{P}\left\{X_1 \leq F_1^{-1}(u_1), \ldots, X_d \leq F_d^{-1}(u_d)\right\}. \end{aligned}$$

Ensuite, d'après l'Équation (3.2) du Théorème 3.1, on déduit que

$$\mathbb{P}\left(U_1 \leq u_1, \ldots, U_d \leq u_d\right) = C\left\{F_1 \circ F_1^{-1}(u_1), \ldots, F_d \circ F_d^{-1}(u_d)\right\} = C(u_1, \ldots, u_d),$$

où $f \circ g$ dénote la composition d'une fonction par une autre, c'est-à-dire que

$$f \circ g(x) = f\left\{g(x)\right\}.$$

3.2.3 Extraction d'une copule

Le Théorème 3.1 dû à Sklar [27] établit qu'une loi F de marges $F_1, \ldots, F_d$ continues et de copule C admet la représentation unique $H(x) = C\{F_1(x_1), \ldots, F_d(x_d)\}$. En posant, dans cette expression, $u_j = F_j(x_j)$ pour chaque $j \in \{1, \ldots, d\}$, cette équation s'écrit de façon équivalente

$$C(\mathbf{u}) = F\left\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right\}. \tag{3.3}$$

Il s'agit de la recette pour extraire l'unique copule d'une loi d-dimensionnelle continue.

3.2.4 Densité d'une copule

Puisqu'une copule est une fonction de répartition, sa densité, si elle existe, est simplement définie par

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

En partant de la formule (3.3) pour l'extraction d'une copule, on a

$$\begin{aligned} c(\mathbf{u}) &= \frac{\partial^d}{\partial u_1 \cdots \partial u_d} F\left\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right\} \\ &= f\left\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right\} \prod_{j=1}^{d} \frac{\partial}{\partial u_j} F_j^{-1}(u_j) \\ &= f\left\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right\} \Big/ \prod_{j=1}^{d} f_j \circ F_j^{-1}(u_j)\,, \end{aligned}$$

où f est la densité de F et $f_1, \ldots, f_d$ sont les densités marginales.

3.2.5 Propriété d'invariance des copules

Soit un vecteur aléatoire continu $\mathbf{X} = (X_1, \ldots, X_d)$ dont les marges sont $F_1, \ldots, F_d$ et la copule est C. Cette copule est invariante sous des transformations monotones croissantes. Autrement dit, la copule des variables aléatoires transformées $\widetilde{X}_1, \ldots, \widetilde{X}_d$, où $\widetilde{X}_j = \gamma_j(X_j)$ pour chaque $j \in \{1, \ldots, d\}$ et où $\kappa_1, \ldots, \kappa_d : \mathbb{R} \to \mathbb{R}$ sont des fonctions monotones croissantes, est la même que celle de $\mathbf{X}$. D'une part, par un calcul direct, la fonction de répartition de $(\widetilde{X}_1, \ldots, \widetilde{X}_d)$ s'écrit

$$\begin{aligned}
\widetilde{F}(x_1, \ldots, x_d) &= \mathbb{P}\left\{\kappa_1(X_1) \leq x_1, \ldots, \kappa_d(X_d) \leq x_d\right\} \\
&= \mathbb{P}\left\{X_1 \leq \kappa_1^{-1}(x_1), \ldots, X_d \leq \kappa_d^{-1}(x_d)\right\} \\
&= F\left\{\kappa_1^{-1}(x_1), \ldots, \kappa_d^{-1}(x_d)\right\} \\
&= C\left\{F_1 \circ \kappa_1^{-1}(x_1), \ldots, F_d \circ \kappa_d^{-1}(x_d)\right\}.
\end{aligned}$$

D'autre part, par un calcul similaire, on établit que la loi marginale de $\widetilde{X}_j$ est

$$\widetilde{F}_j(x_j) = \mathbb{P}\left\{\kappa_j(X_j) \leq x_j\right\} = F_j \circ \kappa_j^{-1}(x_j).$$

On peut donc déduire que $\widetilde{F}(x_1, \ldots, x_d) = C\{\widetilde{F}_1(x_1), \ldots, \widetilde{F}_d(x_d)\}$. Par conséquent, on voit bien que la copule $\widetilde{F}$ est C.

3.2.6 Copule de survie

Supposons maintenant des fonctions $\kappa_1, \ldots, \kappa_d : \mathbb{R} \to \mathbb{R}$ monotones décroissantes et, à nouveau, les variables aléatoires transformées $\widetilde{X}_1, \ldots, \widetilde{X}_d$, où $\widetilde{X}_j = \gamma_j(X_j)$ pour

chaque $j \in \{1, \ldots, d\}$. D'autres part on a $\bar{F}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \geq \mathbf{x})$. De là,

$$\begin{aligned}
\widetilde{F}(x_1, \ldots, x_d) &= \mathbb{P}\left\{\kappa_1(X_1) \leq x_1, \ldots, \kappa_d(X_d) \leq x_d\right\} \\
&= \mathbb{P}\left\{X_1 \geq \kappa_1^{-1}(x_1), \ldots, X_d \geq \kappa_d^{-1}(x_d)\right\} \\
&= \bar{F}\left\{\kappa_1^{-1}(x_1), \ldots, \kappa_d^{-1}(x_d)\right\} \\
&= C^\star\left\{\bar{F}_1 \circ \kappa_1^{-1}(x_1), \ldots, \bar{F}_d \circ \kappa_d^{-1}(x_d)\right\} .
\end{aligned}$$

La copule de $\widetilde{\mathbf{X}}$ dans ce cas s'appelle la *copule de survie* de C ; elle est notée $C^\star$. Les marges de $\widetilde{F}$ sont $\bar{F}_1 \circ \kappa_1^{-1}, \ldots, \bar{F}_d \circ \kappa_d^{-1}$ avec $\bar{F}_j = 1 - F_j$, la copule de survie satisfait

$$\bar{F}(x_1, \ldots, x_d) = C^\star\left\{\bar{F}_1(x_1), \ldots, \bar{F}_d(x_d)\right\} .$$

3.2.7 Indépendance & dépendance positive parfaite

Si des variables aléatoires $X_1, \ldots, X_d$ sont indépendantes, alors

$$F(x_1, \ldots, x_d) = F_1(x_1) \times \cdots \times F_d(x_d) .$$

Dans ce cas, une application directe de la formule (3.3) permet de déduire que la copule associée à l'indépendance entre d variables est $\Pi(\mathbf{u}) = u_1 \times \cdots \times u_d$.

Soient des fonctions monotones croissantes $\kappa_1, \ldots, \kappa_d$ et une certaine variable aléatoire continue W de la loi F_W. On dit que les variables $X_1, \ldots, X_d$ sont parfaitement dépendantes positivement si $X_\ell = \kappa_\ell(W)$. Comme la loi de X_ℓ est $F_W \circ \kappa_\ell^{-1}$, la copule est la loi jointe de $U = (U_1, \ldots, U_d)$, où

$$U_\ell = F_W \circ \kappa_\ell^{-1}(X_\ell) = F_W(W) .$$

Puisque $V = F_W(W) \sim \mathcal{U}[0,1]$, on a directement

$$\mathbb{P}(\mathbf{U} \leq \mathbf{u}) = \mathbb{P}\left(V \leq u_1, \ldots, V \leq u_d\right) = \mathbb{P}\left(V \leq \min_{j \in \{1,\ldots,d\}} u_j\right) = \min_{j \in \{1,\ldots,d\}} u_j.$$

La fonction $M(\mathbf{u}) = \min_{j \in \{1,\ldots,d\}} u_j$ est la copule de la dépendance positive parfaite.

3.3 Régression par copules

3.3.1 Contexte

Supposons que la fonction de répartition de Y est F_0, et que les fonctions de répartition de $X_1, \ldots, X_d$ sont respectivement $F_1, \ldots, F_d$. On notera par la même occasion la densité de Y par f_0 et celles de $X_1, \ldots, X_d$ par $f_1, \ldots, f_d$. Pour $\mathbf{x} = (x_1, \ldots, x_d) \in R^d$, on notera également $\mathbf{F}(\mathbf{x}) = (F_1(x_1), \ldots, F_d(x_d)$ comme le vecteur des fonctions de répartitions marginales de $X_1, \ldots, X_d$ évaluées respectivement aux points $x_1, \ldots, x_d$. D'après le Théorème de Sklar [27], la fonction de répartition conjointe de $(Y, \mathbf{X})$ évaluée en $(y, \mathbf{x})$ peut être exprimée par

$$F_{Y,\mathbf{X}}(y, \mathbf{x}) = C\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\}, \tag{3.4}$$

où C est la copule de $(Y, \mathbf{X})$. En particulier, la fonction de répartition du vecteur $\mathbf{X} = (X_1, \ldots, X_d)$ des covariables s'écrit $F_{\mathbf{X}}(\mathbf{x}) = C_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\}$, où $C_{\mathbf{X}}(\mathbf{u}) = C(1, \mathbf{u})$.

3.3.2 Une expression pour la densité conditionnelle

Dans le contexte décrit à la sous-section 3.3.1, Noh *et coll.* [22] ont étudié une nouvelle approche pour estimer une fonction de régression basée sur les copules. Celle-ci consiste

à écrire la fonction de régression en termes de la copule et des distributions marginales. Pour la décrire, soient d'abord

$$c(u_0, \mathbf{u}) = \frac{\partial^{d+1}}{\partial u_0 \partial u_1 \cdots \partial u_d} C\left(u_0, u_1, \ldots, u_d\right),$$

la densité de la copule C, de même que $c_{\mathbf{X}}$, la densité de $C_{\mathbf{X}}$. En dérivant l'expression de $F_{Y,\mathbf{X}}$ à l'Équation (3.4) par rapport à $y, x_1, \ldots, x_d$, on obtient que la densité conjointe de $(Y, \mathbf{X})$ s'exprime par

$$f_{Y,\mathbf{X}}(y, \mathbf{x}) = c\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\} f_0(y)\, f_1(x_1) \times \cdots \times f_d(x_d).$$

Par le même raisonnement, la densité de $\mathbf{X}$ peut s'écrire

$$f_{\mathbf{X}}(\mathbf{x}) = c_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\} f_1(x_1) \times \cdots \times f_d(x_d).$$

La densité conditionnelle de Y étant donné $\mathbf{X} = \mathbf{x}$ peut alors s'exprimer par

$$\begin{aligned} f_{Y|\mathbf{X}}(y, \mathbf{x}) &= \frac{f_{Y,\mathbf{X}}(y, \mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{c\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\} f_0(y)\, f_1(x_1) \times \cdots \times f_d(x_d)}{c_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\} f_1(x_1) \times \cdots \times f_d(x_d)} \\ &= \frac{c\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\} f_0(y)}{c_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\}}. \end{aligned}$$

3.3.3 L'équation d'une régression par copules

En partant de l'équation (3.1), la moyenne de Y étant donné que $\mathbf{X} = \mathbf{x}$ peut donc s'écrire sous la forme

$$m(\mathbf{x}) = \int_{\mathbb{R}} y\, \frac{c\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\}}{c_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\}}\, \mathrm{d}F_0(y). \tag{3.5}$$

En effectuant le changement de variable $u_0 = F_0(y)$, alors $\mathrm{d}u_0 = f_0(y)\,\mathrm{d}y$ et $y = F_0^{-1}(u_0)$. On a alors la formulation alternative

$$m(\mathbf{x}) = \int_0^1 F_0^{-1}(u_0)\,\frac{c\,\{u_0, \mathbf{F}(\mathbf{x})\}}{c_{\mathbf{X}}\,\{\mathbf{F}(\mathbf{x})\}}\,\mathrm{d}u_0.$$

Enfin, si on suppose que toutes les marges sont uniformes sur $[0,1]$, c'est-à-dire que $F_0(a) = F_1(a) = \cdots = F_d(a) = a$, alors pour $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$,

$$m(\mathbf{u}) = \int_0^1 u_0\,\frac{c\,(u_0, \mathbf{u})}{c_{\mathbf{X}}(\mathbf{u})}\,\mathrm{d}u_0.$$

3.3.4 Cas d'une seule variable explicative

Dans le cas d'une seule variable explicative X_1, c'est-à-dire lorsque $d = 1$, alors $c_{X_1}\{F_1(x_1)\} = 1$. L'équation (3.5) se simplifie alors à

$$m(x_1) = \mathrm{E}(Y|X_1 = x_1) = \int_{\mathbb{R}} y\,c\,\{F_0(y), F_1(x_1)\}\,f_0(y)\,\mathrm{d}y.$$

Dans la situation où les marges de Y et de X_1 sont uniformes sur $[0,1]$, alors

$$m(u_1) = \int_0^1 u_0\,c\,(u_0, u_1)\,\mathrm{d}u_0. \tag{3.6}$$

3.4 Estimation semi-paramétrique

Supposons un échantillon de n observations d'une variable dépendante $\mathbf{X}$ et d'un vecteur de variables explicatives Y, à savoir $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$, où $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$. On suppose que ces vecteurs aléatoires de dimension $p+1$ proviennent d'une distribution conjointe F dont les marges sont F_0 et $\mathbf{F} = (F_1, \ldots, F_p)$, respectivement. L'estimation non paramétrique des courbes de régression va s'appuyer sur des estima-

tions entièrement non-paramértiques de F_0 et de $\mathbf{F}$. D'une part, F_0 sera estimée par la fonction de répartition empirique

$$F_{n0}(y) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\left(Y_i \leq y\right).$$

D'autre part, pour chaque $j \in \{1, \ldots, n\}$, F_j sera estimée via

$$F_{nj}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\left(X_{ij} \leq x\right).$$

Pour une raison qui apparaitra plus claire lors de la decription des estimateurs des courbes de régression, on emploiera plutôt des versions légèrement re-standardisées de F_{n0} et de $F_{n1}, \ldots, F_{np}$, à savoir

$$\widetilde{F}_{n0}(y) = \frac{1}{n+1}\sum_{i=1}^{n} \mathrm{I}\left(Y_i \leq y\right) \quad \text{et} \quad \widetilde{F}_{nj}(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathrm{I}\left(X_{ij} \leq x\right).$$

L'objectif ici est d'estimer la courbe de régression exprimée par l'Équation (3.5), à savoir

$$m(\mathbf{x}) = \int_{\mathbb{R}} y \, \frac{c\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\}}{c_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\}} \, \mathrm{d}F_0(y).$$

On suppose dans la suite que la copule C de la population est connue et admet une densité c. En posant $\widetilde{\mathbf{F}}_n = (\widetilde{F}_{n1}, \ldots, \widetilde{F}_{np})$, une version empirique de m est alors

$$m_n(\mathbf{x}) = \int_{\mathbb{R}} y \, \frac{c\left\{\widetilde{F}_{n0}(y), \widetilde{\mathbf{F}}_n(\mathbf{x})\right\}}{c_{\mathbf{X}}\left\{\widetilde{\mathbf{F}}_n(\mathbf{x})\right\}} \, \mathrm{d}\widetilde{F}_{n0}(y).$$

Puisque $\mathrm{d}\widetilde{F}_{n0}(y)$ accorde un poids de $1/(n+1)$ à $y = Y_i$ pour chaque $i \in \{1, \ldots, n\}$, et s'annule ailleurs, on a l'expression équivalente

$$m_n(\mathbf{x}) = \frac{1}{n+1}\sum_{i=1}^{n} Y_i \, \frac{c\left\{\widetilde{F}_{n0}(Y_i), \widetilde{\mathbf{F}}_n(\mathbf{x})\right\}}{c_{\mathbf{X}}\left\{\widetilde{\mathbf{F}}_n(\mathbf{x})\right\}}.$$

Dans le cas d'une seule variable explicative, c'est-à-dire lorsque $p = 1$, on a

$$m_n(x) = \frac{1}{n+1} \sum_{i=1}^{n} Y_i \, c \left\{ \widetilde{F}_{n0}(Y_i), \widetilde{F}_{n1}(x) \right\}.$$

Ici, l'utilisation de $\widetilde{F}_{n0}, \widetilde{\mathbf{F}}_n$ au lieu de F_{n0}, F_n est pour éviter d'évaluer la densité d'une copule à une composante égale à 1, ce dernier cas de figure pouvant causer des problèmes numériques. On note enfin que les valeurs de $\widetilde{F}_{n0}(Y_1), \ldots, \widetilde{F}_{n0}(Y_n)$ sont directement liées aux rangs des observations $Y_1, \ldots, Y_n$. En effet, si R_{Yi} dénote le rang de Y_i parmi $Y_1, \ldots, Y_n$, alors

$$\widetilde{F}_{n0}(Y_i) = \frac{1}{n+1} \sum_{j=1}^{n} \mathrm{I}\left(Y_j \leq Y_i\right) = \frac{R_{Yi}}{n+1}.$$

On pourrait donc tout aussi bien écrire

$$m_n(x) = \frac{1}{n+1} \sum_{i=1}^{n} Y_i \, c \left\{ \frac{R_{Yi}}{n+1}, \widetilde{F}_{n1}(x) \right\}.$$

En annexe, quelques détails sont donnés concernant la convergence de l'estimateur m_n. Notamment, il est montré que $m_n(x)$ est asymptotiquement sans biais pour $m(x)$. La normalité asymptotique de $\sqrt{n}\{m_n(x) - m(x)\}$ est également discutée.

3.5 Conclusion du Chapitre 3

Dans ce chapitre, un modèle général de régression par copule a été présenté. Cette stratégie est extrêmement porteuse, dans la mesure qu'elle permet de contrôler la structure de dépendance des variables exogènes vis-à-vis la variable dépendante à l'aide d'une copule, et de laisser les marges arbitraires. La question de l'estimation des paramètres dans de tels modèles a aussi été abordée.

CHAPITRE 4

EXEMPLES AVEC DES COPULES POPULAIRES

4.1 Copule de Farlie–Gumbel–Morgenstern

La copule de Farlie–Gumbel–Morgenstern (FGM) est donnée par

$$C(u_1, u_2) = u_1 u_2 + \theta u_1 u_2 (1 - u_1)(1 - u_2),$$

où $|\theta| \leq 1$. Sa densité est donc donnée par

$$c(u_1, u_2) = 1 + \theta(1 - 2u_1)(1 - 2u_2).$$

Comme cas particulier de l'Équation (3.6), on a alors

$$\begin{aligned} m(u_1) &= \int_0^1 u_0 \left\{1 + \theta(1 - 2u_1)(1 - 2u_0)\right\} \mathrm{d}u_0 \\ &= \int_0^1 u_0 \, \mathrm{d}u_0 + \theta(1 - 2u_1) \int_0^1 \left(u_0 - u_0^2\right) \mathrm{d}u_0 \\ &= \frac{1}{2} - \frac{\theta}{6}(1 - 2u_1). \end{aligned}$$

Le cas où $\theta = 0$ correspond à la copule d'indépendance. Il est donc normal d'obtenir $m(u_1) = 1/2$ dans cette situation, car il s'agit de la moyenne d'une loi uniforme sur

$(0,1)$. Par contre, la copule FGM a la particularité de fournir $m(1/2) = 1/2$, et ce pour n'importe quelle valeur de $\theta \in [-1,1]$. On remarque enfin que

$$\left| m(u_1) - \frac{1}{2} \right| \leq \frac{\theta}{6}\,.$$

4.2 Copules Archimédiennes

4.2.1 Formulation générale

On dit qu'une copule multidimensionnelle appartient à la famille Archimédienne si elle peut s'écrire sous la forme

$$C_\phi(\mathbf{u}) = \Psi\left\{\sum_{j=1}^{d} \phi(u_j)\right\},$$

où $\phi : [0,1] \to [0,\infty)$ s'appelle le générateur et $\Psi = \phi^{-1}$. Le générateur doit satisfaire $\phi(1) = 0$ et $(-1)^\ell \psi^{[\ell]}(t) \geq 0$ pour tout $\ell \in \{1,\ldots,d\}$. Cette classe de copules a été introduite par Genest & Mackay [13]. Dans le cas $d = 2$, la densité de C_ϕ s'écrit

$$c_\phi(u_1,u_2) = \Psi''\left\{\phi(u_1) + \phi(u_2)\right\}\phi'(u_1)\phi'(u_2).$$

Ainsi, comme cas particulier de l'Équation (3.6),

$$m(u_1) = \int_0^1 u_0\,\Psi''\left\{\phi(u_1) + \phi(u_0)\right\}\phi'(u_1)\phi'(u_0)\mathrm{d}u_0.$$

En effectuant le changement de variable $s = \phi(u_1) + \phi(u_0)$, on obtient

$$m(u_1) \;=\; \int_\infty^{\phi(u_1)} \psi\left\{s - \phi(u_1)\right\}\Psi''(s)\,\phi'(u_1)\,\mathrm{d}s. \tag{4.1}$$

Quelques cas particuliers de copules Archimédiennes sont traités dans la suite.

4.2.2 Copule de Clayton

La copule de s'extrait d'un modèle initialement étudié par Clayton [6]. Le générateur de cette copule est défini pour $\theta > 0$ par $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$. On montre alors que cette copule s'écrit

$$C_\theta(\mathbf{u}) = \left(\sum_{j=1}^{d} u_j^{-\theta} - d + 1 \right)^{-1/\theta}.$$

Quelques calculs directs permettent de montrer que

$$\Psi(s) = (\theta s + 1)^{-\frac{1}{\theta}}, \quad \Psi'(s) = -(\theta s + 1)^{-\frac{1}{\theta}-1} \quad \text{et} \quad \Psi''(s) = (\theta + 1)(\theta s + 1)^{-\frac{1}{\theta}-2}.$$

Comme cas particulier de l'Équation (4.1),

$$m(u_1) = (\theta + 1) \int_{\infty}^{\phi(u_1)} \left(\theta s - u_1^{-\theta}\right)^{-\frac{1}{\theta}} (\theta s + 1)^{-\frac{1}{\theta}-2} \, \mathrm{d}s.$$

4.2.3 Copule de Gumbel

La copule de Gumbel [16] telle qu'extraite du modèle bidimensionnel logistique considéré par Gumbel et Hougaard est générée par $\phi_\theta(t) = |\ln t|^{1/(1-\theta)}$, où $\theta \in [0, 1]$. La copule de Gumbel est donc de la forme

$$C_\theta(\mathbf{u}) = \exp\left\{ - \left(\sum_{j=1}^{d} |\ln u_j|^{1/(1-\theta)} \right)^{1-\theta} \right\}.$$

Genest & Rivest [14] ont montré qu'elle possède l'unique particularité d'appartenir à la fois aux familles Archimédienne et à valeurs extrêmes. Pour utiliser l'Équation (4.1) de la fonction de régression des copules archimédiennes dans le cas de la copule de Gumbel, on note que $\Psi(s) = \exp\left(s^{1-\theta}\right)$, $\Psi'(s) = (1-\theta)\, s^{-\theta} \exp\left(s^{1-\theta}\right)$ et

$$\Psi''(s) = (1-\theta) \left\{ (1-\theta) s^{-2\theta} - \theta\, s^{-\theta-1} \right\} \exp\left(s^{1-\theta}\right).$$

La fonction de régression de la copule de Gumbel peut alors s'écrire

$$\begin{aligned} m(u_1) &= (1-\theta)\int_{\infty}^{\phi(u_1)} \exp\left\{\left(s-\ln(u_1)^{\frac{1}{1-\theta}}\right)^{1-\theta}\right\}\left\{(1-\theta)s^{-2\theta}-\theta\, s^{-\theta-1}\right\}\exp\left(s^{1-\theta}\right)\mathrm{d}s \\ &= (1-\theta)\int_{\infty}^{\phi(u_1)} \left\{(1-\theta)s^{-2\theta}-\theta\, s^{-\theta-1}\right\}\exp\left\{s^{1-\theta}+\left(s-\ln(u_1)^{\frac{1}{1-\theta}}\right)^{1-\theta}\right\}\mathrm{d}s \end{aligned}$$

4.2.4 Copule de Frank

La copule de Frank [10] a été initialement étudiée par Genest [12]. Son générateur est $\phi_\theta(t) = \ln\left(1-e^{-\theta}\right) - \ln\left(1-e^{-\theta t}\right)$. Ainsi,

$$C_\theta(\mathbf{u}) = -\frac{1}{\theta}\left\{1-\frac{1}{1-e^{-\theta}}\prod_{j=1}^{d}(1-e^{-\theta u_j})\right\}.$$

On montre que

$$\Psi(s) = -\frac{1}{\theta}\ln\left(\frac{e^s-e^{-\theta}-1}{e^s}\right), \quad \Psi'(s) = -\frac{1}{\theta}\left(\frac{1-e^{-\theta}}{e^s-e^{-\theta}-1}\right)$$

et

$$\Psi''(s) = \frac{e^s(1-e^{-\theta})}{\theta(e^s-e^{-\theta}-1)^2}.$$

La fonction de régression induite par la copule de Frank est donc

$$\begin{aligned} m(u_1) &= \int_{\infty}^{\phi(u_1)} -\frac{1}{\theta}\ln\left\{1-\frac{(e^{-\theta}-1)^2}{e^s(1-e^{-\theta u_1})}\right\}\frac{e^s(1-e^{-\theta})}{\theta(e^s-e^{-\theta}-1)^2}\,\mathrm{d}s \\ &= -\frac{1}{\theta^2}\int_{\infty}^{\phi(u_1)} \frac{e^s(e^{-\theta}-1)}{(e^s-e^{-\theta}-1)^2}\ln\left\{1-\frac{(e^{-\theta}-1)^2}{e^s(1-e^{-\theta u_1})}\right\}\mathrm{d}s. \end{aligned}$$

4.3 Copules elliptiques

4.3.1 Lois et copules elliptiques multidimensionnelles

Un vecteur aléatoire $\mathbf{X} \in \mathbb{R}^d$ est de distribution elliptique si et seulement si il admet la représentation stochastique $\mathbf{X} = \boldsymbol{\mu} + R\,A\,\mathbf{U}$, où $\boldsymbol{\mu} \in \mathbb{R}^d$ est le vecteur des moyennes, $\mathbf{U}$ est un vecteur aléatoire uniformément distribué sur la sphère unitaire de $\mathbb{R}^d$, R est une variable aléatoire positive et indépendante de $\mathbf{U}$ et $\mathbf{A} \in \mathbb{R}^{d\times d}$ est telle que $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$ est non singulière. La densité d'une distribution elliptique est de la forme

$$f(\mathbf{x}) = |\Sigma|^{1/2}\, g\left\{(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\},$$

où $g : \mathbb{R}^+ \to \mathbb{R}$ est une fonction génératrice de densité. Une copule elliptique est tout simplement la fonction de dépendance que l'on extrait d'une loi elliptique.

4.3.2 Copule Normale

Puisque la loi Normale multidimensionnelle fait partie des lois elliptiques, la copule Normale est une copule elliptique. Ainsi, si Φ_Σ est la fonction de répartition de la loi Normale d-dimensionnelle de moyennes nulles et dont la matrice de corrélation est Σ, alors la copule Normale s'exprime explicitement via

$$C_\Sigma(u_1,\ldots,u_d) = \Phi_\Sigma\left\{\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right\}.$$

Dans cette dernière expression, Φ^{-1} est l'inverse de la fonction de répartition de la loi Normale standard. En particulier, la copule Normale bivariée s'exprime sous la forme

$$C_\rho(u_1,u_2) = \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_1)} \phi_\rho(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

où pour $\rho \in [-1, 1]$ qui correspond au coefficient de corrélation de Pearson,

$$\phi_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{\frac{-(x^2+y^2-2\rho xy)}{2(1-\rho^2)}\right\}.$$

La densité associée à C_ρ est alors

$$\begin{aligned} c_\rho(u_1, u_2) &= \phi_\rho\left\{\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right\}\left(\Phi^{-1}(u_1)\right)'\left(\Phi^{-1}(u_2)\right)' \\ &= \phi_\rho\left\{\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right\} / \phi\left\{\Phi^{-1}(u_1)\right\}\phi\left\{\Phi^{-1}(u_2)\right\}. \end{aligned}$$

Cette densité est montrée à la Figure 4.1 dans le cas $\rho = 0,75$.

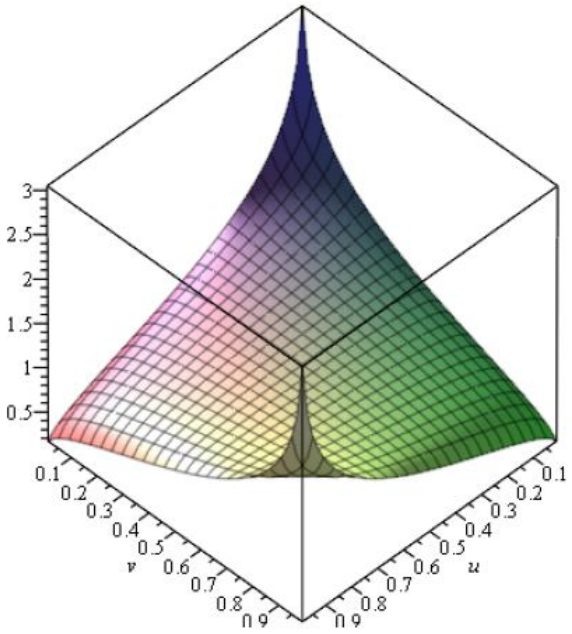

Figure 4.1 – Densité de la copule normale bivariée lorsque $\rho = 0,75$

À noter maintenant que

$$\begin{aligned}
\phi_\rho(x,y) &= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{(y-\rho x)^2+x^2-\rho^2x^2}{2(1-\rho^2)}\right\}\\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{(y-\rho x)^2}{2(1-\rho^2)}-\frac{x^2(1-\rho^2)}{2(1-\rho^2)}\right\}\\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left\{-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right\}\\
&= \phi(x)\,\phi\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right).
\end{aligned}$$

On obtient alors

$$\begin{aligned}
c_\rho(u_1,u_2) &= \phi_\rho\left\{\Phi^{-1}(u_1),\Phi^{-1}(u_2)\right\}/\,\phi\left\{\Phi^{-1}(u_1)\right\}\phi\left\{\Phi^{-1}(u_2)\right\}\\
&= \phi_\rho\left\{\Phi^{-1}(u_1)\right\}\phi\left\{\frac{\Phi^{-1}(u_2)-\rho\,\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right\}\Bigg/\,\phi\left\{\Phi^{-1}(u_1)\right\}\phi\left\{\Phi^{-1}(u_2)\right\}\\
&= \phi\left\{\frac{\Phi^{-1}(u_2)-\rho\,\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right\}\Bigg/\,\phi\left\{\Phi^{-1}(u_2)\right\}.
\end{aligned}$$

Comme cas particulier de l'Équation (3.6),

$$m(u_1)=\int_0^1\frac{u_0}{\phi\left\{\Phi^{-}1(u_0)\right\}}\,\phi\left\{\frac{\Phi^{-1}(u_0)-\rho\,\Phi^{-1}(u_1)}{\sqrt{1-\rho^2}}\right\}\mathrm{d}u_0.$$

Bien évidemment, lorsque $\rho=0$ on retrouve $m(u_1)=1/2$.

4.3.3 Copule de Student

La copule de Student est la fonction de dépendance associée à la loi de Student. Ainsi, si $T_{\Sigma,\nu}$ est la fonction de répartition de la loi de Student à ν degrés de liberté dont la matrice de corrélation est $\Sigma\in\mathbb{R}^{d\times d}$, alors la copule de Student s'écrit implicitement

$$C_{\Sigma,\nu}(\mathbf{u})=T_{\Sigma,\nu}\left\{T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_d)\right\},$$

où T_ν est la fonction de répartition de la loi de Student univariée à ν degrés de liberté. La densité associée à $C_{\Sigma,\nu}$ est alors

$$\begin{aligned} c_{\Sigma,\nu}(\mathbf{u}) &= t_{\Sigma,\nu}\left\{T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_d)\right\}\prod_{j=1}^{d}\left(T_\nu^{-1}(u_j)\right)' \\ &= t_{\Sigma,\nu}\left\{T_\nu^{-1}(u_1),\ldots,T_\nu^{-1}(u_d)\right\}\Bigg/\prod_{j=1}^{d}t_\nu\left\{T_\nu^{-1}(u_j)\right\}, \end{aligned}$$

où $t_{\Sigma,\nu}$ est la densité de Student, c'est-à-dire

$$t_{\Sigma,\nu}(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{|\Sigma|^{-1/2}}{(\nu\pi)^{d/2}}\left(1+\frac{1}{\nu}\,\mathbf{x}\,\Sigma^{-1}\,\mathbf{x}^\top\right)^{-\left(\frac{\nu+d}{2}\right)}.$$

À noter que lorsque $d=1$, on retrouve la densité de Student univariée, à savoir

$$t_\nu(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\nu\pi}}\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}.$$

Dans le cas $d=2$,

$$t_{\rho,\nu}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}\left(1+\frac{x_1^2-2\rho\,x_1\,x_2+x_2^2}{\nu(1-\rho^2)}\right)^{-\left(\frac{\nu}{2}+1\right)}.$$

En posant $a_1 = T_\nu^{-1}(u_1)$ et $a_2 = T_\nu^{-1}(u_2)$, on a donc

$$c_{\rho,\nu}(u_1,u_2) = K_{\rho,\nu}\left(1+\frac{a_1^2-2\rho\,a_1\,a_2+a_2^2}{\nu(1-\rho)^2}\right)^{-\left(\frac{\nu+1}{2}\right)}\left(1+\frac{a_1}{\nu}\right)^{-\left(\frac{\nu}{2}+1\right)}\left(1+\frac{a_2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)},$$

où

$$K_{\rho,\nu} = \left\{\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right\}^2\frac{\nu/2}{\sqrt{1-\rho^2}}.$$

Comme cas particulier de l'Équation (3.6), on a alors

$$m(u_1) = K_{\rho,\nu}\left(1+\frac{a_1}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}\int_0^1 u_2\left(1+\frac{a_1^2-2\rho\,a_1\,a_2+a_2^2}{\nu(1-\rho)^2}\right)^{-\left(\frac{\nu}{2}+1\right)}\left(1+\frac{a_2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}\mathrm{d}u_2.$$

4.4 Régression avec des copules pseudo-linéaires

4.4.1 Définition des modèles pseudo-linéaires

Soient les variables aléatoires $X_1, \ldots, X_d$ dont les fonctions de répartition sont respectivement $F_1, \ldots, F_d$, de même qu'une variable dépendante Y dont la fonction de répartition est F_0. On dit que le vecteur des variables explicatives $\mathbf{X} = (X_1, \ldots, X_d)$ est relié à Y par un modèle pseudo-linéaire s'il existe des fonctions $f_0, f_1, \ldots, f_d : \mathbb{R} \to \mathbb{R}$ strictement croissantes telles que pour $\widetilde{Y} = f_0(Y)$, $\widetilde{\mathbf{X}} = (f_1(X_1), \ldots, f_d(X_d))$ et $\mathrm{E}(\epsilon) = 0$,

$$\widetilde{Y} = \widetilde{\mathbf{X}}\,\boldsymbol{\beta} + \epsilon. \tag{4.2}$$

Si on ramène le tout dans l'échelle de mesure de Y, on a la courbe de régression

$$m^{\mathrm{PL}}(\mathbf{x}) = f_0^{-1}\left(\widetilde{\mathbf{x}}\,\boldsymbol{\beta}\right).$$

4.4.2 Régression pseudo-linéaire Normale

Supposons que la copule de $(Y, \mathbf{X})$ est Normale avec une matrice de corrélation

$$\Sigma_{Y,X} = \begin{pmatrix} 1 & \sigma_{Y\mathbf{X}} \\ \sigma_{Y\mathbf{X}}^{\top} & \Sigma_{\mathbf{XX}} \end{pmatrix},$$

où $\Sigma_{\mathbf{XX}}$ est la matrice de corrélation de $\mathbf{X}$ et $\sigma_{Y\mathbf{X}} = (\mathrm{corr}(Y, X_1), \ldots, \mathrm{corr}(Y, X_d))$. Puisque les marges de $(Y, \mathbf{X})$ sont $(F_0, F_1, \ldots, F_d)$, on sait que la distribution de

$$(V, \mathbf{U}) = (F_0(Y), F_1(X_1), \ldots, F_d(X_d))$$

est la copule Normale de matrice de corrélation $\Sigma_{Y,\mathbf{X}}$. Par conséquent, le vecteur

$$\left(\widetilde{Y}, \widetilde{\mathbf{X}}\right) = \left(\Phi^{-1}\left\{F_0(Y)\right\}, \Phi^{-1}\left\{F_1(X_1)\right\}, \ldots, \Phi^{-1}\left\{F_d(X_d)\right\}\right)$$

est distribué selon la loi Normale à $d+1$ dimensions dont la matrice des variances-covariances est $\Sigma_{Y,\mathbf{X}}$. Par un résultat classique sur les lois conditionnelles Normale, la loi de $\widetilde{Y}$ sachant $\widetilde{\mathbf{X}} = \widetilde{\mathbf{x}}$ est $\mathbb{N}(\mu_{\widetilde{Y}|\widetilde{\mathbf{x}}}, \sigma^2_{\widetilde{Y}|\widetilde{\mathbf{x}}})$, où

$$\mu_{\widetilde{Y}|\widetilde{\mathbf{x}}} = \widetilde{\mathbf{x}}\,\Sigma_{\mathbf{XX}}^{-1}\,\sigma_{Y\mathbf{X}}^{\top} \quad \text{et} \quad \sigma^2_{\widetilde{Y}|\widetilde{\mathbf{x}}} = 1 - \sigma_{Y\mathbf{X}}\,\Sigma_{\mathbf{XX}}^{-1}\sigma_{Y\mathbf{X}}^{\top}.$$

Pour $\varepsilon \sim \mathbb{N}(0,1)$, on a donc la représentation stochastique $\widetilde{Y} = \widetilde{\mathbf{X}}\,\Sigma_{\mathbf{XX}}^{-1}\,\sigma_{Y\mathbf{X}}^{\top} + \varepsilon$. Par conséquent, la copule Normale induit un modèle pseudo-linéaire de la forme (4.2), où $f_j(x) = \Phi^{-1}\{F_j(x)\}$ pour $j \in \{0, 1, \ldots, d\}$. En posant $\boldsymbol{\beta} = \Sigma_{\mathbf{XX}}^{-1}\,\sigma_{Y\mathbf{X}}^{\top}$, on a donc

$$m^{\mathrm{PL}}(\mathbf{x}) = F_0^{-1}\left\{\Phi\left(\widetilde{\mathbf{x}}\,\boldsymbol{\beta}\right)\right\},$$

où $\widetilde{\mathbf{x}} = (\Phi^{-1}\{F_1(x_1)\}, \ldots, \Phi^{-1}\{F_d(x_d)\}\ )$. Dans le cas d'une seule variable explicative, $\Sigma_{\mathbf{XX}} = 1$ et $\sigma_{Y\mathbf{X}} = \rho$ pour un certain $\rho \in (-1, 1)$. Donc, la courbe de régression est

$$m^{\mathrm{PL}}(x) = F_0^{-1}\left\{\Phi\left(\rho\,\Phi^{-1}\{F_1(x)\}\right)\right\}.$$

Dans le cas de marges de loi uniforme sur $[0,1]$, c'est-à-dire que $F_0(a) = F_1(a) = \cdots = F_d(a) = a$, alors pour $\mathbf{u} \in [0,1]^d$, on a $m^{\mathrm{PL}}(\mathbf{u}) = \Phi\left(\widetilde{\mathbf{x}}\,\boldsymbol{\beta}\right)$, où $\widetilde{\mathbf{x}} = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))$.

4.4.3 Régression pseudo-linéaire Elliptique

La régression pseudo-linéaire basée sur la copule Normale peut se généraliser à toute loi elliptique. Supposons donc que le vecteur $(Y, \mathbf{X})$ est distribué selon une distribution elliptique caractérisée par la fonction $g : \mathbb{R}^+ \to \mathbb{R}$ et une matrice de corrélation de la forme $\Sigma_{Y,\mathbf{X}}$. On peut supposer sans perte de généralité que les moyennes sont nulles

et les variances sont unitaires. On sait qu'alors, il existe $\boldsymbol{\beta} \in \mathbb{R}^d$ tel que

$$\mathrm{E}(Y|\mathbf{X} = \mathbf{x}) = \mathbf{x}\,\boldsymbol{\beta}.$$

On peut donc procéder de la même façon que pour la copule Normale en posant $f_j(x) = G^{-1}\{F_j(x)\}$ pour chaque $j \in \{0, 1, \ldots, d\}$, où G est la loi marginale de la loi elliptique générée par g. On peut donc imaginer des modèles de régression pseudo-linéaires basés sur les copules de Student, Laplace et Pearson de type II. On retrouve des détails sur la façon d'estimer les paramètres dans les régressions par copules pseudo-linéaires à l'Annexe B.

4.5 Conclusion du Chapitre 4

Ce chapitre a développé des formes explicites pour les courbes de régression de copules induites par plusieurs familles populaires, incluant les copules de Farlie–Gumbel–Morgenstern, les copules Archimédiennes (Clayton, Gumbel, Frank) et les copules elliptiques (Normal, Student). Une façon alternative de faire de la régression de copules dans la classe des modèles elliptiques a également été décrite.

CHAPITRE 5

COPULES QUI INDUISENT UNE RÉGRESSION NON-MONOTONE

5.1 Failles de la régression par copules

5.1.1 Mauvaise spécification de la copule

La connaissance de la vraie famille de copule conduit à un bon fonctionnement de l'estimateur choisi. Par ailleurs, il y a une indisponibilité de ces informations et la sélection de la forme de la copule se fait à l'aide des données. Il est donc possible que la mauvaise famille de copules soit sélectionnée et l'utilisation d'un modèle de copule mal spécifié conduira à un estimateur incohérent de $m(\mathbf{x})$.

Supposons que $\{c_\theta; \theta \in \Theta\}$ est une famille paramétrique de densités de copules. Dire que la famille de copules est bien spécifiée signifie qu'il existe $\theta_0 \in \Theta$ tel que c_{θ_0} coïncide avec la densité c de la vraie copule sous-jacente aux observations de (Y, X). Par contre, dans un modèle mal spécifié, un tel θ_0 peut ne pas exister. Cependant, même dans une telle situation, le pseudo-vrai paramètre $\theta^\star$ peut se définir la valeur

minimale dans Θ du critère d'information de Kullback–Leibler, c'est-à-dire

$$I(\theta) = \int_{[0,1]^{d+1}} \ln\left(\frac{c(u_0, \mathbf{u})}{c_\theta(u_0, \mathbf{u})}\right) \mathrm{d}C(u_0, \mathbf{u}).$$

Le comportement asymptotique de m sous un modèle de copule mal spécifié est décrit au Théorème A.1 de l'Annexe A.

5.1.2 Liens non-monotones

Il a été remarqué par Dette *et coll.* [7] que toutes les familles de copules paramétriques couramment utilisées produisent nécessairement une fonction de régression $m(\mathbf{x})$ qui est monotone selon le vecteur des variables explicatives $\mathbf{X}$. Dans le cas où le phénomène observé produit des relations non-monotones entre les différentes variables, les régressions basées sur ces copules s'avèrent totalement inadéquates.

5.2 Solution : usage de la copule Khi-deux

5.2.1 Description de la famille des copules Khi-deux

Soit un vecteur aléatoire $\mathbf{Z} = (Z_1, \ldots, Z_d)$ de loi Normale de moyennes nulles, de variances unitaires, et de matrice de corrélation $\Sigma \in \mathbb{R}^{d\times d}$. La copule Khi-deux de paramètre de non-centralité $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{R}^d$ est définie comme la structure de dépendance extraite de la loi de

$$\mathbf{X} = \left((Z_1 + a_1)^2, \ldots, (Z_d + a_d)^2\right).$$

Ce modèle de copules a d'abord été proposé par Bàrdossy [3], alors que ses propriétés ont été étudiées par Quessy *et coll.* [24]. Un de ces résultats est que l'on retrouve la copule normale lorsque $a_j \to \infty$ pour tout $j \in \{1, \ldots, d\}$. Par conséquent, la famille Khi-deux généralise la classe des copules Normales.

Dans le cas bivarié, Quessy *et coll.* [24] montre qu'une expression pour la copule Khi-deux est

$$C^{\chi}_{\rho,a_1,a_2}(u_1,u_2) = \sum_{\epsilon_1,\epsilon_2\in\{-1,1\}} \epsilon_1\,\epsilon_2\,\Phi_\rho\left\{h_{a_1}(\epsilon_1 u_1), h_{a_2}(\epsilon_2 u_2)\right\},$$

où Φ_ρ est la fonction de répartition de la loi Normale bivariée de corrélation ρ et pour $G_a(x) = \Phi(\sqrt{x} - a) + \Phi(\sqrt{x} + a) - 1$,

$$h_a(u) = \text{signe}(u)\sqrt{G_a^{-1}(|u|)} - a.$$

Une expression alternative de la copule Khi-deux en fonction de la copule Normale C^{N}_ρ a également été proposée par Quessy *et coll.* [24]. Spécifiquement, pour $\widetilde{h}_a(u) = \Phi\{h_a(u)\}$, on a

$$C^{\chi}_{\rho,a_1,a_2}(u_1,u_2) = \sum_{\epsilon_1,\epsilon_2\in\{-1,1\}} \epsilon_1\,\epsilon_2\,C^{\mathrm{N}}_\rho\left\{\widetilde{h}_{a_1}(\epsilon_1 u_1), \widetilde{h}_{a_2}(\epsilon_2 u_2)\right\}.$$

De là, on déduit que la densité de la copule Khi-deux peut être exprimée par

$$c^{\chi}_{\rho,a_1,a_2}(u_1,u_2) = \sum_{\epsilon_1,\epsilon_2\in\{-1,1\}} \widetilde{h}'_{a_1}(\epsilon_1 u_1)\,\widetilde{h}'_{a_2}(\epsilon_2 u_2)\,c^{\mathrm{N}}_\rho\left\{\widetilde{h}_{a_1}(\epsilon_1 u_1), \widetilde{h}_{a_2}(\epsilon_2 u_2)\right\},$$

où

$$\widetilde{h}'_a(u) = \frac{\phi\{h_a(u)\}}{\phi\{h_a(u)\} + \phi\{h_a(-u)\}}.$$

5.2.2 Deux cas particuliers d'intérêt

Dans la suite, on s'intéressera aux cas particuliers

$$c^{\chi}_{\rho,a} = \lim_{a_2 \to \infty} c^{\chi}_{\rho,a,a_2} \quad \text{et} \quad c^{\chi}_{\rho} = c^{\chi}_{\rho,0}.$$

On peut montrer que

$$c^{\chi}_{\rho,a}(u_1, u_2) = \widetilde{h}'_a(u_2)\, c^{\mathrm{N}}_{\rho}\left\{u_1, \widetilde{h}_a(u_2)\right\} + \widetilde{h}'_a(-u_2)\, c^{\mathrm{N}}_{\rho}\left\{u_1, \widetilde{h}_a(-u_2)\right\}.$$

De là, on déduit que

$$c^{\chi}_{\rho}(u_1, u_2) = \frac{1}{2}\left\{c^{\mathrm{N}}_{\rho}\left(u_1, \frac{1+u_2}{2}\right) + c^{\mathrm{N}}_{\rho}\left(u_1, \frac{1-u_2}{2}\right)\right\}.$$

5.2.3 Courbes de régression induites par la copule Khi-deux

La fonction de régression de la copule de Khi-deux bivariée peut être construite en utilisant l'Équation (3.6). On obtient alors, pour $c^{\chi}_{\rho,a}$, la courbe de régression non-monotone

$$m_{\rho,a}(u_1) = \int_0^1 u_2 \left[\widetilde{h}'_a(u_2)\, c^{\mathrm{N}}_{\rho}\left\{u_1, \widetilde{h}_a(u_2)\right\} + \widetilde{h}'_a(-u_2)\, c^{\mathrm{N}}_{\rho}\left\{u_1, \widetilde{h}_a(-u_2)\right\}\right] \mathrm{d}u_2.$$

En particulier, avec $m_{\rho} = m_{\rho,0}$, on a

$$m_{\rho}(u_1) = \frac{1}{2}\int_0^1 u_2 \left\{c^{\mathrm{N}}_{\rho}\left(u_1, \frac{1+u_2}{2}\right) + c^{\mathrm{N}}_{\rho}\left(u_1, \frac{1-u_2}{2}\right)\right\} \mathrm{d}u_2.$$

5.2.4 Adaptation à la copule de Fisher

La copule de Fisher est une extension à la Khi-deux proposée et étudiée par Favre *et coll.* [9]. Il s'agit de considérer que $\mathbf{X} = (X_1, \ldots, X_d)$ est de loi de Student à ν degré de liberté. A l'instar de la copule de Khi-deux, la copule de Fisher est la dépendance extraite de $\mathbf{Y} = (X_1^2, \ldots, X_d^2)$. Cette copule s'exprime en fonction de $C^T_{\Sigma,\nu}$, la copule de Student à ν degrés de liberté et de matrice de corrélation Σ

$$C^F_{\Sigma,\nu}(\mathbf{u}) = \sum_{\epsilon\in\{-1,+1\}^d} \left(\prod_{\ell=1}^{d} \epsilon_\ell\right) C^T_{\Sigma,\nu}\left(\frac{1+\epsilon_1 u_1}{2}, \ldots, \frac{1+\epsilon_d u_d}{2}\right)$$

Pour $d = 2$, on retrouve la copule de Fisher bivariée $C^F_{\rho,\nu}$ associée à la copule de Student bivariée de corrélation ρ à ν degré de liberté, à savoir

$$\begin{aligned} C^F_{\rho,\nu}(u_1, u_0) &= C^T_{\rho,\nu}\left(\frac{1+u_1}{2}, \frac{1+u_0}{2}\right) - C^T_{\rho,\nu}\left(\frac{1+u_1}{2}, \frac{1-u_0}{2}\right) \\ &\quad - C^T_{\rho,\nu}\left(\frac{1-u_1}{2}, \frac{1+u_0}{2}\right) + C^T_{\rho,\nu}\left(\frac{1-u_1}{2}, \frac{1-u_0}{2}\right) \end{aligned}$$

La densité de la copule de Fisher bivariée peut être obtenue à partir de la densité de la copule de Student bivariée en différenciant $C^F_{\rho,\nu}$, ce qui amène

$$\begin{aligned} c^F_{\rho,\nu}(u_1, u_0) &= \frac{1}{4}\Bigg\{c^T_{\rho,\nu}\left(\frac{1+u_1}{2}, \frac{1+u_0}{2}\right) + c^T_{\rho,\nu}\left(\frac{1+u_1}{2}, \frac{1-u_0}{2}\right) \\ &\quad + c^T_{\rho,\nu}\left(\frac{1-u_1}{2}, \frac{1+u_0}{2}\right) + c^T_{\rho,\nu}\left(\frac{1-u_1}{2}, \frac{1-u_0}{2}\right)\Bigg\} \end{aligned}$$

La fonction de régression de la copule de Fisher bivariée peut être construite en utilisant l'Équation (3.6), ce qui amène

$$\begin{aligned} m^F_{\rho,\nu}(u_1) &= \frac{1}{4}\int_0^1 u_0 \Bigg\{c^T_{\rho,\nu}\left(\frac{1+u_1}{2}, \frac{1+u_0}{2}\right) + c^T_{\rho,\nu}\left(\frac{1+u_1}{2}, \frac{1-u_0}{2}\right) \\ &\quad + c^T_{\rho,\nu}\left(\frac{1-u_1}{2}, \frac{1+u_0}{2}\right) + c^T_{\rho,\nu}\left(\frac{1-u_1}{2}, \frac{1-u_0}{2}\right)\Bigg\} \mathrm{d}u_0. \end{aligned}$$

5.2.5 Généralisation aux copules *Squared*

On considère un vecteur $\mathbf{X} = (X_1, \ldots, X_d)$ dont les marges $F_1, \ldots, F_d$ sont symétriques en 0. La copule *squared* associée à $\mathbf{X}$ est donc la structure de dépendance de $\mathbf{Y} = (X_1^2, \ldots, X_d^2)$. Un attribut intéressant de la copule *squared* $\widetilde{C}$ est qu'elle dépend uniquement de la copule C de $\mathbf{X}$. En effet , il a été montré par Quessy & Durocher [23] que la copule *squared* peut être donnée par

$$\widetilde{C}(\mathbf{u}) = \sum_{\epsilon \in \{-1,+1\}^d} \left(\prod_{\ell=1}^{d} \epsilon_\ell \right) C\left(\frac{1+\epsilon_1 u_1}{2}, \ldots, \frac{1+\epsilon_d u_d}{2} \right)$$

Cette expression est indépendante des marges initiales $F_1, \ldots, F_d$. La forme de la copule *squared* bivariée devient

$$\begin{aligned} \widetilde{C}(u_1, u_0) &= C\left(\frac{1+u_1}{2}, \frac{1+u_0}{2} \right) - C\left(\frac{1+u_1}{2}, \frac{1-u_0}{2} \right) \\ &\quad - C\left(\frac{1-u_1}{2}, \frac{1+u_0}{2} \right) + C\left(\frac{1-u_1}{2}, \frac{1-u_0}{2} \right). \end{aligned}$$

Après différenciation, on obtient la densité de la copule *squared* bivariée

$$\begin{aligned} \widetilde{c}(u_1, u_0) &= \frac{1}{4} \Bigg\{ c\left(\frac{1+u_1}{2}, \frac{1+u_0}{2} \right) + c\left(\frac{1+u_1}{2}, \frac{1-u_0}{2} \right) \\ &\quad + c\left(\frac{1-u_1}{2}, \frac{1+u_0}{2} \right) + c\left(\frac{1-u_1}{2}, \frac{1-u_0}{2} \right) \Bigg\} \end{aligned}$$

Comme cas particulier de l'Équation (3.6), on obtient

$$\begin{aligned} m(u_1) &= \frac{1}{4} \int_0^1 u_0 \Bigg\{ c\left(\frac{1+u_1}{2}, \frac{1+u_0}{2} \right) + c\left(\frac{1+u_1}{2}, \frac{1-u_0}{2} \right) \\ &\quad + c\left(\frac{1-u_1}{2}, \frac{1+u_0}{2} \right) + c\left(\frac{1-u_1}{2}, \frac{1-u_0}{2} \right) \Bigg\} \mathrm{d}u_0. \end{aligned}$$

5.3 Illustrations sur des données simulées

Dette *et coll.* [7] ont étudié les types de dépendance de régression qui peuvent être obtenus à partir de familles de copules utilisées. Il stipulent que les caractéristiques non monotones de la fonction de régression ne peuvent pas être reproduites par l'estimation de régression basée sur la copule. Nous allons montrer que ces données peuvent être correctement modélisées, mais en autant qu'une copule qui permet la non-monotonie soit employée.

5.3.1 Modélisation de liens monotones

La première illustration concerne des paires de points simulées selon un modèle qui induit de la dépendance monotone. Les $n = 200$ paires de points sur les graphiques de gauche de la Figure 5.1 ont été générées à partir d'une distribution bivariée dont la copule est Clayton avec $\tau_C = 0,7$ et où les marges sont respectivement $\mathbb{N}(250, 2)$ et $\mathbb{N}(100, 5)$. Les graphiques à gauche proviennent d'une loi dont la copule est Gumbel et les marges sont Exponentielles de moyennes respectives $\lambda_X = 1$ et $\lambda_Y = 10$.

Sans surprise, on constate que pour les données provenant d'une copule de Clayton, la relation entre Y et X est mieux modélisée lorsque le modèle sous-jacent est supposé Clayton (voir le graphique en haut à gauche) ; les courbes basées sur les copules Normale et Student reproduisent correctement cette relation au centre du nuage de points, mais divergent énormément des paires de points pour des valeurs plus extrêmes.

L'expérience présentée dans les graphiques à droite de la Figure 5.1 représente un cas où les copules utilisées sont *mal spécifiées*, c'est-à-dire qu'elles ne correspondent pas au processus qui a généré les données. Néanmoins, cette mauvaise spécification n'empêche pas les courbes basées sur les copules Normale et Student de très bien

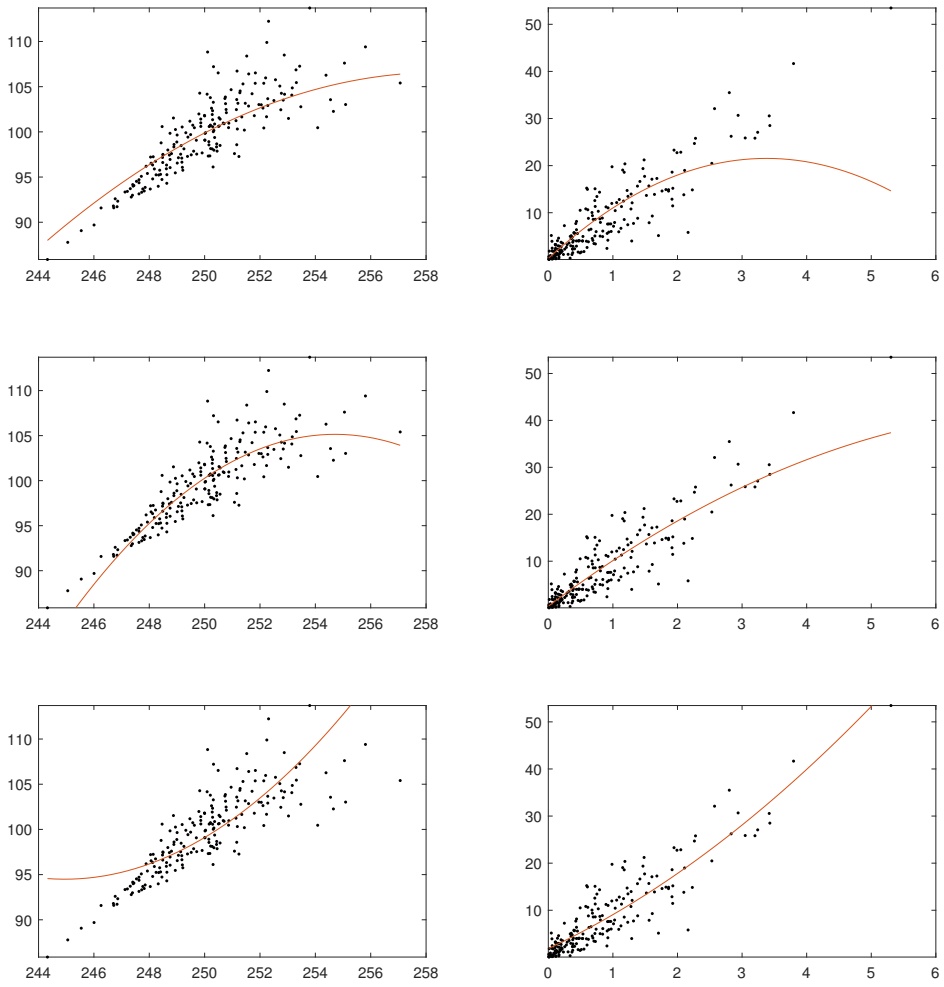

Figure 5.1 – De haut en bas : courbes de régression estimées basées sur les copules Clayton, Normale et Student. À gauche : $n = 200$ paires générées par la copule de Clayton avec des marges Normales ; à droite : $n = 200$ paires générées par la copule de Gumbel avec des marges Exponentielles.

s'ajuster aux données. Ici, la mauvaise performance de la courbe de Clayton peut s'expliquer par le fait que cette copule a des propriétés très différentes de la copule de Gumbel, notamment au niveau des queues inférieure et supérieure.

5.3.2 Modélisation de liens non-monotones

Tel que documenté par Dette *et coll.* [7], la plupart des copules utilisées pour construire des courbes de régression sont tout-à-fait inadéquates pour modéliser des liaisons non-monotones. Par exemple, si on suppose que X est uniformément distribuée sur l'intervalle $[0, 1]$ et que $\epsilon \sim \mathbb{N}(0, \sigma^2)$, alors la relation entre X et

$$Y = \left(X - \frac{1}{2}\right)^2 + \sigma\,\epsilon \tag{5.1}$$

est de nature quadratique, donc non-monotone. Pour illustrer $n = 200$ paires de points $(X_1, Y_1), \ldots, (X_{200}, Y_{200})$ ont été simulées à partir du modèle de l'Équation (5.1) lorsque $\sigma = 1/10$; le nuage de points se retrouve à la Figure 5.2.

Tel qu'anticipé, les copules Clayton, Normale et Student sont incapables de bien saisir ce lien non-monotone (graphiques de gauche de la Figure 5.2). Cependant, les courbes basées sur le copule Khi-deux font un excellent travail (graphiques de droite de la Figure 5.2). À l'oeil, la courbe qui s'ajuste le mieux aux données semble celle qui est basée sur la Khi-deux lorsque $a = 0$.

5.3.3 Distorsion des marges du modèle de Dette *et coll.* [7]

Comme dernière expérience, on considérera à nouveau un nuage de $n = 200$ points générés du modèle (5.1), cette fois avec $\sigma = 1/5$, mais pour lequel on appliquera une transformation sur les marges. L'objectif est d'avoir une idée concernant la robustesse des courbes de régression Khi-deux selon les échelles de mesure. Spécifiquement, pour (X, Y) générée à partir du modèle de Dette *et coll.* [7] décrit à l'Équation (5.1), on considérera le couple aléatoire (Z, W), où $Z = \log(X)$ et $W = -\exp(-Y)$. Le nuage de points avant cette transformation se retrouve dans les graphiques de gauche de la

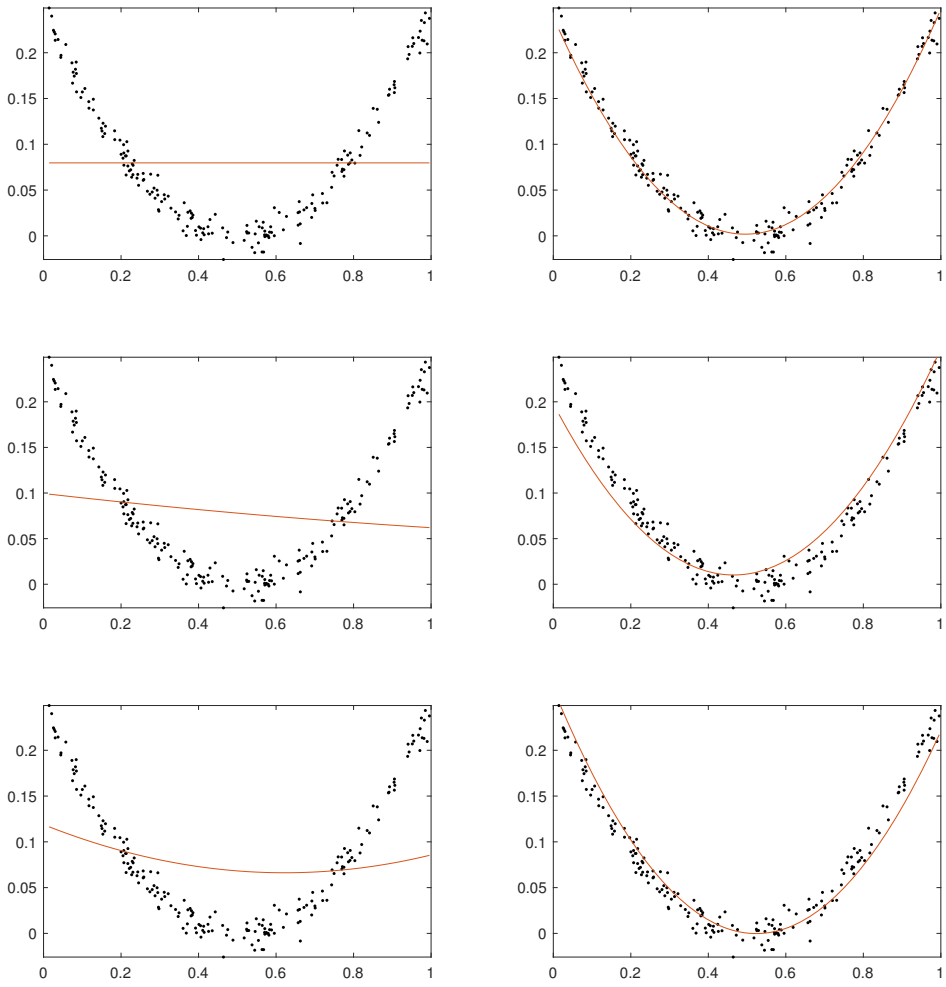

Figure 5.2 – $n = 200$ paires de points simulées à partir du modèle de l'Équation (5.1). À gauche, de haut en bas : courbes de régression estimées basées sur les copules Clayton, Normale et Student ; à droite, de haut en bas : courbes de régression estimées basées sur le copule Khi-deux avec $a = 0$, $a = 0,15$ et $a = -0,15$.

Figure 5.3, alors que le nuage de points qui résulte de cette transformation se retrouve aux graphiques de droite.

On constate que le lien entre X et Y pour les données à gauche, qui proviennent de la version originale du modèle (5.1), est bien modélisé par les trois courbes de régression Khi-deux ($a = 0$, $a = 0,15$ et $a = -0,15$). Par contre, ce n'est plus vraiment le cas une

fois que les échelles des données ont été transformées. Dans ce cas, la régression Khideux a plus de difficulté à capturer correctement le lien entre les deux variables. Des simulations supplémentaires seraient nécessaires pour bien comprendre ce phénomène.

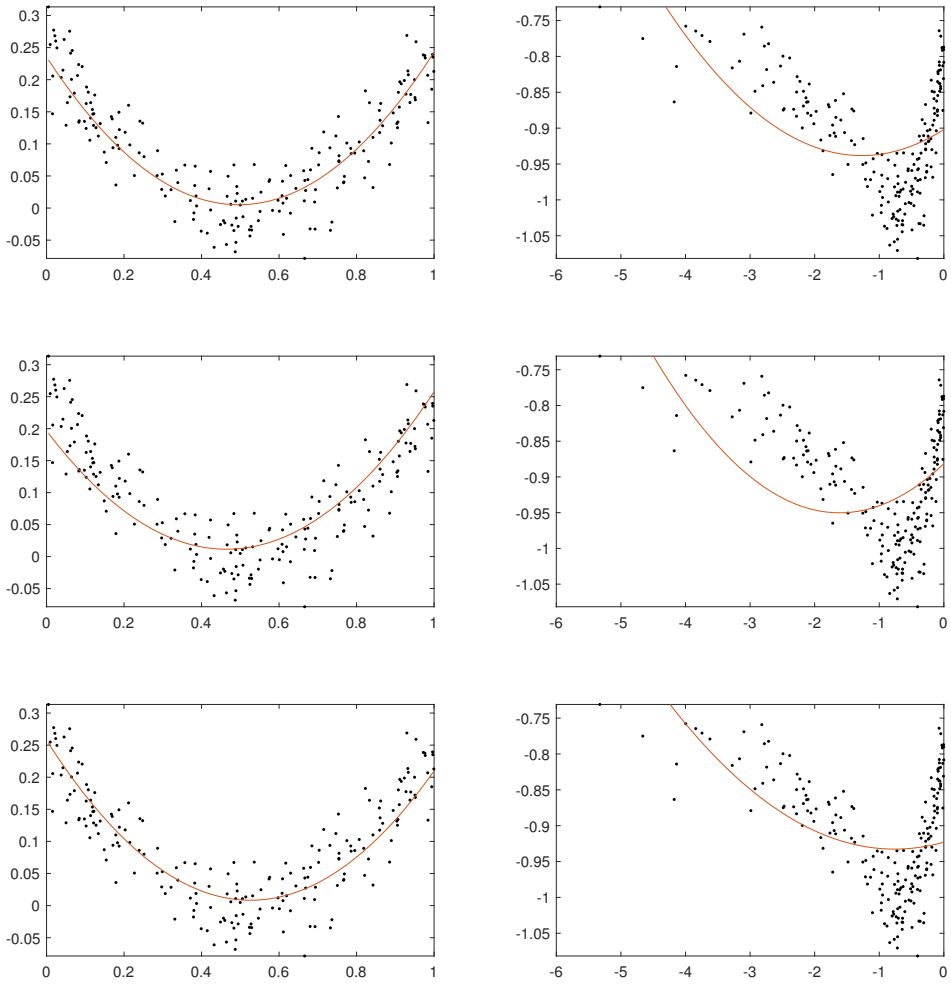

Figure 5.3 – De haut en bas : courbes de régression estimées basées sur le copule Khideux avec $a = 0$, $a = 0,15$ et $a = -0,15$. À gauche : $n = 200$ paires de points simulées à partir du modèle (5.1) ; à droite : paires transformées $(Z, W) = (\log(X), -e^{-Y})$.

5.4 Conclusion du Chapitre 5

Ce chapitre a d'abord exposé la faille principale de l'approche de régression par copules, à savoir que pour la plupart des modèles, la courbe de régression induite est nécessairement monotone. Cette limitation a été élégamment résolue en définissant une nouvelle famille de courbes de régression basée sur la famille des copules Khi-deux. Des illustrations avec des données simulées présentant une structure non monotone convainquent de la viabilité et de la pertinence de l'approche proposée. Elle répond en outre à une critique formulée par Dette *et coll.* [8] au sujet de l'approche introduite par Noh *et coll.* [22], ces premiers auteurs ayant justement noté que pour les copules usuelles, la régression est nécessairement monotone.

CHAPITRE 6

CONCLUSION ET PERSPECTIVES

L'approche par les copules dans le domaine de la régression s'impose de plus en plus au cours de ces dernières décennies. Au tout début de ce papier, nous avions parlé de la régression dans son sens général notamment la régression linéaire simple et multiple, la régression non linéaire ainsi que la régression non paramétrique. Des rappels ont été faits dans ce sens en montrant les limites de ces méthodes qui pour les contrer, d'autres ont vu le jour. La régression basée sur les copules se profile donc comme une bonne alternative à la régression linéaire pour remédier à certains de ses limites. Elle permet d'avoir une description plus précise de la distribution d'une variable conditionnelle à ses déterminants, contrairement à la régression linéaire. Dans ce mémoire, l'approche de Noh *et coll.* [22] a été adoptée pour modéliser la distribution conjointe de la réponse Y et de ses covariables $\mathbf{X}$ à partir d'une famille de copules paramétriques et les marginales à partir des méthodes non paramétriques. Cette méthode flexible, facile à mettre en œuvre, consiste donc à écrire la fonction de régression en termes de la copule et des distributions marginales.

Après avoir construit le modèle de régression de la copule nous avons pu déterminer les fonctions de régression de différents copules notamment les copules archimédiennes (Clayton, Frank, Gumbel), la copule FGM, les copules normale, Student, les copules elliptiques de façon générale. Puis nous avons établi un lien avec les modèles pseudo-linéaires. Nous avons proposé une estimation non paramétrique des courbes de

régressions basée sur des estimations non paramétriques des marginales pour obtenir un estimateur convergent.

Comme pour toutes méthodes de régression, il existe des failles sur la régression par copules. D'une part si la véritable structure de la copule a été mal spécifiée, l'approche de Noh *et coll.* [22] ne produit souvent pas d'estimations fiables de la fonction de régression. D'autres part si la fonction de régression n'est pas monotone, les estimations de régression basées sur la copule ne reproduisent pas les caractéristiques non monotones de la fonction de régression. Cette dernière a bien pu être contourner avec l'usage de la copule de Khi-deux. Des simulations ont été faites en ce sens en utilisant le jeu de données de Dette *et coll.* [7] pour montrer qu'en faisant l'ajustement avec la Khi-deux, on parvient à obtenir une estimation fiable de la fonction de régression.

Comme perspectives, nous envisageons donc d'étudier les propriétés théoriques des nouveaux estimateurs proposés, de comparer par simulations l'efficacité des régressions par copules selon différentes copules Squared (Normale, Student, Laplace, Pearson de type II) et aussi d'améliorer l'ajustement des courbes de régression en estimant le paramètre de telle sorte que le R-carré de la courbe versus les données soit maximisé.

Bibliographie

[1] I. Ahamada and E. Flachaire. *Économétrie Non-Paramétrique*. Economica, Sep 2008.

[2] M. Avriel. *Nonlinear Programming : Analysis and Methods*. Dover Books on Computer Science Series. Dover Publications, 2003.

[3] A. Bàrdossy. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11) :1–12, 2006.

[4] A. Cameron, P. Trivedi, P. Trivedi, P. Trivedi, E. Library, C. U. Press, and E. Corporation. *Microeconometrics : Methods and Applications*. Cambridge University Press, 2005.

[5] G. Casella and R. L. Berger. *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.

[6] 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.

[7] H. Dette, R. V. Hecke, and S. Volgushev. Some comments on copula-based regression. *Journal of the American Statistical Association*, 109(507) :1319–1324, 2014. doi : 10.1080/01621459.2014.916577.

[8] H. Dette, R. Van Hecke, and S. Volgushev. Some comments on copula-based regression. *J. Amer. Statist. Assoc.*, 109(507) :1319–1324, 2014. ISSN 0162-1459.

[9] A.-C. Favre, J.-F. Quessy, and M.-H. Toupin. The new family of Fischer copulas to model upper tail dependence and radial asymmetry : properties and application to high-dimensional rainfall data. *Environmetrics*, 29(3) :e2494, 17, 2018.

[10] M. J. Frank. On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Math.*, 19(2-3) :194–226, 1979.

[11] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Analytical Methods for Social Research. Cambridge University Press, 2006.

[12] C. Genest. Frank's family of bivariate distributions. *Biometrika*, 74(3) :549–555, 1987. ISSN 0006-3444.

[13] C. Genest and J. MacKay. The joy of copulas : bivariate distributions with uniform marginals. *Amer. Statist.*, 40(4) :280–283, 1986.

[14] C. Genest and L.-P. Rivest. A characterization of Gumbel's family of extreme value distributions. *Statist. Probab. Lett.*, 8(3) :207–211, 1989.

[15] 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.

[16] E. J. Gumbel. Bivariate logistic distributions. *J. Amer. Statist. Assoc*, 56 : 335–349, 1961.

[17] G. Kim, M. J. Silvapulle, and P. Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.*, 51(6) : 2836–2850, 2007.

[18] I. Kojadinovic and J. Yan. A non-parametric test of exchangeability for extreme-value and left-tail decreasing bivariate copulas. *Scandinavian Journal of Statistics*, 2012.

[19] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2) :164–168, 1944.

[20] C. F. Manski. Regression. *Journal of Economic Literature*, 29(1) :34–50, 1991.

[21] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2) : 431–441, 1963.

[22] H. Noh, A. El Ghouch, and T. Bouezmarni. Copula-based regression estimation and inference. *J. Amer. Statist. Assoc.*, 108(502) :676–688, 2013.

[23] J.-F. Quessy and M. Durocher. The class of copulas arising from squared distributions : Properties and inference. *Econom. Stat.*, 12 :148–166, 2019.

[24] J.-F. Quessy, L.-P. Rivest, and M.-H. Toupin. On the family of multivariate chi-square copulas. *J. Multivariate Anal.*, 152 :40–60, 2016.

[25] G. Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, 2011.

[26] J. Shao. *Mathematical statistics*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2003.

[27] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231, 1959.

[28] R. Tibshirani. Regression shrinkage and selection via the lasso : a retrospective. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73 (3) :273–282, 2011.

[29] H. Tsukahara. Semiparametric estimation in copula models. *Canad. J. Statist.*, 33 :357–375, 2005.

[30] L. Wasserman and L. Wasserman. *All of Statistics : A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.

Annexe A

Propriétés de l'estimateur dans les modèles pseudo-linéaires

A.1 Résultats de convergence de l'estimateur

Hypothèse A. *Pour tout* $x \in \mathbb{R}^d$,

$$\widetilde{F}_{nj}(x_j) = \frac{1}{n}\sum_{i=1}^{n} I(X_{ij} \leq x_j) + o_p\left(n^{-1/2}\right), \qquad j = 1, \ldots, d.$$

On note la famille de copules à laquelle appartient $c : \mathcal{C} = \{c(.;\theta), \theta \in \Theta\}$, où Θ est un sous-ensemble compact de $\mathbb{R}^p$. On définit θ_0 comme le paramètre de copule vrai (mais inconnu), qui se trouve à l'intérieur de Θ de sorte que $c(.)$ coïncide avec $c(.;\theta_0)$. Cette approche permet d'estimer θ_n par θ_0 , ce qui satisfait l'hypothèse suivante.

Hypothèse B. *Pour* $\eta_i = \eta\left(U_{0,i}, U_i; \theta_0\right)$*, un vecteur aléatoire p-dimensionnel tel que*

$\mathbb{E}_\eta = 0$ *et* $\mathbb{E}_{\eta^\top \eta} < \infty$ *et* $U_i = (U_{1,i}, \ldots, U_{d,i})^\top$,

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{i=1}^{n} \eta_i + o_p\left(n^{-1/2}\right). \tag{A.1}$$

Un estimateur prometteur de θ est l'estimateur de pseudo-vraisemblance maximum (semiparamétrique) (PL) $\widehat{\theta}_n^{PL}$, qui est défini comme le maximiseur de

$$L(\theta) = \sum_{i=1}^{n} \log c\left(\widehat{F}_{n0}\left(Y_i\right), \widehat{F}_n\left(X_i\right); \theta\right).$$

où $\widehat{F}_n\left(X_i\right) = \left(\widehat{F}_{n1}\left(X_{i1}\right), \ldots, \widehat{F}_{nd}\left(X_{id}\right)\right)^\top$. $\widehat{\theta}_n^{PL}$ a été étudié par Genest *et coll.* [15], Kim *et coll.* [17], Tsukahara [29] et Kojadinovic & Yan [18]. Pour l'estimateur PL, la fonction η est donnée par $\eta\left(U_0, \mathbf{U}; \theta\right) = J^{-1}(\theta) K\left(U_0, \mathbf{U}; \theta\right)$, où

$$J(\theta) = \int_{[0,1]^{d+1}} \left(\frac{\partial^2}{\partial\theta\partial\theta^\top} \log c\left(u_0, \mathbf{u}; \theta_0\right)\right) \mathrm{d}C\left(u_0, \mathbf{u}; \theta_0\right),$$

alors que $K\left(U_0, \mathbf{U}; \theta\right)$ est un vecteur p-dimensionnel dont le k-ième élément est

$$\begin{aligned} &\frac{\partial}{\partial\theta^k} \log c\left(U_0, \mathbf{U}; \theta\right) \\ &+ \sum_{j=1}^{d} \int_{[0,1]^{d+1}} \left(I\left(U_j \leq \upsilon_j\right) - \upsilon_j\right) \times \left(\frac{\partial^2}{\partial\theta_k \partial\upsilon_j} \log c\left(\upsilon_0, \upsilon; \theta\right)\right) \mathrm{d}C\left(\upsilon_0, \upsilon; \theta\right). \end{aligned}$$

Considérons les notations suivantes pour la suite :

- $\partial_j = \frac{c}{u_j}$ pour $j = 1, \ldots, d$ et $\partial_j c_X = \frac{c_X}{u_j}$ et $\partial_j e = \frac{e}{u_j}$ pour $j = 1, \ldots, d$, où $e\left(u; \theta\right) = \mathbb{E}\left(Y c\left(F_0(Y), u; \theta\right)\right)$.
- $\dot{c} = \left(\frac{\partial c}{\partial\theta_1}, \ldots, \frac{\partial c}{\partial\theta_p}\right)^\top$, $\dot{c}_X = \left(\frac{\partial c_X}{\partial\theta_1}, \ldots, \frac{\partial c_X}{\partial\theta_p}\right)^\top$ et $\dot{e} = \left(\frac{\partial e}{\partial\theta_1}, \ldots, \frac{\partial e}{\partial\theta_p}\right)^\top$

Hypothèse C. *Soit* g *et* $\dot{c}$ *tel que* $\partial_j c, j = 1, \ldots, d$ *et* $x \in \mathbb{R}^d$ *soit un point d'intérêt donné qui satisfait* $\mathbf{F}(\mathbf{x}) \in (0,1)^d$.

(C1) $(u, 0) \mapsto g_{u_0}\left(u, \theta\right) \equiv g\left(u_0, \mathbf{u}; \theta\right)$ *est continu à* $\mathbf{F}(\mathbf{x}), \theta_0$ *uniformément dans* $u_0 \in [0,1]$.

(C2) $u_0 \mapsto g\left(u_0, \mathbf{F}(\mathbf{x}); \theta_0\right)$ *est continu dans* $[0,1]$.

Il sera question de montrer la représentation asymptotique indépendant et identiquement distribué de l'estimateur proposé ce qui implique que l'estimateur suit une distribution normale asymptotiquement. On commence par considérer le cas univarié ; soit en présence de la covariable X_1. Dans ce cas,

$$m(x_1) = e\left(F_1(x_1); \theta_0\right) = \mathbb{E}\left[Y c\left(F_0(Y), F_1(x_1); \theta_0\right)\right]$$

peut être estimée par

$$\widehat{m}(x_1) = \widehat{e}\left(\widetilde{F}_{1,n}(x_1); \widehat{\theta}_n\right) \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i c\left(\widehat{F}_{0,n}(Y_i), \widetilde{F}_{1,n}(X_i); \widehat{\theta}_n\right).$$

Le théorème suivant donne une représentation asymptotique de cet estimateur. Sa preuve est donnée en annexe.

Théorème A.1. *Supposons que* $\widetilde{F}_{1,n}(.), \widehat{\theta}_n$ *et* $c(.)$ *satisfont respectivement aux hypothèses A, B et C. Si* $\mathrm{E}(Y^2) < \infty$*, alors*

$$\begin{aligned}\sqrt{n}\left(\widehat{m}(x_1) - m(x_1)\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[\left\{I\left(X_{1,i} \leq x_1\right) - F_1(x_1)\right\}\partial_1 e\left(F_1(x_1); \theta_0\right) \\ &\quad - \int \left\{I\left(Y_i \leq y\right) - F_0(y)\right\} c\left(F_0(y), F_1(x_1); \theta_0\right) \mathrm{d}y \\ &\quad + \eta_i^{\top} \dot{e}\left(F_1(x_1); \theta_0\right)\Bigg] + o_p(1). \qquad \text{(A.2)}\end{aligned}$$

Le Théorème A.1 implique que $\sqrt{n}(\widehat{m} - m)$ est asymptotiquement Normal de moyenne zéro et de variance $\sigma^2 = \mathrm{var}\left(E_i(\theta_0)\right)$ avec $E_i(\theta_0)$ le terme entre crochets du Théorème A.1. Si on suppose que toutes les marges sont uniformes sur $[0, 1]$, c'est-à-dire

que $F_0(a) = F_1(a) = \cdots = F_d(a) = a$, alors pour $\mathbf{u} = (u_1, \ldots, u_d) \in [0,1]^d$,

$$\begin{aligned}\widehat{m}(u_1) - m(u_1) &= \frac{1}{n}\sum_{i=1}^{n}\Bigg[\left\{I\left(X_{1,i} \leq u_1\right) - u_1\right\}\partial_1 e\left(u_1;\theta_0\right) \\ &\quad - \int \left\{I\left(Y_i \leq u_0\right) - u_0\right\} c\left(u_0, u_1;\theta_0\right)\mathrm{d}y \\ &\quad + \eta_i^{\top}\dot{e}\left(u_1;\theta_0\right)\Bigg] + o_p(1).\end{aligned}$$

A.2 Prolongement au cas multivarié

Dans le cas général $d \geq 2$, la fonction de régression est donnée par

$$m(\mathbf{x}) = \int_{\mathbb{R}} y\,\frac{c\left\{F_0(y), \mathbf{F}(\mathbf{x})\right\} f_0(y)}{c_{\mathbf{X}}\left\{\mathbf{F}(\mathbf{x})\right\}}\mathrm{d}y = \frac{e(F(\mathbf{x});\theta_0)}{c_X(F(\mathbf{x});\theta_0)}. \tag{A.3}$$

D'une part, $e(F(\mathbf{x});\theta_0)$ est estimée par

$$\widehat{e}\left(\widetilde{F}_n(\mathbf{x});\theta_n\right) \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i c\left(\widehat{F}_{0,n}(Y_i), \widetilde{F}_n(\mathbf{x});\widehat{\theta}_n\right),$$

où $\widetilde{F}_n(\mathbf{x}) = (\widetilde{F}_{1,n}(x_1), \ldots, \widetilde{F}_{d,n}(x_d))$. D'après le Théorème A.1,

$$\widehat{e}\left(\widetilde{F}_n(\mathbf{x});\theta_n\right) - e\left(F(\mathbf{x});\theta_0\right) = \frac{1}{n}\sum_{i=1}^{n} E_i(x;\theta_0) + O_p(n^{-1/2}), \tag{A.4}$$

où

$$\begin{aligned}E_i\left(x;\theta_0\right) &= \sum_{j=1}^{d}\left\{I\left(X_{j,i} \leq x_j\right) - F_j(x_j)\right\}\partial_j e\left(F(\mathbf{x});\theta_0\right) \\ &\quad - \int \left\{I\left(Y_i \leq y\right) - F_0(y)\right\} c\left(F_0(y), F(\mathbf{x});\theta_0\right)\mathrm{d}y \\ &\quad + \eta_i^{\top}\dot{e}\left(F(\mathbf{x});\theta_0\right).\end{aligned}$$

D'autre part, à partir de l'Équation (3.6), il existe deux estimateurs possibles pour $c_X(F(\mathbf{x});\theta_0)$, à savoir

$$\begin{aligned}\widetilde{c}_X\left(\widetilde{F}_n(\mathbf{x});\widehat{\theta}_n\right) &= \int_0^1 c\left(u_0,\widetilde{F}_n(\mathbf{x});\widehat{\theta}_n\right)\mathrm{d}u_0,\\ \widehat{c}_X\left(\widetilde{F}_n(\mathbf{x});\widehat{\theta}_n\right) &= \frac{1}{n}\sum_{i=1}^{n} c\left(\widehat{F}_{0,n},\widetilde{F}_n(\mathbf{x});\widehat{\theta}_n\right).\end{aligned}$$

Ceci conduit à deux estimateurs différents pour $m(\mathbf{x})$, c'est-à-dire $\widetilde{m}(\mathbf{x})$ et $\widehat{m}(\mathbf{x})$. Cependant, la différence entre $\widetilde{m}(\mathbf{x})$ et $\widehat{m}(\mathbf{x})$ est asymptotiquement négligeable. A partir du Théorème A.1, la représentation asymptotique de $\widehat{c}_X(\widetilde{F}_n(\mathbf{x});\widehat{\theta}_n)$ donne

$$\widehat{c}_X\left(\widetilde{F}_n(\mathbf{x});\widehat{\theta}_n\right) - c_{\mathbf{X}}\left(F(\mathbf{x});\theta_0\right) = n^{-1}\sum_{i=1}^{n} D_i(c;\theta_0) + o_p(n^{-1/2}), \tag{A.5}$$

où

$$D_i\left(c;\theta_0\right) = \sum_{d=1}^{d}\left\{I\left(X_{j,i}\leq x_j\right) - F_j(x_j)\right\}\partial_j c_{\mathbf{X}}\left(F(\mathbf{x});\theta_0\right) + \eta_i^{\top}\dot{c}_X\left(F(\mathbf{x});\theta_0\right).$$

La combinaison des résultats (A.4) et (A.5) conduit au résultat suivant.

Théorème A.2. *Supposons que chaque composante de $\widetilde{F}_n$ satisfait l'hypothèse A,$\widehat{\theta}_n$ et $c(.)$ satisfassent les hypothèses B et C, respectivement. Si $\mathbb{E}[Y^2] < \infty$, alors*

$$\sqrt{n}\left(\widehat{m}(\mathbf{x}) - m(\mathbf{x})\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{c_{\mathbf{X}}\left(F(\mathbf{x});\theta_0\right)} \times \left[E_i(x;\theta_0) - m(\mathbf{x})D_1(x;\theta_0)\right] + o_p(1).$$

Annexe B

Estimation dans les modèles pseudo-linéaires

B.1 Le cas d'une copule Normale

B.1.1 Estimateur de la matrice des corrélations

L'approche consiste à utiliser le tau de Kendall, qui peut être bien estimé à partir de $\mathbf{Z} = (Y, \mathbf{X})$. Soit $n = d + 1$, si $\widetilde{\mathbf{Z}} \sim \mathcal{N}_n(0, \Sigma)$ avec $\Sigma = (\sigma_{k\ell})_{1 \leq k, \ell \leq n}$, alors

$$\sigma_{k\ell} = \sin\left(\frac{\pi}{2}\tau_{k\ell}\right), \tag{B.1}$$

où $\tau_{k\ell}$ est appelé tau de Kendall et est défini par

$$\tau_{k\ell} = \mathbb{E}\left[\mathrm{sgn}(\widetilde{z}_{1k} - \widetilde{z}_{2k})\mathrm{sgn}(\widetilde{z}_{1\ell} - \widetilde{z}_{2\ell})\right]. \tag{B.2}$$

τ_{jk} donné en (B.2) est invariant sous des transformations marginales strictement croissantes. Cela conduit à une estimation de τ_{ij} basée sur les données observées $\mathbf{Z}_1, \ldots, \mathbf{Z}_d$ sous la copule gaussienne modèle de régression.

$$\begin{aligned} \widehat{\tau}_{k\ell} &= \frac{2}{n(n-1)} \sum_{1 \leqslant i_1 \leqslant i_2 \leqslant d} \operatorname{sgn}(\widetilde{Z}_{i_1 k} - \widetilde{Z}_{i_2 k}) \operatorname{sgn}(\widetilde{Z}_{i_1 \ell} - \widetilde{Z}_{i_2 \ell}) \\ &= \frac{2}{n(n-1)} \sum_{1 \leqslant i_1 \leqslant i_2 \leqslant d} \operatorname{sgn}(Z_{i_1 k} - Z_{i_2 k}) \operatorname{sgn}(Z_{i_1 \ell} - Z_{i_2 \ell}) \end{aligned} \tag{B.3}$$

Sur la base du tau de Kendall, (B.1) conduit immédiatement à l'estimateur suivant pour la corrélation matrice ,

$$\widehat{\Sigma} = (\widehat{\sigma}_{k\ell})_{n \times n} \text{ avec } \widehat{\sigma}_{k\ell} = \sin\left(\frac{\pi}{2} \widehat{\tau}_{k\ell}\right) . \tag{B.4}$$

B.1.2 Estimation des coefficients de régression

Nous présentons maintenant la procédure d'estimation du vecteur β en (4.2). Si les transformations marginales $f_i, i = 1, \ldots, p$ ont été données, puis $(\widetilde{Y}, \widetilde{\mathbf{X}})$ sont disponibles et, dans ce cas une approche naturelle de l'estimation consiste à utiliser l'estimateur Lasso :

$$\begin{aligned} \widehat{\beta}_{Lasso} &= \operatorname{argmin} \left\{ \frac{1}{2d} \parallel \widetilde{Y} - \widetilde{X}\beta \parallel_2^2 + \lambda \parallel \beta \parallel_1 \right\} \\ &= \operatorname{argmin} \left\{ \frac{1}{2d} (\beta^\top \widetilde{X}^\top \widetilde{X} \beta - 2\widetilde{Y}^\top \widetilde{X}) + \lambda \parallel \beta \parallel_1 \right\} . \end{aligned} \tag{B.5}$$

Puisque $(\widetilde{Y}_i, \widetilde{\mathbf{X}}_i)$ ne sont pas directement accessibles car les transformations f_i sont inconnues, l'estimateur donné en (B.5) ne peut pas être utilisé. Les quantités $\widetilde{X}^\top \widetilde{X}/n$ et $\widetilde{Y}^\top \widetilde{X}/n$ dans (B.5) peuvent être considérées comme des estimateurs des covariances Σ_{XX} et Σ_{YX} respectivement. Dans cette perspective, il est naturel de remplacer $\widetilde{X}^\top \widetilde{X}/n$ et $\widetilde{Y}^\top \widetilde{X}/n$ dans (B.5) par les estimateurs alternatifs de covariance $\widehat{\Sigma}_{XX}$ et $\widehat{\Sigma}_{YX}$ basé sur celui de Kendall. Nous proposons donc la procédure de minimisation

pénalisée ℓ_1 suivante pour estimer β.

Exemple B.1 (Estimateur adaptatif de β)**.** *Pour un certain $\lambda > 0$, on calcule d'abord les estimateurs $\widehat{\Sigma}_{\mathbf{XX}}$ et $\sigma_{Y,\mathbf{X}}$ basés sur les tau de Kendall, et ensuite on obtient l'estimateur régularisé*

$$\widehat{\beta}(\lambda) = \operatorname{argmin} \left\{ \frac{1}{2} \left(\beta^\top \widehat{\Sigma}_{\mathbf{XX}} \, \beta - 2\widehat{\sigma}_{Y,\mathbf{X}} \, \beta \right) + \lambda \, \|\beta\| \right\}. \tag{B.6}$$

B.1.3 Prédiction

Après l'estimation de β, il restera à prédire la réponse $Y^\star$ pour une valeur donnée des covariables $\mathbf{x}^\star$ basé sur le modèle de régression à copule gaussienne (B.2). Dans le cas où les transformations $f_0, \ldots, f_p$ et le vecteur coefficient β sont connus, la prédiction optimale de la réponse est :

$$m(\mathbf{x}^\star) = f_0^{-1} \left(\sum_{i=1}^{p} f_i(x_i^*) \beta_i \right).$$

L' objectif est de construire un prédicteur $\widehat{m(\mathbf{x})}^\star$, basé uniquement sur les données observées $(Y_1, X_1), \ldots, (Y_d, X_d)$, qui est proche de l'oracle prédicteur $m(\mathbf{x})^\star$. Soit F_0 la fonction de distribution cumulative de Y et soit F_i la fonction de distribution cumulative de X_i pour $i = 1, \ldots, p$. Comme pour la version de l'échantillon, Soit $\widehat{F_0}$ la fonction de distribution cumulative empirique de $Y_1, \ldots, Y_d$ et $\widehat{F}_i$ la fonction de distribution cumulative empirique de $X_{i1}, \ldots, X_{id}$.

$$\widehat{f}_0(t) = \Phi^{-1} \left(\widetilde{F}_0(t) \right); i = 1, \ldots, d \tag{B.7}$$

$$\widehat{f}_i(t) = \Phi^{-1} \left(\widehat{F}_i(t) \right); i = 1, \ldots, d. \tag{B.8}$$

Où

$$\widetilde{F}_0(t) = \frac{1}{n^2} I\left(\widehat{F}_0(t) < \frac{1}{n^2}\right) + \widehat{F}_0(t) I \widehat{F}_0(t) \in \left[\frac{1}{n^2}, 1 - \frac{1}{n^2}\right] + \frac{n^2-1}{n^2} I\left(\widehat{F}_0(t) > 1 - \frac{1}{n^2}\right).$$

Pour une valeur donnée des covariables $x^\star = (x_1^\star, \ldots, x_p^\star)$, on déduit le prédicteur

$$\widehat{m}(\mathbf{x}^\star) = \widehat{f}_0^{-1}\left(\sum_{i=1}^{p} \widehat{f}_i(x_i^\star)\widehat{\beta}(\lambda)_i\right) \tag{B.9}$$

Où $\widehat{\beta}$ est l'estimateur donné en (B.6) et $\widehat{f}_0^{-1}$ est l'inverse généralisé de $\widehat{f}_0$:

$$\widehat{f}_0^{-1} = inf\left\{x \in \mathbb{R} : \widehat{f}_0(x) \geq t\right\}.$$

B.1.4 Estimation de la courbe de régression

À partir de l'équation (B.9) nous pouvons construire l'estimateur de la courbe de régression de la copule gaussienne. Ainsi ; on a pour Φ la fonction de répartition de la distribution normale,

$$\widehat{m}(\mathbf{x}^\star) = \Phi\left(\widetilde{F}_0\left(\sum_{i=1}^{p} \Phi^{-1}\left(\widehat{F}_i(x_i^\star)\right)\widehat{\beta}(\lambda)_i\right)\right)$$

Enfin, si on suppose que toutes les marges sont uniformes sur $[0,1]$, c'est-à-dire que $\widehat{F}_0(a) = \widehat{F}_1(a) = \cdots = \widehat{F}_p(a) = a$, alors pour $\mathbf{u} = (u_1, \ldots, u_p) \in [0,1]^p$,

$$\widehat{m}(\mathbf{u}) = \Phi\left(\sum_{i=1}^{p} \Phi^{-1}(\mathbf{u})\widehat{\beta}(\lambda)_i\right)$$

B.2 Extension aux copules elliptiques

B.2.1 Estimation de la matrice des corrélations

Étant donné que $\mathbf{f}$ et f_0 sont strictement croissantes, les vecteurs $\mathbf{Z} = (\mathbf{X}, Y)$ et $(\mathbf{f}(\mathbf{X}), f_0(Y))$ ont la même copule elliptique. Pour cette raison, le modèle est appelé un modèle de régression d'une copule elliptique à réponse multivariée . La matrice de corrélation de la copule commune $\mathbf{\Sigma}$ de $\mathbf{Z}$ et $(f(\mathbf{X}), f_0(Y))$ coïncide avec la matrice de corrélation de cette dernière. En conséquence, $\mathbf{\Sigma}$ peut alors être estimée à partir de l'échantillon observé de $\mathbf{Z}$ par inversion du tau de Kendall.

Soit $Z_1 = (\mathbf{X}_1, Y_1), \ldots, Z_d = (\mathbf{X}_d, Y_d)$ un échantillon aléatoire de taille $d \geq 2$ de $\mathbf{Z}$. Soit aussi $\mathbf{Z} = (Z_1, \ldots, Z_{p+1})$ et pour $i \in [d]$ soit $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{i(p+1)})$. Pour $k, \ell \in [p+1]$, la valeur du tau de Kendall entre le k-ième et le ℓ-ième coordonnée de $\mathbf{Z}$ est

$$\tau_{k\ell} = \mathrm{E}\{sgn(Z_{1k} - Z_{2k})sgn(Z_{1\ell} - Z_{2\ell})\}.$$

Soit $\mathbf{T} \in \mathbb{R}^{(p+1)\times(p+1)}$ la matrice dont la (k,ℓ)-ième entrée est $\tau_{k\ell}$. Lorsque $\mathbf{Z}$ est elliptique, on a $\Sigma = \sin(\pi\mathbf{T}/2)$. Considérons maintenant l'analogue empirique $\widehat{\mathbf{T}}$ de $\mathbf{T}$, dont la (k,ℓ)-ième entrée $\widehat{\tau}_{k\ell}$ est la version empirique du tau de Kendall entre les k-ième et ℓ-ième coordonnées de $\mathbf{Z}$, à savoir

$$\widehat{\tau}_{k\ell} = \frac{2}{d(d+1)} \sum_{1 \leqslant i \leqslant j \leqslant d} \{sgn(Z_{ik} - Z_{jk})sgn(Z_{i\ell} - Z_{j\ell})\}\,.$$

Un estimateur plug-in de $\mathbf{\Sigma}$ est alors donné par

$$\widehat{\mathbf{\Sigma}} = \sin\Big(\pi\widehat{\mathbf{T}}/2\Big)\,. \tag{B.10}$$

On sait également que $\widehat{\mathbf{T}}$ est une matrice, dont la distribution limite est gaussienne

et centrée sur $\mathbf{T}$. En conséquence, $\widehat{\mathbf{\Sigma}}$ est un estimateur cohérent de :

$$\begin{aligned} \mathbf{\Sigma} = \operatorname{cov}\{(f(\mathbf{X}), f_0(Y))\} &= \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{XX}} & \mathbf{\Sigma}_{\mathbf{X}Y} \\ \mathbf{\Sigma}_{Y\mathbf{X}}^\top & \mathbf{\Sigma}_{YY} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{XX}} & \mathbf{\Sigma}_{\mathbf{XX}}\mathbf{B}^* \\ \mathbf{B}^\top\mathbf{\Sigma}_{\mathbf{XX}} & 1 \end{pmatrix}. \end{aligned} \tag{B.11}$$

Ici, la dernière égalité de (B.11) découle de l'ellipticité conjointe et de la non-corrélation de $\mathbf{X}$ et ϵ. L'estimateur basé sur les rangs $\widehat{\mathbf{\Sigma}}$ de $\mathbf{\Sigma}$ peut être utilisé pour estimer la matrice des coefficients $\mathbf{B}^\star$ par le biais de l'équation (B.11). Dans cette partie, nous étudions une estimation colonne par colonne, de $\mathbf{B}^\star$ dont on ignore toute information sur la matrice de précision $\Omega_{\epsilon\epsilon}$. On considère un programme Lasso basé sur la matrice de conception $\widehat{\mathbf{\Sigma}}^+_{\mathbf{XX}}$ défini par

$$\widehat{\mathbf{\Sigma}}^+_{\mathbf{XX}} = \underset{\mathbf{M}\in\mathbb{R}^{p\times p}:\mathbf{M}\geqslant 0}{\operatorname{argmin}} \parallel \mathbf{M} - \widehat{\mathbf{\Sigma}}_{\mathbf{XX}} \parallel_{\ell\infty}.$$

B.2.2 L'approche Lasso

Pour chaque ℓ, on définit la perte $\mathcal{L}_\ell : \mathbb{R}^p \to \mathbb{R}$ la ℓ-ième colonne de $\mathbf{B}^\star = (\beta_1, \ldots, \beta_n)$ en fixant pour tout $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\mathcal{L}_\ell(\boldsymbol{\beta}) = \boldsymbol{\beta}\widehat{\mathbf{\Sigma}}^+_{\mathbf{XX}}\boldsymbol{\beta}/2 - (\widehat{\mathbf{\Sigma}}_{\mathbf{X}Y})_{\bullet\ell}\boldsymbol{\beta}. \tag{B.12}$$

$\mathcal{L}_\ell$ est motivé par la fonction de perte standard des moindres carrés carrés pour la ℓ-ième colonne de $\mathbf{B}^\star$ dans le modèle (4.2), mais avec les quantités transformées marginalement (mais non observées) $\sum_{i\in[d]} f(\mathbf{X}_i)f(\mathbf{X}_i)/n$ et $\sum_{i\in[d]} f(\mathbf{X}_i)f_0(Y_i)/n$ remplacés par les estimateurs plug-in $\widehat{\mathbf{\Sigma}}^+_{\mathbf{XX}}$ et $\widehat{\mathbf{\Sigma}}^+_{\mathbf{X}Y}$ respectivement. La perte $\mathcal{L}_\ell$ est convexe car $\widehat{\mathbf{\Sigma}}^+_{\mathbf{XX}}$ est semi-définie positive. L'estimateur Lasso de $\boldsymbol{\beta}^*_\ell$ avec λ_ℓ est un paramètre

d'ajustement est alors donné par

$$\widehat{\boldsymbol{\beta}}_\ell = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ \mathcal{L}_\ell(\boldsymbol{\beta}) + \lambda_\ell \mathcal{R}(\boldsymbol{\beta}) \right\} . \tag{B.13}$$

L'estimateur Lasso de $\mathbf{B}^\star$ est alors $\widetilde{\mathbf{B}} = \widehat{\beta}_\ell$

B.2.3 Estimation de la courbe de régression

Par extension de la copule gaussienne, nous pouvons construire l'estimateur de la courbe de régression de la copule elliptique. Ainsi, on a pour Ω la fonction de répartition de la distribution elliptique ,

$$\widehat{m}(\mathbf{x}^\star) = \Omega \left(\widetilde{F}_0 \left(\sum_{i=1}^{p} \Omega^{-1} \left(\widehat{F}_i(x_i^\star) \right) \widetilde{\mathbf{B}}_i \right) \right)$$

Enfin, si on suppose que toutes les marges sont uniformes sur $[0,1]$, c'est-à-dire que $\widehat{F}_0(a) = \widehat{F}_1(a) = \cdots = \widehat{F}_p(a) = a$, alors pour $\mathbf{u} = (u_1, \ldots, u_p) \in [0,1]^p$,

$$\widehat{m}(\mathbf{u}) = \Omega \left(\sum_{i=1}^{p} \Omega^{-1}(\mathbf{u}) \widetilde{\mathbf{B}}_i \right)$$