

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÉLISATION PAR COPULES DES RÉPONSES BINAIRES : RÉGRESSION MULTIVARIÉE, EN PRÉSENCE DE
COVARIABLES MIXTES ET INFÉRENCE ROBUSTE POUR DONNÉES DÉSÉQUILBRÉES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN MATHÉMATIQUES

PAR

YOUSSEF HANDI

MARS 2026

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

La réalisation de cette thèse n'aurait pu aboutir sans le soutien précieux de nombreuses personnes, auxquelles j'exprime ma profonde gratitude.

Tout d'abord, je tiens à exprimer ma profonde gratitude à mon directeur de recherche, le professeur Karim Oualkacha. Sa patience, sa disponibilité et la richesse de ses conseils ont été une source précieuse d'inspiration et ont grandement nourri ma réflexion.

Je remercie également chaleureusement mon co-directeur, le professeur Mhamed Mesfioui, pour son encadrement et les outils méthodologiques qu'il m'a apportés, essentiels à la réussite de ce travail.

Un immense merci aux professeurs et au personnel administratif du département de mathématiques, dont l'accompagnement a été précieux tout au long de mon parcours.

Je tiens également à exprimer ma profonde gratitude aux membres du jury, les professeurs Taoufik Bouezmarni, Juli Atherton et Michaël Lalancette, pour le temps qu'ils ont consacré à l'évaluation de cette thèse. Leurs remarques éclairées et leur engagement ont été d'une valeur inestimable.

Je tiens également à exprimer ma profonde gratitude à ma famille, pilier indéfectible tout au long de cette aventure. J'adresse une pensée particulière à mes parents et à mes frères, dont l'affection et les encouragements ont été ma plus grande force, ainsi qu'à mon père, qui m'a transmis sa passion pour les mathématiques et l'informatique.

Enfin, je remercie mes amis et collègues pour leur précieux soutien moral et intellectuel, qui a marqué mon parcours et enrichi cette expérience.

TABLE DES MATIÈRES

TABLE DES FIGURES	viii
LISTE DES TABLEAUX	ix
RÉSUMÉ	xiii
INTRODUCTION	1
CHAPITRE 1 CHAPITRE PRÉLIMINAIRE	6
1.1 Modèles linéaires généralisés vectoriels	6
1.1.1 Modèle Log-linéaire	7
1.1.2 Estimation des paramètres dans le modèle VGLM.....	10
1.1.3 Conclusion	13
1.2 Introduction aux copules.....	14
1.2.1 Définition mathématique	14
1.2.2 Lien entre la fonction de répartition et la copule	15
1.2.3 Extraction d'une copule	16
1.2.4 Bornes de Fréchet-Hoeffding et copule d'indépendance	18
1.2.5 Distributions conditionnelles bivariées exprimées en termes de leurs copules	19
1.2.6 Mesure de dépendance	20
1.2.7 Tau de Kendall	23
1.2.8 Estimation des paramètres de la fonction de copule	24
1.2.9 Copules archimédiennes (bidimensionnelle et multidimensionnelle)	26
1.2.10 Copules elliptiques	27
1.2.11 Décomposition des copules par les vignes	30
1.2.12 Estimation des paramètres dans une décomposition vigne	37

1.3	Fonctions de perte asymétriques utilisées : Huber-type, pseudo-Huber, expectile	42
1.3.1	Introduction	42
1.3.2	Notations et pondération asymétrique.....	42
1.3.3	Perte de Huber-type (piècewise quadratique-linéaire)	42
1.3.4	Fonction de perte pseudo-Huber	44
1.3.5	Perte expectile (moindres carrés asymétriques)	45
1.3.6	Extension Huber-type via un minimum	46
1.3.7	Synthèse des compromis (robustesse vs convenance analytique)	47
CHAPITRE 2 ON MULTIVARIATE BINARY OUTCOMES COPULAS-REGRESSION PROBLEM		48
2.1	Abstract	48
2.2	Introduction	49
2.3	Theoretical model	51
2.3.1	Theoretical Predictive Probability (PP) based on copulas	51
2.3.2	Example of predictive probability in some common copula families	54
2.4	Estimation and inference	59
2.5	Asymptotic Properties	61
2.5.1	Notations	61
2.5.2	Assumptions	61
2.6	Simulation study.....	63
2.6.1	Data generation	63
2.6.2	Simulation results	65
2.7	Real data illustration	69
2.7.1	Data description	69

2.7.2	Results	70
2.8	Conclusion	74
CHAPITRE 3 EXTENSION OF MULTIVARIATE BINARY OUTCOMES COPULAS-REGRESSION PROBLEM FOR MIXED COVARIABLES		76
3.1	Abstract	76
3.2	Introduction	77
3.3	Model formulation	78
3.3.1	Predictive probability based on copulas with mixed variables	78
3.4	Predictive probability based on common copulas families	81
3.4.1	Archimedean case	81
3.4.2	Gaussian case	84
3.5	Estimation	86
3.6	Simulation study.....	87
3.6.1	Results	91
3.7	Conclusion	93
CHAPITRE 4 COPULAS BASED INFERENCE FOR ASYMMETRIC BINARY REGRESSION		94
4.1	Abstract	95
4.2	Introduction	95
4.3	Model and Inference	98
4.3.1	Model Formulation	98
4.3.2	Illustration of $\lambda_0(\mathbf{x})$ based on some common loss functions	101
4.3.3	Copula-based representation of $\mathbb{P}(Y = y \mathbf{X} = \mathbf{x})$	104
4.3.4	Semi-parametric estimator of $\lambda_0(\cdot)$	107

4.4	Asymptotic properties	108
4.4.1	Assumption	108
4.5	Illustration and simulation of the optimal predictor based on copulas	112
4.5.1	Data generation process	112
4.5.2	Solution based on Huber-type loss function	113
4.5.3	Optimal predictor based on expectile loss	117
4.5.4	Simulation study	119
4.6	Real data illustration	121
4.6.1	Data description	121
4.6.2	Results	121
4.7	Discussion and extensions	123
	CONCLUSION	125
	ANNEXE A APPENDICES	129
A.1	Chapter 2 Appendices	129
A.1.1	Proof for Lemma 2.2	129
A.1.2	Proof of the Theorem 2.7	130
A.1.3	Evaluation of the variance of $\hat{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$	132
A.1.4	Parameters copula	133
A.1.5	Plots of $\hat{M}_{y_1,y_2}^{1,2}$ versus the covariables, based on Western Collaborative Group Study data	133
A.1.6	IMSE result of additional setting	134
A.1.7	Results for bias, SD, SE and coverage probability based on different scenarios from Settings 1 and 2	144
A.1.8	Selection of pair-copula families for real-data application	147

A.2	Chapter 3 Appendices	148
A.2.1	IMSE Results case $d=3, p=2$	148
A.2.2	Result of IMSE based on imbalanced data scenario	149
A.3	Chapter 4 Appendices.....	150
A.3.1	Proof of the Representation of $\lambda_0(\boldsymbol{x})$ given by (4.19)	150
A.3.2	Graphic illustration of $\lambda_0(\cdot)$ and $\hat{\lambda}_n(\cdot)$ based on Huber-type loss function	153
A.3.3	Algorithm.....	154
A.3.4	Monte Carlo Simulation Procedure	154
A.3.5	Accuracy metrics	155
A.3.6	Simulation Results	156
A.3.7	Result of real data illustration case $d = 2$	163
A.3.8	Accuracy metric results	163
A.3.9	Configuration of conditional copula based on vine decomposition.....	164
	BIBLIOGRAPHIE	165

TABLE DES FIGURES

Figure 1.1 Estimation séquentielle des paramètres dans une décomposition de vigne régulière 41

Figure 4.1 Comparison of quantile and Huber-type loss defined in (4.8) functions for $\tau = 0.25$ 102

Figure 4.2 Comparison of quantile and pseudo Huber-type loss functions for $\tau = 0.25$ 102

Figure 4.3 Graphical representation of $\lambda_0(\cdot)$ across expectile levels τ based on equation (4.17), with the Kendall parameter fixed at 0.5 106

Figure 4.4 Representation of $\lambda_0(\mathbf{x})$, and $\hat{\lambda}_n(\mathbf{x})$ on different quantile level based on pseudo Huber loss function with $p = 0.3$ 114

Figure 4.5 Representation of $\lambda_0(\mathbf{x})$, and $\hat{\lambda}_n(\mathbf{x})$ on different quantile level based on pseudo Huber loss function, where $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$ 115

Figure 4.6 Representation of $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\cdot)$ based on Huber-type loss function defined in (4.8) with $k = 0.05$ and $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$ 116

Figure 4.7 The representations of $\lambda_0(\cdot)$ and $\hat{\lambda}_n(\cdot)$ are derived based on the expectile loss function and Joe copula, evaluated at different expectile levels τ , where $X_1 \sim \mathcal{N}(0, 1)$ 118

Figure 4.8 The representations of $\lambda_0(\cdot)$ and $\hat{\lambda}_n(\cdot)$ are derived based on the expectile loss function and Joe copula, evaluated at different expectile levels τ , with $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$ 119

Figure A.1 Setting 1: Confidence interval of the estimator $\mathbb{M}_{0,0}^{1,2}(\mathbf{x})$ with $X \sim \mathcal{N}(0, 1)$ and sample size $n = 300$ 133

Figure A.2 The left panel shows the probability $\hat{\mathbb{M}}_{y_1, y_2}$ versus age. The right panel illustrates $\hat{\mathbb{M}}_{y_1, y_2}$ versus age and BMI index. For $\{y_1, y_2\} \in \{0, 1\}^2$ 133

Figure A.3 Estimation of the probability of success on different quantile levels based on the Huber loss function defined in Eq. (4.8) for different values of k , dashed lined and solid lines represent $\lambda_0(x_1)$ and $\hat{\lambda}_n(x_1)$ respectively. 153

LISTE DES TABLEAUX

Table 2.1	Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications, for $d = 2, p = 1$.	66
Table 2.2	Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications where $d = 2, p = 2$.	67
Table 2.3	Setting 2 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications where $d = 2, p = 1$.	67
Table 2.4	Setting 2 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications where where $d = 2, p = 2$.	68
Table 2.5	Observed frequencies of dependent variables and the average of estimated probability of each class in presence of $d = 2, p = 1$ (Age).	71
Table 2.6	Observed frequencies of dependent variables and the average of estimated probability of each class $d = 2, p = 2$ (Age, BMI).	71
Table 2.7	Observed and model-based joint state estimates for (Y_1, Y_2) with $d = 2, p = 2$.	73
Table 2.8	Error rates of classification for each joint state (Y_1, Y_2) .	73
Table 3.1	Algorithm 2 : Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, for $d = 2$ and $p = 2$.	92
Table 3.2	Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, using the correlation matrices defined in equation (3.35).	92
Table 4.1	Accuracy metrics for each method.	123
Table A.1	Setup of the data generation utilized in simulations.	133
Table A.2	Setting 1 (Unbalanced data) : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications, for $d = 2, p = 2$.	134
Table A.3	Setting 1 (Unbalanced data) : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications, for $d = 2, p = 1$.	135
Table A.4	Setting 1 (Unbalanced data) : Results of the IMSE ($\times 100$), based on $K=1000$ replications, for $d = 3, p = 2$.	136

Table A.5	Setting 1 (Unbalanced data) : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications, for $d = 3$, $p = 3$, with $p_1 = 0.05$, $p_2 = 0.3$, and $p_3 = 0.4$.	137
Table A.6	Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications, for $d = 3$, $p = 3$.	138
Table A.7	Setting 2 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications where $d = 3$, $p = 3$.	138
Table A.8	Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K = 1000$ replications where $d = 3$, $p = 2$.	139
Table A.9	Setting 2 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications where $d = 3$, $p = 2$.	139
Table A.10	Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications case $d = 3$, $p = 3$.	140
Table A.11	Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications case $d = 3$, $p = 3$.	141
Table A.12	Setting 1 : Results of IMSE ($\times 100$), $K=1000$ replications case $d = 3$, $p = 2$.	142
Table A.13	Setting 1 : Result of IMSE ($\times 100$) based on $K=1000$ replications where $d = 3$, $p = 2$.	143
Table A.14	Results of Bias, empirical SD, average SE, and 95% coverage $n = 300$, where $d = 2$, $p = 2$, based on setting 1.	144
Table A.15	Results of Bias, empirical SD, average SE, and 95% coverage probability for $n = 500$, where $d = 2$, $p = 2$, based on setting 1.	144
Table A.16	Results of Bias, empirical SD, average SE, and 95% coverage ($\times 100$) $n = 300$, where $d = 2$, $p = 2$, based on setting 2.	145
Table A.17	Results of Bias, empirical SD, average SE, and 95% coverage ($\times 100$) $n = 300$, with $d = 2$, $p = 2$ $p_1 = 0.05$, $p_2 = 0.4$ based on setting 1.	145
Table A.18	Results of Bias, empirical SD, average SE, and 95% coverage ($\times 100$) $n = 500$, where $d = 2$, $p = 2$ $p_1 = 0.05$, $p_2 = 0.4$ based on setting 1.	146
Table A.19	Results of Bias, empirical SD, average SE ($\times 100$) and 95% coverage, with $d = 3$, $p = 3$, $p_1 = 0.3$, $p_2 = 0.4$, $p_3 = 0.2$ based on setting 1, $n = 300$.	146

Table A.20 Specification of the copula family for each tree in the vine construction involving coronary heart disease Y_1 , Age and BMI.	147
Table A.21 Specification of the copula family for each tree in the vine construction involving corneal arcus disease Y_2 , Age and BMI.	147
Table A.22 Copula selection between the conditional marginal distributions of corneal arcus and coronary heart disease based on the proposed method.	147
Table A.23 Algorithm 2 : Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, for $d = 3$ and $p = 2$	148
Table A.24 Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, using the correlation matrices defined in equation (A.7).	148
Table A.25 Algorithm 2 : Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, for $d = 2$ and $p = 2$	149
Table A.26 Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, using the correlation matrices defined in equation (3.35) for $d = 2$ and $p = 2$	149
Table A.27 Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$).	156
Table A.28 Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$).	157
Table A.29 Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$).	157
Table A.30 Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$).	158
Table A.31 Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$).	158
Table A.32 Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$).	159
Table A.33 Mean of the accuracy metrics for each method based on Algorithm 3 (Gaussian dependence with $\rho_{Y,X} = 0.4$, $n = 1000$, and $p = 0.1$).	159

Table A.34 Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.5$, $n = 1000$ and $p = 0.1$).	160
Table A.35 Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.7$, $n = 1000$ and $p = 0.1$).	160
Table A.36 Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$)	161
Table A.37 Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$)	161
Table A.38 Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.5$, $n = 1000$ and $p = 0.1$).	162
Table A.39 Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.7$, $n = 1000$ and $p = 0.1$).	162
Table A.40 Mean of the accuracy metrics for each method based on Algorithm 3 (Gaussian dependence with the following correlation matrix defined in (A.9), $p = 0.1$ and $n = 1000$).	163
Table A.41 Accuracy metrics for each method with $d = 2$	163
Table A.42 C-vine copula configuration.	164

RÉSUMÉ

Dans plusieurs domaines d'application, tels que la médecine ou la finance, il est courant de chercher à modéliser un vecteur de réponses binaires en fonction de variables explicatives (covariables) afin de mieux prédire les résultats associés binaires à ces dernières. Une telle modélisation se fait, en général, à l'aide de modèles linéaires généralisés multidimensionnels (VGLM, de l'anglais "vector generalized linear models"). Toutefois, ces modèles présentent certaines limites lorsqu'il s'agit de capturer des dépendances non linéaires et complexes entre les variables étudiées. En effet, la relation entre les variables explicatives et les réponses ne peut souvent pas être décrite adéquatement par une simple structure de dépendance linéaire. Pour remédier à cette limite, nous faisons appel à la modélisation via les copules qui offrent la possibilité de modéliser une grande variété de formes de dépendances. Cette approche permet non seulement d'améliorer la précision des modèles, mais aussi d'affiner l'estimation des probabilités prédictives. Dans cette thèse, nous explorons de nouvelles approches pour la régression fondée sur les copules, afin de prédire simultanément plusieurs variables réponses binaires. Nous abordons d'abord le cas où les covariables sont continues, avant de généraliser notre méthodologie au cas plus complexe des covariables mixtes. Nous étudions également la régression binaire asymétrique dans le cadre particulier d'une seule variable réponse binaire.

Premièrement, nous introduisons un modèle basé sur les copules afin d'estimer la probabilité conjointe d'un vecteur de variables binaires, conditionnellement à un vecteur de covariables continues. Dans notre approche, la distribution marginale de chaque variable réponse est modélisée à l'aide d'une approche de régression univariée basée sur les copules. Ces distributions marginales servent ensuite à représenter la dépendance entre les variables réponses par l'intermédiaire d'une deuxième copule appropriée. Pour pallier aux limitations associées à l'utilisation d'un paramètre unique, qui contrôlerait simultanément la dépendance entre les variables réponses et la relation entre une variable réponse binaire et les covariables, nous proposons une décomposition de copules en vignes. Cette dernière procure une flexibilité accrue dans la modélisation des structures de dépendance complexes. La méthode proposée repose sur l'estimation distincte des paramètres de la copule décrivant la relation entre chaque variable réponse et les covariables à l'aide d'une fonction de vraisemblance, séparément des paramètres de la copule caractérisant la dépendance entre les distributions marginales. Cette distinction garantit une robustesse accrue de l'estimation et permet un développement asymptotique rigoureux de l'estimateur proposé pour la probabilité conjointe de prédiction.

Deuxièmement, nous développons une extension de notre première approche adaptée aux covariables mixtes, combinant à la fois des variables continues et discrètes. Cette généralisation implique la modélisation explicite des distributions marginales en tenant compte des covariables discrètes, avant d'appliquer les principes méthodologiques établis dans notre première approche.

Finalement, nous proposons une nouvelle méthode pour estimer un prédicteur optimal dans le cadre de la régression binaire basée sur les copules. Cette méthode s'appuie sur l'adaptation de fonctions de perte asymétriques, telles que la fonction de Huber appliquée à la régression quantile et expectile, particulièrement utile en présence de données déséquilibrées. Ces fonctions de perte asymétriques attribuent des poids différents aux observations, permettant ainsi de balancer la contribution de la classe minoritaire par rapport à celle de la classe majoritaire dans l'ensemble du modèle. Nous introduisons ensuite un estimateur pour ce prédicteur optimal et établissons ses propriétés asymptotiques. Enfin, nous évaluons l'efficacité de l'estimateur à travers une application sur des données réelles.

INTRODUCTION

Dans les approches classiques de la régression, l'objectif principal est d'expliquer un phénomène, représenté par une variable réponse, à partir de plusieurs variables explicatives (covariables). Autrement dit, il s'agit d'étudier la relation ou l'association entre une variable réponse et un ensemble de covariables. Plusieurs méthodes sont disponibles pour traiter ce type de problématique (McCullagh et Nelder, 1989; Hastie *et al.*, 2009; Agresti, 2010) . Cependant, dans de nombreux contextes, notamment en médecine, l'intérêt porte souvent sur l'analyse de la relation entre la présence simultanée de plusieurs maladies et divers indicateurs, tels que l'état de santé général d'un individu, son âge et son sexe, ainsi que ses habitudes de consommation d'alcool, entre autres.

Parmi les méthodes utilisées pour aborder ce type de problématique figure l'approche des modèles linéaires généralisés vectoriels (de l'anglais : *Vector Generalized Linear Models* (VGLM)). Cette approche repose sur le même principe que les modèles linéaires généralisés. Elle modélise l'effet individuel des covariables sur chaque variable réponse binaire à l'aide d'une fonction de lien linéaire, puis décrit la dépendance entre les variables réponses binaires en utilisant une fonction de masse. Plus précisément, ce modèle se caractérise par deux types de paramètres : d'une part, ceux représentant l'association entre chaque variable réponse binaire et les covariables via la fonction de lien, et d'autre part, ceux qui quantifient la dépendance entre les différentes variables réponses binaires. Dans ce contexte, un exemple bien connu est l'approche proposée par Zhao et Prentice (1990). Cette méthode repose sur l'utilisation du modèle quadratique exponentiel afin de modéliser les interactions (les dépendances) entre les variables réponses. Ce modèle permet d'intégrer explicitement les relations entre les réponses binaires, facilitant ainsi la prise en compte des influences réciproques existantes. De plus, il améliore la précision ainsi que la fiabilité des résultats obtenus dans le cadre de régressions binaires multiples. Toutefois, cette approche présente certaines limitations. Elle se restreint à l'étude de liens linéaires entre chaque variable réponse binaire et les covariables, tandis que les structures d'interaction demeurent limitées à des formes quadratiques, réduisant ainsi sa flexibilité pour capturer des dépendances plus complexes entre les différentes variables réponses.

Pour remédier à cette problématique, plusieurs auteurs ont exploré la modélisation statistique fondée sur les copules. Introduite par Sklar (1959), la théorie des copules est un outil mathématique largement utilisé en statistique pour modéliser la dépendance entre plusieurs variables aléatoires, indépendamment de leurs distributions marginales. Elle offre une flexibilité notable dans la construction des modèles multivariés. En

effet, les copules permettent de capturer des structures de dépendance très complexes entre les variables, incluant notamment des relations non linéaires et asymétriques qui échappent généralement aux modèles traditionnels reposant sur la corrélation qui est une mesure d'association linéaire. Cette propriété rend les copules particulièrement précieuses pour l'analyse des dépendances, expliquant ainsi leur adoption croissante par les chercheurs dans des contextes variés.

À titre d'exemple, Nikoloulopoulos et Karlis (2010) ont introduit un modèle de régression fondé sur les copules pour traiter le cas spécifique de deux variables réponses de type comptage. Leur approche consiste tout d'abord à modéliser la distribution marginale de chaque variable à l'aide d'un modèle de régression adéquat, tel que la régression de Poisson. Ces marges sont ensuite combinées afin de déterminer la dépendance entre les deux variables réponses par le biais d'une copule. Cette technique procure une flexibilité considérable dans le choix des structures de dépendance, dépassant ainsi les approches traditionnelles basées sur des relations linéaires. Par ailleurs, les paramètres du modèle sont estimés selon une approche de maximum de vraisemblance, ce qui permet de distinguer clairement l'effet des covariables sur chaque variable de comptage de la dépendance intrinsèque entre ces dernières.

Dans une perspective similaire, De Leon et Wu (2011) ont étendu cette approche en considérant une variable réponse continue et une variable réponse discrète binaire. Leur méthodologie repose également sur la spécification des distributions marginales conditionnelles pour chacune des variables réponses. Ainsi, la variable réponse continue est modélisée par une régression linéaire, tandis que la variable réponse binaire est modélisée par une régression logistique, chacune intégrant les covariables pertinentes. Une copule est alors employée afin de capturer la structure de dépendance entre ces deux types de variables réponses. Cette méthode présente l'avantage d'analyser des variables réponses de nature mixte.

Enfin, Genest *et al.* (2013) ont proposé une approche complémentaire permettant de prédire les résultats d'un vecteur binaire, en combinant la régression logistique avec des copules elliptiques. Dans cette méthode, chaque variable réponse est modélisée à l'aide d'une régression logistique intégrant des covariables afin d'estimer la probabilité de succès conditionnelle. Ensuite, une copule elliptique est utilisée pour modéliser la dépendance entre les marges estimées par la régression logistique, permettant ainsi de capturer des liens non linéaires et complexes entre les variables binaires.

Les travaux mentionnés précédemment ont eu recours aux copules pour modéliser la dépendance entre les

variables réponses, dont les distributions marginales sont généralement spécifiées au moyen de modèles linéaires ou de modèles linéaires généralisés. Récemment, Mesfioui *et al.* (2023) ont proposé une approche alternative basée sur une fonction de lien calculée à l'aide des copules, permettant de capturer la relation entre une variable réponse binaire et un ensemble de covariables. Cette méthode vise particulièrement à fournir une estimation précise de la probabilité conditionnelle de succès :

$$\mathbb{P}(Y = 1|\mathbf{X} = \mathbf{x}) = 1 - \mathbb{P}(Y = 0|\mathbf{X} = \mathbf{x}), \quad \text{où } \mathbf{x} \in \mathbb{R}^p.$$

La représentation de cette probabilité peut être obtenue de façon explicite à l'aide des copules, comme le montre le travail de Bouyé et Salmon (2009). Ce qui permet d'explorer des liens complexes entre une variable binaire et les covariables. En outre, dans le travail de Mesfioui *et al.* (2023), les auteurs ont fourni les propriétés asymptotiques de leur estimateur, fournissant ainsi le comportement asymptotique de cette probabilité estimée. Cette représentation asymptotique est cruciale car elle permet de déduire la distribution limite de l'estimateur, facilitant ainsi l'établissement d'intervalles de confiance pour évaluer la précision des estimations obtenues.

Dans le cadre de cette thèse, nous développons une méthodologie novatrice visant à modéliser la dépendance au sein d'un vecteur de réponses binaires, en recourant à un modèle de régression structuré sur deux niveaux et fondé sur l'utilisation des copules. Dans un premier temps, les marges conditionnelles des réponses, sachant les covariables, sont estimées selon l'approche proposée par Mesfioui *et al.* (2023). Cette stratégie offre une flexibilité accrue dans la représentation de la relation entre chaque variable réponse binaire et l'ensemble des covariables continues, en s'affranchissant des contraintes imposées par les approches traditionnelles reposant sur des liens linéaires. Dans un second temps, ces marges conditionnelles sont mobilisées pour modéliser la structure de dépendance entre les différentes composantes du vecteur réponse à l'aide d'une copule appropriée, sélectionnée de manière rigoureuse. Par ailleurs, l'estimateur semi-paramétrique proposé s'appuie sur une procédure d'inférence garantissant ses propriétés asymptotiques, assurant ainsi la robustesse et la validité statistique des résultats obtenus. Cette méthodologie a également été étendue au contexte des covariables mixtes, intégrant conjointement des variables continues et discrètes.

Dans un second volet, nous nous intéressons à la régression binaire univariée dans un contexte où l'une des modalités de la variable réponse est nettement moins représentée que l'autre, une problématique largement connue dans la littérature sous l'appellation de données déséquilibrées (de l'anglais "imbalanced data"). Dans ce cadre, nous proposons d'intégrer des fonctions de perte asymétriques, telles que celles de type Huber et expectile, à la fonction de lien pour la régression binaire introduite par Mesfioui *et al.* (2023).

L'objectif est d'identifier un prédicteur optimal de la variable d'intérêt, $Y \in \{0, 1\}$, en présence d'une forte disparité de représentation entre les classes. Cette approche présente l'avantage d'accroître la flexibilité de la modélisation de la dépendance entre la réponse binaire et les covariables grâce à une représentation par copules, tout en s'écartant des méthodes heuristiques couramment utilisées dans la littérature qui utilisent des méthodes de rééchantillonnage afin d'équilibrer les catégories. En effet, notre approche repose sur la résolution directe d'une équation explicite et sur une procédure d'inférence robuste, assurant des propriétés asymptotiques appropriées de l'estimateur associé au prédicteur optimal proposé.

Cette thèse est structurée en quatre chapitres. Le premier chapitre constitue une introduction préliminaire. Il commence par une présentation détaillée de la méthode VGLM, en mettant l'accent sur les méthodes d'estimation des paramètres associés à ce modèle. Ensuite, nous introduisons la notion de copules, en définissant rigoureusement leurs fondements mathématiques et leurs propriétés essentielles. Cette section se conclut par une discussion sur le théorème de Sklar (1959), qui établit la possibilité de décomposer une distribution multivariée en ses marges univariées et une copule. Dans la suite du chapitre, nous passons en revue les principales familles de copules utilisées en statistique, tout en soulignant leurs caractéristiques distinctives. Enfin, nous abordons des méthodes modernes d'estimation des paramètres des modèles basés sur les copules, visant à dépasser la contrainte selon laquelle un seul paramètre contrôle la dépendance. À cet égard, nous explorons des approches telles que les décompositions de copules en vignes (notamment C-vine et D-vine), initialement introduites par Joe (1996) et développées par Bedford et Cooke (2001) et Bedford et Cooke (2002). Les Chapitres 2, 3 et 4 présenteront nos travaux réalisés au cours de cette thèse, rédigés en anglais. Ces chapitres présenteront des résultats à la fois théoriques et empiriques portant sur la modélisation de la dépendance entre variables binaires à l'aide des copules, ainsi que sur la régression binaire basée sur les copules en présence de données déséquilibrées.

Le Chapitre 2, soumis à la revue "Japanese Journal of Statistics and Data Science", a reçu des évaluations favorables qui ont motivé des révisions mineures substantielles, désormais transmises. Il présente un cadre de régression multivariée pour un vecteur de réponses binaires. Dans ce contexte, l'association entre chaque variable réponse et les covariables est spécifiée au moyen de distributions conditionnelles induites par des copules, conformément à l'approche proposée par Mesfioui *et al.* (2023). Ces marges conditionnelles sont ensuite utilisées pour représenter la structure de dépendance entre les composantes du vecteur de réponses. La contribution principale du Chapitre consiste à introduire une fonction de lien non linéaire, fondée sur les copules, pour modéliser les marges de chaque réponse binaire via le modèle de Mesfioui *et al.* (2023), puis à recourir à une seconde copule afin de capturer la dépendance entre les réponses, ce qui ac-

croît substantiellement la flexibilité du modèle. Cette flexibilité découle de la capacité du lien induit par les copules à capturer des relations non linéaires et potentiellement asymétriques entre une variable réponse et les covariables. Elle provient également de la dissociation entre la modélisation des marges et celle de la structure de dépendance, laquelle peut être spécifiée, le cas échéant, à l'aide d'une décomposition en vignes, permettant ainsi d'adapter finement la forme de la dépendance en dimension élevée. En outre, nous avons établi les propriétés asymptotiques de notre estimateur de la probabilité prédictive multivariée, permettant ainsi de caractériser sa loi limite. Ces résultats facilitent la construction d'intervalles de confiance et offrent des moyens appropriés pour évaluer l'efficacité de la méthode proposée. Par la suite, dans le Chapitre 3, nous étendons la méthodologie introduite précédemment au cas de covariables mixtes. Nous adaptons ainsi la formulation mathématique des marges conditionnelles des réponses binaires, toujours basées sur la régression binaire par copules proposée par Mesfioui *et al.* (2023), en intégrant explicitement les covariables mixtes. Afin de valider l'efficacité de cette adaptation, nous avons mené une étude par simulations. Les résultats obtenus sont comparés à ceux de méthodes de référence dans la littérature, notamment le modèle C-logit introduit par Genest *et al.* (2013) ainsi que la méthode (VGLM) proposée par Zhao et Prentice (1990).

Le Chapitre 4, actuellement en révision à la revue "Journal of Statistical Theory and Practice" à la suite d'évaluations favorables, développe un cadre d'inférence du prédicteur optimal en régression binaire en présence de données déséquilibrées. L'approche repose sur l'utilisation de copules combinées à des fonctions de perte asymétriques, telles que la fonction de Huber adaptée à la régression quantile et la fonction de perte expectile. Ce travail propose une nouvelle approche pour estimer le meilleur prédicteur dans le cas où la variable réponse binaire est caractérisée par une distribution asymétrique (c'est-à-dire des données déséquilibrées). De plus, nous avons dérivé les propriétés asymptotiques du meilleur prédicteur estimé par notre méthode. Ensuite, nous avons testé l'efficacité de la classification basée sur notre approche sur des données réelles, concernant la détection de transactions frauduleuses en présence de données déséquilibrées.

CHAPITRE 1

CHAPITRE PRÉLIMINAIRE

Dans le domaine des statistiques, l'analyse de la dépendance joue un rôle central. Elle est également essentielle dans divers domaines d'application, tels que l'assurance et l'épidémiologie. Traditionnellement, cette analyse repose sur des modèles de régression linéaire, qui s'avèrent toutefois insuffisants pour capturer des structures de dépendance complexes ou non linéaires. C'est dans ce contexte que les copules, introduites par Sklar (1959), offrent une solution particulièrement intéressante.

Dans ce Chapitre, nous analysons en profondeur la méthode Vector Generalized Linear Model (VGLM), en mettant en lumière les techniques utilisées pour l'estimation de ses paramètres ainsi que les limitations auxquelles cette approche est confrontée. Par la suite, nous proposons une introduction détaillée aux copules, afin de fournir une vue d'ensemble complète des approches méthodologiques qui seront déployées tout au long de cette thèse.

1.1 Modèles linéaires généralisés vectoriels

Supposons que l'on observe un vecteur de réponses binaires $\mathbf{Y} = (Y_1, \dots, Y_d)^\top$, ainsi qu'un vecteur de covariables \mathbf{X} composé de p éléments, $(X_1, \dots, X_p) \in \mathbb{R}^p$. Pour modéliser la relation entre \mathbf{Y} et \mathbf{X} , le modèle VGLM est formulé en supposant que la distribution conditionnelle de \mathbf{Y} sachant \mathbf{x} est donnée par l'expression suivante :

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\Gamma}) = \mathcal{F}(\mathbf{y}, \eta_1, \dots, \eta_d; \boldsymbol{\sigma}^2, \boldsymbol{\Gamma}), \quad (1.1)$$

où $\eta_j = \eta_j(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}_j$, avec $\boldsymbol{\beta}_j = (\beta_{j0}, \dots, \beta_{jp})^\top$, représente le vecteur des coefficients de régression associé au j -ième prédicteur linéaire. La fonction $\mathcal{F}(\cdot; \eta_1, \dots, \eta_d, \boldsymbol{\sigma}^2, \boldsymbol{\Gamma})$ désigne une fonction de masse jointe appartenant à la famille exponentielle quadratique, paramétrée par le vecteur des dispersions marginales $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_d^2)^\top$ et par la matrice d'association $\boldsymbol{\Gamma} = (\gamma_{ij}), i, j = 1, \dots, d$ qui représente la dépendance entre les composantes du vecteur de réponses \mathbf{Y} .

La fonction \mathcal{F} doit satisfaire les deux propriétés suivantes :

1. **Reproductibilité ou fermeture marginale** : Un sous-ensemble des composantes du vecteur \mathbf{Y} et leurs modèles de régression correspondants devraient conserver la même forme de distribution que le modèle complet. En d'autres termes, ajuster un modèle de régression pour un sous-ensemble de réponses maintient les propriétés de la distribution des erreurs, qui restent cohérentes avec celles

du modèle global.

2. **Caractérisation des associations** : Les composantes de la matrice $\mathbf{\Gamma}$ doivent refléter les associations positives et négatives entre les composantes de \mathbf{Y} .

Le modèle VGLM présente une importante similarité avec les modèles de dispersion introduits par (Jorgensen, 1997). En effet, ces deux approches introduisent des paramètres supplémentaires afin de capturer des aspects complexes des données. Dans les modèles de dispersion, des paramètres spécifiques sont utilisés pour modéliser la variabilité non constante des erreurs et la dépendance entre les observations. De manière similaire, dans la méthode VGLM, les paramètres de dispersion σ^2 et la matrice d'association $\mathbf{\Gamma}$ sont introduits pour modéliser la dépendance entre plusieurs réponses ainsi que la variabilité entre une réponse binaire Y_j et les covariables \mathbf{X} .

De plus, la méthode VGLM peut être vue comme une extension naturelle des modèles de dispersion dans un cadre multivarié, où plusieurs variables réponses sont modélisées simultanément tout en prenant en compte la complexité de la dépendance entre elles. Ces deux approches partagent également l'idée de relaxer les hypothèses de variance constante et d'indépendance, permettant ainsi une modélisation plus flexible et réaliste des données.

Après avoir présenté le modèle VGLM dans ses grandes lignes, la section suivante se concentre sur un modèle largement adopté pour la distribution \mathcal{F} , connu sous le nom de modèle log-linéaire. Fréquemment utilisé dans divers contextes, ce modèle constitue une approche à la fois flexible et efficace pour représenter la relation entre les variables réponses et les covariables. Sa structure log-linéaire, en particulier, le rend adapté à la modélisation de relations complexes, tout en permettant une interprétation claire et intuitive des coefficients estimés.

1.1.1 Modèle Log-linéaire

Le modèle log-linéaire est largement utilisé pour modéliser les corrélations entre les composantes d'un vecteur de variables binaires (Bishop *et al.*, 1977). La fonction de masse associée à ce modèle est donnée par :

$$\mathbb{P}(\mathbf{y}) = c(\boldsymbol{\theta}) \exp \left\{ \sum_{j=1}^d \theta_j y_j + \sum_{j_1 < j_2} \theta_{j_1 j_2} y_{j_1} y_{j_2} + \dots + \theta_{1\dots d} y_1 \dots y_d \right\}. \quad (1.2)$$

Dans cette expression, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d, \theta_{12}, \dots, \theta_{d-1d}, \dots, \theta_{1\dots d})^T$ représente un vecteur de paramètres canoniques contenant $(2^d - 1)$ éléments, et $c(\boldsymbol{\theta})$ désigne le terme de normalisation. L'équation (1.2) correspond au modèle saturé, qui est contraint par une seule relation : $\sum_{\mathbf{y}} \mathbb{P}(\mathbf{y}) = 1$. Il est important de noter que cette forme saturée du modèle log-linéaire est en fait une paramétrisation complète ; elle peut représenter n'importe quelle distribution de probabilité sur l'espace des variables binaires $\{0, 1\}^d$, pourvu que les $2^d - 1$ paramètres canoniques $\boldsymbol{\theta}$ soient choisis de manière appropriée. Le vecteur $\boldsymbol{\theta}$ capture ainsi toutes les associations et interactions possibles entre les variables binaires constituant \mathbf{Y} .

Pour simplifier le modèle et réduire le nombre de paramètres à estimer, nous supposons que les termes d'ordre supérieur au second sont nuls. Cette hypothèse conduit à une version réduite du modèle exponentiel quadratique, telle qu'introduite dans le travail de (Zhao et Prentice, 1990). La fonction de masse associée s'exprime alors comme suit :

$$\begin{aligned} \mathbb{P}(\mathbf{y}; \boldsymbol{\theta}) &= c(\boldsymbol{\theta}) \exp \left\{ \sum_{j=1}^d \theta_j y_j + \sum_{j < k} \theta_{jk} y_j y_k \right\} \\ &= c(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \exp \{ \mathbf{y}^T \boldsymbol{\theta}_1 + \mathbf{w}^T \boldsymbol{\theta}_2 \}, \end{aligned} \quad (1.3)$$

avec $\boldsymbol{\theta}_1 = (\theta_1, \dots, \theta_d)^\top$, $\boldsymbol{\theta}_2 = (\theta_{12}, \dots, \theta_{d-1d})^\top$, $\mathbf{y} = (y_1, \dots, y_d)^\top$, et $\mathbf{w} = (y_1 y_2, \dots, y_{d-1} y_d)^\top$.

De plus, $\theta_j = \log \left(\frac{\mathbb{P}(Y_j=1|Y_k=0, k \neq j)}{\mathbb{P}(Y_j=0|Y_k=0, k \neq j)} \right)$, où θ_j représente l'effet individuel de chaque variable sur la probabilité de succès par rapport à la probabilité d'échec, en ignorant les effets des autres variables binaires. La probabilité $\mathbb{P}(Y_j = 1)$ représente l'espérance marginale notée μ_j . Dans le cadre d'un modèle de régression, le paramètre μ_j est relié aux covariables via un lien logit, avec $\text{logit}(\mu_j) = \mathbf{x}^\top \boldsymbol{\beta}_j$. De la même manière :

$$\log \left\{ \frac{\mathbb{P}(Y_j = 1, Y_k = 1 | Y_l = y_l, l \neq j, k) \mathbb{P}(Y_j = 0, Y_k = 0 | Y_l = y_l, l \neq j, k)}{\mathbb{P}(Y_j = 1, Y_k = 0 | Y_l = y_l, l \neq j, k) \mathbb{P}(Y_j = 0, Y_k = 1 | Y_l = y_l, l \neq j, k)} \right\} = \theta_{jk}.$$

Remarque 1.1 Notons qu'il existe une relation entre les paramètres $(\boldsymbol{\mu}, \boldsymbol{\theta}_2)$ et la paramétrisation canonique $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ dans le cadre de la représentation donnée par l'équation (1.3), comme l'énonce la proposition suivante.

Proposition 1.2 Rappelons l'équation (1.3)

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = c(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \exp \{ \mathbf{y}^\top \boldsymbol{\theta}_1 + \mathbf{w}^\top \boldsymbol{\theta}_2 \},$$

en fixant le vecteur θ_2 , il existe une correspondance bijective entre $(\theta_1, \theta_2) \longleftrightarrow (\mu, \theta_2)$ où $\mu = \mathbb{E}_\theta[\mathbf{Y}]$.

Preuve. Commençons par montrer que la fonction de masse (1.3) appartient à une famille exponentielle multivariée.

La fonction (1.3) est donnée par :

$$\mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \mathbf{x}; \theta_1, \theta_2) = c(\theta_1, \theta_2) \exp \left\{ \mathbf{y}^\top \theta_1 + \mathbf{w}(\mathbf{y})^\top \theta_2 \right\}, \quad \mathbf{y} \in \{0, 1\}^d.$$

Pour qu'une fonction appartienne à une famille exponentielle, elle doit s'écrire sous la forme canonique :

$$p(\mathbf{y}; \theta) = h(\mathbf{y}) \exp \left\{ \theta^\top T(\mathbf{y}) - A(\theta) \right\},$$

où $T(\mathbf{y})$ est la statistique exhaustive, θ est le vecteur des paramètres naturels, $A(\theta)$ est la fonction log-normale, et $h(\mathbf{y})$ est une fonction de base indépendante des paramètres. Pour déterminer les composantes de la fonction de masse, on doit établir une expression de notre fonction de normalisation $c(\theta_1, \theta_2)$

Comme notre fonction de masse définie dans l'équation (1.3) est une densité donc on a que :

$$\sum_{\mathbf{y} \in \{0, 1\}^d} \mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \mathbf{x}; \theta_1, \theta_2) = 1.$$

Donc :

$$c(\theta_1, \theta_2) \sum_{\mathbf{y} \in \{0, 1\}^d} \exp \left\{ \mathbf{y}^\top \theta_1 + \mathbf{w}(\mathbf{y})^\top \theta_2 \right\} = 1,$$

avec

$$\mathbf{W}(\mathbf{y}) = (y_1 y_2, y_1 y_3, \dots, y_1 y_d, y_2 y_3, \dots, y_{d-1} y_d)^\top, \quad \mathbf{y} \in \{0, 1\}^d.$$

Ce qui donne :

$$c(\theta_1, \theta_2) = \left(\sum_{\mathbf{y} \in \{0, 1\}^d} \exp \left\{ \mathbf{y}^\top \theta_1 + \mathbf{w}(\mathbf{y})^\top \theta_2 \right\} \right)^{-1}.$$

En posant :

$$\psi(\theta_1, \theta_2) = \log \left(\sum_{\mathbf{y} \in \{0, 1\}^d} \exp \left\{ \mathbf{y}^\top \theta_1 + \mathbf{w}(\mathbf{y})^\top \theta_2 \right\} \right),$$

on obtient :

$$c(\theta_1, \theta_2) = \exp \left\{ -\psi(\theta_1, \theta_2) \right\}.$$

Ainsi, la fonction log-normale $A(\boldsymbol{\theta})$ est :

$$A(\boldsymbol{\theta}) = \psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Par la propriété générale des familles exponentielles voir (Peter et Song, 2007, Proposition 2.8), on sait que :

$$\mathbb{E}_{\boldsymbol{\theta}}[T(\mathbf{Y})] = \nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta}),$$

où $\mathbf{T}(\mathbf{Y}) = (\mathbf{Y}^\top, \mathbf{w}(\mathbf{Y})^\top)$. En particulier :

$$\boldsymbol{\mu} = \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{Y}] = \nabla_{\boldsymbol{\theta}_1}\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \quad \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{w}(\mathbf{Y})] = \nabla_{\boldsymbol{\theta}_2}\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

où

$$\nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta}) = \begin{pmatrix} \nabla_{\boldsymbol{\theta}_1}\psi(\boldsymbol{\theta}) \\ \nabla_{\boldsymbol{\theta}_2}\psi(\boldsymbol{\theta}) \end{pmatrix},$$

$$\nabla_{\boldsymbol{\theta}_1}\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \left(\frac{\partial\psi}{\partial\theta_1}, \dots, \frac{\partial\psi}{\partial\theta_d} \right)^\top,$$

$$\nabla_{\boldsymbol{\theta}_2}\psi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \left(\frac{\partial\psi}{\partial\theta_{12}}, \dots, \frac{\partial\psi}{\partial\theta_{(d-1)d}} \right)^\top.$$

Pour établir la relation entre $\boldsymbol{\theta}_1$ et $\boldsymbol{\mu}$, fixons $\boldsymbol{\theta}_2$. Comme la fonction ψ est strictement convexe et de classe C^1 , l'application $\nabla_{\boldsymbol{\theta}_1}\psi$ est une bijection. Ainsi, la relation entre $\boldsymbol{\mu}$ et $\boldsymbol{\theta}_1$ est bijective via $\nabla_{\boldsymbol{\theta}_1}\psi$, ce qui achève la preuve. \square

Après avoir présenté une définition complète du modèle VGML, la section suivante expose les méthodes d'estimation des paramètres, en s'appuyant sur la fonction de masse définie dans l'équation (1.3).

1.1.2 Estimation des paramètres dans le modèle VGML

L'estimation des paramètres d'un modèle linéaire généralisé vectoriel (VGML), particulièrement dans le cas de variables réponse de type binaire, repose sur la maximisation de la vraisemblance jointe. Cette section présente en détail les différentes méthodes utilisées pour estimer les paramètres du modèle VGML, dans lequel la fonction de masse de la distribution conditionnelle $\mathbf{Y} \mid \mathbf{X} = \mathbf{x}$ est spécifiée à l'aide du modèle proposé par Zhao et Prentice (1990).

1.1.2.1 Estimation par maximum de vraisemblance

L'estimation des paramètres $\boldsymbol{\theta}$ dans un modèle VGML, où les variables réponses sont binaires, suit le principe du *maximum de vraisemblance*. La fonction de vraisemblance (ou fonction de likelihood) mesure la

probabilité (pour des variables discrètes) ou la densité (pour des variables continues) d'observer les données effectivement observées, considérée comme une fonction des paramètres inconnus du modèle.

1.1.2.2 Fonction de vraisemblance :

Pour un échantillon de taille n , soit $\mathbf{y}_i = (y_{i1}, \dots, y_{id})^\top$ les observations des variables de réponse binaires pour l'individu i , et \mathbf{x}_i le vecteur des covariables associé à cet individu. La vraisemblance jointe du modèle s'exprime comme suit :

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \mathbb{P}(\mathbf{Y} = \mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta}),$$

où $\mathbb{P}(\mathbf{Y} = \mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})$ représente la probabilité d'observer \mathbf{y}_i , étant donné les covariables \mathbf{x}_i et les paramètres $\boldsymbol{\theta}$, comme défini par l'équation (1.3). En pratique, on maximise plutôt le logarithme de la fonction de vraisemblance pour simplifier les calculs. Cela permet de transformer le produit des probabilités en une somme :

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \ln(\mathbb{P}(\mathbf{Y} = \mathbf{y}_i \mid \mathbf{x}_i; \boldsymbol{\theta})).$$

Cette fonction de log-vraisemblance est ensuite maximisée pour obtenir les estimations des paramètres $\boldsymbol{\theta}$.

1.1.2.3 Méthodes d'estimation

Maximiser la log-vraisemblance pour des modèles VGLM n'a généralement pas de solution analytique simple, surtout dans le cas de modèles avec des variables de réponse corrélées et des termes d'interaction. Par conséquent, des méthodes numériques sont nécessaires pour obtenir les estimations des paramètres. Voici les méthodes couramment utilisées.

1.1.2.3.1 Algorithme de Newton-Raphson

L'algorithme de Newton-Raphson (Christopher et Nasrabadi, 2006, chap. 4 et annexe sur l'optimisation)) est une méthode itérative utilisée pour maximiser la fonction de vraisemblance. Il repose sur le calcul du gradient (la dérivée première) et de la matrice hessienne (la dérivée seconde) de la fonction de log-vraisemblance.

L'idée est de commencer par une estimation initiale $\boldsymbol{\theta}^{(0)}$ des paramètres et de la mettre à jour itérativement

selon la formule :

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \mathbf{H}^{-1}(\boldsymbol{\theta}^{(t)})\nabla l(\boldsymbol{\theta}^{(t)}),$$

où :

- $\nabla l(\boldsymbol{\theta}^{(t)})$ est le vecteur gradient de la log-vraisemblance au point $\boldsymbol{\theta}^{(t)}$,
- $\mathbf{H}(\boldsymbol{\theta}^{(t)})$ est la matrice hessienne (matrice des dérivées secondes) au point $\boldsymbol{\theta}^{(t)}$.

1.1.2.3.2 Algorithme basé sur l'information de Fisher

Cet Algorithme (McCulloch et Searle, 2004, chap. 7) est une variante de l'algorithme de Newton-Raphson, mais au lieu d'utiliser la matrice hessienne, il utilise une approximation de celle-ci basée sur l'information de Fisher espérée (Fisher, 1925). La mise à jour des paramètres est donnée par :

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \mathbf{I}^{-1}(\boldsymbol{\theta}^{(t)})\nabla l(\boldsymbol{\theta}^{(t)}),$$

où $\mathbf{I}(\boldsymbol{\theta}^{(t)})$ est la matrice d'information de Fisher espérée, qui est une approximation de la matrice hessienne et est toujours définie positive. Cela rend cet algorithme plus stable que celui de Newton-Raphson.

Remarque 1.3 *Les algorithmes présentés dans les sections 1.1.2.3.1 et 1.1.2.3.2 constituent des méthodes itératives conçues pour déterminer en pratique les estimations des paramètres d'intérêt. Ces approches itératives reposent sur des critères d'arrêt rigoureusement définis, assurant ainsi la convergence vers les solutions optimales recherchées. L'algorithme ci-dessous détaille précisément les différentes étapes nécessaires à la mise en œuvre effective de ces estimateurs.*

Algorithm 1 Estimation du maximum de vraisemblance par Newton-Raphson ou Fisher-scoring

Require: Données \mathcal{D} , log-vraisemblance $\ell(\boldsymbol{\theta})$ tolérances $(\varepsilon_\theta, \varepsilon_g, \varepsilon_\ell)$ nombre maximal d'itérations T_{\max}

choix du *mode* $\in \{\text{Newton, Fisher}\}$

1: Initialiser $\boldsymbol{\theta}^{(0)}$ (valeur de départ)

2: **for** $t = 0$ **to** $T_{\max} - 1$ **do**

3: Calculer le **score** $u^{(t)} \leftarrow \nabla \ell(\boldsymbol{\theta}^{(t)})$

4: **if** *mode* **is** Newton **then**

5: Calculer la **hessienne observée** $H^{(t)} \leftarrow \nabla^2 \ell(\boldsymbol{\theta}^{(t)})$

6: $\Delta^{(t)} \leftarrow -(H^{(t)})^{-1} u^{(t)}$

7: **else**

▷ Fisher-scoring

8: Calculer l'**information de Fisher espérée** $I^{(t)} \leftarrow \mathbb{E}[-H^{(t)}]$

9: $\Delta^{(t)} \leftarrow +(I^{(t)})^{-1} u^{(t)}$

10: **end if**

11: Mettre à jour le paramètre $\boldsymbol{\theta}^{(t+1)} \leftarrow \boldsymbol{\theta}^{(t)} + \Delta^{(t)}$

▷ — Vérification des critères d'arrêt —

12: **if** $\frac{\|\Delta^{(t)}\|_2}{1 + \|\boldsymbol{\theta}^{(t)}\|_2} < \varepsilon_\theta$ **or** $\|u^{(t)}\|_\infty < \varepsilon_g$ **or** $|\ell(\boldsymbol{\theta}^{(t+1)}) - \ell(\boldsymbol{\theta}^{(t)})| < \varepsilon_\ell$ **then**

13: **break**

▷ Convergence atteinte

14: **end if**

15: **end for**

16: **Sortie :** $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}^{(t+1)}$

▷ Estimé final

17: Calculer la matrice variance-covariance $\widehat{\text{Var}}(\hat{\boldsymbol{\theta}}) \leftarrow \begin{cases} -H^{(t+1)^{-1}}, & \text{si mode = Newton} \\ I^{(t+1)^{-1}}, & \text{si mode = Fisher} \end{cases}$

1.1.3 Conclusion

L'estimation par maximisation de la vraisemblance dans les modèles VGLM présente plusieurs limitations notables. Tout d'abord, elle est associée à une complexité computationnelle importante, en particulier lorsque le modèle intègre un grand nombre de variables réponses ou d'interactions. Cette méthode est également sensible aux valeurs initiales des paramètres, ce qui peut entraîner des problèmes de convergence ou conduire à des solutions sous-optimales locales. En outre, la colinéarité entre les covariables peut provoquer des estimations instables, augmenter les erreurs standards, et compliquer l'interprétation des résultats. Par ailleurs, des hypothèses incorrectes sur la distribution ou la fonction de lien peuvent intro-

duire des biais significatifs dans les estimations. Cette méthode est également vulnérable au surajustement, notamment lorsque le modèle comprend un nombre excessif de paramètres, ce qui compromet sa capacité à se généraliser sur de nouvelles données. Enfin, les interactions complexes entre les variables réponses peuvent rendre difficile le calcul précis de la matrice hessienne, ajoutant une couche supplémentaire de complexité au processus d'estimation. Ces problèmes sont abordés dans l'ouvrage de Yee (2015) notamment aux chapitres 3 et 4.

Pour pallier ces défis, nous proposons une approche alternative dans laquelle la distribution $\mathbb{P}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x})$ est modélisée à l'aide d'une copule. Cette méthode permet de capturer des relations non linéaires de manière plus efficace, tout en réduisant le nombre de paramètres à estimer, offrant ainsi une solution potentiellement plus flexible et robuste.

Afin d'illustrer l'utilité des copules dans la modélisation des dépendances entre variables aléatoires, il est nécessaire d'en introduire le cadre théorique. Cette section a pour objet de définir formellement le concept de copule, d'en exposer les propriétés fondamentales et de dresser un panorama des familles les plus couramment utilisées pour caractériser la structure de dépendance stochastique.

1.2 Introduction aux copules

Dans cette section, nous donnons une définition mathématique de la fonction copule. Nous présentons ensuite ses propriétés, puis nous expliquons comment extraire la copule à partir de la loi jointe et des lois marginales d'un vecteur de variables aléatoires, conformément au théorème de Sklar (1959).

1.2.1 Définition mathématique

Soit $\mathbf{u} = (u_1, \dots, u_d)$ un vecteur de variables aléatoires à valeurs dans le cube unitaire d -dimensionnel $[0, 1]^d$. On appelle copule toute fonction

$$\mathcal{C} : [0, 1]^d \longrightarrow [0, 1],$$

où $d \geq 2$, qui satisfait les trois propriétés fondamentales énoncées dans l'ouvrage de Nelsen (2006) :

- (C_1) : $\mathcal{C}(\mathbf{u}) = 0$ si $\min(u_1, \dots, u_d) = 0$, c'est-à-dire si au moins une des composantes de \mathbf{u} est nulle.
- (C_2) : Si toutes les composantes de \mathbf{u} sont égales à 1, sauf la j -ème composante (qui vaut alors u_j), on a $\mathcal{C}(\mathbf{u}) = u_j$.

- (C_3) : La fonction \mathcal{C} possède la propriété de d -croissance, 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 $C_1 \implies \mathcal{C}(0, u) = \mathcal{C}(u, 0) = 0$, tandis que $C_2 \implies \mathcal{C}(1, u) = \mathcal{C}(u, 1) = u$, $\forall u \in [0, 1]$. Pour terminer, la propriété C_3 assure que, pour n'importe quels $u_1 \leq u'_1 \in [0, 1]$ et $u_2 \leq u'_2 \in [0, 1]$, on a :

$$\mathcal{C}(u'_1, u'_2) - \mathcal{C}(u'_1, u_2) - \mathcal{C}(u_1, u'_2) + \mathcal{C}(u_1, u_2) \geq 0.$$

1.2.2 Lien entre la fonction de répartition et la copule

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

Théorème 1.4 (Théorème de Sklar (1959))

Soit $\mathbf{F}(\cdot)$ une fonction de répartition de dimension $d \geq 2$ dont les lois marginales sont F_1, \dots, F_d . Par conséquent, il existe une copule $\mathcal{C} : [0, 1]^d \rightarrow [0, 1]$ telle que

$$\mathbf{F}(x_1, \dots, x_d) = \mathcal{C}\{F_1(x_1), \dots, F_d(x_d)\}, \quad (1.4)$$

pour tout $x_1, \dots, x_d \in \mathcal{X}$, où \mathcal{X} définit le support de la variable aléatoire x_i , $i = 1, \dots, d$.

Si les marges F_1, \dots, F_d sont continues, alors \mathcal{C} dans la représentation (1.4) est unique.

La démonstration du Théorème 1.4 est donnée dans le livre de (Nelsen, 2006, Chapitre 2, page 21).

Pour bien illustrer l'unicité de la représentation donnée par l'équation (1.4) dans le Théorème 1.4, considérons un vecteur aléatoire (X_1, \dots, X_d) suivant une loi conjointe \mathbf{F} avec des marges continues F_1, \dots, F_d .

L'unique copule associée à \mathbf{F} correspond alors à la fonction de répartition du vecteur transformé :

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

En effet, si F_j^{-1} désigne l'inverse de la fonction de répartition F_j , on a :

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

D'après le Théorème 1.4, on déduit que :

$$\mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d) = \mathcal{C}\{F_1 \circ F_1^{-1}(u_1), \dots, F_d \circ F_d^{-1}(u_d)\} = \mathcal{C}(u_1, \dots, u_d), \quad (1.5)$$

où la notation $f \circ g(x) = f(g(x))$ désigne la composition des fonctions. Ainsi, cette construction illustrée par l'équation (1.5) met en évidence l'unicité de la copule associée à une loi conjointe à marges continues, comme le garantit le théorème 1.4 de Sklar (1959). Elle montre également que la structure de dépendance entre les composantes du vecteur aléatoire est entièrement capturée par la copule, indépendamment des lois marginales.

1.2.3 Extraction d'une copule

Le Théorème 1.4 (Sklar, 1959) stipule qu'une loi de probabilité \mathbf{F} , dont les marges F_1, \dots, F_d sont continues et associées à une copule \mathcal{C} , peut être exprimée de manière unique sous la forme suivante :

$$\mathbf{F}(x_1, \dots, x_d) = \mathcal{C} \{F_1(x_1), \dots, F_d(x_d)\}. \quad (1.6)$$

En posant $u_j = F_j(x_j)$ pour chaque $j \in \{1, \dots, d\}$, l'équation (1.6) peut également s'écrire sous une forme équivalente :

$$\mathcal{C}(u_1, \dots, u_d) = \mathbf{F} \{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\}.$$

Cette formulation permet d'extraire de manière unique la copule associée à une loi d -dimensionnelle continue, illustrant ainsi le lien fondamental entre une loi conjointe et ses marges par l'intermédiaire de la copule.

Exemple 1.5 Soit \mathbf{F} une fonction de répartition définie pour le couple $(x, y) \in \mathbb{R}^2$ et un paramètre $\theta \geq 1$, exprimée comme suit :

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

Les fonctions de répartition marginales de X et Y s'écrivent alors de la manière suivante :

$$F(x) = \lim_{y \rightarrow \infty} \mathbf{F}(x, y) = \exp(-e^{-x}),$$

$$G(y) = \lim_{x \rightarrow \infty} \mathbf{F}(x, y) = \exp(-e^{-y}).$$

De plus, pour $u, v \in [0, 1]$, les fonctions quantiles inverses sont données par :

$$F_1^{-1}(u) = -\ln\{-\ln(u)\}, \quad F_2^{-1}(v) = -\ln\{-\ln(v)\}.$$

La copule correspondante est définie pour tout $u, v \in [0, 1]$ par :

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

Proposition 1.6 La densité d'une copule bivariee est donnee pour tout $u, v \in [0, 1]$ par

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} \mathcal{C}(u, v) = \frac{\mathbf{f} \{F_1^{-1}(u), F_2^{-1}(v)\}}{f_1 \{F_1^{-1}(u)\} f_2 \{F_2^{-1}(v)\}},$$

où f_1, f_2, \mathbf{f} representent respectivement les densites de F_1, F_2 et \mathbf{F} .

Dans un contexte multivarie, en appliquant le Theoreme 1.4, la densite se definit de la maniere suivante :

$$c(\mathbf{u}) = \frac{\partial^d}{\partial u_1 \dots \partial u_d} \mathcal{C}(\mathbf{u}),$$

avec $\mathbf{u} = (u_1, \dots, u_d)$. La densite du vecteur aleatoire (X_1, \dots, X_d) peut egalement etre exprimee en fonction de la densite de la copule, comme suit :

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \dots f_d(x_d). \quad (1.7)$$

L'extraction de la densite de la copule est donnee par :

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

où f est la densite de \mathbf{F} et f_1, \dots, f_d sont les densites marginales.

Exemple 1.7 Soit \mathbf{H} la fonction de repartition definie pour tout $x, y \in [0, 1]$ par

$$\mathbf{F}(x, y) = \sqrt{\frac{xy}{x+y-xy}}.$$

La copule correspondante est definie pour tout $u, v \in [0, 1]$ par

$$\mathcal{C}(u, v) = (u^{-2} + v^{-2} - 1)^{-1/2}.$$

La densite correspondante s'exprime pour tout $u, v \in [0, 1]$ comme suit :

$$c(u, v) = 3(uv)^{-3} (u^{-2} + v^{-2} - 1)^{-5/2}.$$

1.2.4 Bornes de Fréchet–Hoeffding et copule d'indépendance

La copule d'indépendance est l'une des copules les plus utilisées pour modéliser l'absence de dépendance entre deux variables aléatoires. Cette indépendance est caractérisée par la factorisation de la loi conjointe \mathbf{F} en le produit de ses marges F_1 et F_2 . En appliquant le théorème 1.4 de Sklar (1959), la copule associée est définie pour tout $u, v \in [0, 1]$ comme suit :

$$\Pi(u, v) = uv.$$

Dans le cas où $d > 2$, la copule d'indépendance s'exprime de la manière suivante :

$$\Pi(\mathbf{u}) = \prod_{i=1}^d u_i.$$

La copule d'indépendance est aussi appelée copule de produit.

La proposition suivante donne un intervalle de valeurs qu'une copule arbitraire peut prendre pour un point arbitraire $(u, v) \in [0, 1]^2$.

Proposition 1.8 Soit \mathcal{C} une copule. Pour tout $u, v \in [0, 1]$, on a

$$M(u, v) = \max(u + v - 1, 0) \leq \mathcal{C}(u, v) \leq m(u, v) = \min(u, v).$$

Par conséquent, si \mathbf{F} est une fonction de répartition bivariée, alors pour tout $x, y \in \mathbb{R}$ on a

$$\max\{F_1(x) + F_2(y) - 1, 0\} \leq \mathbf{F}(x, y) \leq \min\{F_1(x), F_2(y)\}.$$

Preuve. Soient A et B deux événements. On peut écrire que

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(A \cap B) \leq \mathbb{P}(B). \tag{1.8}$$

Donc, on a

$$\mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\}.$$

En supposant les événements suivants, $A = \{X \leq x\}$ et $B = \{Y \leq y\}$ pour $x, y \in \mathbb{R}$, en appliquant le résultat de l'équation (1.8), on a

$$\mathbb{P}(X \leq x, Y \leq y) \leq \min\{\mathbb{P}(X \leq x), \mathbb{P}(Y \leq y)\},$$

ce qui signifie que $\mathbf{F}(x, y) \leq \min\{F(x), G(y)\}$.

De plus, en appliquant les lois de De Morgan, on a

$$\begin{aligned}\mathbb{P}(A \cap B) &= 1 - \mathbb{P}(A^c \cup B^c) \\ &= 1 - \{\mathbb{P}(A^c) + \mathbb{P}(B^c) - \mathbb{P}(A^c \cap B^c)\} \\ &= \mathbb{P}(A) + \mathbb{P}(B) - 1 + \mathbb{P}(A^c \cap B^c) \\ &\geq \mathbb{P}(A) + \mathbb{P}(B) - 1.\end{aligned}$$

Considérant que $\mathbb{P}(A \cap B) \geq 0$, on a donc

$$\mathbb{P}(A \cap B) \geq \max\{\mathbb{P}(A) + \mathbb{P}(B) - 1, 0\}.$$

Par définition des événements $A = \{X \leq x\}$ et $B = \{Y \leq y\}$, on déduit que

$$\mathbf{F}(x, y) \geq \max\{F_1(x) + F_2(y) - 1, 0\}.$$

Par conséquent, nous avons montré le résultat désiré. \square

Remarque 1.9 En vertu de la Proposition 1.8, on en conclut que :

- Les quantités $M(u, v)$ et $m(u, v)$ représentent respectivement la borne inférieure et supérieure de la copule bivariable \mathcal{C} .
- La borne supérieure $m(\cdot)$ représente une dépendance positive, tandis que la borne inférieure $M(\cdot)$ traduit une dépendance négative.

1.2.5 Distributions conditionnelles bivariées exprimées en termes de leurs copules

Cette Section présente la définition de la distribution conditionnelle associée à une copule dans le cadre bivarié, largement utilisée en modélisation de régression via les copules. Dans ce qui suit, nous utiliserons le fait que nous pouvons exprimer les densités conditionnelles en termes d'un produit d'une copule et d'une densité marginale.

Lemme 1.10 (Densités conditionnelles et distribution bivariées en termes de leur copule)

La densité et la fonction de répartition conditionnelles peuvent être exprimées à l'aide des copules en appli-

quant le Théorème 1.4 de Sklar (1959). La densité conditionnelle est exprimée par :

$$\begin{aligned} f_{1|2}(x_1 | x_2) &= c_{1,2}(F_1(x_1), F_2(x_2)) f_1(x_1) \\ F_{1|2}(x_1 | x_2) &= \frac{\partial}{\partial u_2} \mathcal{C}_{1,2}(F_1(x_1), u_2) \Big|_{u_2=F_2(x_2)} \\ &= \frac{\partial}{\partial F_2(x_2)} \mathcal{C}_{1,2}(F_1(x_1), F_2(x_2)). \end{aligned}$$

Où $\mathcal{C}_{1,2}$ désigne la copule associée au vecteur (X_1, X_2) , caractérisant leur structure de dépendance, tandis que $c_{1,2}$ représente sa densité.

Preuve. En utilisant la définition d'une densité conditionnelle et l'équation (1.7), nous avons

$$\begin{aligned} f_{1|2}(x_1 | x_2) &= \frac{f_{1,2}(x_1, x_2)}{f_2(x_2)} \\ &= \frac{c_{1,2}(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2)}{f_2(x_2)} \\ &= c_{1,2}(F_1(x_1), F_2(x_2)) f_1(x_1) \\ &= \frac{\partial^2 \mathcal{C}_{1,2}(u_1, u_2)}{\partial u_1 \partial u_2} \Big|_{u_1=F_1(x_1), u_2=F_2(x_2)} \frac{\partial u_1}{\partial x_1} \\ &= \frac{\partial}{\partial u_2} \left(\frac{\partial}{\partial x_1} \mathcal{C}_{1,2}(F_1(x_1), u_2) \right) \Big|_{u_2=F_2(x_2)}. \end{aligned}$$

Cela implique que

$$\begin{aligned} F_{1|2}(x_1 | x_2) &= \int_{-\infty}^{x_1} \frac{\partial}{\partial u_2} \left(\frac{\partial}{\partial z_1} \mathcal{C}_{1,2}(F_1(z_1), u_2) \right) \Big|_{u_2=F_2(x_2)} dz_1 \\ &= \frac{\partial}{\partial u_2} \left(\int_{-\infty}^{x_1} \frac{\partial}{\partial z_1} \mathcal{C}_{1,2}(F_1(z_1), u_2) dz_1 \right) \Big|_{u_2=F_2(x_2)} \\ &= \frac{\partial}{\partial u_2} \mathcal{C}_{1,2}(F_1(x_1), u_2) \Big|_{u_2=F_2(x_2)}. \end{aligned}$$

□

Le Lemme 1.10 peut également être appliqué à la copule bivariable $\mathcal{C}_{1,2}$. Dans ce cas-ci, la fonction de répartition conditionnelle et la densité conditionnelle sont notées respectivement par $\mathcal{C}_{1|2}$ et $c_{1|2}$.

1.2.6 Mesure de dépendance

À la suite de la définition mathématique de la fonction copule, qui mesure la dépendance entre les variables aléatoires indépendamment de leurs marges, cette section présente les différents types de mesures de dépendance.

1.2.6.1 Mesure de concordance

La mesure de concordance est largement utilisée en statistique et en théorie des probabilités. Elle permet de quantifier le degré de dépendance entre deux variables aléatoires, offrant ainsi une indication à la fois de la force et de la direction de leur relation.

Définition 1.11 *Considérons (x_i, y_i) et (x_j, y_j) comme deux observations du vecteur (X, Y) de variables aléatoires continues. On énonce que (x_i, y_i) et (x_j, y_j) sont en concordance si*

$$\left\{ \begin{array}{l} x_i < x_j, \\ y_i < y_j, \end{array} \right. \text{ ou } \left\{ \begin{array}{l} x_j < x_i, \\ y_j < y_i. \end{array} \right.$$

De façon similaire, on dit que (x_i, y_i) et (x_j, y_j) sont en discordance si

$$\left\{ \begin{array}{l} x_i > x_j, \\ y_i < y_j, \end{array} \right. \text{ ou } \left\{ \begin{array}{l} x_i < x_j, \\ y_i > y_j. \end{array} \right.$$

Nous allons formuler notre mesure de concordance d'une manière mathématique. Supposons qu'on veut mesurer la concordance entre les deux couples (X_1, Y_1) et (X_2, Y_2) . La mesure de leur concordance est définie comme suit :

$$Q = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Par conséquent, la mesure de concordance se définit comme la différence entre les probabilités de concordance et de discordance entre ces deux couples. Dans le cas particulier où les lois \mathbf{F}_1 et \mathbf{F}_2 , associées à (X_1, Y_1) et (X_2, Y_2) respectivement, ont les mêmes marges, on peut montrer que :

$$Q = Q(\mathcal{C}_1, \mathcal{C}_2) = 4 \int_0^1 \int_0^1 \mathcal{C}_2(u, v) d\mathcal{C}_1(u, v) - 1, \quad (1.9)$$

où \mathcal{C}_1 et \mathcal{C}_2 représentent respectivement les couples de (X_1, Y_1) et (X_2, Y_2) . Les mesures d'association de Kendall et de Spearman sont des cas particuliers de mesures de concordance. En effet, en prenant (X_1, Y_1) et (X_2, Y_2) indépendants et provenant de la même loi \mathbf{F} de copule \mathcal{C} , le tau de Kendall est défini par

$$\tau(\mathcal{C}) = Q(\mathcal{C}, \mathcal{C}) = 4 \int_0^1 \int_0^1 \mathcal{C}(u, v) d\mathcal{C}(u, v) - 1.$$

1.2.6.2 Rhô de Spearman théorique

La mesure de dépendance de Spearman découle en supposant que la copule de (X_1, Y_1) est \mathcal{C} , tandis que celle de (X_2, Y_2) correspond à l'indépendance, telle que $\mathcal{C}_2(u, v) = \Pi(u, v) = uv$. Par l'équation (1.9), le Rhô de Spearman s'exprime de la manière suivante :

$$\rho_S(\mathcal{C}) = 3Q(\mathcal{C}, \Pi) = 12 \int_0^1 \int_0^1 uv \, d\mathcal{C}(u, v) - 3. \quad (1.10)$$

Le coefficient de corrélation de Spearman, également appelé rhô de Spearman, est une mesure non paramétrique de la corrélation entre deux variables aléatoires. Contrairement à la corrélation de Pearson, qui évalue uniquement les relations linéaires entre des variables continues, le coefficient de Spearman se base sur les rangs des observations, ce qui le rend adapté à des données non normales ou à des relations non linéaires. Cette approche est particulièrement utile pour analyser des relations monotones ou lorsque les hypothèses de normalité et de linéarité ne sont pas respectées.

La définition mathématique du rhô de Spearman repose sur la proportion de différence entre la probabilité de concordance et celle de discordance des couples aléatoires. Par exemple, considérons les couples $(X_1, Y_1), (X_2, Y_2)$ et (X_3, Y_3) le rhô de Spearman est donné par :

$$\rho_{S_{X,Y}} = 3\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0].$$

En prenant la probabilité de l'événement suivant

$$\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0] = 1 - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0],$$

notre rhô de Spearman devient :

$$\rho_{S_{X,Y}} = 6\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3.$$

Si \mathbf{F} représente la distribution de (X_1, Y_1) et Π celle de (X_2, Y_3) (car les variables X_2 et Y_3 sont indépendantes), alors le Rhô de Spearman s'exprime par l'équation (1.10).

En pratique, lorsque la distribution \mathbf{F} est inconnue, on utilise un estimateur non paramétrique pour évaluer le rhô de Spearman.

Définition 1.12 (Rhô de Spearman empirique) Soit $\{(X_i, Y_i)\}_{i=1, \dots, n}$ une famille de vecteurs bivariés. Le rhô de Spearman empirique, noté ρ_n , est défini par

$$\rho_n = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - S_i)^2,$$

où R_i est le rang de X_i et S_i celui de Y_i .

Remarque 1.13 Le coefficient de corrélation de Spearman empirique, noté ρ_n , prend ses valeurs dans l'intervalle $[-1, 1]$. En particulier :

- $\rho_n = 1$ si et seulement si, les rangs des deux variables sont parfaitement concordants (relation monotone strictement croissante).
- $\rho_n = -1$ si, et seulement si, les rangs sont parfaitement inversés (relation monotone strictement décroissante).
- Une valeur $\rho_n \approx 0$ suggère l'absence d'association monotone entre les rangs, sans pour autant garantir l'indépendance au sens probabiliste.

1.2.7 Tau de Kendall

Le tau de Kendall est une mesure de corrélation non paramétrique qui quantifie la dépendance entre deux variables aléatoires. Introduite par Kendall (1938), cette méthode se distingue par sa capacité à détecter d'autres dépendances que la dépendance linéaire entre les variables.

Définition 1.14 Soit $(X_i, Y_i)_{i=1,2}$ un échantillon bivarié tiré d'un couple de variables aléatoires continues (X, Y) , supposées indépendantes et identiquement distribuées (i.i.d.), dont la fonction de répartition conjointe est notée \mathbf{F} . Le tau de (Kendall, 1938), associé à ce couple de variables aléatoires et désigné par $\tau_{X,Y}$, est défini comme suit :

$$\tau_{X,Y} = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Définition 1.15 (Version empirique du tau de Kendall) Soit un échantillon de données bivariées de taille n d'un couple de variables aléatoires (X, Y) continues. On définit le tau de Kendall empirique (Genest et Favre, 2007) par

$$\tau_n = \frac{[\text{Nombre de paires concordantes}] - [\text{Nombre de paires discordantes}]}{\text{Nombre total de paires}}.$$

Remarque 1.16 Le tau de Kendall vérifie les propriétés suivantes :

- Le tau de Kendall est une quantité prenant des valeurs entre -1 et 1, comme le coefficient de Pearson.
- Un coefficient de Kendall proche de 1 signifie qu'une dépendance positive existe entre nos variables, et plus il tend vers -1, plus la dépendance est négative.
- Un coefficient de Kendall proche de zéro signifie qu'il n'existe aucune dépendance entre nos variables, avec une probabilité très élevée.

Après avoir présenté une définition approfondie des fonctions de copule, il convient désormais de préciser comment en estimer les paramètres dans un contexte appliqué. Parmi les méthodes les plus couramment utilisées figure l'approche par pseudo-vraisemblance. Le principe consiste à ne pas imposer de modèle paramétrique pour les distributions marginales des variables, mais à les estimer d'abord de manière non paramétrique, puis à construire la fonction de vraisemblance de la copule à partir des pseudo-observations ainsi obtenues. Cette stratégie, introduite notamment dans Genest *et al.* (1995), permet d'isoler l'estimation de la dépendance de la modélisation des marges. La section suivante décrit cette procédure en détail.

1.2.8 Estimation des paramètres de la fonction de copule

Soit $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$, $i = 1, \dots, n$, un échantillon i.i.d. de loi jointe F . Lorsque les marges F_1, \dots, F_d sont continues, le théorème de Sklar permet d'écrire

$$\mathbf{F}(\mathbf{x}) = \mathcal{C}(F_1(x_1), \dots, F_d(x_d); \boldsymbol{\theta}),$$

où $\mathcal{C}(\cdot; \boldsymbol{\theta})$ est une copule paramétrique (densité $c(\cdot; \boldsymbol{\theta})$) et $\boldsymbol{\theta} \in \Theta$ désigne le paramètre de dépendance. Au niveau densité,

$$f(\mathbf{x}) = c(F_1(x_1), \dots, F_d(x_d); \boldsymbol{\theta}) \prod_{j=1}^d f_j(x_j).$$

La vraisemblance complète dépend donc à la fois de $\boldsymbol{\theta}$ et des marges, ces dernières pouvant être considérées comme des paramètres de nuisance. La méthode de pseudo-vraisemblance proposée par Genest *et al.* (1995) consiste à éliminer les marges en les remplaçant par des estimateurs non paramétriques, conduisant à une estimation semiparamétrique de $\boldsymbol{\theta}$.

1.2.8.0.1 Pseudo-observations (transformation par les rangs).

Pour chaque marge $j \in \{1, \dots, d\}$, on considère la fonction de répartition empirique

$$\widehat{F}_{n,j}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_{ij} \leq x).$$

On construit ensuite les *pseudo-observations*

$$\widehat{U}_{ij} = \widehat{F}_{n,j}(X_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, d.$$

En pratique, afin d'éviter les valeurs 0 et 1 (problématiques pour $\log c$), on utilise souvent une version à base de rangs R_{ij} :

$$\widehat{U}_{ij} = \frac{R_{ij}}{n+1} \quad (\text{où } \widehat{U}_{ij} = \frac{R_{ij} - 1/2}{n} \text{ et } R_{ij} = \sum_{k=1}^n \mathbb{1}(X_{kj} \leq X_{ij})).$$

Intuitivement, lorsque F_j est continue, la transformation de probabilité $U_{ij} = F_j(X_{ij})$ donne des variables uniformes sur $[0, 1]$, et \widehat{U}_{ij} est une approximation par rangs de U_{ij} .

1.2.8.0.2 Pseudo-log-vraisemblance de copule (MPL / CML).

En traitant $\widehat{\mathbf{U}}_i = (\widehat{U}_{i1}, \dots, \widehat{U}_{id})$ comme des observations approximatives issues de la copule $\mathcal{C}(\cdot; \boldsymbol{\theta})$, on définit la pseudo-log-vraisemblance

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^n \log c(\widehat{\mathbf{U}}_i; \boldsymbol{\theta}),$$

et l'estimateur MPL (ou "canonical" ML) par

$$\widehat{\boldsymbol{\theta}}_{\text{MPL}} \in \arg \max_{\boldsymbol{\theta} \in \Theta} \ell_n(\boldsymbol{\theta}).$$

Cette procédure est *invariante* par transformations strictement croissantes appliquées individuellement aux marges (car fondée sur les rangs), et ne requiert aucune spécification paramétrique des distributions marginales. Des résultats de consistance et de normalité asymptotique Les conditions régulières sont données dans l'article de Genest *et al.* (1995) ainsi que dans le livre de Joe (2014), au chapitre 5.

Après une présentation approfondie des copules et des différentes mesures de dépendance, nous introduirons, dans la section suivante, quelques familles de copules et leurs propriétés. Les deux familles les plus connues et couramment utilisées sont les copules elliptiques et les copules archimédiennes.

1.2.9 Copules archimédiennes (bidimensionnelle et multidimensionnelle)

Définition 1.17 Soit une fonction $\varphi : [0, 1] \rightarrow [0, \infty)$ décroissante et convexe, avec $\varphi(1) = 0$. La copule archimédienne bivariee de générateur φ est définie par

$$\mathcal{C}(u, v) = \varphi^{-1}\{\varphi(u) + \varphi(v)\}. \quad (1.11)$$

L'équation (1.11) reste valide même lorsque l'on considère un pseudo-inverse de φ . De plus, si la copule \mathcal{C} est archimédienne, son générateur peut être déduit de la manière suivante :

$$\frac{\varphi'(u)}{\varphi'(v)} = \frac{\partial \mathcal{C}(u, v) / \partial u}{\partial \mathcal{C}(u, v) / \partial v}.$$

Afin d'illustrer la définition 1.17, prenons l'exemple de la copule d'indépendance $\mathcal{C}(u, v) = uv$. Pour trouver son générateur, il faudra résoudre l'équation suivante :

$$\frac{\varphi'(u)}{\varphi'(v)} = \frac{v}{u}.$$

Pour certaines constantes K_1, K_2 , on déduit que $\varphi'(u) = K_1/u$, et $\varphi(u) = K_1 \ln u + K_2$. Sachant que $\varphi(1) = 0$, on trouve que $K_2 = 0$, et de plus, la décroissance de φ permet de déduire que $K_1 < 0$. Ainsi, la fonction $\varphi(u) \propto -\ln u$ est un générateur de la copule d'indépendance.

Plus généralement, une copule \mathcal{C} à d dimensions est archimédienne si elle peut s'écrire sous la forme $\mathcal{C}_\varphi(u_1, \dots, u_d) = \varphi^{-1}\{\varphi(u_1) + \dots + \varphi(u_d)\}$ en termes d'un générateur $\varphi : [0, 1] \rightarrow [0, \infty)$ tel que $\varphi(1) = 0$ et pour tout $j \in \{1, \dots, d-2\}$

$$(-1)^j \frac{d^j}{dt^j} \varphi^{-1}(t) > 0.$$

Cette dernière condition implique que φ est complètement monotone. De façon générale, la densité d'une copule, si elle existe, est définie par

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

Dans le cas d'une copule archimédienne de générateur φ , on a pour $\psi = \varphi^{-1}$:

$$c_\varphi(u_1, \dots, u_d) = \frac{\psi^{(d)}\{\varphi(u_1) + \dots + \varphi(u_d)\}}{\psi'\{\varphi(u_1)\} \times \dots \times \psi'\{\varphi(u_d)\}}.$$

Ici, la fonction ψ est le pseudo-inverse du générateur φ , et $\psi^{(d)}$ désigne la dérivée d'ordre d de la fonction ψ .

Exemple 1.18

Copule de Frank

Parmi les modèles des copules archimédiennes, on retrouve la copule de Frank :

$$\mathcal{C}_\theta(u, v) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right\},$$

dont le générateur est :

$$\varphi_\theta(t) = -\ln \left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right), \quad \theta \in \mathbb{R} \setminus \{0\}.$$

Copule de Gumbel

Lorsque le générateur est $\varphi_\theta(t) = |\ln t|^{1/(1-\theta)}$, $\theta \in [0, 1]$, on obtient la copule de Gumbel-Hougaard, à savoir :

$$\mathcal{C}_\theta(u, v) = \exp \left\{ - \left(|\ln u|^{1/(1-\theta)} + |\ln v|^{1/(1-\theta)} \right)^{1-\theta} \right\}.$$

Copule de Clayton

Enfin, le modèle de Clayton découle du générateur $\varphi_\theta(t) = (t^{-\theta} - 1) / \theta$, où $\theta > 0$. La copule est alors :

$$\mathcal{C}_\theta(u, v) = \left(u^{-\theta} + v^{-\theta} - 1 \right)^{-1/\theta}.$$

1.2.9.1 Relation entre le paramètre de la copule et le tau de Kendall

Il existe une relation bijective entre le tau de Kendall et le paramètre de la copule. Par exemple, dans le cas de la copule de Clayton, cette relation s'exprime par :

$$\tau = \frac{\theta}{\theta + 2}.$$

On en déduit que :

$$\theta = \frac{2\tau}{1 - \tau}.$$

1.2.10 Copules elliptiques

Les copules elliptiques sont définies à partir des distributions elliptiques en s'appuyant sur le théorème de (Sklar, 1959). Cette famille inclut principalement les copules gaussienne et de Student. Dans la section suivante, nous détaillerons les propriétés essentielles des principales copules elliptiques.

1.2.10.1 Distribution elliptique

Considérons un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)^\top$ issu d'une distribution elliptique. Ce vecteur peut être exprimé sous la forme :

$$\mathbf{X} = \boldsymbol{\mu} + R\mathbf{A}\mathbf{U},$$

où :

- $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^\top$ est le vecteur représentant la moyenne de \mathbf{X} ,
- \mathbf{U} est un vecteur uniforme sur la sphère unité dans \mathbb{R}^d ,
- R est une variable aléatoire indépendante de \mathbf{U} et suivant une distribution arbitraire,
- \mathbf{A} est une matrice carrée de dimension d telle que $\boldsymbol{\Sigma} := \mathbf{A}\mathbf{A}^\top$, avec $\boldsymbol{\Sigma}$ une matrice de corrélation, supposée inversible.

Définition 1.19 La densité d'une distribution elliptique est définie par

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

où $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est une fonction génératrice de densité. C'est-à-dire une fonction non négative telle que f soit une densité sur \mathbb{R}^d .

1.2.10.2 Copule Gaussienne

Étant donné que la loi normale multidimensionnelle appartient à la famille des lois elliptiques, la copule normale est par définition, une copule elliptique. Par conséquent, si $\Phi_{\boldsymbol{\Sigma}}$ désigne la fonction de répartition d'une loi normale d -dimensionnelle à moyennes nulles et de matrice de corrélation $\boldsymbol{\Sigma}$, alors la copule normale peut être explicitement exprimée comme suit :

$$\mathcal{C}_{\boldsymbol{\Sigma}}(u_1, \dots, u_d) = \Phi_{\boldsymbol{\Sigma}}\left\{\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)\right\}.$$

Dans cette dernière expression, Φ^{-1} est l'inverse de la fonction de répartition de la loi normale standard.

Par exemple, la copule normale bivariée s'exprime sous la forme :

$$\mathcal{C}_{\rho}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_1)} \phi_{\rho}(x, y) dx dy,$$

où ρ désigne le coefficient de corrélation de Pearson entre les scores des variables X_1 et X_2 et ϕ_{ρ} représente la densité d'une loi normale bivariée.

Exemple 1.20 Pour $\rho \in [-1, 1]$, qui correspond au coefficient de corrélation de Pearson, nous avons

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

La densité associée à \mathcal{C}_ρ est alors :

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

De plus,

$$\begin{aligned} \mathcal{C}_\rho(u_1, u_2) &= \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left(-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)} \right) dx dy. \end{aligned}$$

Donc, la densité bivariée associée à la copule gaussienne est :

$$c_\rho(u_1, u_2) = \frac{1}{\phi(x_1)\phi(x_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left(-\frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2 - 2\rho\Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\rho^2)} \right),$$

où $\Phi^{-1}(u_1) = x_1$ et $\Phi^{-1}(u_2) = x_2$.

Tau de Kendall (Copule Gaussienne)

La relation entre le paramètre de la copule gaussienne et les mesures de dépendance (rhô de Spearman ρ_S et tau de Kendall τ) est définie comme suit :

$$\tau = \frac{2}{\pi} \arcsin \rho, \quad (1.12)$$

$$\rho_S = \frac{6}{\pi} \arcsin \frac{\rho}{2}, \quad (1.13)$$

où $\rho \in [-1, 1]$ est le paramètre de la copule Gaussienne.

1.2.10.3 Copule de Student

La copule de Student est la fonction de dépendance associée à la loi de Student multidimensionnelle. Ainsi, si $T_{\Sigma, \nu}$ désigne la fonction de répartition d'une loi de Student à ν degrés de liberté, avec une matrice de corrélation $\Sigma \in \mathbb{R}^{d \times d}$, alors la copule de Student peut être exprimée comme suit :

$$\mathcal{C}_{\Sigma, \nu}(\mathbf{u}) = \mathbf{T}_{\Sigma, \nu} \{ T_\nu^{-1}(u_1), \dots, T_\nu^{-1}(u_d) \},$$

où $\mathbf{T}_{\Sigma, \nu}$ est la fonction de répartition de la loi de Student multivariée à ν degrés de liberté. La densité associée à $\mathcal{C}_{\Sigma, \nu}$ est alors :

$$\begin{aligned} c_{\Sigma, \nu}(\mathbf{u}) &= \mathbf{t}_{\Sigma, \nu} \{T_{\nu}^{-1}(u_1), \dots, T_{\nu}^{-1}(u_d)\} \prod_{j=1}^d (T_{\nu}^{-1}(u_j))' \\ &= \mathbf{t}_{\Sigma, \nu} \{T_{\nu}^{-1}(u_1), \dots, T_{\nu}^{-1}(u_d)\} \Big/ \prod_{j=1}^d t_{\nu} \{T_{\nu}^{-1}(u_j)\}, \end{aligned}$$

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

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

où $\Gamma(z+1) = z\Gamma(z)$, avec $z \in \mathbb{R}$.

À noter que lorsque $d = 1$, la densité multivariée de Student se réduit à la densité univariée de Student, donnée par :

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

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 :

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

où :

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

Remarque 1.21 La copule de Student a le même tau de Kendall donné par l'équation (1.12).

1.2.11 Décomposition des copules par les vignes

Comme nous l'avons vu dans les sections précédentes, les copules sont des outils puissants pour modéliser des relations complexes et non linéaires entre les composantes d'un vecteur de variables aléatoires.

Cependant, dans le cas des copules archimédiennes, la dépendance entre ces composantes est souvent caractérisée par un seul paramètre θ . Dans un contexte de régression, où l'objectif est de comprendre l'effet de chaque variable explicative sur la variable réponse. Pour surmonter les limitations liées à la représentation par un seul paramètre, Joe (1996) a introduit une approche novatrice reposant sur la factorisation des distributions multivariées à l'aide de copules bivariées. Cette méthode consiste à décomposer la densité multivariée en une série de produits successifs de densités conditionnelles basées sur des copules bivariées, ouvrant ainsi la voie à une modélisation plus flexible de la dépendance en grande dimension. Par la suite, Bedford et Cooke (2001, 2002) ont approfondi cette approche en développant une construction systématique des densités, connue sous le nom de décomposition de densité par les vignes régulières. Ce cadre général permet d'identifier toutes les constructions possibles dans ce contexte. Pour mieux comprendre cette décomposition, nous allons l'illustrer dans un premier temps à travers un exemple impliquant trois variables aléatoires.

1.2.11.1 Illustration

Supposons trois variables aléatoires X_1, X_2 et X_3 . On peut exprimer leur densité trivariée de la manière suivante :

$$f_{1,2,3}(x_1, x_2, x_3) = f_{3|1,2}(x_3 | x_1, x_2) f_{2|1}(x_2 | x_1) f_1(x_1). \quad (1.14)$$

Maintenant, examinons la densité suivante $f_{3|1,2}(x_3 | x_1, x_2) = \frac{f_{1,3|2}(x_1, x_3 | x_2)}{f_{1|2}(x_1 | x_2)}$. Une application du Théorème 1.4 Sklar (1959) nous donne :

$$f_{3|1,2}(x_3 | x_1, x_2) = c_{1,3;2}(F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2); x_2) f_{3|2}(x_3 | x_2), \quad (1.15)$$

où

$$c_{1,3;2}(u_1, u_3; x_2) = \frac{\partial^2}{\partial v_1 \partial v_3} F_{1,3|2}(v_1, v_3 | x_2) \Big|_{\substack{v_1 = F_{1|2}^{-1}(u_1 | x_2) \\ v_3 = F_{3|2}^{-1}(u_3 | x_2)}} \frac{dv_1}{du_1} \frac{dv_3}{du_3},$$

et

$$f_{2|1}(x_2 | x_1) = c_{1,2}(F_1(x_1), F_2(x_2)) f_2(x_2) \quad (1.16)$$

$$f_{3|2}(x_3 | x_2) = c_{2,3}(F_2(x_2), F_3(x_3)) f_3(x_3). \quad (1.17)$$

En remplaçant les quantités (1.15), (1.16) et (1.17) dans (1.14), cela engendre une décomposition de la densité jointe en fonction des copules bivariées.

Définition 1.22 La décomposition d'une densité tridimensionnelle arbitraire, basée sur les densités de copules par paires, est exprimée par :

$$f_{1,2,3}(x_1, x_2, x_3) = c_{1,3;2}(F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2); x_2) \times c_{2,3}(F_2(x_2), F_3(x_3)) \\ \times c_{1,2}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1). \quad (1.18)$$

La décomposition introduite dans l'équation (1.18) n'est pas unique. En effet, la densité jointe $f(x_1, x_2, x_3)$ peut s'exprimer de la manière suivante :

$$f_{1,2,3}(x_1, x_2, x_3) = c_{1,2;3}(F_{1|3}(x_1 | x_3), F_{2|3}(x_2 | x_3); x_3) \times c_{1,3}(F_1(x_1), F_3(x_3)) \\ \times c_{2,3}(F_2(x_2), F_3(x_3)) f_3(x_3) f_2(x_2) f_1(x_1),$$

ou bien

$$f_{1,2,3}(x_1, x_2, x_3) = c_{2,3;1}(F_{2|1}(x_2 | x_1), F_{3|1}(x_3 | x_1); x_1) \times c_{1,3}(F_1(x_1), F_3(x_3)) \\ \times c_{1,2}(F_1(x_1), F_2(x_2)) f_3(x_3) f_2(x_2) f_1(x_1).$$

Ces deux décompositions découlent d'un réarrangement (réorganisation) des variables de l'équation (1.14).

Remarque 1.23 Afin de simplifier l'analyse, nous supposons que la copule conditionnelle $c_{1,3;2}(\cdot; x_2)$, apparaissant dans l'équation (1.15), est indépendante de la variable de conditionnement x_2 , conformément à l'hypothèse de simplification. Cette hypothèse facilite l'estimation et l'interprétation des modèles de copules en vignes, car elle permet de considérer $c_{1,3;2}(\cdot)$ comme une fonction uniquement des arguments (u_1, u_3) , indépendamment de x_2 .

Définition 1.24 (L'hypothèse de simplification (Stoeber et al., 2013)) On dit qu'une construction de copules par paires basée sur la décomposition (1.15) satisfait l'hypothèse de simplification si

$$c_{1,3;2}(u_1, u_3; x_2) = c_{1,3;2}(u_1, u_3) \text{ pour } u_1 \in [0, 1], u_3 \in [0, 1].$$

En admettant l'hypothèse de simplification, cela nous facilite la tâche pour spécifier les familles de copules bivariées. En se basant sur l'équation (1.18), les paramètres $\theta_{1,2;3}$, $\theta_{2,3}$ et $\theta_{1,2}$ représentent respectivement les paramètres arbitraires pour les copules $c_{1,2;3}$, $c_{2,3}$ et $c_{1,2}$.

Définition 1.25 Une construction de copules par paires (1.18) est paramétrée par un vecteur de dimension trois $\theta = (\theta_{12}, \theta_{23}, \theta_{12;3})$. L'équation (1.18) devient :

$$\begin{aligned} f_{1,2,3}(x_1, x_2, x_3; \theta) &:= c_{1,3;2}(F_{1|2}(x_1 | x_2), F_{3|2}(x_3 | x_2); \theta_{1,3;2}) \times c_{2,3}(F_2(x_2), F_3(x_3); \theta_{2,3}) \\ &\times c_{1,2}(F_1(x_1), F_2(x_2); \theta_{1,2}) f_3(x_3) f_2(x_2) f_1(x_1). \end{aligned} \quad (1.19)$$

1.2.11.2 Construction de copules par paires sous la forme D-vigne et C-vigne

Jusqu'à présent, nous avons présenté une construction de copules par paires pour modéliser une densité tridimensionnelle. Dans cette section, nous introduisons les décompositions C-vigne et D-vigne, qui offrent une méthode systématique pour construire des copules par paires. Pour illustrer ces deux approches, considérons un vecteur de variables aléatoires (X_1, \dots, X_d) , doté d'une fonction de répartition $\mathbf{F}_{1,\dots,d}$ et d'une densité $f_{1,\dots,d}$. La densité multivariée de ce vecteur peut être formulée comme suit :

$$\begin{aligned} f_{1,\dots,d}(x_1, \dots, x_d) &= f_{d|1,\dots,d-1}(x_d | x_1, \dots, x_{d-1}) f_{1,\dots,d-1}(x_1, \dots, x_{d-1}) \\ &= f_{d|1,\dots,d-1}(x_d | x_1, \dots, x_{d-1}) f_{d-1|1,\dots,d-2}(x_{d-1} | x_1, \dots, x_{d-2}) \\ &\quad f_{1,\dots,d-2}(x_1, \dots, x_{d-2}) \\ &= \dots = \left[\prod_{t=2}^d f_{t|1,\dots,t-1}(x_t | x_1, \dots, x_{t-1}) \right] \times f_1(x_1), \end{aligned} \quad (1.20)$$

avec $f(\cdot | \cdot)$ et $F(\cdot | \cdot)$ représentant respectivement les densités et les fonctions de répartition conditionnelles des variables. Afin de simplifier notre écriture, nous introduisons quelques notations. Par la suite, nous utiliserons l'équation (1.20) pour trouver les décompositions C-vigne et D-vigne.

Définition 1.26 (Copules associées aux distributions conditionnelles bivariées)

- Supposons que (X_1, \dots, X_d) soit un ensemble de variables aléatoires, et D l'ensemble des indices de $\{1, \dots, d\}$ sans inclure i et j . La copule associée aux distributions conditionnelles bivariées (X_i, X_j) , sachant l'ensemble $\mathbf{X}_D = \mathbf{x}_D$, est définie par $\mathcal{C}_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$.
- En revanche, la distribution conditionnelle du couple (U_i, U_j) étant donnée $\mathbf{U}_D = \mathbf{u}_D$ est définie par $\mathcal{C}_{i,j|D}(\cdot, \cdot; \mathbf{u}_D)$ avec la densité $c_{i,j|D}(\cdot, \cdot; \mathbf{u}_D)$.
- Pour des indices distincts i, j et $D := \{i_1, \dots, i_k\}$, où $|D| = d - 2$, avec $i < j$ et $i_1 < \dots < i_k$,

on utilise l'abréviation suivante :

$$c_{i,j;D} := c_{i,j;D} (F_{i|D} (x_i | \mathbf{x}_D), F_{j|D} (x_j | \mathbf{x}_D); \mathbf{x}_D). \quad (1.21)$$

Remarque 1.27 (Différence entre $C_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$ et $C_{i,j|D}(\cdot, \cdot; \mathbf{u}_D)$)

En général, la copule conditionnelle $C_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$, associée à la distribution jointe conditionnelle de (X_i, X_j) sachant $\mathbf{X}_D = \mathbf{x}_D$, n'est pas équivalente à la fonction de répartition conditionnelle bivariée $C_{i,j|D}(\cdot, \cdot; \mathbf{u}_D)$, même lorsque $\mathbf{u}_D = \mathbf{F}_D(\mathbf{x}_D)$.

Cette distinction provient du fait que $C_{i,j|D}$ peut présenter des marginales non uniformes, alors que $C_{i,j;D}$ doit, par définition, satisfaire la propriété fondamentale des copules imposant des marginales uniformes sur $[0, 1]$.

Cette différence conceptuelle est mise en évidence dans l'exemple qui suit.

Exemple 1.28 Considérons l'extension trivariée de la copule de Frank définie par :

$$C_{1,2,3}(u_1, u_2, u_3; \delta) = \frac{1}{\ln(\alpha)} \ln \left[1 - (1 - \alpha) \prod_{i=1}^3 \frac{1 - \alpha^{u_i}}{1 - \alpha} \right], \quad \text{où } \alpha = e^{-\delta}, \quad \delta \in \mathbb{R} \setminus \{0\}.$$

L'objectif est de déterminer la fonction de répartition conditionnelle du couple (U_1, U_3) sachant $U_2 = v_2$.

Par dérivation directe, on obtient :

$$C_{1,3|2}(u_1, u_3|v_2; \delta) = \frac{\partial}{\partial u_2} C_{1,2,3}(u_1, u_2, u_3; \delta) \Big|_{u_2=v_2} = \frac{-\alpha^{v_2}(1 - \alpha^{u_1})(1 - \alpha^{u_3})}{(1 - \alpha)^2 - \prod_{i=1}^3 (1 - \alpha^{u_i}) \Big|_{u_2=v_2}}.$$

Pour obtenir la copule conditionnelle associée $C_{13;2}(u_1, u_3|u_2 = v_2; \delta)$, il est nécessaire de déterminer les marges suivantes :

$$C_{1|2}(u_1|v_2; \delta) = \frac{\partial}{\partial u_2} C_{1,2}(u_1, u_2; \delta) \Big|_{u_2=v_2} = \frac{-\alpha^{v_2}(1 - \alpha^{u_1})}{(1 - \alpha) - (1 - \alpha^{u_1})(1 - \alpha^{v_2})},$$

et de manière analogue pour $C_{3|2}(u_3|v_2; \delta)$.

En inversant ces marges par la relation $p_i = C_{i|2}(u_i|v_2; \delta)$, $i = 1, 3$, on exprime la copule conditionnelle :

$$C_{1,3;2}(p_1, p_3|v_2; \delta) = \frac{p_1 p_3}{(1 - \alpha^{v_2})[1 - (1 - p_1)(1 - p_3)] - \alpha^{v_2}}.$$

L'exemple 1.28 illustre concrètement la Remarque 1.27 concernant la distinction entre $\mathcal{C}_{i,j;D}(\cdot, \cdot; \mathbf{x}_D)$ et $\mathcal{C}_{i,j|D}(\cdot, \cdot; \mathbf{u}_D)$.

En effet, à partir de l'extension trivariée de la copule de Frank, on dérive d'une part la fonction de répartition conditionnelle $C_{13|2}(u_1, u_3 | v_2)$ et, d'autre part, la copule conditionnelle $\mathcal{C}_{1,3;2}(p_1, p_3 | v_2)$, après un changement de variables tenant compte des marges conditionnelles.

Cela met en évidence que $\mathcal{C}_{1,3|2}$ n'est pas une copule au sens strict, car ses marginales ne sont pas uniformes, tandis que $\mathcal{C}_{1,3;2}$, obtenue par normalisation, satisfait la définition d'une copule.

Ainsi, cet exemple illustre la nécessité de distinguer soigneusement entre fonction de répartition conditionnelle et copule conditionnelle.

Après avoir défini toutes les quantités nécessaires à la décomposition d'une densité multivariée basée sur la représentation en vigne à l'aide des copules, nous introduisons maintenant les deux décompositions utilisées.

Théorème 1.29 (Densité basée sur D-vigne (Drawable vine)) *Toute densité $f_{1,\dots,d}$ s'écrit de la manière suivante :*

$$f_{1,\dots,d}(x_1, \dots, x_d) = \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j);(i+1),\dots,(i+j-1)} \right] \times \left[\prod_{k=1}^d f_k(x_k) \right]. \quad (1.22)$$

La distribution associée à la décomposition de densité précédente est appelée D-vigne. Nous avons utilisé l'abréviation de l'équation (1.21) afin d'exprimer les densités conditionnelles bivariées.

Preuve. En appliquant le Lemme 1.10 aux distributions conditionnelles (X_1, X_t) , sachant X_2, \dots, X_{t-1} , de manière récursive sur $f_{t|1,\dots,t-1}(x_t | x_1, \dots, x_{t-1})$, nous obtenons :

$$\begin{aligned} f_{t|1,\dots,t-1}(x_t | x_1, \dots, x_{t-1}) &= c_{1,t|2,\dots,t-1} \times f_{t|2,\dots,t-1}(x_t | x_2, \dots, x_{t-1}) \\ &= \left[\prod_{s=1}^{t-2} c_{s,t;s+1,\dots,t-1} \right] \times c_{(t-1),t} \times f_t(x_t). \end{aligned} \quad (1.23)$$

En utilisant (1.23) dans la décomposition donnée dans (1.20) et en posant $s = i, t = i + j$, nous obtenons :

$$\begin{aligned} f_{1,\dots,d}(x_1, \dots, x_d) &= \left[\prod_{t=2}^d \prod_{s=1}^{t-2} c_{s,t;s+1,\dots,t-1} \right] \times \left[\prod_{t=2}^d c_{(t-1),t} \right] \times \left[\prod_{k=1}^d f_k(x_k) \right] \\ &= \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,(i+j);(i+1),\dots,(i+j-1)} \right] \times \left[\prod_{k=1}^d f_k(x_k) \right]. \end{aligned} \quad (1.24)$$

□

Remarque 1.30 La décomposition (1.22) de la densité jointe est composée de densités de copules par paires $c_{i,j|D}(\cdot, \cdot; \mathbf{x}_D)$ évaluées aux marges conditionnelles $F_{i|D}(x_i | x_{i_1}, \dots, x_{i_k})$ et $F_{j|D}(x_j | x_{i_1}, \dots, x_{i_k})$ pour des indices prédéfinis i, j, i_1, \dots, i_k et les densités f_k . C'est pour cette raison que l'on appelle cette décomposition des copules par paires. L'appellation a été donnée par Bedford et Cooke (2001, 2002).

Une autre décomposition est possible si l'on commence par la distribution conditionnelle des dernières et avant-dernières variables, que nous définirons dans la prochaine définition.

Théorème 1.31 (Densité basée sur C-vigne (Canonical-vine)) Pour la décomposition C-vigne, la densité multivariée s'exprime de la manière suivante :

$$f_{1,\dots,d}(x_1, \dots, x_d) = \left[\prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i;1,\dots,j-1} \right] \times \left[\prod_{k=1}^d f_k(x_k) \right]. \quad (1.25)$$

Nous avons utilisé l'abréviation de l'équation (1.21) afin d'exprimer les densités conditionnelles bivariées. D'après Bedford et Cooke (2002), cette décomposition en paires de copules est appelée la C-vigne.

Preuve. Pour la preuve, nous suivons le même raisonnement que celui utilisé dans la preuve du Théorème 1.29. Appliquons le Lemme 1.10 à la distribution conditionnelle de (X_{t-1}, X_t) , donnée X_1, \dots, X_{t-2} , pour exprimer $f_{t|1,\dots,t-1}(x_t | x_1, \dots, x_{t-1})$ de manière récursive. \square

Exemple 1.32 (Densité simplifiée de D-vigne et C-vigne d=4) La densité simplifiée D-vigne a la forme suivante :

$$\begin{aligned} f_{1,2,3,4}(x_1, x_2, x_3, x_4) &= \left[\prod_{i=1}^4 f_i(x_i) \right] \times c_{1,2}(x_1, x_2) \times c_{2,3}(x_2, x_3) \times c_{3,4}(x_3, x_4) \\ &\times c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \times c_{2,4;3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)) \\ &\times c_{1,4;2,3}(F_{1|2,3}(x_1|x_2, x_3), F_{4|2,3}(x_4|x_2, x_3)). \end{aligned}$$

Pour la C-vigne simplifiée, nous avons :

$$\begin{aligned} f_{1,2,3,4}(x_1, x_2, x_3, x_4) &= \left[\prod_{i=1}^4 f_i(x_i) \right] \times c_{12}(x_1, x_2) \times c_{1,3}(x_1, x_3) \times c_{1,4}(x_1, x_4) \\ &\times c_{2,3;1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \times c_{2,4;1}(F_{2|1}(x_2|x_1), F_{4|1}(x_4|x_1)) \\ &\times c_{3,4;1,2}(F_{3|1,2}(x_3|x_1, x_2), F_{4|1,2}(x_4|x_1, x_2)). \end{aligned}$$

Jusqu'à présent, nous avons donné une définition de la décomposition d'une densité jointe par une structure de dépendance de type vigne et de ses deux décompositions les plus utilisées. Dans la section suivante, nous étudions l'estimation des paramètres des copules correspondantes pour les copules par paires dans une structure de vigne donnée.

1.2.12 Estimation des paramètres dans une décomposition vigne

L'estimation des paramètres de dépendances dans une décomposition en vignes se base sur une procédure séquentielle de maximisation de la vraisemblance, comme le décrit le travail de Dissmann *et al.* (2013). Cette approche est essentielle pour rendre les modèles de vigne opérationnels et appropriés à l'analyse de données multivariées complexes. Dans la section suivante, nous présenterons en détail la méthode d'estimation des paramètres dans une décomposition en vignes.

1.2.12.1 Décomposition vigne par les arbres

Bedford et Cooke (2002) ont introduit la *vigne régulière* (R-vine), c'est-à-dire une factorisation d'une copule multivariée en copules de paires organisées selon une hiérarchie d'arbres imbriqués. Soit donc une suite d'arbres

$$T_i = (N_i, E_i), \quad i = 1, \dots, d - 1,$$

où

- N_i est l'ensemble des **nœuds** de T_i .
- E_i est l'ensemble des **arêtes** (non orientées) de T_i .
- T_1 porte les indices des variables $\{1, \dots, d\}$, et pour $i \geq 2$ l'ensemble des nœuds N_i est, par construction, égal à E_{i-1} (chaque *nœud* de T_i est donc une arête du niveau précédent).

À chaque arête $e \in E_i$ on associe trois ensembles

$$e = (a_e, b_e; D_e),$$

a_e, b_e ensembles conditionnés (singletons dans T_1 , puis arêtes aux niveaux supérieurs);

D_e ensemble de conditionnement (intersection des *complete unions* de a_e et b_e , de cardinal $|D_e| = i - 1$).

À cet ensemble, on associe une copule bivariée.

$$\mathcal{C}_{a_e, b_e; D_e} : [0, 1]^2 \longrightarrow [0, 1],$$

et de densité $c_{a_e, b_e; D_e}$ et (le cas échéant) de paramètre θ_e . Ainsi, pour tout vecteur $\mathbf{x} = (x_1, x_2, \dots, x_d)$, la densité de la copule conjointe peut s'écrire comme suit :

$$c(\mathbf{u}) = \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{i_e, j_e; D_e} \left(\mathcal{C} \left(u_{i_e} \mid \mathbf{u}_{D_e} \right), \mathcal{C} \left(u_{j_e} \mid \mathbf{u}_{D_e} \right); \mathbf{u}_{D_e} \right), \quad (1.26)$$

où $c(\mathbf{u})$ est la densité de copule associée à la distribution conditionnelle de $(u_{i_e}, u_{j_e}; D_e) \mid \mathbf{u}_{D_e}$, évaluée à $\mathcal{C} \left(u_{i_e} \mid \mathbf{u}_{D_e} \right)$ et $\mathcal{C} \left(u_{j_e} \mid \mathbf{u}_{D_e} \right)$ représentant l'intégrale de la distribution conditionnelle de la copule de $u_{i_e} \mid \mathbf{u}_{D_e}$ et $u_{j_e} \mid \mathbf{u}_{D_e}$.

Dans le cadre d'un modèle de copule R-vigne simplifiée, la densité de vraisemblance est exprimée comme une combinaison de copules bivariées associées aux différentes paires de variables. La vraisemblance totale correspond au produit des densités de copules bivariées conditionnelles, conformément à l'équation (1.26), et est formalisée dans la définition suivante.

Définition 1.33 (Vraisemblance d'une copule de vigne régulière simplifiée)

La vraisemblance d'une copule de vigne régulière simplifiée avec des paramètres de copule $\theta = \{\theta_e, e \in E = \bigsqcup_{i=1}^{d-1} E_i\}$ et des données observées u est définie comme

$$\ell(\theta; \mathbf{u}) = \prod_{k=1}^n \prod_{j=1}^{d-1} \prod_{e \in E_j} c_{a_e, b_e; D_e} \left(\mathcal{C}_{a_e | D_e}(u_{k, a_e} | u_{k, D_e}), \mathcal{C}_{b_e | D_e}(u_{k, b_e} | u_{k, D_e}) \right). \quad (1.27)$$

Exemple 1.34 (Vraisemblance d'une copule paramétrique D-vine et C-vine)

Nous supposons que toutes les familles de copules par paires nécessaires pour une copule C-vine ou D-vine sont paramétriques. En particulier pour une D-vine, le(s) paramètre(s) de la copule par paire $\mathcal{C}_{i, i+j; i+1: i+j-1}$ sont notés $\theta_{i, i+j; i+1: i+j-1}$ pour $j = 1, \dots, d-1$ et $i = 1, \dots, d-j$, et tous les paramètres sont regroupés dans θ . De même, nous utilisons $\theta_{j, i+j; i-1: j-1}$ pour le(s) paramètre(s) de la copule par paire $\mathcal{C}_{j, i+j; i-1: j-1}$ et regroupons à nouveau tous les paramètres dans θ . En utilisant (1.22) et (1.25), respectivement, la vraisemblance de la copule D-vigne est donnée par

$$\ell(\theta, \mathbf{u}) = \prod_{k=1}^n \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i, i+j; i+1: i+j-1} \left(\mathcal{C}_{i | i+1: i+j-1}(u_{k, i} | \mathbf{u}_{k, i+1: i+j-1}), \mathcal{C}_{i+j | i+1: i+j-1}(u_{k, i+j} | \mathbf{u}_{k, i+1: i+j-1}) \right),$$

tandis que la vraisemblance d'une copule C-vigne peut être écrite comme

$$\ell(\theta, \mathbf{u}) = \prod_{k=1}^n \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j, i+j; i: j-1} \left(\mathcal{C}_{i | i+1: i+j-1}(u_{k, i} | \mathbf{u}_{k, 1: j-1}), \mathcal{C}_{i+j | i: j-1}(u_{k, i+j} | \mathbf{u}_{k, 1: j-1}) \right).$$

En particulier pour les copules D-vine, la fonction de distribution conditionnelle $C_{i|i+1:i+j-1}$ dépend d'abord de $\theta_{i,i+j-1;i+j-2}$, tandis que $C_{i+j|i+1:i+j-1}$ de $\theta_{i+2,i+j-1;i+j-1}$. Cela peut être utilisé de manière récursive pour déterminer la dépendance complète des paramètres, ce qui est expliqué dans la sous-section suivante.

1.2.12.2 Estimation Séquentielle

Avant de donner notre définition de l'estimation séquentielle introduite par Dissmann *et al.* (2013), nous allons introduire quelques notations que nous allons utiliser :

- Soit θ_e le(s) paramètre(s) de copule correspondant à l'arête $e = (a_e, b_e; D_e)$ dans une séquence de vignes régulières dans l'arbre T_i .
- Les paramètres de copule associés aux arêtes dans l'arbre T_i sont notés par $\theta(T_i)$ et leurs estimations par $\hat{\theta}(T_i)$.
- Supposons que tous les paramètres de copules par paires identifiés dans les arbres T_1 à T_{i-1} sont déjà estimés. Nous désignons cette collection d'estimations de paramètres par $\hat{\theta}(T_1, \dots, i-1)$.

Définition 1.35 L'estimation séquentielle de θ_e pour l'arête $e = (a_e, b_e; D_e)$ dans l'arbre T_i est basée sur les pseudo-observations

$$u_{k,a_e|D_e,\hat{\theta}(T_1,\dots,i-1)} := C_{a_e|D_e} \left(u_{k,a_e} | u_{k,D_e}, \hat{\theta}(T_1,\dots,i-1) \right)$$

$$u_{k,b_e|D_e,\hat{\theta}(T_1,\dots,i-1)} := C_{b_e|D_e} \left(u_{k,b_e} | u_{k,D_e}, \hat{\theta}(T_1,\dots,i-1) \right),$$

pour $k = 1, \dots, n$. En particulier, $\hat{\theta}_e$ est estimé en maximisant

$$\prod_{k=1}^n C_{a_e,b_e;D_e} \left(u_{k,a_e|D_e,\hat{\theta}(T_1,\dots,i-1)}, u_{k,b_e|D_e,\hat{\theta}(T_1,\dots,i-1)}; \theta_e \right),$$

ou par inversion du τ de Kendall empirique basé sur les pseudo-observations.

L'approche d'estimation séquentielle des paramètres d'une décomposition en vignes, telle que décrite par Dissmann *et al.* (2013), repose principalement sur l'estimation des copules bivariées des marges, lesquelles sont ensuite utilisées pour calculer les pseudo-observations nécessaires aux étapes suivantes. Pour illustrer cette méthode, rappelons la décomposition définie par l'équation (1.19). Les étapes à suivre pour estimer les paramètres de la décomposition en vignes, spécifiée par l'équation (1.19), sont les suivantes :

- **Illustration pour Dimensions $d = 3$**

1. Première Étape :

- Estimez les paramètres des copules bivariées des marges simples $(u_{k,1}, u_{k,2})$ et $(u_{k,2}, u_{k,3})$ en utilisant la méthode du maximum de vraisemblance ou l'inversion de τ de Kendall. Cela donne les estimations $\hat{\theta}_{12}$ et $\hat{\theta}_{23}$.

2. Calcul des Pseudo-Observations :

- Utilisez les estimations pour calculer les pseudo-observations :

$$u_{k,1|2,\hat{\theta}_{12}} = C_{1|2}(u_{k,1}|u_{k,2}, \hat{\theta}_{12})$$

$$u_{k,3|2,\hat{\theta}_{23}} = C_{3|2}(u_{k,3}|u_{k,2}, \hat{\theta}_{23}).$$

3. Estimation des Copules Conditionnelles :

- Les pseudo-observations sont ensuite utilisées pour estimer les paramètres des copules conditionnelles $(u_{k,1|2,\hat{\theta}_{12}}, u_{k,3|2,\hat{\theta}_{23}})$.

— Extension aux Dimensions Supérieures $d \geq 4$:

- La procédure décrite pour $d = 3$ s'étend aux copules de paires conditionnelles en dimensions supérieures. Pour chaque arbre de la structure R-vine, les paramètres sont estimés itérativement à l'aide de pseudo-observations dérivées des arbres antérieurs de la séquence.

La figure suivante illustre les étapes d'estimation des paramètres des copules dans le cadre d'une décomposition R-vine pour une densité à d variables.

Algorithm 1 Estimation Séquentielle des Paramètres dans une décomposition de vigne régulière

Require: Matrice de données X (n observations \times d variables)

Ensure: Modèle Regular Vine Copula estimé avec structure et paramètres sélectionnés

- 1: **Étape 1 : Transformation des Marges**
 - 2: $U \leftarrow$ Appliquer la fonction de distribution empirique à chaque colonne de X pour obtenir des variables uniformes sur $[0, 1]$
 - 3: **Étape 2 : Sélection de la Structure de la Vine**
 - 4: vine_structure \leftarrow Sélectionner la structure de la vine en maximisant la dépendance (par exemple, information mutuelle)
 - 5: **Étape 3 : Sélection et Estimation des Copulas**
 - 6: **for** chaque arbre T_i dans vine_structure **do**
 - 7: **for** chaque arête $e = (a_e, b_e; D_e)$ dans T_i **do**
 - 8: Calculer les pseudo-observations conditionnelles :
 - 9: $u_{k,a_e|D_e} \leftarrow C_{a_e|D_e} \left(u_{k,a_e} \mid u_{k,D_e}, \hat{\theta}(T_1, \dots, T_{i-1}) \right)$ $\triangleright \forall k = 1, \dots, n$
 - 10: $u_{k,b_e|D_e} \leftarrow C_{b_e|D_e} \left(u_{k,b_e} \mid u_{k,D_e}, \hat{\theta}(T_1, \dots, T_{i-1}) \right)$ $\triangleright \forall k = 1, \dots, n$
 - 11: Sélectionner la famille de copula appropriée pour l'arête e (par exemple, Gaussian, t, Clayton, Gumbel, etc.)
 - 12: Estimer les paramètres $\hat{\theta}_e$ de la copula $C_{a_e,b_e;D_e}$ en maximisant la vraisemblance :
 - 13: $\hat{\theta}_e = \arg \max_{\theta_e} \prod_{k=1}^n C_{a_e,b_e;D_e} \left(u_{k,a_e|D_e}, u_{k,b_e|D_e}; \theta_e \right)$ \triangleright Ou par inversion du τ de Kendall empirique
 - 14: Stocker la copula et ses paramètres dans vine_structure
 - 15: **end for**
 - 16: **end for**
 - 17: **Étape 4 : Validation du Modèle**
 - 18: Effectuer des diagnostics pour vérifier la qualité de l'ajustement du modèle (tests de vraisemblance, graphiques Q-Q, etc.)
 - 19: **if** le modèle n'est pas adéquat **then**
 - 20: Réviser la sélection des copulas ou la structure de la vine
 - 21: **end if**
 - 22: **Étape 5 : Application du Modèle**
 - 23: Utiliser le modèle estimé pour des applications telles que la gestion du risque, la diversification de portefeuille ou la simulation de scénarios
-

Figure 1.1 – Estimation séquentielle des paramètres dans une décomposition de vigne régulière

Jusqu'à présent, nous avons introduit les outils nécessaires utilisés dans les manuscrits développés au cours de cette thèse. Il reste toutefois un outil fondamental, utilisé dans le manuscrit présenté au Chapitre 4 : les fonctions de perte asymétriques. La section suivante en donne une définition et présente leurs propriétés mathématiques.

1.3 Fonctions de perte asymétriques utilisées : Huber-type, pseudo-Huber, expectile

1.3.1 Introduction

Cette section présente, dans un cadre général, les fonctions de perte asymétriques considérées dans le manuscrit, à savoir : (i) la perte de Huber-type (piècewise quadratique-linéaire), (ii) la perte pseudo-Huber (approximation lisse de la Huber), (iii) la perte expectile (moindres carrés asymétriques), ainsi que (iv) une extension Huber-type définie par une combinaison *quadratique/linéaire* via un opérateur \min . L'objectif est de rappeler leurs définitions en fonction d'un résidu général $u \in \mathbb{R}$, puis d'établir leurs propriétés mathématiques essentielles : non-négativité, convexité, régularité (dérivabilité, classe C^1 , non- C^2), et comportement asymptotique (robustesse via saturation de la dérivée). Les propriétés présentées ci-dessous servent ensuite à justifier l'existence, l'unicité et la stabilité numérique des solutions obtenues dans le manuscrit.

1.3.2 Notations et pondération asymétrique

Soit $u \in \mathbb{R}$ un résidu (par exemple $u = y - \hat{y}$). Une fonction de perte est une application $L : \mathbb{R} \rightarrow \mathbb{R}$, généralement telle que $L(u) \geq 0$ et $L(0) = 0$. L'asymétrie est contrôlée par $\tau \in (0, 1)$ au moyen du poids

$$w_\tau(u) = |\tau - \mathbb{1}_{\{u < 0\}}| = \begin{cases} \tau, & u \geq 0, \\ 1 - \tau, & u < 0. \end{cases}$$

Dans ce cadre, on construit souvent une perte asymétrique à partir d'une fonction de base $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$ via

$$\rho_\tau(u) = w_\tau(u) \rho(u).$$

Lorsque $\tau > 1/2$, la pénalité associée aux résidus positifs ($u > 0$) est plus forte; lorsque $\tau < 1/2$, ce sont les résidus négatifs ($u < 0$) qui sont davantage pénalisés.

1.3.3 Perte de Huber-type (piècewise quadratique-linéaire)

Dans cette sous-section, nous présentons la fonction huber introduit par Huber (1964) la definition est présenté dans la definition suivante :

Definition 1.36 Pour un seuil $k > 0$, la fonction de Huber-type utilisée dans le manuscrit peut s'écrire

$$\rho(u) = \begin{cases} \frac{1}{2k} u^2, & |u| \leq k, \\ |u| - \frac{k}{2}, & |u| > k. \end{cases} \quad (k > 0),$$

et la version asymétrique associée est

$$\rho_\tau(u) = w_\tau(u) \rho(u).$$

Cette perte est quadratique au voisinage de 0 (petites erreurs) et devient linéaire dans les queues (grosses erreurs), ce qui induit une robustesse face aux observations influentes (outliers).

1.3.3.1 Convexité, continuité et minimum

La fonction ρ_k est continue sur \mathbb{R} (raccord en $|u| = k$) et convexe (car quadratique convexe sur $[-k, k]$ et affine en $|u|$ hors de cet intervalle, avec pentes compatibles au raccord). De plus, $\rho_k(u) \geq 0$ et $\rho_k(0) = 0$, donc ρ_k admet un minimum en $u = 0$. Comme $w_\tau(u) > 0$ pour $\tau \in (0, 1)$, les mêmes propriétés valent pour $\rho_{\tau,k}$.

1.3.3.2 Dérivabilité et fonction d'influence

La dérivée première (au sens classique) de ρ_k est, pour $u \neq \pm k$,

$$\rho'(u) = \begin{cases} \frac{u}{k}, & |u| < k, \\ \text{sign}(u), & |u| > k. \end{cases}$$

La fonction d'influence (ou score) associée est

$$\psi_\tau(u) = \frac{d}{du} \rho_\tau(u) = w_\tau(u) \rho'(u),$$

soit explicitement

$$\psi_\tau(u) = \begin{cases} \tau \frac{u}{k}, & 0 < u < k, \\ \tau, & u > k, \\ (1 - \tau) \frac{u}{k}, & -k < u < 0, \\ -(1 - \tau), & u < -k. \end{cases}$$

En particulier, ψ_τ est bornée :

$$|\psi_{\tau,k}(u)| \leq \max(\tau, 1 - \tau),$$

ce qui traduit la robustesse de la perte de Huber-type (saturation de l'influence).

1.3.3.3 Seconde dérivée et non- C^2

La dérivée seconde existe pour tout $u \neq \pm k$ et vaut

$$\rho''(u) = \begin{cases} \frac{1}{k}, & |u| < k, \\ 0, & |u| > k. \end{cases} \Rightarrow \rho''_\tau(u) = w_\tau(u)\rho''(u) \quad \text{pour } u \neq 0, \pm k.$$

Ainsi, ρ_τ est de classe C^1 mais n'est pas C^2 (discontinuité de ρ'' aux points $\pm k$). Néanmoins, comme ces points sont de mesure nulle, ρ_τ est deux fois dérivable *presque partout*, ce qui suffit dans de nombreux arguments asymptotiques (dérivée faible au sens de Sobolev/distributions).

1.3.4 Fonction de perte pseudo-Huber

Dans cette sous-section, nous présentons un autre type de fonction de perte de type Huber, appelée pseudo-Huber

Definition 1.37 (Perte pseudo-Huber (approximation lisse de Huber)) *la pseudo-Huber (sans paramètre de seuil) est définie par :*

$$\rho(u) = \sqrt{1 + u^2} - 1, \quad \rho_\tau(u) = w_\tau(u) \rho(u).$$

Près de 0, $\rho(u) \approx \frac{1}{2}u^2$; pour $|u|$ grand, $\rho(u) \approx |u| - 1$. Elle reproduit donc le schéma quadratique/linéaire, mais de manière lisse.

1.3.4.1 Dérivées, régularité et convexité de la fonction Pseudo Huber

Les dérivées s'écrivent

$$\rho'(u) = \frac{u}{\sqrt{1 + u^2}}, \quad \rho''(u) = \frac{1}{(1 + u^2)^{3/2}} > 0.$$

Donc ρ est strictement convexe et $\rho_{PH} \in C^\infty(\mathbb{R})$. La fonction d'influence associée est

$$\psi_\tau(u) = \frac{d}{du} \rho_\tau(u) = w_\tau(u) \frac{u}{\sqrt{1 + u^2}}.$$

Comme $\left| \frac{u}{\sqrt{1 + u^2}} \right| < 1$, on obtient la borne

$$|\psi_{\tau,PH}(u)| \leq \max(\tau, 1 - \tau),$$

ce qui conserve la robustesse (influence bornée) tout en assurant une grande lissité. Noter que $\rho''_{PH}(u) \rightarrow 0$ quand $|u| \rightarrow \infty$, donc la forte convexité n'est pas globale sur \mathbb{R} (mais elle tient sur tout compact).

1.3.5 Perte expectile (moindres carrés asymétriques)

Cette section présente une fonction de perte fréquemment utilisée dans la littérature : la fonction de perte des expectiles (Newey et Powell, 1987). Elle permet d'estimer les expectiles d'une distribution. En d'autres termes, l'expectile généralise la moyenne d'une distribution de la même manière que le quantile (Koenker et Bassett, 1978) généralise la médiane. De plus, les expectiles d'une distribution sont obtenus comme les minimiseurs de la fonction de perte définie ci-dessous.

Definition 1.38 (Perte expectile) *La perte expectile est la perte quadratique asymétrique*

$$\rho_\tau(u) = w_\tau(u) u^2 = \begin{cases} \tau u^2, & u \geq 0, \\ (1 - \tau) u^2, & u < 0, \end{cases} \quad \tau \in (0, 1).$$

Elle généralise la moyenne de manière analogue à la façon dont les quantiles généralisent la médiane, mais avec une fonction objective lisse et quadratique.

Les expectiles sont largement utilisés en régression en raison de leurs propriétés théoriques et de leur interprétation en termes de minimisation d'une perte quadratique asymétrique. Nous présentons donc ci-dessous une définition formelle de l'expectile conditionnel associé à la loi de $Y \mid \mathbf{X} = \mathbf{x}$.

Definition 1.39 *Soient $Y \in \mathcal{Y}$ une variable reponse, où \mathcal{Y} represents le domaine de definition de la variable reponse, et $\mathbf{X} \in \mathbb{R}^p$ un vecteur de covariables. Pour un niveau $\tau \in (0, 1)$, l'expectile conditionnel $\mu_\tau(\mathbf{X})$ est défini comme la solution du problème de minimisation conditionnelle :*

$$\mu_\tau(\mathbf{x}) = \arg \min_{m \in \mathcal{Y}} \mathbb{E}[\rho_\tau(Y - m) \mid \mathbf{X} = \mathbf{x}].$$

Ceci équivaut à résoudre l'équation des moments conditionnels suivante pour presque toute réalisation \mathbf{x} :

$$\tau \mathbb{E}[(Y - \mu_\tau(\mathbf{x}))_+ \mid \mathbf{X} = \mathbf{x}] = (1 - \tau) \mathbb{E}[(\mu_\tau(\mathbf{x}) - Y)_+ \mid \mathbf{X} = \mathbf{x}], \quad (1.28)$$

où $(u)_+ = \max(u, 0)$.

Dans l'équation (1.28) $\mu_\tau(\mathbf{x})$ représente une mesure de localisation conditionnelle asymétrique, pondérant différemment les écarts positifs et négatifs d'autre mots l'expectile de la distribution conditionnelle $Y \mid \mathbf{X} = \mathbf{x}$.

1.3.5.1 Convexité, forte convexité, dérivabilité

La fonction ρ_τ^{ex} est convexe et même fortement convexe, car

$$\rho_\tau''(u) = \begin{cases} 2\tau, & u > 0, \\ 2(1-\tau), & u < 0, \end{cases} \Rightarrow (\rho_\tau)''(u) \geq 2 \min(\tau, 1-\tau) > 0$$

La dérivée première (fonction d'influence) vaut

$$\rho_\tau'(u) = \begin{cases} 2\tau u, & u > 0, \\ 2(1-\tau)u, & u < 0, \\ 0, & u = 0, \end{cases}$$

ce qui montre que ρ_τ est C^1 (la dérivée est continue en 0), mais pas C^2 en général (saut de la courbure en 0 si $\tau \neq 1/2$). Contrairement à Huber/pseudo-Huber, l'influence n'est pas bornée d'où une robustesse plus faible aux valeurs extrêmes, mais une excellente stabilité d'optimisation grâce à la forte convexité.

1.3.6 Extension Huber-type via un minimum

Dans notre manuscrit introduit dans le Chapitre 4 propose une forme combinant une partie quadratique et une partie linéaire via

$$\rho_\tau(u) = w_\tau(u) \min\left(\frac{u^2}{2}, |u| - \frac{1}{2}\right).$$

On remarque toutefois que, pour tout $u \in \mathbb{R}$,

$$\frac{u^2}{2} - \left(|u| - \frac{1}{2}\right) = \frac{(|u| - 1)^2}{2} \geq 0,$$

donc

$$|u| - \frac{1}{2} \leq \frac{u^2}{2} \quad \text{pour tout } u,$$

ce qui implique

$$\min\left(\frac{u^2}{2}, |u| - \frac{1}{2}\right) = |u| - \frac{1}{2} \quad (\text{partout, avec égalité au point } |u| = 1).$$

Ainsi, telle qu'écrite, cette fonction se réduit à une expression essentiellement linéaire (décalée), et peut prendre des valeurs négatives pour $|u| < 1/2$. Si l'intention recherchée est "quadratique près de 0 et linéaire au-delà d'un seuil", la forme standard correspondante est plutôt une définition *piecewise* (Huber à

seuil 1) :

$$\rho(u) = \begin{cases} \frac{u^2}{2}, & |u| \leq 1, \\ |u| - \frac{1}{2}, & |u| > 1, \end{cases} \quad \rho_\tau(u) = w_\tau(u) \tilde{\rho}(u),$$

qui est non-négative, convexe, et C^1 .

1.3.7 Synthèse des compromis (robustesse vs convenance analytique)

Les pertes Huber-type et pseudo-Huber offrent une robustesse accrue grâce à une influence bornée (saturation de la pente), au prix d'une forme non strictement quadratique dans les queues (pas de forte convexité globale). La pseudo-Huber présente en plus une régularité C^∞ , utile numériquement. La perte expectile, entièrement quadratique (avec poids asymétriques), est fortement convexe et génère des problèmes d'optimisation très stables, mais demeure plus sensible aux valeurs extrêmes, car son influence croît linéairement avec $|u|$.

CHAPITRE 2

ON MULTIVARIATE BINARY OUTCOMES COPULAS-REGRESSION PROBLEM

Dans ce Chapitre, nous présentons une méthode innovante pour modéliser les dépendances entre variables binaires multivariées dans un cadre de régression en utilisant les copules. Cette approche flexible permet non seulement de capturer les relations entre les variables binaires via une copule, mais aussi d'utiliser une fonction de lien basée sur les copules, comme investigué par (Mesfioui *et al.*, 2023), pour modéliser le lien entre chaque variable binaire $Y_j \in \{0, 1\}$, $j = 1, \dots, d$ et les covariables $\mathbf{X} \in \mathbb{R}^p$. De plus, nous explorons l'utilisation de la décomposition par les vignes (telles que les C-vines et D-vines) pour assouplir la contrainte imposant qu'un seul paramètre contrôle la dépendance entre les variables. Notamment, notre méthode englobe des modèles déjà existants, comme le modèle de régression probit pour deux variables binaires (Chib et Greenberg, 1998). En outre, les simulations montrent que notre méthode surpasse des approches classiques telles que le C-logit (Genest *et al.*, 2013) et le VGLM (Yee, 2015), particulièrement lorsque les dépendances sont non linéaires ou fortes.

En conclusion, une application aux données du Western Collaborative Group Study illustre l'efficacité de notre modèle pour évaluer les probabilités conjointes de maladies telles que la maladie coronarienne et l'arc cornéen, en fonction de l'âge et de l'indice de masse corporelle (IMC). Ce travail constitue donc une avancée méthodologique prometteuse pour l'analyse des données binaires dans divers domaines. Ce travail a été accepté à la revue "Japanese Journal of Statistics and Data Science".

2.1 Abstract

Multivariate binary data, often encountered in various fields including epidemiology, biology, and finance, pose unique challenges due to their discrete nature and complex dependencies. Copulas, which provide a flexible framework for capturing and modelling dependencies between random variables, offer a promising approach for analyzing such data. In this work, we propose a copula-based joint regression approach to investigate the relationship between a vector of binary response variables and a set of covariates, with aim to provide a unified and flexible modelling of the predictive probability of success for multivariate binary outcomes. To gain modelling flexibility, the proposed approach assumes that i) the marginal distributions of the binary response variables depend on the covariates through copula models, and ii) the outcomes residual dependence is also captured through a copula model. Several joint regression models fall within the proposed framework, including multivariate latent probit and logistic models. We provide a closed-form ex-

pression for the estimator of the predictive probability of success for multivariate binary outcomes, and derive its asymptotic properties (weak convergence and i.i.d. representation). We validate the performance of our method by conducting simulation studies and analyzing the Western Collaborative Group Study (WCGS) cohort to evaluate the risk of coronary heart disease and corneal arcus (two binary outcomes) and a set of covariates, including age, BMI,

2.2 Introduction

In various practical contexts, we aim to study the relationship between multiple binary variables $\mathbf{Y} = (Y_1, \dots, Y_d) \in \{0, 1\}^d$ given a vector of covariates $\mathbf{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$. In other words, we want to study the probability distribution of these binary outcomes given the covariates. A widely employed technique is the Vector Generalized Linear Model (VGLM). The VGLM offers a framework for analyzing multiple outcomes simultaneously, using a single joint distribution. They extend the familiar generalized linear model (GLM) approach by allowing each outcome to have its own linear predictor while also accommodating potential correlations or interactions among the outcomes. For more details, see (Zhao et Prentice, 1990) and (Yee, 2015). However, this technique is limited to linear relationships between each binary variable and the covariates. Moreover, the interaction structures between the outcomes are restricted to quadratic forms, which reduces flexibility in capturing more complex dependencies among the response variables.

To address this issue, many authors have employed copulas (Sklar, 1959). Copulas are used in statistics to model dependence among a vector of random variables, independently of their marginal distributions. This flexibility allows us to capture complex relationships, including nonlinear and asymmetric dependencies. Consequently, by accurately modeling the dependence between the outcomes and the covariates, one can achieve more precise estimations and accurate results. For example, Nikoloulopoulos et Karlis (2010) proposed a flexible regression model based on copulas to handle a bivariate regression in the presence of two count outcomes given a set of covariates. Their approach involves modeling the margins of each of the two outcomes with an appropriate regression model (e.g., Poisson regression). Then, they used these margins to express the dependence between the two response variables through an appropriate copula. Furthermore, De Leon et Wu (2011) investigated a regression model based on copulas for mixed outcomes. This approach modeled bivariate mixed outcomes through elliptical class of copulas, where the margins of each outcome is modeled by a suitable regression model. Moreover, Genest *et al.* (2013) proposed an approach to predict outcomes of a vector of binary response variables by combining logistic regression and elliptical copulas. In this approach, each response variable is modeled through a logistic regression with the

covariates to estimate the conditional probability of success. Subsequently, an elliptical copula is employed to model the conditional margins to capture the relationship among the outcomes.

Recently, Yamaguchi et Maruo (2019) proposed a fully parametric approach for conducting bivariate meta analysis, particularly suited for rare binary outcomes. Their methodology employs beta binomial marginal distributions for modeling event counts in each study arm, effectively addressing overdispersion and accommodating zero-event studies. The marginal distributions are connected through a Gaussian copula, with a single correlation parameter quantifying the association between outcomes. This formulation facilitates direct maximization of the closed-form joint likelihood, thereby providing estimates of pooled log-odds ratios and corresponding valid standard errors. In subsequent research, Briseño Sanchez *et al.* (2024) introduced a component-wise gradient boosting framework specifically tailored for distributional copula regression. Their method enables flexible modeling of bivariate outcomes regardless of their nature : binary, discrete, continuous, or mixed. The primary objective of their approach is the simultaneous estimation of model parameters and implicit data-driven variable selection. Consequently, their framework efficiently addresses complex dependency structures commonly encountered in biomedical and epidemiological studies.

Note that all the methodologies discussed above leverage copulas to model complex and nonlinear dependencies among outcomes, while typically presupposing a linear relationship between each individual outcome and associated covariates. Moreover, these approaches commonly treat covariates as deterministic. In contrast, Mesfioui *et al.* (2023) proposed a novel copula-based link function specifically designed to characterize the dependence structure between a univariate binary outcome and a vector of continuous covariates, explicitly considering these covariates as random regressors. The authors provided model parameters estimators using both parametric and semi-parametric approaches, outperforming traditional methods such as logistic and probit models. Hence, this method offers great flexibility when the dependence between the outcome and the explanatory variables is not linear.

The primary objective of this work is to provide a unified framework for assessing the probability of predicting multiple binary outcomes by learning and examining the conditional distribution function, defined as

$$\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x}) := \mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x}), \quad (2.1)$$

where $\mathbf{y} = (y_1, \dots, y_d)$, with $y_j \in \{0, 1\}$, $j = 1, \dots, d$, and $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$.

To this end, we propose a strategy based on copulas to model the probability defined in equation (2.1). Specifically, we establish a relationship between each outcome and the corresponding covariates through a novel copula-based link function for the conditional margins ($Y \mid \mathbf{X} = \mathbf{x}$), as introduced by Mesfioui *et al.*

(2023). Subsequently, we characterize the dependence structure among these conditional margins using an appropriate copula. Further elaboration of this model formulation is provided in Section 2.3. Additionally, we explore recent advances in vine copula construction methods to address limitations associated with representing the entire dependence structure among variables by a single parameter. Such techniques are particularly beneficial for analyzing mixed-type variables under simplifying assumptions, as comprehensively discussed by Stöber (2013), and Stöber *et al.* (2015). A notable application within regression analysis was presented by Kraus et Czado (2017), who extended the quantile regression methodology by incorporating a D-vine decomposition to model dependencies between response and explanatory variables. The authors further demonstrated the effectiveness of this approach, particularly when dealing with mixed variable types.

The remainder of this work is organized as follows. In Section 2.3, we discuss the proposed methodology and mathematical expression of equation (2.1) in the presence of d binary outcomes. Then, we provide a closed theoretical form of the probability defined in (2.1), for various classes of copulas such as Archimedean, Elliptical classes, and vine copulas. In Section 2.4, we describe the proposed estimator. Then, we provide the asymptotic behavior of our estimator in Section 2.5. Section 2.6 summarizes the results of our simulation by comparing the performance of the proposed approach with the competitive methods C-logit (Genest *et al.*, 2013) and VGLM. Finally, Section 2.7 presents an application of our methodology through analyzing a real data set, called Western Collaborative Group Study.

2.3 Theoretical model

This section provides a mathematical representation of Equation (2.1) in terms of copulas. Further, we give explicit form of the predictive probability, focusing primarily on useful classes of copulas, such as Elliptical and Archimedean copula families (e.g., Gaussian, Clayton, and Frank copulas).

2.3.1 Theoretical Predictive Probability (PP) based on copulas

In order to explain the predictive probability based on copulas, the proposed approach models the dependence between the conditional marginal distributions via a copula. More precisely, we assume

$$\mathbb{P}\left(Y_1 \leq y_1, \dots, Y_d \leq y_d \mid \mathbf{X} = \mathbf{x}\right) = C_\gamma\left(F_1(y_1 | \mathbf{X} = \mathbf{x}), \dots, F_d(y_d | \mathbf{X} = \mathbf{x}) \mid \mathbf{X} = \mathbf{x}\right), \quad (2.2)$$

where γ represents the copula parameter, which governs the dependence between the conditional marginal distributions. Additionally, $F(y_j | \mathbf{X} = \mathbf{x})$ denotes the conditional distribution of the j^{th} outcome given

the covariates, for $j = 1, \dots, d$. This formulation relies on the simplifying assumption that the dependence structure, as described by the chosen copula, remains constant across all values of \mathbf{x} . In the literature, this hypothesis is referred to as the simplifying assumption and was investigated and tested in (Derumigny et Fermanian, 2017; Acar *et al.*, 2012; Gijbels *et al.*, 2017; Stoeber *et al.*, 2013).

To complete the specification of the proposed method, one needs to specify $F_j(y_j | \mathbf{X} = \mathbf{x})$, $j = 1, \dots, d$. To this end, we rely on the approach developed in (Mesfioui *et al.*, 2023), which models the conditional margins using conditional copulas as follows :

$$F_j(y_j | \mathbf{X} = \mathbf{x}) = \mathcal{C}_{\theta_j}^c \left(G_j(y_j) | \mathbf{F}(\mathbf{x}) \right), \quad (2.3)$$

where $G_j(\cdot)$ represents the marginal distribution function of the j^{th} outcome, and $\mathbf{F}(\mathbf{x}) := (F_1(x_1), F_2(x_2), \dots, F_p(x_p)) \in [0, 1]^p$ represents the vector of marginal distributions of the covariates. Here, θ_j is the parameter vector that captures the dependence between the j^{th} binary response and the explanatory variables. More precisely, we link each dichotomous response variable and the set of covariates via a copula model (Bouyé et Salmon, 2009) as follows :

$$\mathcal{C}_{\theta}^c \left(G_j(y_j) | \mathbf{F}(\mathbf{x}) \right) := \frac{\partial^p \mathcal{C}_{\theta_j} (u_j, \mathbf{v})}{\partial v_1 \dots \partial v_p} \bigg/ \frac{\partial^p \mathcal{C}_{\theta_j} (1, \mathbf{v})}{\partial v_1 \dots \partial v_p}, \quad (2.4)$$

where $u_j = G_j(y_j)$ $j = 1, \dots, d$, and $\mathbf{v} = \mathbf{F}(\mathbf{x})$.

Substituting (2.4) in (2.2) leads to the following result

$$\mathbb{P} \left(Y_1 \leq y_1, \dots, Y_d \leq y_d \mid \mathbf{X} = \mathbf{x} \right) = \mathcal{C}_{\gamma} \left\{ \mathcal{C}_{\theta_1}^c \left(G_1(y_1) \mid \mathbf{F}(\mathbf{x}) \right), \dots, \mathcal{C}_{\theta_d}^c \left(G_d(y_d) \mid \mathbf{F}(\mathbf{x}) \right) \right\}. \quad (2.5)$$

Based on (2.5), the probability of the event $\{\mathbf{Y} = \mathbf{0}\}$, for instance, can be expressed as :

$$\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \mathcal{C}_{\gamma} \left\{ \mathcal{C}_{\theta_1}^c \left(1 - p_1 \mid \mathbf{F}(\mathbf{x}) \right), \dots, \mathcal{C}_{\theta_d}^c \left(1 - p_d \mid \mathbf{F}(\mathbf{x}) \right) \right\}, \quad (2.6)$$

where p_j is the marginal probability of observing the event $\{Y_j = 1\}$; i.e., $p_j = P(Y_j = 1)$, $j = 1, \dots, d$.

Remark 2.1 (Copula identifiability issues with Bernoulli margins) *In the case where the margins are Bernoulli distributions, the copula uniqueness may not be guaranteed with binary outcomes. However, for $d = 2$, Genest et Nešlehová (2007) demonstrate, in their Example 1 (Section 2), that for a Bernoulli pair, every copula compatible with a given 2×2 contingency table takes identical values at the four cumulative-probability points corresponding to the cells :*

$$(G_1(0), G_2(0)), \quad (G_1(0), G_2(1)), \quad (G_1(1), G_2(0)), \quad (G_1(1), G_2(1)).$$

This result is a direct consequence of (Sklar, 1959) Theorem, which guarantees that the copula \mathcal{C}_γ is unique on the support $(G_1(Y_1), G_2(Y_2), \mathbf{F}(\mathbf{X}))$. Therefore, the predictive probability computed on this support is unaffected by any potential non-uniqueness of the copula outside these specific points. Consequently, the joint probabilities of the observed cells are entirely determined by the data and do not depend on the chosen copula. Any possible indeterminacy only pertains to interpolation points between these observed cells. This clarification emphasizes that the joint probabilities reported remain identifiable despite theoretical non-uniqueness.

Notice that to model the conditional distribution $F_j(y_j|\mathbf{X} = \mathbf{x})$, $j = 1, \dots, d$, in (2.2), several candidate models have been proposed in the literature (e.g., the logistic and probit regression methods). Here, we rely on the model defined in (2.4) given its flexibility for describing the dependence between each outcome Y_j and the covariates. This is well demonstrated in (Mesfioui *et al.*, 2023), which shows the effectiveness of this method compared to probit and logistic models.

Illustration of the proposed model is given next for the case of two binary outcomes (i.e., $d = 2$). In this case, one can write

$$\begin{aligned}\mathbb{P}\left(Y_1 = 1, Y_2 = 1 | \mathbf{X} = \mathbf{x}\right) &= 1 - \mathbb{M}_0^1(\mathbf{x}) - \mathbb{M}_0^2(\mathbf{x}) + \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}), \\ \mathbb{P}\left(Y_1 = 1, Y_2 = 0 | \mathbf{X} = \mathbf{x}\right) &= \mathbb{M}_0^1(\mathbf{x}) - \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}), \\ \mathbb{P}\left(Y_1 = 0, Y_2 = 1 | \mathbf{X} = \mathbf{x}\right) &= \mathbb{M}_0^2(\mathbf{x}) - \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}),\end{aligned}\tag{2.7}$$

where $\mathbb{M}_0^j(\mathbf{x}) = \mathbb{P}\left(Y_j = 0 \mid \mathbf{X} = \mathbf{x}\right) = \mathcal{C}_{\theta_j}^c\left(1 - p_j \mid \mathbf{F}(\mathbf{x})\right)$, $j = 1, 2$, and

$$\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}) = \mathbb{P}\left(Y_1 = 0, Y_2 = 0 \mid \mathbf{X} = \mathbf{x}\right) = \mathcal{C}_\gamma\left\{\mathcal{C}_{\theta_1}^c\left(1 - p_1 \mid \mathbf{F}(\mathbf{x})\right), \mathcal{C}_{\theta_2}^c\left(1 - p_2 \mid \mathbf{F}(\mathbf{x})\right)\right\}.$$

In a similar way, in the presence of three binary outcomes (i.e., $d = 3$), the conditional distribution can be

evaluated in terms of the probabilities $\mathbb{M}_0^j(\mathbf{x})$, $\mathbb{M}_{(0,0)}^{j,k}(\mathbf{x})$, and $\mathbb{M}_{(0,0,0)}^{1,2,3}(\mathbf{x})$, where $\mathbb{M}_{(0,0)}^{j,k}(\mathbf{x}) = \mathbb{P}\left(Y_j = 0, Y_k = 0 \mid \mathbf{X} = \mathbf{x}\right)$, $(j, k) \in \{1, 2, 3\}$, and $\mathbb{M}_{(0,0,0)}^{1,2,3}(\mathbf{x}) = \mathcal{C}_\gamma\left(\mathcal{C}_{\theta_1}^c(1 - p_1 | \mathbf{F}(\mathbf{x})), \mathcal{C}_{\theta_2}^c(1 - p_2 | \mathbf{F}(\mathbf{x})), \mathcal{C}_{\theta_3}^c(1 - p_3 | \mathbf{F}(\mathbf{x}))\right)$.

For instance, one can verify that for the event $\{\mathbf{Y} = \mathbf{1}\}$, one has

$$\begin{aligned}
\mathbb{M}_{(1,1,1)}^{1,2,3}(\mathbf{x}) &= 1 - \mathcal{C}_{\theta_1}^c(1 - p_1 \mid \mathbf{F}(\mathbf{x})) - \mathcal{C}_{\theta_2}^c(1 - p_2 \mid \mathbf{F}(\mathbf{x})) - \mathcal{C}_{\theta_3}^c(1 - p_3 \mid \mathbf{F}(\mathbf{x})) \\
&\quad + \mathcal{C}_\gamma(1, \mathcal{C}_{\theta_2}^c(1 - p_2 \mid \mathbf{F}(\mathbf{x})), \mathcal{C}_{\theta_3}^c(1 - p_3 \mid \mathbf{F}(\mathbf{x}))) \\
&\quad + \mathcal{C}_\gamma(\mathcal{C}_{\theta_1}^c(1 - p_1 \mid \mathbf{F}(\mathbf{x})), 1, \mathcal{C}_{\theta_3}^c(1 - p_3 \mid \mathbf{F}(\mathbf{x}))) \\
&\quad + \mathcal{C}_\gamma(\mathcal{C}_{\theta_1}^c(1 - p_1 \mid \mathbf{F}(\mathbf{x})), \mathcal{C}_{\theta_2}^c(1 - p_2 \mid \mathbf{F}(\mathbf{x})), 1) - \mathbb{M}_{(0,0,0)}^{1,2,3}(\mathbf{x}). \tag{2.8}
\end{aligned}$$

Based on equation (2.8), one can write $\mathbb{M}_{(1,1,1)}^{1,2,3}(\mathbf{x})$ in terms of the following probabilities ($\mathbb{M}_0^1(\mathbf{x}), \mathbb{M}_0^2(\mathbf{x}), \mathbb{M}_0^3(\mathbf{x}), \mathbb{M}_{(0,0)}^{2,3}(\mathbf{x}), \mathbb{M}_{(0,0)}^{1,3}(\mathbf{x}), \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}), \mathbb{M}_{(0,0,0)}^{1,2,3}(\mathbf{x})$) as follows :

$$\mathbb{M}_{(1,1,1)}^{1,2,3}(\mathbf{x}) = 1 - \mathbb{M}_0^1(\mathbf{x}) - \mathbb{M}_0^2(\mathbf{x}) - \mathbb{M}_0^3(\mathbf{x}) + \mathbb{M}_{(0,0)}^{2,3}(\mathbf{x}) + \mathbb{M}_{(0,0)}^{1,3}(\mathbf{x}) + \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}) - \mathbb{M}_{(0,0,0)}^{1,2,3}(\mathbf{x}). \tag{2.9}$$

The result in (2.9) can be generalized to the case of a d -dimensional binary outcome, as outlined in the following Lemma.

Lemma 2.2 *For every combination of binary outcomes represented by the vector (y_1, \dots, y_d) , the probability $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$ is fully and uniquely determined by the following probabilities*

$$\left\{ \left\{ \mathbb{M}_0^j(\mathbf{x}), \right\}_{j=1, \dots, d}, \left\{ \mathbb{M}_{(0,0)}^{j,k}(\mathbf{x}) \right\}_{j < k}, \left\{ \mathbb{M}_{(0,0,0)}^{j,k,e}(\mathbf{x}) \right\}_{j < k < e}, \dots, \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) \right\},$$

where $\mathbb{M}_0^j = \mathbb{P}(Y_j = 0 \mid \mathbf{X} = \mathbf{x})$, $\mathbb{M}_{(0,0)}^{j,k} = \mathbb{P}(Y_j = 0, Y_k = 0 \mid \mathbf{X} = \mathbf{x})$, $\mathbb{M}_{(0,0,0)}^{j,k,e} = \mathbb{P}(Y_j = 0, Y_k = 0, Y_e = 0 \mid \mathbf{X} = \mathbf{x})$, \dots , $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{X} = \mathbf{x}) = \mathbb{P}(Y_1 = 0, \dots, Y_d = 0 \mid \mathbf{X} = \mathbf{x})$.

The proof of Lemma 2.2 is postponed to Appendix A.1. Of note, as stated in Lemma 2.2, because all probability events can be obtained from $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$, throughout the rest of this work, we exclusively focus on the estimation of $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$. Thus, we derive the theoretical properties of the proposed estimator of $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ and validate its performance through a simulation study.

2.3.2 Example of predictive probability in some common copula families

In this section, we express the probability $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ under common copula families, such as the Gaussian and Archimedean classes (e.g., Clayton and Frank).

2.3.2.1 Predictive probability based on Gaussian copula

Elliptical copulas, such as the Gaussian and Student-t copulas, provide an adaptable framework capable of capturing both linear and nonlinear dependence structures, making them useful to model broad range of dependence structures. These advantages are crucial for several applications, such as in health research, finance, and hydrology.

In our context, if we assume the dependence structure between the conditional marginal distributions of $Y_j | \mathbf{X} = \mathbf{x}, j = 1, \dots, d$, is modeled by a Gaussian copula, then equation (2.6) can be expressed as follows :

$$\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \Phi_{\Gamma} \left\{ \Phi^{-1} \left(\mathcal{C}_{\theta_1}^c \left(1 - p_1 \mid \mathbf{F}(\mathbf{x}) \right) \right), \dots, \Phi^{-1} \left(\mathcal{C}_{\theta_d}^c \left(1 - p_d \mid \mathbf{F}(\mathbf{x}) \right) \right) \right\}, \quad (2.10)$$

Here, Φ_{Γ} represents the multivariate normal cumulative distribution function, with mean vector $\mathbf{0}$ and correlation matrix Γ , which measures the dependence structure between the conditional marginal distributions. Furthermore, Φ denotes the univariate standard normal cumulative distribution function.

Assuming also that the copula \mathcal{C}_{θ_j} , which models the dependence between the j -th outcome and the covariates, to be a Gaussian copula leads to

$$\mathcal{C}_{\Gamma_j}^c(1 - p_j \mid \mathbf{F}(\mathbf{x})) = \Phi \left(\frac{\Phi^{-1}(1 - p_j) - \boldsymbol{\rho}_j \mathcal{R}^{-1} \Phi^{-1}(\mathbf{F}(\mathbf{x}))}{\sqrt{1 - \boldsymbol{\rho}_j \mathcal{R}^{-1} \boldsymbol{\rho}_j^{\top}}} \right), \quad (2.11)$$

where

$$\Gamma_j = \begin{pmatrix} 1 & \boldsymbol{\rho}_j \\ \boldsymbol{\rho}_j^{\top} & \mathcal{R} \end{pmatrix},$$

with $\Phi^{-1}(\mathbf{F}(\mathbf{x})) = \left(\Phi^{-1}(F_1(x_1)), \dots, \Phi^{-1}(F_p(x_p)) \right)^{\top}$, and $\boldsymbol{\rho}_j = (\rho_{j1}, \dots, \rho_{jp})$ represent the correlation coefficients between the normal scores of j^{th} response variable and p covariates, where the normal score of a random variable X is obtained via the transformation $q(x) := \Phi^{-1}(F(x))$, so that $q(X)$ is standard normally distributed. Finally, \mathcal{R} represents the correlation matrix between the normal scores of the covariates. Then, one can write equation (2.10) as follows

$$\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \Phi_{\Gamma} \left(\Phi^{-1} \left(\Phi \left(\frac{\Phi^{-1}(1 - p_1) - \boldsymbol{\rho}_1 \mathcal{R}^{-1} \Phi^{-1}(\mathbf{F}(\mathbf{x}))}{\sqrt{1 - \boldsymbol{\rho}_1 \mathcal{R}^{-1} \boldsymbol{\rho}_1^{\top}}} \right) \right), \dots, \Phi^{-1} \left(\Phi \left(\frac{\Phi^{-1}(1 - p_d) - \boldsymbol{\rho}_d \mathcal{R}^{-1} \Phi^{-1}(\mathbf{F}(\mathbf{x}))}{\sqrt{1 - \boldsymbol{\rho}_d \mathcal{R}^{-1} \boldsymbol{\rho}_d^{\top}}} \right) \right) \right).$$

Example 2.3 Consider the case $d = 2$ and $p = 1$, equation (2.11) becomes

$$C_{\Gamma_j}^c(1 - p_j \mid F_1(x_1)) = \Phi \left(\frac{\Phi^{-1}(1 - p_j) - \rho_j \Phi^{-1}(F_1(x_1))}{\sqrt{1 - \rho_{j1}^2}} \right). \quad (2.12)$$

where $\rho_{j1} = \text{Corr}(\Phi^{-1}(G_j(Y_j)), \Phi^{-1}(F_1(X_1)))$, $j = 1, 2$.

Thus, substituting (2.12) into (2.10) leads to

$$\mathbb{M}_{(0,0)}^{1,2}(x_1) = \Phi_{\Gamma} \left(\frac{\Phi^{-1}(1 - p_1) - \rho_{11} \Phi^{-1}(F_1(x_1))}{\sqrt{1 - \rho_{11}^2}}, \frac{\Phi^{-1}(1 - p_2) - \rho_{21} \Phi^{-1}(F_1(x_1))}{\sqrt{1 - \rho_{21}^2}} \right), \quad (2.13)$$

where ρ_{j1} is the Pearson correlation between the normal scores of Y_j , for $j = 1, 2$, and X_1 ; measuring the association between Y_j and X_1 . Assuming the covariate X_1 follows a normal distribution, $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, equation (2.13) becomes

$$\mathbb{M}_{(0,0)}^{1,2}(x_1) = \mathbb{P}(Y_1 = 0, Y_2 = 0 \mid X_1 = x_1) = \Phi_{\Gamma} \left(\beta_{01} + \beta_{11}x_1, \beta_{02} + \beta_{12}x_1 \right), \quad (2.14)$$

where $\beta_{0j} = \frac{\Phi^{-1}(1-p_j) + \rho_{j1}\mu}{\sigma\sqrt{1-\rho_{j1}^2}}$ and $\beta_{1j} = \frac{-\rho_{j1}}{\sigma\sqrt{1-\rho_{j1}^2}}$, for $j = 1, 2$. This leads to the multivariate probit model for bivariate binary outcomes (Chib et Greenberg, 1998).

Remark 2.4 Of note, if the logit link function is preferred to model $F(Y_j \mid \mathbf{X} = \mathbf{x})$ in (2.4), then one gets the C-logit model described in (Genest et al., 2013). In addition, assume that the outcomes follow a Multivariate Exponential Distribution (MED) family (Song et al., 2009), which integrates separate generalized linear models (GLMs) for the single-binary outcomes into a joint regression model and uses the Gaussian copula to model the residual dependence. In this case, one can express (2.6) using an MED model by replacing $C_{\theta_j}^c$ in (2.10) by marginal GLM link functions (i.e., GLM models) as investigated by Xue-Kun Song (2000) and Song et al. (2009).

2.3.2.2 Predictive probability based on Archimedean copulas

The Archimedean class of copulas, well known for their exchangeable properties, is among the most widely employed families in dependence modeling (Genest et MacKay, 1986).

An Archimedean copula is defined by a generator function $\varphi(\cdot)$, which is continuous, strictly decreasing, and convex, where the mapping of its pseudo inverse φ^{-1} is defined as $\varphi^{-1} : [0, \infty] \rightarrow [0, 1]$, and its expression is given as follows :

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

with $\psi(\cdot)$ representing the true inverse of φ on its domain, is differentiable up to the order d such that $(-1)^i \psi^{(i)} \geq 0$. Then the copula is expressed as :

$$\mathcal{C}(u_1, u_2, \dots, u_d) = \varphi^{-1} \left(\sum_{i=1}^d \varphi(u_i) \right). \quad (2.15)$$

Let the p -th derivative of ψ be denoted by $\psi^{(p)} = \frac{d^p \psi}{dv^p}$. Then, equation (2.15) can be rewritten as :

$$\mathcal{C}(u_1, u_2, \dots, u_d) = \psi \left(\sum_{i=1}^d \varphi(u_i) \right). \quad (2.16)$$

In our context, if we assume that both dependence between the vector of conditional margins (e.g., $(F(y_1 | \mathbf{x}), \dots, F(y_d | \mathbf{x}))$) and the dependence between the j -th outcome and the covariates are expressed using Archimedean copulas, then one can express the following probability $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ as :

$$\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d} = \psi_\gamma \left(\sum_{j=1}^d \varphi_\gamma(\pi_j(\mathbf{x})) \right), \quad (2.17)$$

where

$$\pi_j(\mathbf{x}) = \mathcal{C}_{\theta_j}^c(1 - p_j | \mathbf{F}(\mathbf{x})) = \frac{\psi_{\theta_j}^{(p)} [\varphi_{\theta_j}(1 - p_j) + \sum_{i=1}^p \varphi_{\theta_j}(F_i(x_i))]}{\psi_{\theta_j}^{(p)} [\sum_{i=1}^p \varphi_{\theta_j}(F_i(x_i))]} \quad (2.18)$$

Here, the parameters γ and θ_j are scalars, where φ_{θ_j} and φ_γ represent the generators of the following copulas $\mathcal{C}_{\theta_j}^c(\cdot)$ and $\mathcal{C}_\gamma(\cdot)$, respectively. For example, considering the presence of a single covariate ($p = 1$), the expression given in (2.18) simplifies to

$$\pi_j(x_1) = \mathcal{C}_{\theta_j}^c(1 - p_j | F_1(x_1)) = \psi_{\theta_j}'(\varphi_{\theta_j}(1 - p_j) + \varphi_{\theta_j}(F_1(x_1))) \varphi_{\theta_j}'(F_1(x_1)).$$

Substituting the vector $\boldsymbol{\pi}(x_1) = (\pi_1(x_1), \dots, \pi_d(x_1))$ into $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(x_1)$ given in (2.17), one can express $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(x_1)$ as follows :

$$\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(x_1) = \psi_\gamma \left(\sum_{j=1}^d \varphi_\gamma \left(\psi_{\theta_j}'(\varphi_{\theta_j}(1 - p_j) + \varphi_{\theta_j}(F_1(x_1))) \varphi_{\theta_j}'(F_1(x_1)) \right) \right). \quad (2.19)$$

2.3.2.3 Examples

1. Clayton Copula

In this case, we assume that the dependence between the conditional margins (*i.e.*, $(F_1(y_1 | \mathbf{F}(\mathbf{x})), \dots, F_d(y_d | \mathbf{F}(\mathbf{x})))$) and the dependence between each outcome Y_j and the covariates $\mathbf{X} \in \mathbb{R}^p$ are modeled by a Clayton copula, whose generator is defined as :

$$\varphi_{\theta_j}(z) = \frac{z^{-\theta_j} - 1}{\theta_j}, \quad \theta_j > 0, \quad z \in (0, 1), \quad (2.20)$$

and

$$\psi_{\theta_j}(t) = (\theta_j t + 1)^{-\frac{1}{\theta_j}}. \quad (2.21)$$

Based on the Clayton generator and its inverse defined in Equations (2.20) and (2.21), Equation (2.19) simplifies to

$$\mathbb{M}_{(0,0)}^{1,2}(x_1) = \left(\sum_{j=1}^2 \left[\left((1-p_j)^{-\theta_j} + F_1(x_1)^{-\theta_j} - 1 \right)^{\gamma \left(\frac{1}{\theta_j} + 1 \right)} F_1(x_1)^{\gamma(1+\theta_j)} \right] - 1 \right)^{-\frac{1}{\gamma}}.$$

2. Frank Copula

Suppose that both the dependence among the conditional marginal distributions ($(F_1(y_1 | \mathbf{F}(\mathbf{x})), \dots, F_d(y_d | \mathbf{F}(\mathbf{x})))$) and the dependence between a particular outcome Y_j and the covariates $\mathbf{x} \in \mathbb{R}^p$ are modeled using a Frank copula, whose generator is defined as :

$$\varphi_{\theta_j}(z) = -\ln \left(\frac{e^{-\theta_j z} - 1}{e^{-\theta_j} - 1} \right), \quad \theta_j \in \mathbb{R}^* \quad z \in (0, 1), \quad (2.22)$$

and

$$\psi_{\theta_j}(t) = -\frac{1}{\theta_j} \ln \left(e^{-t} (e^{\theta_j} - 1) + 1 \right). \quad (2.23)$$

Based on the definition of ψ_{θ_j} and φ_{θ_j} in Equations (2.22) and (2.23) , one can express Equation (2.19) as follows :

$$\mathbb{M}_{(0,0)}^{1,2}(x_1) = \psi_{\gamma} \left(\sum_{j=1}^2 \varphi_{\gamma} \left(\frac{1}{1 + \exp \left(-\ln \left(\exp [\varphi_{\theta_j}(1-p_j)] - 1 \right) - \theta_j F_1(x_1) \right)} \right) \right).$$

2.3.2.4 Mixed vine decomposition for predictive probability

Referring back to Equation (2.6), the parameters γ and θ_j , respectively, capture the dependence among the outcomes and between the j^{th} outcome and the covariates. However, in a situation where either p or

d is relatively large, this assumption becomes overly restrictive. To address this limitation, we employ the vine decomposition technique, which decomposes a multivariate copula density into a product of bivariate conditional copula distributions (Joe, 1996). Later, (Bedford et Cooke, 2001) and (Bedford et Cooke, 2002) introduced regular vine decomposition, which constructs bivariate copulas based on a set of trees. To illustrate this mechanism within our context, consider a random vector (Y_j, X_1, X_2) , where X_1 and X_2 are continuous covariates, and Y_j represents the j^{th} outcome, with $j = 1, \dots, d$. Assume specifically that the dependence structure in the first tree follows the order $Y_j - X_2 - X_1$. Under this structure, the conditional copula distribution is expressed as follows :

$$\mathcal{C}_{\theta_j}^c(1-p_j|\mathbf{F}(\mathbf{x})) = \int_0^{1-p_j} c_{\theta_{Y_j, X_2}}(z, F_{X_2}(x_2)) \times c_{\theta_{Y_j, X_1; X_2}}(\mathcal{C}_{\theta_{Y_j|X_2}}(z|F_{X_2}(x_2)), \mathcal{C}_{\theta_{X_1|X_2}}(F_{X_1}(x_1)|F_{X_2}(x_2))) dz. \quad (2.24)$$

As a second illustration if the first tree follows the order $Y_j - X_1 - X_2$, then the conditional distribution of the copula is given by

$$\mathcal{C}_{\theta_j}^c(1-p_j|\mathbf{F}(\mathbf{x})) = \int_0^{1-p_j} c_{\theta_{Y_j, X_1}}(z, F_{X_1}(x_1)) \times c_{\theta_{Y_j, X_2; X_1}}(\mathcal{C}_{\theta_{Y_j|X_1}}(z|F_{X_1}(x_1)), \mathcal{C}_{\theta_{X_2|X_1}}(F_{X_2}(x_2)|F_{X_1}(x_1))) dz. \quad (2.25)$$

Remark 2.5 Note that once we obtain the conditional distribution of $\mathcal{C}_{\theta_j}^c(1-p_j | \mathbf{F}(\mathbf{x}))$ via the vine construction, we can subsequently apply the same process to relax the assumption on γ as it is illustrated in (2.24) and (2.25) for the joint copula distribution, which governs the dependence across the conditional margins in (2.6). As an alternative to the proposed vine-based approach, it is possible to directly obtain the probability $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ by applying a vine decomposition to the joint density of the vector (\mathbf{Y}, \mathbf{X}) . This method can handle both continuous and discrete variables and can be implemented using the R package `rvinecopulib`; for more details, see (Nagler et Vatter, 2018). However, in the case of binary outcomes, the conditional pair-copulas induced by this vine decomposition might include several binary components of \mathbf{Y} in the conditioning sets. As each binary outcome splits the data into only two levels, many of the resulting strata might contain too few observations, especially in the presence of rare-event binary outcomes. This can harm both the precision and the stability of the probability estimates.

2.4 Estimation and inference

This section describes the estimation procedure for the proposed model parameters used to assess the probability $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$ defined in Equation (2.1). To accomplish this, consider that $(\mathbf{Y}_1, \mathbf{X}_1), (\mathbf{Y}_2, \mathbf{X}_2), \dots, (\mathbf{Y}_n, \mathbf{X}_n)$

are mutually independent copies of (\mathbf{Y}, \mathbf{X}) , where $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$ and $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})$ for $i = 1, \dots, n$.

The estimation procedure follows a semiparametric approach and consists of three main steps :

Step 1 : Compute the empirical distribution functions for both the covariates and the outcomes as follows :

$$\widehat{F}_{nk}(x_k) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(X_{ik} \leq x_k), \quad \text{for } k = 1, \dots, p, \quad (2.26)$$

and the probability of success :

$$\widehat{p}_{nj} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(Y_{ij} = 1), \quad \text{for } j = 1, \dots, d. \quad (2.27)$$

Step 2 : Estimate the dependence parameters for each outcome given the covariates using the following pseudo-likelihood :

$$\widehat{\boldsymbol{\theta}}_j = \operatorname{argmax}_{\boldsymbol{\theta}_j \in \mathbb{R}^r} \prod_{i=1}^n \ell_i(\boldsymbol{\theta}_j, \widehat{p}_{nj}, \widehat{\mathbf{F}}_n(\mathbf{x})), \quad (2.28)$$

where $\widehat{\mathbf{F}}_n(\mathbf{x}) = (\widehat{F}_{n1}(x_1), \dots, \widehat{F}_{np}(x_p))$, and

$$\ell_i(\boldsymbol{\theta}_j, \widehat{p}_{nj}, \widehat{\mathbf{F}}_n(\mathbf{x})) = \left(1 - \mathcal{C}_{\boldsymbol{\theta}_j}^c(1 - \widehat{p}_{nj} \mid \widehat{\mathbf{F}}_n(\mathbf{x}))\right)^{Y_{ij}} \mathcal{C}_{\boldsymbol{\theta}_j}^c(1 - \widehat{p}_{nj} \mid \widehat{\mathbf{F}}_n(\mathbf{x}))^{1-Y_{ij}}. \quad (2.29)$$

This method follows the approach investigated by (Mesfioui *et al.*, 2023). After obtaining the estimator $\widehat{\boldsymbol{\theta}}_j$, compute $\widehat{\pi}_j(\mathbf{x}) = \mathcal{C}_{\widehat{\boldsymbol{\theta}}_j}^c(1 - \widehat{p}_{nj} \mid \widehat{\mathbf{F}}_n(\mathbf{x}))$ before proceeding to the next step.

Step 3 : Estimate the parameter γ , using the following maximum pseudo-likelihood approach :

$$\widehat{\gamma}_n = \operatorname{argmax}_{\gamma \in \mathbb{R}^q} \prod_{i=1}^n h_{\mathcal{C}_\gamma}(Y_{i1}, \dots, Y_{id} \mid \mathbf{x}), \quad (2.30)$$

where

$$h_{\mathcal{C}_\gamma}(Y_{ij}, \dots, Y_{id} \mid \mathbf{x}) = \sum_{j_1=1}^2 \dots \sum_{j_d=1}^2 (-1)^{j_1 + \dots + j_d} \mathcal{C}_\gamma(u_{1j_1}, \dots, u_{dj_d}),$$

with $u_{j1} = \mathcal{C}_{\widehat{\boldsymbol{\theta}}_j}^c(\widehat{G}_j(Y_{ij}) \mid \widehat{\mathbf{F}}_n(\mathbf{x}))$ and $u_{j2} = \mathcal{C}_{\widehat{\boldsymbol{\theta}}_j}^c(\widehat{G}_j(Y_{ij}^-) \mid \widehat{\mathbf{F}}_n(\mathbf{x}))$, as modeled in equation (2.4). Here, $\widehat{G}_j(Y_{ij}^-)$ represents the left-hand limit of \widehat{G}_j at Y_{ij} , which is equal to $\widehat{G}_j(Y_{ij} - 1)$.

Remark 2.6 The proposed estimation method assumes that the copulas $\mathcal{C}_{\boldsymbol{\theta}_j}^c$'s and \mathcal{C}_γ are known. However, in real data situations, we have no prior knowledge about the dependence structure between each j^{th} outcome and the covariates, as well as between the outcomes. To overcome this limitation, we propose a more flexible alternative estimation procedure, which is based on vine copulas, to construct both the $\mathcal{C}_{\boldsymbol{\theta}_j}^c$'s and \mathcal{C}_γ copulas, as demonstrated in section 2.3.2.4.

2.5 Asymptotic Properties

This section presents the asymptotic properties of the estimator $\hat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d}(\cdot)$. Before presenting the theoretical results, the following notation is assumed :

2.5.1 Notations

Let $\partial_j \mathcal{C}_\gamma = \partial \mathcal{C}_\gamma / \partial \pi_j$, for $j = \{1, \dots, d\}$, $\Delta_\gamma \mathcal{C}_\gamma(\cdot) = \left(\frac{\partial \mathcal{C}_\gamma(\cdot)}{\partial \gamma_1}, \dots, \frac{\partial \mathcal{C}_\gamma(\cdot)}{\partial \gamma_d} \right)^\top$,
 $\partial_u \mathcal{C}_{\theta_j}^c(u | \mathbf{v}) = \frac{\partial_u \mathcal{C}_{\theta_j}^c(u | \mathbf{v})}{\partial u}$, $\boldsymbol{\pi}(\mathbf{x}) = (\pi_1(\mathbf{x}), \dots, \pi_d(\mathbf{x}))$,
 $\Delta_{\theta_j} \mathcal{C}_{\theta_j}^c(u | \mathbf{v}) = \left(\frac{\partial \mathcal{C}_{\theta_j}^c(u | \mathbf{v}; \boldsymbol{\theta})}{\partial \theta_{1j}}, \dots, \frac{\partial \mathcal{C}_{\theta_j}^c(u | \mathbf{v})}{\partial \theta_{rj}} \right)^\top$, and $\Delta_{\mathbf{v}} \mathcal{C}_{\theta_j}^c(u | \mathbf{v}) = \left(\frac{\partial \mathcal{C}_{\theta_j}^c(u | \mathbf{v})}{\partial v_1}, \dots, \frac{\partial \mathcal{C}_{\theta_j}^c(u | \mathbf{v})}{\partial v_p} \right)^\top$.
For some quantities $\tau_j \in [0, 1]$, $j \in \{1, \dots, d+1\}$, one can express :

$$\begin{aligned} \tilde{\pi}_j(\mathbf{x}) &:= \pi_j(\mathbf{x}) + \tau_j (\hat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})), \\ \tilde{\boldsymbol{\gamma}}_n &:= \boldsymbol{\gamma} + \tau_{(d+1)} (\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}). \end{aligned}$$

Also, the following assumptions are necessary for the theoretical development to establish our asymptotic normality of the proposed estimator of $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$.

2.5.2 Assumptions

A1- The hypothesis **B** cited bellow and presented in (Mesfioui *et al.*, 2023) holds for the estimator $\hat{\pi}_j(\mathbf{x})$ developed in Step 2 in Section 2.4, for $j = 1, \dots, d$. This assumption guaranty the following result

$$\hat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathcal{Z}_{ij}(\mathbf{x}) + o_p(n^{-1/2}),$$

where

$$\begin{aligned} \mathcal{Z}_{ij}(\mathbf{x}) &= (p_j - \mathbb{1}(Y_{ij} = 1)) \partial_u \mathcal{C}_{\theta_j}^c(1 - p_j | \mathbf{F}(\mathbf{x})) + \boldsymbol{\zeta}_{ij}^\top \Delta_{\theta_j} \mathcal{C}_{\theta_j}^c(1 - p_j | \mathbf{F}(\mathbf{x})) \\ &\quad + (\mathbb{1}(\mathbf{X}_i \leq \mathbf{x}) - \mathbb{P}(\mathbf{X} \leq \mathbf{x}))^\top \Delta_{\mathbf{v}} \mathcal{C}_{\theta_j}^c(1 - p_j | \mathbf{F}(\mathbf{x})), \end{aligned}$$

with $\mathbb{P}(\mathbf{X} \leq \mathbf{x}) = (\mathbb{P}(X_1 \leq x_1), \dots, \mathbb{P}(X_p \leq x_p))$, $\mathbb{1}(\mathbf{X}_i \leq \mathbf{x}) = (\mathbb{1}(X_{i1} \leq x_1), \dots, \mathbb{1}(X_{ip} \leq x_p))$, $\mathbf{X} = (X_{i1}, \dots, X_{ip})$ and its theoretical vector $\mathbf{X} = (X_1, \dots, X_p)$. Here $\boldsymbol{\zeta}_{ij}$ is a random quantity with zero mean and $\mathbb{E}\|\boldsymbol{\zeta}_{ij}\|^2 < \infty$, as defined in Lemma 1 of (Mesfioui *et al.*, 2023).

— Hypothesis B

(B.1) $\mathcal{C}_{\theta_j}^c(\cdot)$ is twice continuously differentiable in a neighborhood of the true parameter $\boldsymbol{\theta}_j$, $j = 1, \dots, d$ and it is bounded away from zero.

(B.2) The information matrix $I(\boldsymbol{\theta}_j) = \mathbb{E} \left(-\frac{\partial^2 \log \ell(\boldsymbol{\theta}_j, p, \mathbf{F}(\mathbf{x}))}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j^\top} \right)$ exists and is non-singular, where

$\ell(\cdot)$ is defined in (2.29).

(B.3) $\frac{\partial^2 \log C_{\boldsymbol{\theta}}^c(u|\mathbf{v})}{\partial u \partial \theta_{jt}}$ and $\frac{\partial^2 \log C_{\boldsymbol{\theta}}^c(u|\mathbf{v})}{\partial v_k \partial \theta_{jt}}$ are continuous on $\Theta \times (0, 1) \times (0, 1)^p$, for $t = 1, \dots, r$, $k = 1, \dots, p$, and $\Theta \subset \mathbb{R}^r$.

A2- For any $j \in \{1, \dots, d\}$ and $l \in \{1, \dots, q\}$, the second derivative

$$\frac{\partial^2 \log \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x}); \gamma)}{\partial \pi_j(\mathbf{x}) \partial \gamma_l}$$

is continuous on $(0, 1)^d \times \mathbb{R}^q$, which leads to the following results :

- $\partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) - \partial_j \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x})) = o_p(1)$, $\forall j = 1, \dots, d$;
- $\Delta_\gamma \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x})) - \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) = o_p(1)$.

A3- The information matrix

$$I(\gamma) = -\mathbb{E} \left(\frac{\partial^2 \log \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x}); \gamma)}{\partial \gamma \partial \gamma^\top} \right)$$

exist.

A4- Under Assumption **A3**, the maximum pseudo-likelihood estimator of Step 3, $\hat{\gamma}_n$, given in (2.30), satisfies

$$\hat{\gamma}_n - \gamma \equiv \frac{1}{n} \sum_{i=1}^n \boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma) + o_p\left(n^{-\frac{1}{2}}\right),$$

where $\boldsymbol{\nu}_i$ is a vector score, with $\mathbb{E}(\boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma)) = \mathbf{0}$, and $\mathbb{E}\|\boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma)\|^2 < \infty$. This estimator has been studied by (Genest *et al.*, 1995).

Assumptions **A1** to **A4** represent standard conditions for copula-based regression models. Conditions **A1** and **A3**, as detailed by Mesfioui *et al.* (2023), are satisfied by various copula families, as demonstrated by (Omelka *et al.*, 2009). These assumptions are essential to ensure the existence of the estimator $\hat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ near the boundaries of the unit square $[0, 1]^d$. Additionally, Condition **A4** is satisfied for the pseudo-likelihood estimator studied by Genest *et al.* (1995). Notably, **A4** remains valid even when the copula family is misspecified, as investigated by Noh *et al.* (2013).

Theorem 2.7 Under Assumptions **A1** to **A4**, one has :

$$\hat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n Z_i(\mathbf{x}) + o_p\left(n^{-\frac{1}{2}}\right),$$

where

$$Z_i(\mathbf{x}) = \sum_{j=1}^d \partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) \mathcal{Z}_{ij}(\mathbf{x}) + \boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \boldsymbol{\gamma})^\top \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})), \quad (2.31)$$

and $\mathcal{Z}_{ij}(\mathbf{x})$ and $\boldsymbol{\nu}_i$ are defined in Assumptions **A1** and **A4**, respectively.

The proof of Theorem 2.7 is given in Appendix A.1.2.

Theorem 2.7 states that the random variable $\sqrt{n} \left(\widehat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) \right)$ converges to a normal distribution with mean 0 and variance $\text{Var}(Z(\mathbf{x}))$, where $Z(\mathbf{x})$ is defined in equation (2.31). In practice, the variance, $\text{Var}(Z(\mathbf{x}))$, can be estimated using the sample variance of $Z_i(\mathbf{x})$, defined as follows :

$$\widehat{\text{Var}}(Z(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n \left(\widehat{Z}_i(\mathbf{x}) - \frac{1}{n} \sum_{i=1}^n \widehat{Z}_i(\mathbf{x}) \right)^2, \quad (2.32)$$

where $\widehat{Z}_i(\mathbf{x})$ has the same form as $Z_i(\mathbf{x})$ but evaluated with $\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\gamma}}_n$ instead of $\boldsymbol{\pi}, \boldsymbol{\gamma}$, respectively.

In practice, evaluation of the variance defined in (2.32) poses significant computational challenges. A widely employed solution is the bootstrap method, outlined explicitly in Algorithm A.1, provided in Appendix A.1.

This technique facilitates the construction of confidence intervals of the quantity $\left(\widehat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) \right)$, thereby enabling the calculation of the coverage probabilities, reported in Appendix A.1.7, associated with this parameter difference.

2.6 Simulation study

Simulations were carried out to evaluate and compare our proposed approach. Comparisons were conducted with respect to the elliptic copula combined with the logistic regression model (C-logit), developed in (Genest *et al.*, 2013), and the vector generalized linear model (VGLM) (Zhao et Prentice, 1990). As outlined before, the C-logit method links each binary outcome and the covariates through a (marginal) logistic model and then uses a Gaussian copula to capture the residual dependence between the marginal models (i.e., the marginal distributions). The VGLM method uses a quadratic exponential form parameterization to model the marginal means and pairwise correlations for the regression analysis of correlated binary outcomes.

2.6.1 Data generation

In this section, we provide a comprehensive overview of the steps taken to generate simulated data for the evaluation of the methods.

Data were generated according to two settings that are defined by varying combinations of values for the parameters (n, τ, p) , for a total of $3 \times 3 \times 2 = 18$ scenarios. Table A.1 in Appendix A.1.4 outlines the parameters combination of the simulation setup. In all scenarios, we set the number of the outcomes to $d = 2$. For a given scenario, we simulate the data according to our proposed model given in equation (2.6) (i.e., Setting 1). For the model misspecification concerns, we also considered data generation under the C-logit model (i.e., Setting 2). Detail of the simulation mechanisms under these two settings is given next.

Setting 1 : for each scenario and for each individual independently, we simulated the data according to the following mechanisms :

1. Generate $\mathbf{X} = (X_1, X_2) \in \mathbb{R}^2$ from marginal parametric distributions : $X_k \sim F_k(\cdot)$, for $k = 1, 2$, where $F_k(\cdot)$ is either the standard normal or the Gamma(2, 3) distribution.

2. For $j = 1, 2$, set

$$\pi_j(\mathbf{x}) = \mathcal{C}_{\theta_j}^c \left(1 - p_j \mid \mathbf{F}(\mathbf{x}) \right),$$

with $1 - \pi_j(\mathbf{x}) = \mathbb{P}(Y_j = 1 \mid \mathbf{X} = \mathbf{x})$ and $\mathcal{C}_{\theta_j}(\cdot)$ is a specific copula. The marginal probabilities of success, $p_j = \mathbb{P}(Y_j = 1)$, for $j = 1, 2$, are population-level parameters, and thus, they do not vary for each subject. For all scenarios of our simulation study, we set $p_1 = 0.3$ and $p_2 = 0.4$.

3. Generate (U_1, U_2) from a copula \mathcal{C}_γ , where $\gamma \in \mathbb{R}^q$.

4. For $j = 1, 2$, set $Y_j = \mathbb{1}(U_j \geq \pi_j(\mathbf{x}))$, where $\mathbb{1}(A)$ denotes the indicator function of event A .

Setting 2 : all steps of this setting are similar to the steps of Setting 1, except step 2 where the (marginal) conditional probabilities, $\pi_j(\mathbf{x})$, for $j = 1, 2$, are calculated based on the logistic regression model as follows :

$$\pi_j(\mathbf{x}) = 1 - \left(1 + \exp\{-\beta_{j0} - \mathbf{x}^\top \boldsymbol{\beta}_j\} \right)^{-1},$$

where $\boldsymbol{\beta}_j$ are regression coefficients that link the covariates to each j -th binary outcome, and β_{j0} 's are intercept parameters.

For both settings, the dependence structure between (U_1, U_2) was induced using a selected Kendall's τ value and either a Gaussian, Clayton, Gumbel or Joe copula. Kendall's τ has a one-to-one relationship with the copula parameter γ . In Setting 1, for $j = 1, 2$, the dependence structure between each Y_j and the covariates was also induced using a selected Kendall's τ_j value and either a Gaussian, Clayton, Gumbel or Joe copula. Our selected τ and τ_j values $\tau = \tau_j = (0.3, 0.5, 0.7)$, correspond to weak, moderate, and strong dependency, respectively.

For Setting 2, we fixed the intercept coefficients β_{j0} at 0.3 in all scenarios. The association between each Y_j and the covariates was induced through the regression vector parameter $\boldsymbol{\beta}_j$. For the single-covariate

scenarios (i.e., $p = 1$), we set $\beta_{11} = 0.5$ and $\beta_{21} = 0.3$, and for scenarios with $p = 2$, we set $\beta_1 = (\beta_{11}, \beta_{12}) = (0.5, 0.2)$ and $\beta_2 = (\beta_{21}, \beta_{22}) = (0.3, 0.7)$.

To evaluate the finite-sample performance of the proposed estimators of $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\cdot)$, we generate $K = 1000$ independent datasets $\{(Y_i^{(k)}, \mathbf{X}_i^{(k)})_{i=1}^n\}_{k=1}^K$ under identical parameter configurations. Additionally, we draw a fixed grid of $m = 100$ covariate vectors $\mathbf{x}_1, \dots, \mathbf{x}_m$ independently from the distribution of \mathbf{X} . We repeat this analysis for three different sample sizes $n \in \{100, 300, 500\}$ to investigate finite-sample convergence. The estimator's accuracy is assessed by computing an empirical version of the integrated mean-squared error (IMSE) (Parzen, 1962) evaluated over the m covariate points $\mathbf{x}_1, \dots, \mathbf{x}_m$.

$$\widehat{\text{IMSE}} = \frac{1}{m} \sum_{i=1}^m \left\{ \frac{1}{K} \sum_{k=1}^K \left(\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x}_i) - \widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d(k)}(\mathbf{x}_i) \right)^2 \right\}, \quad (2.33)$$

where $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x}_i)$ denotes the true probability at covariates point \mathbf{x}_i and $\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d(k)}(\mathbf{x}_i)$ its estimate from the k -th replication.

The estimator $\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d(k)}(\cdot)$ is calculated based on the two proposed methods: the plug-in estimator (*plug-in*) described in Section 2.4 and the estimator based on vine copulas (*vine-based*) described in Subsection 2.3.2.4. The *plug-in* estimator was fitted with the “true” copula that was used for the data generation process. However, for the *vine-based* estimator, a suitable vine copula decomposition is chosen based on AIC criterion, and then the parameters of the vine decomposition are estimated. To compute the proposed estimators, we relied on `bicop` and `vinecop` functions from R package software `rvinecopulib` (Nagler et Vatter, 2018), which is suitable for vine decomposition in the presence of mixed variables. Our estimators are implemented in R software. For both settings, results were additionally obtained for the VGLM and the C-logit approaches. The VGLM method is implemented in R (Yee, 2015) in the `VGAM` library. The C-logit approach was implemented using our own R code. All the simulation studies are implemented in R software.

2.6.2 Simulation results

As a first illustration of the simulation study, we contrast in Figure A.1 the “true” probability parameter of interest, $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$, and the plug-in estimator, $\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$, based on the proposed method.

Figure A.1 results are based on a single run of data generated under Setting 1, with the following parameters configuration: $n = 300$, $\tau = \tau_j \in \{0.3, 0.5, 0.7\}$, $d = 2$ and $p = 1$. Figure A.1 shows the confidence interval of the proposed estimator, which is obtained based on the parametric bootstrap. Algorithm A.1 gives more details concerning the bootstrap technique that was used for the confidence interval calculation. As one can see from Figure A.1, the proposed estimator approximates well the true probability, and the latter lies

within the confidence interval band, all along the support of the x -axis. This preliminary result demonstrates the good performance of the proposed approach on estimating the true probability in the presence of one covariate (i.e., $p = 1$).

The simulation results of Settings 1 and 2 are presented in Tables 2.1 and 2.3 when $p = 1$, and in Tables 2.4 and 2.2 for $p = 2$. As one can see from these tables, for all scenarios, the IMSE decreases as the sample size, increases for all estimators, as expected. Notably, the proposed copula-based approach clearly outperforms the C-logit and VGML models, particularly when the data is generated from the proposed copula models (i.e., Settings 1, tables 2.1 and 2.2). In the scenarios where data is generated from C-logit model (i.e., setting 2 tables 2.3 and 2.4), the proposed estimator slightly under performs compared to the C-logit estimator. However, in cases where the data is generated from the proposed copula model, both the vine-based and the plug-in estimators consistently outperform the C-logit and VGML models across all scenarios. This difference is particularly significant when Kendall's tau is moderate or large ($\tau = 0.5$ or 0.7). Finally, one can notice that, interestingly, the vine-based estimator exhibits valid results for all scenarios.

Table 2.1 – Setting 1 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications, for $d = 2, p = 1$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.417	0.114	0.055	0.229	0.115	0.064	0.23	0.277	0.089
vine-based	1.988	1.923	1.666	0.999	1.030	0.900	1.528	1.425	1.250
VGML	4.123	3.683	3.787	8.723	8.593	8.487	11.779	11.503	12.08
C-logit	2.309	2.208	2.110	1.561	1.610	1.420	2.688	2.498	2.591
Joe									
plug-in	1.353	1.122	1.073	1.499	1.161	1.071	1.069	0.813	0.186
vine-based	2.244	1.986	2.158	4.453	4.405	3.869	4.105	4.124	4.304
VGML	3.256	3.654	4.728	9.526	7.995	9.435	18.640	14.860	15.920
C-logit	6.345	3.013	3.337	5.333	5.145	5.149	5.536	5.915	5.583
Gumbel									
plug-in	0.296	0.106	0.057	0.197	0.087	0.043	0.258	0.073	0.049
vine-based	1.568	1.546	1.500	2.348	2.195	2.145	2.326	2.124	2.084
VGML	3.156	4.248	3.530	7.616	8.783	8.187	14.189	14.348	14.065
C-logit	2.768	2.726	2.600	3.456	3.284	3.234	3.313	3.106	3.067
Normal									
plug-in	1.285	0.903	0.834	1.642	1.236	1.845	1.713	1.060	1.084
vine-based	2.968	2.715	2.811	2.548	2.953	2.923	2.138	2.404	2.517
VGML	6.757	5.395	5.265	9.071	10.497	10.404	18.478	18.470	18.420
C-logit	3.597	3.790	3.643	3.948	3.466	3.652	4.664	4.114	3.755

Table 2.2 – Setting 1 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications where $d = 2, p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.695	0.577	0.568	1.015	0.895	0.869	2.289	1.844	1.850
<i>vine-based</i>	2.197	2.046	1.903	1.897	1.857	1.834	4.413	4.656	8.530
<i>VGLM</i>	3.761	4.001	3.894	9.443	9.018	8.989	14.122	15.524	13.994
<i>C – logit</i>	2.114	2.110	2.002	6.140	5.592	5.756	6.432	9.794	9.014
Joe									
plug-in	1.340	1.221	1.179	2.839	2.586	2.162	2.455	2.374	2.209
<i>vine – based</i>	2.792	2.767	3.717	3.140	2.781	3.128	2.761	2.949	2.499
<i>VGLM</i>	2.268	3.131	2.008	3.213	3.22	4.934	11.362	12.405	11.452
<i>C-logit</i>	4.298	4.822	4.304	3.67	3.315	3.133	4.963	4.865	4.537
Gumbel									
plug-in	1.322	1.321	1.249	2.970	2.109	1.900	4.376	3.406	2.698
<i>vine-based</i>	3.419	3.430	3.394	3.528	4.200	3.411	6.287	6.408	6.694
<i>VGLM</i>	2.136	2.086	2.120	5.115	5.339	5.241	10.892	11.570	10.944
<i>C – logit</i>	3.252	1.180	1.184	2.475	2.359	2.355	5.921	6.002	5.866
Normal									
plug-in	0.648	0.088	0.170	0.221	0.065	0.036	0.194	0.0530	0.033
<i>vine-based</i>	1.490	1.052	0.976	0.964	0.644	0.683	0.884	0.829	0.785
<i>VGLM</i>	6.280	5.997	5.920	11.73	11.955	11.948	16.46	16.442	18.331
<i>C-logit</i>	2.065	1.0811	1.084	1.312	0.926	0.944	1.402	1.548	1.265

Table 2.3 – Setting 2 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications where $d = 2, p = 1$.

	$\tau = 0.3$			$\tau = 0.5$			$\tau = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
<i>vine-based</i>	2.050	2.180	1.924	1.320	1.980	1.950	3.710	3.950	3.910
<i>VGLM</i>	3.921	4.152	4.033	9.871	9.421	9.382	7.452	12.082	8.362
<i>C-logit</i>	2.352	2.331	2.212	6.472	5.892	6.052	6.752	10.282	9.462
Joe									
<i>vine-based</i>	2.850	2.720	2.680	3.110	2.850	2.880	2.920	2.850	2.790
<i>VGLM</i>	2.492	3.382	2.172	3.462	3.472	5.312	6.782	6.822	6.882
<i>C-logit</i>	4.622	5.182	4.632	3.952	3.572	3.372	5.332	5.232	4.882
Gumbel									
<i>vine-based</i>	1.952	1.852	1.782	2.452	2.322	2.182	3.892	3.722	3.612
<i>VGLM</i>	2.352	2.292	2.332	5.422	5.652	5.552	11.422	12.152	11.492
<i>C-logit</i>	3.522	2.552	2.562	2.822	2.692	2.682	6.222	6.302	6.162
Normal									
<i>vine-based</i>	0.922	0.952	0.782	1.092	0.612	0.492	0.522	0.412	0.322
<i>VGLM</i>	2.752	3.462	6.342	2.552	2.052	1.942	2.552	5.522	4.642
<i>C-logit</i>	2.752	2.732	1.722	2.242	2.052	2.042	1.902	1.242	1.012

Table 2.4 – Setting 2 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications where where $d = 2, p = 2$.

	$\tau = 0.3$			$\tau = 0.5$			$\tau = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
<i>vine-based</i>	3.675	2.581	2.526	5.788	6.021	7.162	1.852	3.937	2.565
<i>VGLM</i>	7.279	5.453	4.93	8.959	10.375	13.434	4.163	8.272	4.979
<i>C-logit</i>	6.140	4.388	4.185	8.145	10.76	13.451	3.116	7.453	4.296
Joe									
<i>vine-based</i>	1.88	1.258	2.189	2.60	2.536	2.551	5.706	5.5462	5.456
<i>VGLM</i>	2.32	1.934	1.895	2.794	2.696	3.1279	6.032	5.790	5.840
<i>C-logit</i>	2.295	2.230	2.189	3.586	3.36	3.432	6.80	6.272	6.678
Gumbel									
<i>vine-based</i>	4.032	4.379	3.550	4.379	5.143	3.926	3.951	3.538	3.226
<i>VGLM</i>	4.723	3.955	3.818	4.926	5.835	6.026	7.183	7.44	8.017
<i>C-logit</i>	3.252	4.379	3.937	4.821	4.667	3.933	5.848	5.043	5.673
Normal									
<i>vine-based</i>	1.223	0.635	0.508	1.335	0.640	0.323	1.012	0.642	0.511
<i>VGLM</i>	1.234	1.66	1.526	1.470	0.665	0.370	1.640	0.666	0.417
<i>C-logit</i>	2.874	2.967	2.760	4.513	2.995	3.055	3.170	3.001	2.859

Additionally, to test robustness of the proposed methods under low event probabilities, we explored additional scenarios derived from Settings 1, with a particular emphasis on unbalanced data, where the binary outcome, Y_1 , exhibits a notably low probability of success; i.e., we set $p_1 = 0.05$ as probability of a rare event in Setting 1.

Moreover, we examined two additional scenarios involving three binary outcomes ($d = 3$) and either two or three covariates ($p = 2, 3$), drawing from the data-generation structures outlined in both Settings 1 and 2. In both scenarios, we added an additional outcome variable, Y_3 , with an associated probability of occurrence $p_3 = 0.2$, and a covariate $X_3 \sim \mathcal{N}(0, 1)$. For Setting 2, we set the regression coefficients as $\beta_1 = (0.5, 0.2, 0.4)$, $\beta_2 = (0.3, 0.7, 0.6)$, and $\beta_3 = (0.1, 0.4, 0.2)$. Finally, we also considered the following configuration scenario of $(\tau, \tau_j) \in \{0.3, 0.5, 0.7\} \times \{0.3, 0.5, 0.7\}$ for Setting 1. For all these scenarios, data replications are generated following the same steps described in Setting 1 and Setting 2.

The resulting $\widehat{\text{IMSE}}$ values for these additional scenarios are detailed separately in the Appendices A.1.6.1, A.1.6.2, and A.1.6.3. From these results, we observe that the proposed plug-in and vine-based methods perform effectively in terms of the empirical integrated mean squared error. Specifically, both the plug-in and vine-based approaches consistently yield significantly lower $\widehat{\text{IMSE}}$ values, confirming that our methods maintain superior performance and robustness, even in scenarios with low event probabilities. Further-

more, we report the bias, empirical standard deviation (SD), average standard error (SE), and coverage probabilities in Tables A.14 to A.19. These tables indicate that, for the proposed plug-in and vine-based approaches, across all considered scenarios, the bias consistently remains below 0.001, empirical SD and average SE differ by less than 0.001, and coverage probabilities range between 90.4% and 95.3%. These results compare favorably with competitive estimators such as VGLM and C-logit. Consequently, we conclude that our proposed plug-in and vine-based estimators outperform the competing methods in terms of bias, variance, and coverage probability.

2.7 Real data illustration

To demonstrate the utility of our approach, we conducted an analysis of data from the Western Collaborative Group Study (WCGS). In our context, we modeled the probability describing the relationship between the corneal arcus and Coronary heart disease given the age and body mass index, using our approach.

2.7.1 Data description

The Western Collaborative Group Study (WCGS) (Rosenman *et al.*, 1975) was established as a forward-looking cohort investigation, focusing on middle-aged men (ages 39 to 59) who were employees of ten different companies based in California. The aim of the study was to investigate the risk factors of cardiovascular diseases (e.g., coronary heart disease, CHD), such as age, body mass index (BMI), systolic and diastolic blood pressures, and smoking habits. The data consists of 3,154 individuals that were enrolled in 1960 and 1961, for which both the CHD disease status and risk factors were measured. In addition, corneal arcus status, which was also reported for the subjects, indicates that an association between corneal arcus and potential cardiovascular risk factors in middle-aged men has been demonstrated to be strongly significant in several studies like (Fernández *et al.*, 2007). Thus, our analysis aims to illustrate the use of the proposed multiple binary regression approach to evaluate the joint probability between Coronary heart disease (Y_1) and corneal arcus (Y_2), conditional on two potential risk factors : age (X_1) and BMI (X_2). Formally, we aim to evaluate the probability of having both Coronary heart disease and corneal arcus as a function of age and BMI as predictors. Comparison is made again with the C-logit and VGLM approaches to investigate which method fits accurately the WCGS dataset. To this end, we use the Pearson statistic, which compares the observed frequencies (e.g., number of subjects with $(y_1, y_2) = (0, 0)$) to their corresponding estimated frequencies using the three approaches, to show which method estimates correctly $\mathbb{M}_{(y_1, y_2)}^{1,2}(\mathbf{x})$. We also conducted a classification analysis to assess the methods' performance in classifying the subjects correctly

to their true bivariate state (Y_1, Y_2) . The results are presented next in section 2.7.2.

2.7.2 Results

First illustration ($d = 2, p = 1$): we use age as a single covariate to calculate the probabilities of the events $\{(y_1, y_2)\} \in \{0, 1\}^2$; i.e., calculate the probabilities $\hat{M}_{(0,0)}^{1,2}, \hat{M}_{(0,1)}^{1,2}, \hat{M}_{(1,0)}^{1,2}$, and $\hat{M}_{(1,1)}^{1,2}$, based on the three competitive methods. For our method, we use the AIC criterion to choose the copula models that fit well the data in this illustration example. The left panel of Figure A.2 in Appendix A.1.5 shows $\hat{M}_{(y_1, y_2)}^{1,2}$ as a function of age. From this panel, we observe that the probability of having both diseases, $\hat{M}_{(1,1)}^{1,2}$, increases with age. Similar behavior can be observed for $\hat{M}_{(1,0)}^{1,2}$. This result is in concordance with the main finding in Rosenman *et al.* (1975) that heart disease is positively correlated with age. On the other hand, $\hat{M}_{(0,0)}^{1,2}$ decreases with age while $\hat{M}_{(0,1)}^{1,2}$ seems to be less affected by age.

To assess the proposed method's ability to estimate the probability $M_{(y_1, y_2)}^{1,2}(\mathbf{x})$, we conduct goodness-of-fit tests to evaluate how well the three models fit the WCGS data. Thus, we compare the empirical relative frequencies $\mathcal{O}_{(0,0)} = n_{00}/n, \mathcal{O}_{(0,1)} = n_{01}/n, \mathcal{O}_{(1,0)} = n_{10}/n$, and $\mathcal{O}_{(1,1)} = n_{11}/n$, to their corresponding estimates using each method, where n_{00} is the number of subjects without both diseases and n is the total number of subjects in the study. The other quantities, n_{01}, n_{11} , and n_{11} , are defined in a similar way. The estimated relative frequencies are defined as, $\mathcal{E}_{(0,0)} = \frac{\sum_{i=1}^n \hat{M}_{(0,0)}^{1,2}(\mathbf{x}_i)}{n}$, $\mathcal{E}_{(0,1)} = \frac{\sum_{i=1}^n \hat{M}_{(0,1)}^{1,2}(\mathbf{x}_i)}{n}$, $\mathcal{E}_{(1,0)} = \frac{\sum_{i=1}^n \hat{M}_{(1,0)}^{1,2}(\mathbf{x}_i)}{n}$, and $\mathcal{E}_{(1,1)} = \frac{\sum_{i=1}^n \hat{M}_{(1,1)}^{1,2}(\mathbf{x}_i)}{n}$, where \mathbf{x}_i is the observed covariates for subject i . Then, we calculate the following Pearson type statistic :

$$\chi^2 = \sum_{\mathbf{y}} \frac{(\mathcal{O}_{\mathbf{y}} - \mathcal{E}_{\mathbf{y}})^2}{\mathcal{E}_{\mathbf{y}}}.$$

Table 2.5 shows a significant decrease of the statistic χ^2 when using our method compared to both C-logit and VGLM methods.

Table 2.5 – Observed frequencies of dependent variables and the average of estimated probability of each class in presence of $d = 2, p = 1$ (Age).

chd	arcus	obs	proposed method	C-logit	VGLM
0	0	0.652	0.671	0.684	0.752
1	0	0.266	0.243	0.233	0.048
0	1	0.048	0.0247	0.016	0.266
1	1	0.032	0.059	0.060	0.022
		χ^2	0.0264	0.051	0.817

Table 2.6 – Observed frequencies of dependent variables and the average of estimated probability of each class $d = 2, p = 2$ (Age, BMI).

chd	arcus	obs	Proposed method	C-logit	VGLM
0	0	0.652	0.644	0.682	0.69
1	0	0.266	0.274	0.236	0.06
0	1	0.048	0.056	0.018	0.225
1	1	0.032	0.024	0.064	0.025
		χ^2	0.0041	0.083	0.812

Second illustration ($d = 2, p = 2$) : we conduct a second analysis, where we add BMI as a second covariate to Age, to assess the joint probabilities $\hat{M}_{(y_1, y_2)}^{1,2}$ for the four classes. In this case, we use the vine decomposition for copula model selection. The C-vine structure is selected as a suitable vine decomposition to model the joint distribution of both (Y_1, X_1, X_2) and (Y_2, X_1, X_2) , based on AIC criterion. Table 2.6 reports the results of this analysis. From this table, one can see that the incorporation of BMI into the model improves the accuracy of our method based on χ^2 statistic. In contrast, both the C-logit and the VGLM approaches still exhibit unsatisfactory performance compared to our proposed method. Also, Table 2.6 shows analysis based on our approach using D-vine decomposition to model the joint distribution of both (Y_1, X_1, X_2) and (Y_2, X_1, X_2) . Moreover, Tables A.20, A.21, and A.22, provided in Appendix A.1.8, detail the selected pair-copula families employed in the vine construction associated with the density vectors $(Y_1, \text{Age}, \text{BMI})$,

$(Y_2, \text{Age}, \text{BMI})$, and the optimal copula chosen for modeling the conditional marginals $Y_1 | (\text{Age}, \text{BMI})$ and $Y_2 | (\text{Age}, \text{BMI})$, respectively.

Additionally, we perform a classification analysis for the second illustrative example, using Age and BMI as covariates. Specifically, after estimating the probabilities $\hat{M}_{(y_{i1}, y_{i2})}^{1,2}(\mathbf{x}_i)$ by each method, we classify each subject i into one of the four categories : $\{(0, 0), (1, 1), (1, 0), (0, 1)\}$. This classification is accomplished using the inversion method, which generates binary outcomes based on cumulative probabilities, serving as the classification criteria.

The approach involves calculating the joint probabilities for Y_1 and Y_2 , using each method, and defining cumulative probability thresholds as follows :

$$\begin{aligned} C_{00} &= \mathbb{P}(Y_1 = 0, Y_2 = 0 \mid \mathbf{X} = \mathbf{x}), \\ C_{01} &= C_{00} + \mathbb{P}(Y_1 = 0, Y_2 = 1 \mid \mathbf{X} = \mathbf{x}), \\ C_{10} &= C_{01} + \mathbb{P}(Y_1 = 1, Y_2 = 0 \mid \mathbf{X} = \mathbf{x}), \\ C_{11} &= C_{10} + \mathbb{P}(Y_1 = 1, Y_2 = 1 \mid \mathbf{X} = \mathbf{x}) = 1. \end{aligned}$$

Once these cumulative probabilities are evaluated, using our method as well as by the competing methods (i.e., C-logit and VGLM), we draw a uniformly distributed random number $u \in [0, 1]$. We then compare the value of u to the cumulative thresholds to determine the outcome pair $\{y_{i1}, y_{i2}\}$:

$$\begin{aligned} (Y_1, Y_2) &= (0, 0), & \text{if } u \leq C_{00}, \\ (Y_1, Y_2) &= (0, 1), & \text{if } C_{00} < u \leq C_{01}, \\ (Y_1, Y_2) &= (1, 0), & \text{if } C_{01} < u \leq C_{10}, \\ (Y_1, Y_2) &= (1, 1), & \text{if } C_{10} < u \leq 1. \end{aligned} \tag{2.34}$$

To reduce the variability induced by random sampling, we repeat the random selection of the auxiliary variable, $u \sim \text{Unif}(0, 1)$, for a total of $R = 100$ times. For each iteration r , we assign the provisional class $(\hat{Y}_{1r}, \hat{Y}_{2r})$ according to the classification rule described in (2.34). The final class assignment corresponds to the most frequently occurring class among the R replications. This process ensures that the binary outcomes Y_1 and Y_2 are generated in a way that respects both their marginal probabilities and their joint dependence. For each method, the resulting matrices of Y_1 and Y_2 , representing coronary heart disease and corneal arcus respectively, are presented in Table 2.7.

Table 2.7 - Observed and model-based joint state estimates for (Y_1, Y_2) with $d = 2, p = 2$.

(a) Observed state			(b) Estimated state (Vine based)		
$Y_1 \backslash Y_2$	0	1	$Y_1 \backslash Y_2$	0	1
0	1987	863	0	1983	870
1	53	251	1	50	246

(c) Estimated state (C-logit)			(d) Estimated state (VGLM)		
$Y_1 \backslash Y_2$	0	1	$Y_1 \backslash Y_2$	0	1
0	1969	871	0	2003	850
1	46	259	1	44	242

Table 2.8 - Error rates of classification for each joint state (Y_1, Y_2) .

(a) $Y_1 = 0, Y_2 = 0$		(b) $Y_1 = 1, Y_2 = 1$	
Method	Error	Method	Error
Vine based	0.002	Vine based	0.019
C-logit	0.009	C-logit	0.031
VGLM	0.008	VGLM	0.035

(c) $Y_1 = 1, Y_2 = 0$		(d) $Y_1 = 0, Y_2 = 1$	
Method	Error	Method	Error
Vine based	0.056	Vine based	0.081
C-logit	0.075	C-logit	0.093
VGLM	0.170	VGLM	0.150

As shown by the results of Tables 2.7 and 2.8 , all methods are highly accurate. However, the proposed

method, based on vine decomposition, outperforms both the VGLM and C-logit approaches. Moreover, the proposed approach demonstrates good accuracy in classifying less frequent events, such as $(1, 0)$ and $(1, 1)$, compared to C-logit and VGLM. This improvement is important in predictive modeling.

In general, the error rates presented in Table 2.8 indicate that our proposed method performs better than C-logit and VGLM across all event states of Y_1 and Y_2 . In contrast, the VGLM method exhibits the highest error rates, especially for the events $Y_1 = 1, Y_2 = 0$ and $Y_1 = 0, Y_2 = 1$. While the C-logit method performs better than VGLM but worse than our proposed method based on vine. In conclusion, the proposed method seems to better capture these rarer occurrences compared to the other candidates.

2.8 Conclusion

In this work, we developed a flexible and unified approach for modeling the dependence structure between binary outcomes in the multivariate regression framework. We modeled the conditional distribution of each outcome given the set of covariates using a link function that is derived based on copulas. Then, we modeled the dependence between conditional margins (i.e., dependence between residuals) via an appropriate copula. The proposed method, under specific copula families, generalizes several existing approaches, such as the methods developed in Song *et al.* (2009), Genest *et al.* (2013), and Chib et Greenberg (1998).

Subsequently, we develop a semi-parametric estimator of $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$, that assesses the probability of dependent binary outcomes based on a given set of covariates. We also derive the asymptotic properties of $\hat{\mathbb{M}}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$. Finally, both the simulation study and the real data analysis demonstrate the performance of our proposed method compared to competing approaches. Both the simulation study and the real data analysis demonstrate the performance of our proposed method compared to competing approaches. Interestingly, our vine-based approach shows similar performance as the “true” plug-in estimator. However, vine construction needs to be chosen carefully. Choosing between a *C-vine* and a *D-vine* should reflect the underlying dependence structure, and prior knowledge of such construction can decisively tip the balance. A *C-vine* is preferable when a single “hub” variable drives most pairwise relationships, because its star-shaped first tree concentrates information and leads to accurate estimation of the dependence parameters. By contrast, a *D-vine* is better suited to naturally ordered variables such as time series, spatial grids, or biochemical pathways because its chain topology captures nearest-neighbor effects more efficiently. In practice, one fits both structures and selects the one with the lower information criterion (e.g., AIC); if the scores are similar, a *D-vine* is slightly advantageous in low dimensions ($d \leq 5$), whereas a *C-vine* scales better for $d > 5$ as investigated by Czado et Nagler (2022). While this study operates under the assumption that covariates

are continuous, we are actively exploring extensions to accommodate mixed covariates.

CHAPITRE 3

EXTENSION OF MULTIVARIATE BINARY OUTCOMES COPULAS-REGRESSION PROBLEM FOR MIXED COVARIABLES

Ce Chapitre étend le travail présenté au Chapitre 2. Plus précisément, nous généralisons le cadre initial en tenant compte de covariables mixtes. Cette démarche s'inspire des recherches récentes de Aldahmani *et al.* (2024), qui ont enrichi la régression basée sur les copules proposée par Noh *et al.* (2013), en y intégrant des covariables mixtes. Dans un premier temps, nous exprimons la probabilité $\mathbb{P}(Y = 0 \mid \mathbf{X} = \mathbf{x})$ en présence des covariables mixtes, en adoptant la méthode développée par Aldahmani *et al.* (2024). Ensuite, la dépendance entre les variables réponses est modélisée à l'aide d'une copule appropriée, ce qui permet d'obtenir la probabilité jointe $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ telle que définie dans l'équation (2.6). Enfin, nous illustrons notre approche en nous appuyant sur des familles de copules spécifiques telles que les copules archimédiennes et la copule gaussienne. Ces exemples mettent en évidence les propriétés distinctes de ces copules ainsi que leurs avantages dans le cadre de notre analyse. Une étude de simulation est également menée afin d'évaluer les performances de l'approche proposée.

3.1 Abstract

This paper extends the framework introduced in Chapter 2 to model the dependence structure between multivariate binary outcomes and mixed covariates, encompassing both continuous and discrete variables. Building upon the approach of Aldahmani *et al.* (2024), we adapt the copula-based link function originally proposed by Mesfioui *et al.* (2023) for binary regression to enhance the estimation of the conditional cumulative distribution function (CDF) for each binary outcome in the presence of both continuous and discrete covariates. To jointly model multiple outcomes, we utilize appropriate copulas capable of capturing their complex dependencies. Our framework incorporates flexible copula families, such as Archimedean copulas (e.g., Clayton, Frank, and Gumbel) and elliptical copulas (e.g., Gaussian), to effectively model conditional dependencies between binary outcomes and associated mixed covariates. Estimation is carried out through a semi-parametric approach, combining parametric copula models with non-parametric estimation of marginal distributions. We establish the theoretical properties of the proposed estimators and assess their performance through extensive simulation studies. The results show that our method consistently outperforms competitive approaches, such as C-logit (Genest *et al.*, 2013) and VGLM (Zhao et Prentice, 1990; Yee, 2015), across various types of dependence structures and particularly in scenarios involving rare events often re-

ferred to in the literature as imbalanced data.

3.2 Introduction

In numerous scientific disciplines, researchers are frequently tasked with examining the relationship between multiple binary outcomes and a set of explanatory variables. While conventional regression techniques are widely utilized across different fields such as medicine, biology, agriculture, economics, and sociology, their applicability can be limited when the focus shifts to model multiple binary responses, particularly in the presence of nonlinear dependencies or deviations from standard distributional assumptions. This paper presents a methodological strategy designed to effectively capture dependency structures among multiple binary responses, particularly accommodating both continuous and discrete components. By explicitly separating marginal effects from joint dependencies, our approach facilitates a more precise and nuanced analysis of complex data environments.

In this context, we emphasize methodologies that represent multivariate distributions through their univariate marginals by employing copulas (Sklar, 1959). Copulas offer considerable flexibility for modeling multivariate distributions, enabling the dependency structure between variables to be analyzed independently from their marginal distributions. By selecting an appropriate copula, one can effectively capture complex dependencies among the marginals. For a comprehensive review, see (Nelsen, 2006). This methodology is particularly powerful because it isolates the dependence structure from marginal behaviors, thereby offering exceptional versatility. Additionally, copulas are adept at modeling both linear and nonlinear dependencies, handling arbitrary marginal distributions, and capturing extreme value behavior effectively. A further advantage of copula-based regression, particularly highlighted for continuous variables in (Noh *et al.*, 2013), is the absence of constraints regarding the specification of probability distributions, thus ensuring remarkable flexibility in statistical modeling. Recently, (Mesfioui *et al.*, 2023) introduced a novel copula-based link function tailored specifically for modeling dependencies between a univariate binary outcome and a vector of continuous covariates. This approach generalizes traditional methods such as probit and logit models; notably, the probit model emerges from the Gaussian copula, whereas the logit model corresponds to the Frank copula. In this Chapter, we extend the framework introduced in Chapter 2 by incorporating discrete variables into the set of covariates X_1, \dots, X_p . This is achieved by proposing an alternative formulation for the following equation :

$$\mathbb{P}(Y_1 \leq y_1, \dots, Y_d \leq y_d | \mathbf{X} = \mathbf{x}) := \mathcal{C}_\gamma \left(F_1(y_1 | \mathbf{x}), \dots, F_d(y_d | \mathbf{x}) \right), \quad (3.1)$$

where γ represents the dependence between the outcomes. Additionally, we investigate this relationship for specific copula families, such as Archimedean copulas and the Gaussian copula, emphasizing their properties and advantages that are particularly relevant to our analysis.

The remainder of this work is organized as follows : Section 3.3 provides the mathematical representation of equation (3.1) in the presence of mixed covariates including both continuous and discrete variables. Section 3.4 illustrates the predictive probabilities in the presence of mixed variables using well-known copula families, such as Archimedean and Elliptical copulas. Section 3.5 outlines the estimation procedure for the proposed methods. Section 3.6 presents a simulation study to assess the performance of the proposed method. Finally, conclusions and remarks are provided in Section 3.7.

3.3 Model formulation

In this section, we present a mathematical formulation of Equation (3.1) that accommodates mixed covariates. Additionally, we derive a closed-form expression for this representation using several widely adopted copula families. Before presenting the complete model specification, we introduce additional notation utilized throughout this work. This preliminary step allows us to rigorously define the copula density in the presence of mixed variables.

3.3.1 Predictive probability based on copulas with mixed variables

Consider the random vector (\mathbf{Y}, \mathbf{X}) , where $\mathbf{Y} = (Y_1, \dots, Y_d)$ is a collection of binary response variables and $\mathbf{X} = (X_1, \dots, X_p)$ is a vector of mixed covariates. For each response variable Y_j , $j = 1, \dots, d$, let G_j denotes its cumulative distribution function (CDF). Likewise, for each covariate X_k , $k = 1, \dots, p$, let F_k denotes its CDF. The vector of marginal CDFs of the covariates is therefore $\mathbf{F}(\mathbf{x}) = (F_1(x_1), \dots, F_p(x_p))$. We further partition the covariate vector as $\mathbf{X} = (\mathbf{X}_q, \mathbf{X}_{p-q})$, where $\mathbf{X}_q = (X_1, \dots, X_q)$ contains the continuous covariates, and $\mathbf{X}_{p-q} = (X_{q+1}, \dots, X_p)$ the discrete covariates. Throughout, we retain the notation $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x}) = \mathbb{P}(\mathbf{Y} = \mathbf{y} \mid \mathbf{X} = \mathbf{x})$, where $\mathbf{y} = (y_1, \dots, y_d)$. To evaluate the predictive probability defined in equation (3.1) in the presence of mixed variables, one must first specify the conditional marginal distribution $F_j(y_j \mid \mathbf{x})$ using copulas. However, assessing this conditional marginal distribution in the context of mixed variables requires defining a method to explicitly evaluate the copula density involving mixed variables. To this end, consider the following illustrative case where $y_j = 0$. In this scenario, the

conditional distribution is expressed as follows :

$$F_j(0|\mathbf{x}) = \pi_j(\mathbf{x}) = \frac{\mathbb{P}(Y_j = 0, \mathbf{X} = \mathbf{x})}{\mathbb{P}(\mathbf{X} = \mathbf{x})} := \frac{f^m(Y_j = 0, \mathbf{X} = \mathbf{x})}{f^m(\mathbf{X} = \mathbf{x})}, \quad (3.2)$$

here f^m symbolizes the density of a mixed random vector. Note that to evaluate the numerator and denominator in equation (3.2) based on copula, it is useful to introduce some notation to facilitate the definition of copula densities involving mixed variables. Let $(1 - p_j, \mathbf{F}(\mathbf{x})) = (1 - p_j, v_1, \dots, v_p) \in (0, 1)^{p+1}$, where $p_j = \mathbb{P}(Y_j = 1)$. In addition, we define the following operators :

$$\partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{v}) := \frac{\partial^q}{\partial v_1 \dots \partial v_q} \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{v}), \quad (3.3)$$

and

$$\partial_q \mathcal{C}_{\theta_j}(1, \mathbf{v}) := \frac{\partial^q}{\partial v_1 \dots \partial v_q} \mathcal{C}_{\theta_j}(1, \mathbf{v}), \quad (3.4)$$

where Equations (3.3) and (3.4) represent the q -th order derivative with respect to the first continuous variable of the copula distribution between $(1 - p_j, \mathbf{v})$ and among the variables $\mathbf{v} = (v_1, \dots, v_p)$, respectively.

Additionally, we define the forward difference operator along discrete coordinates, which serves to evaluate the mass function corresponding to the discrete covariates. Specifically, for each $l = q + 1, \dots, p$, let $\delta_l = (\delta_{l1}, \dots, \delta_{lp})^\top$, with $\delta_{ll} = 1$ and $\delta_{lh} = 0$ for all $l \neq h$. The forward difference operator is given by

$$\Delta_l \partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x})) := \partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x})) - \partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x} - \delta_l)).$$

Let $\Delta_{q+1}^p := \prod_{l=q+1}^p \Delta_l$ denote the $(p-q)$ -fold forward difference operator, with these necessary operators established, we can now rigorously define the copula density in the presence of mixed covariates.

Throughout this section, the discrete covariates X_{q+1}, \dots, X_p are assumed to take values in subsets of the integers, i.e. $\text{supp}(X_\ell) \subseteq \mathbb{Z}$, $\ell = q + 1, \dots, p$. This assumption (which can always be met after an appropriate integer coding of the discrete levels) is required for the definition of the unit forward difference operator Δ_ℓ based on $\mathbf{x} - \delta_\ell$.

Definition 3.1 (Definition of the mixed copula density) *The copula density of (Y_j, \mathbf{X}) , in the presence of mixed variables is specified as :*

$$c_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x})) := \Delta_{q+1}^p \partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x})), \quad (1 - p_j, \mathbf{F}(\mathbf{x})) \in (0, 1)^{p+1}. \quad (3.5)$$

After introducing the notation and describing how to evaluate the copula density in the presence of mixed covariates, one can reexpress (3.2), as specified in the following proposition :

Proposition 3.2 For all $\mathbf{x} \in \mathbb{R}^q \times \prod_{l=q+1}^p \text{Ran}(X_l)$, where $\text{Ran}(X_l)$ represents all possible realizations of the random variable X_l . The conditional probability of a specific outcome equal to zero $\mathbb{P}(Y_j = 0 | \mathbf{X} = \mathbf{x})$, is expressed as :

$$\pi_j(\mathbf{x}) = \mathcal{C}_{\theta_j}^{cm}(1 - p_j | \mathbf{F}(\mathbf{x})) := \frac{\Delta_{q+1}^p \partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x}))}{\Delta_{q+1}^p \partial_q \mathcal{C}_{\theta_j}(1, \mathbf{F}(\mathbf{x}))}. \quad (3.6)$$

The general form is expressed as follows :

$$F_j(y_j | \mathbf{x}) = \mathcal{C}_{\theta_j}^{cm}(G_j(y_j) | \mathbf{F}(\mathbf{x})) := \frac{\Delta_{q+1}^p \partial^q \mathcal{C}_{\theta_j}(G_j(y_j), \mathbf{F}(\mathbf{x}))}{\Delta_{q+1}^p \partial^q \mathcal{C}_{\theta_j}(1, \mathbf{F}(\mathbf{x}))}, \quad j = 1, \dots, d, \quad (3.7)$$

where $\mathcal{C}_{\theta_j}^{cm}(\cdot)$ represents the conditional distribution of copula in the presence of mixed covariates.

Now the equation (3.1) is fully specified with a simplified assumption, and its expression is given as follows :

$$\mathbb{P}(Y_1 \leq y_1, \dots, Y_d \leq y_d | \mathbf{X} = \mathbf{x}) := \mathcal{C}_{\gamma} \left(\mathcal{C}_{\theta_1}^{cm}(G_1(y_1) | \mathbf{F}(\mathbf{x})), \dots, \mathcal{C}_{\theta_j}^{cm}(G_d(y_d) | \mathbf{F}(\mathbf{x})) \right). \quad (3.8)$$

Referring to expression (3.8), the probability of observing the event $\mathbf{Y} = \mathbf{0}$ in the presence of mixed covariates can be written as follows :

$$\begin{aligned} \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) &= \mathcal{C}_{\gamma} \left\{ F_1(Y_1 = 0 | \mathbf{X} = \mathbf{x}), \dots, F_d(Y_d = 0 | \mathbf{X} = \mathbf{x}) \right\} \\ &= \mathcal{C}_{\gamma} \{ \pi_1(\mathbf{x}), \dots, \pi_d(\mathbf{x}) \}, \end{aligned} \quad (3.9)$$

where the expression of the $\pi_j(\mathbf{x})$ is provided in the Proposition 3.2. Note that relying on the result of Lemma 2.2. For every combination of binary outcomes vector $\mathbf{y} = (y_1, \dots, y_d)$ the following probability $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$ is fully expressed in terms of the following probabilities :

$$\left\{ \left\{ \mathbb{M}_0^j(\mathbf{x}), \right\}_{j=1, \dots, d}, \left\{ \mathbb{M}_{(0,0)}^{j,k}(\mathbf{x}) \right\}_{j < k}, \left\{ \mathbb{M}_{(0,0,0)}^{j,k,e}(\mathbf{x}) \right\}_{j < k < e}, \dots, \mathbb{M}_{(0,0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) \right\}.$$

Throughout the remainder of this work, we focus exclusively on estimating and evaluating the following probability $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ where its expression is provided in equation (3.9).

After specifying the expression of $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$, in the context of mixed covariates, we proceed to provide a representation of predictive probabilities based on useful families of copulas, such as Archimedean and Elliptical copulas.

3.4 Predictive probability based on common copulas families

3.4.1 Archimedean case

Assume that the dependence between each response Y_j , the p -dimensional covariate vector \mathbf{X} , and among the binary responses themselves, is modeled using an Archimedean copula (Genest et MacKay, 1986), characterized by its generator φ . Let φ^{-1} denotes the corresponding pseudo-inverse generator, which can be expressed explicitly as follows :

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

Here, ψ denotes the true inverse of the generator function φ , then the d -variate copula distribution is given by

$$\mathcal{C}(v_1, v_2, \dots, v_p) = \psi \left(\sum_{k=1}^p \varphi(v_k) \right). \quad (3.10)$$

Before expressing our predictive probability in the case where $\mathbf{Y} = \mathbf{0}$, we clarify that throughout this section the parameters γ and θ_j are treated as scalars, representing the dependence between the outcome and Y_j , and between the outcome and \mathbf{X} , respectively. In our context, based on Equation (3.10), the expression for $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x})$ can be explicitly written as follows :

$$\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \psi_\gamma \left(\sum_{j=1}^d \varphi_\gamma(\pi_j(\mathbf{x})) \right), \quad (3.11)$$

where

$$\pi_j(\mathbf{x}) = \frac{\Delta_q^p \psi_{\theta_j}^{(q)} \left(\varphi_{\theta_j}(1 - p_j) + \sum_{k=1}^q \varphi_{\theta_j}(F_k(x_k)) \right)}{\Delta_q^p \psi_{\theta_j}^{(q)} \left(\sum_{k=1}^q \varphi_{\theta_j}(F_k(x_k)) \right)}, \quad (3.12)$$

with φ_γ and ψ_γ representing, respectively, the generator function and its inverse associated with the copula \mathcal{C}_γ , and φ_{θ_j} and ψ_{θ_j} representing, respectively, the generator function and its pseudo-inverse associated with the copula $\mathcal{C}_{\theta_j}^{cm}$. To illustrate the computation of predictive probabilities using Archimedean copulas, consider the scenario where $p = 2$, with X_1 being continuous, X_2 being discrete, and $d = 2$. One can write

$$\pi_j(\mathbf{x}) = \frac{\Delta_1^2 \psi'_{\theta_j} \left(\varphi(1 - p_j) + \sum_{k=1}^2 \varphi_{\theta_j}(F_k(x_k)) \right)}{\Delta_1^2 \psi'_{\theta_j} \left(\sum_{k=1}^2 \varphi_{\theta_j}(F_k(x_k)) \right)}, \quad (3.13)$$

where the numerator in (3.13) is given as

$$\begin{aligned} \Delta_1^2 \psi'_{\theta_j} \{ \varphi_{\theta_j}(1 - p_j) + \varphi_{\theta_j}(F_1(x_1)) + \varphi_{\theta_j}(F_2(x_2)) \} &= \psi'_{\theta_j} \{ \varphi_{\theta_j}(1 - p_j) + \varphi_{\theta_j}(F_1(x_1)) + \varphi_{\theta_j}(F_2(x_2)) \} \\ &\quad - \psi'_{\theta_j} \{ \varphi_{\theta_j}(1 - p_j) + \varphi_{\theta_j}(F_1(x_1)) + \varphi_{\theta_j}(F_2(x_2 - 1)) \}, \end{aligned}$$

and the denominator in (3.13) can be written as

$$\begin{aligned} \Delta_1^2 \psi'_{\theta_j} \{ \varphi_{\theta_j}(F_1(x_1)) + \varphi_{\theta_j}(F_2(x_2)) \} &= \psi'_{\theta_j} \{ \varphi_{\theta_j}(F_1(x_1)) + \varphi_{\theta_j}(F_2(x_2)) \} \\ &\quad - \psi'_{\theta_j} \{ \varphi_{\theta_j}(F_1(x_1)) + \varphi_{\theta_j}(F_2(x_2 - 1)) \}. \end{aligned}$$

3.4.1.1 Clayton copula

To illustrate the predictive probability $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$ within the Clayton copula framework, we first define the copula generator and its inverse as follows :

$$\varphi_\gamma(z) = \frac{z^{-\gamma} - 1}{\gamma}, \quad \gamma > 0, \quad z \in (0, 1), \quad (3.14)$$

and

$$\psi_\gamma(t) = (\gamma t + 1)^{-\frac{1}{\gamma}}. \quad (3.15)$$

Based on Equations (3.14) and (3.15), the d -variate Clayton copula distribution \mathcal{C}_γ , modeling the dependence structure between outcomes, is expressed as :

$$\mathcal{C}_\gamma(\pi_1, \dots, \pi_d) = \left(\sum_{i=1}^d \pi_i^{-\gamma} - d + 1 \right)^{-\frac{1}{\gamma}}. \quad (3.16)$$

To complete the formulation of $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$, we assume that the dependence structure between a specific outcome Y_j and covariates \mathbf{X} , for $j = 1, 2$ and $p = 1, 2$, follows from expressions (3.16) and (3.13). Thus, $\pi_j(\mathbf{x})$ is explicitly given by :

$$\pi_j(\mathbf{x}) = \frac{\Delta_1^2 \left((1 - p_j)^{-\theta_j} + F_1(x_1)^{-\theta_j} + F_2(x_2)^{-\theta_j} - 2 \right)^{-\frac{1}{\theta_j} - 1}}{\Delta_1^2 \left(F_1(x_1)^{-\theta_j} + F_2(x_2)^{-\theta_j} - 1 \right)^{-\frac{1}{\theta_j} - 1}}. \quad (3.17)$$

Consequently, $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$ can be expressed as :

$$\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}) = \left(\pi_1(\mathbf{x})^{-\gamma} + \pi_2(\mathbf{x})^{-\gamma} - 1 \right)^{-\frac{1}{\gamma}}. \quad (3.18)$$

3.4.1.2 Frank copula

Illustrating our predictive probability under the Frank copula, we first define the generator and its inverse functions associated with this copula as follows :

$$\varphi_\gamma(z) = -\ln\left(\frac{e^{-\gamma z} - 1}{e^{-\gamma} - 1}\right), \quad \gamma \in \mathbb{R}^*, \quad z \in (0, 1). \quad (3.19)$$

$$\psi_\gamma(t) = -\frac{1}{\gamma} \ln(e^{-t}(e^\gamma - 1) + 1). \quad (3.20)$$

Using equations (3.19) and (3.20), the d -variate Frank copula modeling the dependence among the outcomes is given by :

$$\mathcal{C}_\gamma(\pi_1, \dots, \pi_d) = -\frac{1}{\gamma} \ln\left(1 + \frac{\prod_{i=1}^d (e^{-\gamma \pi_i} - 1)}{(e^{-\gamma} - 1)^{d-1}}\right), \quad \gamma > 0. \quad (3.21)$$

Furthermore, if the dependence between a specific outcome and the covariates (Y_j, X_1, X_2) , for $j = 1, 2$, is modeled using a Frank copula, then the function $\pi_j(\mathbf{x})$, based on Equations (3.19) and (3.20), is expressed as :

$$\pi_j(\mathbf{x}) = \frac{\Delta_1^2 K_{1,\theta_j}(1 - p_j, F_1(x_1), F_2(x_2))}{\Delta_1^2 K_{2,\theta_j}(F_1(x_1), F_2(x_2))}, \quad (3.22)$$

with

$$K_{1,\theta_j}(1 - p_j, F_1(x_1), F_2(x_2)) = \frac{(e^{-\theta_j(1-p_j)} - 1)(e^{-\theta_j F_2(x_2)} - 1)e^{-\theta_j F_1(x_1)}}{(e^{-\theta_j} - 1)^2 + (e^{-\theta_j(1-p_j)} - 1)(e^{-\theta_j F_1(x_1)} - 1)(e^{-\theta_j F_2(x_2)} - 1)},$$

and

$$K_{2,\theta_j}(v_1, v_2) = \frac{e^{-\theta_j v_1}(e^{-\theta_j v_2} - 1)}{(e^{-\theta_j} - 1) + (e^{-\theta_j v_1} - 1)(e^{-\theta_j v_2} - 1)}.$$

In our context, the expression for $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$ is given by :

$$\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}) = -\frac{1}{\gamma} \ln\left(1 + \frac{(e^{-\gamma \pi_1(\mathbf{x})} - 1)(e^{-\gamma \pi_2(\mathbf{x})} - 1)}{e^{-\gamma} - 1}\right), \quad \gamma \in \mathbb{R} \setminus \{0\}. \quad (3.23)$$

3.4.1.3 Gumbel Copula

In this case, we assume that both dependence across the outcomes and between each outcome Y_j ($j = 1, 2$) and the covariates (X_1, X_2) follows a Gumbel copula. In a general case the generator is expressed as follows :

$$\varphi_\gamma(t) = (-\ln(t))^\gamma, \quad t \in (0, 1), \quad \gamma \geq 1, \quad (3.24)$$

and the inverse generator is given by :

$$\psi_\gamma(u) = \exp(-u^{1/\gamma}), \quad u \geq 0, \quad (3.25)$$

where ϕ_γ , and φ_γ represent the generator and its inverse of \mathcal{C}_γ that holds the dependence between the conditional margin $\pi_j, j = 1, \dots, d$. Based on (3.24) and (3.25) the d-variate distribution of a Gumbel copula of \mathcal{C}_γ is given by

$$\mathcal{C}_\gamma(\pi_1, \dots, \pi_d) = \exp \left\{ - \left[\sum_{j=1}^d (-\ln(\pi_j))^\gamma \right]^{1/\gamma} \right\}, \quad \text{for } \gamma \geq 1. \quad (3.26)$$

Moreover, based on Equations (3.13) and (3.26), the conditional margin $\pi_j(\mathbf{x})$ can be explicitly expressed as follows :

$$\pi_j(\mathbf{x}) = \frac{\Delta_1^2 T_{1,\theta_j}(1 - p_j, F_1(x_1), F_2(x_2))}{\Delta_1^2 T_{2,\theta_j}(F_1(x_1), F_2(x_2))}, \quad j = 1, 2,$$

where

$$\begin{aligned} T_{1,\theta_j}(u_1, u_2, u_3) &= \exp \left(- \left[(-\ln(u_1))^{\theta_j} + (-\ln(u_2))^{\theta_j} + (-\ln(u_3))^{\theta_j} \right]^{\frac{1}{\theta_j}} \right) \frac{(-\ln(u_2))^{\theta_j-1}}{u_2} \\ &\times \left[\left((-\ln(u_1))^{\theta_j} + (-\ln(u_2))^{\theta_j} + (-\ln(u_3))^{\theta_j} \right)^{\frac{1}{\theta_j}} \right]^{1-\theta_j}, \end{aligned}$$

and

$$\begin{aligned} T_{2,\theta_j}(u_1, u_2) &= \exp \left(- \left[(-\ln(u_1))^{\theta_j} + (-\ln(u_2))^{\theta_j} \right]^{\frac{1}{\theta_j}} \right) \frac{(-\ln(u_1))^{\theta_j-1}}{u_1} \\ &\times \left[\left((-\ln(u_1))^{\theta_j} + (-\ln(u_2))^{\theta_j} \right)^{\frac{1}{\theta_j}} \right]^{1-\theta_j}, \end{aligned}$$

where ϕ_{θ_j} and φ_{θ_j} denote the generator and its inverse, respectively, associated with the conditional copula $\mathcal{C}_{\theta_j}^{cm}$, capturing the dependence structure between Y_j and \mathbf{X} . Consequently, the predictive probability $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$ can be represented as follows :

$$\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}) = \exp \left(- \left[(-\ln(\pi_1(\mathbf{x})))^\gamma + (-\ln(\pi_2(\mathbf{x})))^\gamma \right]^{\frac{1}{\gamma}} \right).$$

3.4.2 Gaussian case

This section derives the expression of $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$ (e.g., $d = 2$ and $p = 2$) when the copula \mathcal{C}_γ is set to be Gaussian. More precisely, the copula \mathcal{C}_γ is defined based on the standardized bivariate normal distribution Φ_Σ and the non-singular correlation matrix Σ . The relationship is described by the following Equation :

$$\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}) = \Phi_\Sigma \left(\Phi^{-1}(\pi_1(\mathbf{x})), \Phi^{-1}(\pi_2(\mathbf{x})) \right). \quad (3.27)$$

Furthermore, the dependence between (Y_j, \mathbf{X}_p) is expressed through a Gaussian copula. This means that :

$$\mathcal{C}(u, \mathbf{v}; \Sigma_j) = \Phi_{\Sigma_j}(\Phi^{-1}(u), \Phi^{-1}(\mathbf{v})) \quad \forall (u, \mathbf{v}) \in [0, 1]^{p+1},$$

where $\Phi^{-1}(\mathbf{v}) = (\Phi^{-1}(v_1), \dots, \Phi^{-1}(v_p))$, and

$$\Sigma_j = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{with sizes} \quad \begin{pmatrix} q \times q & q \times (p - q + 1) \\ (p - q + 1) \times q & (p - q + 1) \times (p - q + 1) \end{pmatrix}.$$

Here, Σ_{11} and Σ_{22} denote the correlation matrices of the normal scores of the continuous covariates \mathbf{X}_q and the normal scores of the remaining discrete random vector (Y_j, \mathbf{X}_{p-q}) , respectively. Additionally, Σ_{12} represents the correlation matrix between the normal scores of the continuous covariates \mathbf{X}_q and the normal scores of the discrete vector (Y_j, \mathbf{X}_{p-q}) , with the relationship $\Sigma_{21} = \Sigma_{12}^\top$. A normal score of a random variable X is defined as $\Phi^{-1}(F_X(x))$, where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution. Subsequently, the representation of $\pi_j(\mathbf{x})$ is given by :

$$\begin{aligned} \pi_j(\mathbf{x}) &= \frac{\Delta_q^p \partial_q \mathcal{C}_{\theta_j}(1 - p_j, \mathbf{F}(\mathbf{x}))}{\Delta_q^p \partial_q \mathcal{C}_{\theta_j}(1, \mathbf{F}(\mathbf{x}))} \\ &= \frac{\Delta_q^p \mathcal{C}_{p-q+1|q}^c(1 - p_j, \mathbf{F}_{p-q}(\mathbf{x}_{p-q}) | (\mathbf{F}_q(\mathbf{x}_q)) c_q(\mathbf{F}_q(\mathbf{x}_q))}{\Delta_q^p \mathcal{C}_{p-q|q}^c(\mathbf{F}_{p-q}(\mathbf{x}_{p-q}) | \mathbf{F}_q(\mathbf{x}_q)) c_q(\mathbf{F}_q(\mathbf{x}_q))} \\ &= \frac{\Delta_q^p \mathcal{C}_{p-q+1|q}^c(1 - p_j, \mathbf{F}_{p-q}(\mathbf{x}_{p-q}) | \mathbf{F}_q(\mathbf{x}_q))}{\Delta_q^p \mathcal{C}_{p-q|q}^c(\mathbf{F}_{p-q}(\mathbf{x}_{p-q}) | \mathbf{F}_q(\mathbf{x}_q))}, \end{aligned} \quad (3.28)$$

where $\partial_q \mathcal{C}(u, \mathbf{v}_p) = \mathcal{C}_{p-q+1|q}^c(u, \mathbf{v}_{p-q} | \mathbf{v}_q) c(\mathbf{v}_q)$ and $\partial_q \mathcal{C}(1, \mathbf{v}_p) = \mathcal{C}_{p-q|q}^c(\mathbf{v}_{p-q} | \mathbf{v}_q) c(\mathbf{v}_q)$. Using the fact that $(\Phi^{-1}(u), \Phi^{-1}(\mathbf{v}_p))$ is normally distributed, then the following conditional vector

$$(\Phi^{-1}(u), \Phi^{-1}(\mathbf{v}_{p-q}) | \Phi^{-1}(\mathbf{v}_q)) \sim \mathcal{N}\left(\Sigma_{21} \Sigma_{11}^{-1} \Phi^{-1}(\mathbf{v}_q), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{21}^\top\right),$$

with $\Phi^{-1}(\mathbf{v}_q) = (\Phi^{-1}(v_1), \dots, \Phi^{-1}(v_q))$, $\Phi^{-1}(\mathbf{v}_{p-q}) = (\Phi^{-1}(v_{q+1}), \dots, \Phi^{-1}(v_p))$ and c is the density copula of the random vector (X_1, \dots, X_p) . The next step is to compute $\mathcal{C}_{p-q|q}^c$, where the copula distribution corresponding to the covariates is represented by a Gaussian copula with the following correlation matrix :

$$\tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{pmatrix}, \quad \text{with sizes} \quad \begin{pmatrix} q \times q & q \times (p - q) \\ (p - q) \times q & (p - q) \times (p - q) \end{pmatrix},$$

where $\tilde{\Sigma}_{11}$ and $\tilde{\Sigma}_{22}$ correspond to the correlation matrices of the q -continuous random vector \mathbf{X}_q and the $(p-q)$ -discrete random vector \mathbf{X}_{p-q} , respectively. Likewise, $\tilde{\Sigma}_{12}$ represents the correlation matrix between the random vectors \mathbf{X}_{p-q} and \mathbf{X}_q , with $\tilde{\Sigma}_{21} = \tilde{\Sigma}_{12}^\top$. Consequently, for all $(\mathbf{v}_q, \mathbf{v}_{p-q}) \in (0, 1)^q \times (0, 1)^{p-q}$, the conditional random vector $\Phi^{-1}(\mathbf{V}_{p-q}) | \mathbf{V}_q = \mathbf{v}_q$ is distributed as

$$\mathcal{N}\left(\tilde{\Sigma}_{21} \tilde{\Sigma}_{11}^{-1} \Phi^{-1}(\mathbf{v}_q), \tilde{\Sigma}_{22} - \tilde{\Sigma}_{21} \tilde{\Sigma}_{11}^{-1} \tilde{\Sigma}_{21}^\top\right).$$

Example 3.3 Consider the case where $p = 2$ and $q = 1$, meaning that X_1 is continuous and X_2 is discrete. Assume additionally that the copula of (X_1, Y_j, X_2) is Gaussian and characterized by its correlation matrix

$$\Sigma_j = \begin{pmatrix} 1 & \rho_{Y_j1} & \rho_{12} \\ \rho_{Y_j1} & 1 & \rho_{Y_j2} \\ \rho_{12} & \rho_{Y_j2} & 1 \end{pmatrix},$$

where $\rho_{Y_j,k}$, (i.e, $j = 1, 2$ and $k = 1, 2$) represents the correlation coefficient between the j^{th} outcome and the k^{th} covariate, where $\Sigma_{11} = 1$, $\Sigma_{21} = \Sigma_{12}^\top = (\rho_{Y_j1}, \rho_{12})^\top$, $\Sigma_{2,2} = \begin{pmatrix} 1 & \rho_{Y_j2} \\ \rho_{Y_j2} & 1 \end{pmatrix}$. Moreover $\tilde{\Sigma}_{21} = \tilde{\Sigma}_{12}^\top = \rho_{12}$ and $\tilde{\Sigma}_{11} = \tilde{\Sigma}_{22} = 1$ with simple calculation the following conditional distribution of copula $C_{2|1}^c$ is distributed normally with the following mean $\mu_j(x_1) = \left(\rho_{Y_j1} \Phi^{-1}(F_1(x_1)), \rho_{12} \Phi^{-1}(F_1(x_1)) \right)^\top$, and with the following correlation matrix $\Sigma_{cond} = \begin{pmatrix} 1 - \rho_{Y_j1}^2 & \rho_{Y_j2} - \rho_{Y_j1} \rho_{12} \\ \rho_{Y_j2} - \rho_{12} \rho_{Y_j1} & 1 - \rho_{12}^2 \end{pmatrix}$. Then the conditional copula $C_{2|1}^c$ is expressed as follows :

$$C_{2|1}^c \left((\Phi^{-1}(1 - p_j), \Phi(F_2(x_2))) \mid \Phi^{-1}(F_2(x_1)) \right) = \Phi \left(\Sigma_{cond}^{-\frac{1}{2}} (1 - p_j, \Phi^{-1}(F_2(x_2)))^\top - \mu_j(x_1) \right),$$

and for

$$C_{1|1}^c \left(\Phi^{-1}(F_2(x_2)) \mid \Phi^{-1}(F_1(x_1)) \right) = \Phi \left(\frac{\Phi^{-1}(F_2(x_2)) - \rho_{12} \Phi^{-1}(F_1(x_1))}{\sqrt{1 - \rho_{12}^2}} \right).$$

Then, the probability given in (3.28), in the context of the example 3.3, can be written as :

$$\pi_j(\mathbf{x}) = \frac{\Delta_1^2 \Phi \left(\Sigma_{cond}^{-\frac{1}{2}} \left((\Phi^{-1}(1 - p_j), \Phi^{-1}(F_2(x_2)))^\top - \mu_j(x_1) \right) \right)}{\Delta_1^2 \Phi \left(\frac{\Phi^{-1}(F_2(x_2)) - \rho_{12} \Phi^{-1}(F_1(x_1))}{\sqrt{1 - \rho_{12}^2}} \right)}. \quad (3.29)$$

Substituting the expressions for $\pi_j(\mathbf{x})$, $j = 1, 2$, defined in (3.29) into Equation (3.27), we obtain a complete specification of the predictive probability based on the Gaussian copula in the presence of mixed covariates.

3.5 Estimation

This section outlines the estimation procedure for the proposed approach parameters used to evaluate the probability $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$. To achieve this, consider a sample of n observations $(\mathbf{Y}_1, \mathbf{X}_1), (\mathbf{Y}_2, \mathbf{X}_2), \dots, (\mathbf{Y}_n, \mathbf{X}_n)$, which are mutually independent realizations of the random pair (\mathbf{Y}, \mathbf{X}) , where $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$ and

$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})$ for $i = 1, \dots, n$. To estimate the probability $\mathbb{M}_{(y_1, \dots, y_d)}^{1, \dots, d}(\mathbf{x})$ first we estimate the marginal distribution of the covariates F_1, \dots, F_p as well as the probability of observing the j^{th} outcome, where their estimator is given by

$$\hat{p}_{nj} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(Y_{ij} = 1), \quad j = 1, \dots, d \text{ and } \hat{F}_{nk}(x_k) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(X_{ik} \leq x_k), \quad k = 1, \dots, p,$$

The second step involves estimating the copula parameter θ_j using a parametric approach. Precisely, assume that the copula $C_{\theta_j}^{cm}$ belongs to a parametric family $\mathcal{F} = \{C_{\theta_j}^{cm}, \theta_j \in \Theta\}$, where $\Theta \subset \mathbb{R}$. Consequently, the copula $C_{\theta_j}^{cm}$ is estimated by $C_{\hat{\theta}_j}^{cm}$, $j = 1, \dots, d$. The estimator $\hat{\theta}_j$ is typically obtained by maximizing the following pseudo-likelihood function with respect to θ_j :

$$\hat{\theta}_j = \operatorname{argmax}_{\theta_j \in \Theta} \prod_{i=1}^n \ell_i(\theta_j, \hat{p}_{nj}, \hat{\mathbf{F}}_n(\mathbf{x})), \quad (3.30)$$

where

$$\ell_i(\theta_j, \hat{p}_{nj}, \hat{\mathbf{F}}_n(\mathbf{x})) = \left(1 - C_{\theta_j}^{cm} \left(1 - \hat{p}_{nj} \mid \hat{\mathbf{F}}_n(\mathbf{x})\right)\right)^{Y_{ij}} C_{\theta_j}^{cm} \left(1 - \hat{p}_{nj} \mid \hat{\mathbf{F}}_n(\mathbf{x})\right)^{1-Y_{ij}},$$

and $\hat{\mathbf{F}}_n(\mathbf{x}) = (\hat{F}_{n1}(x_1), \dots, \hat{F}_{np}(x_p))$.

The third step consists of estimating the parameter $\gamma \in \mathbb{R}^q$ via the following pseudo-likelihood :

$$\hat{\gamma}_n = \operatorname{argmax}_{\gamma \in \mathbb{R}^q} \prod_{i=1}^n h_{C_\gamma}(Y_{i1}, \dots, Y_{id} \mid \mathbf{x}), \quad (3.31)$$

where

$$h_{C_\gamma}(Y_{i1}, \dots, Y_{id} \mid \mathbf{x}) = \sum_{j_1=1}^2 \dots \sum_{j_d=1}^2 (-1)^{j_1 + \dots + j_d} C_\gamma(u_{1j_1}, \dots, u_{dj_d}),$$

with $u_{j1} = C_{\theta_j}^{cm}(\hat{G}_j(Y_{ij}) \mid \hat{\mathbf{F}}(\mathbf{x}))$ and $u_{j2} = C_{\theta_j}^{cm}(\hat{G}_j(Y_{ij}^-) \mid \hat{\mathbf{F}}(\mathbf{x}))$. Here, $\hat{G}_j(Y_{ij}^-)$ represents the left-hand limit of \hat{G}_j at Y_{ij} , which is equal to $\hat{G}_j(Y_{ij} - 1)$.

3.6 Simulation study

This section presents simulation studies designed to evaluate the performance of the predictive probability estimator proposed in this paper. Comparisons are made with established methods from the literature, notably the C-logit approach introduced by Genest *et al.* (2013) and the Vector Generalized Linear Model (VGLM), developed by Zhao et Prentice (1990) and implemented in the R package VGAM by Yee (2015). The analysis specifically considers the scenario where $d = 2$, involving mixed covariates, with X_1 continuous

and X_2 discrete. Under these conditions, the proposed estimator is explicitly derived from its general formulation, detailed in Equation below.

$$\widehat{\mathbb{M}}_{(0,0)}^{1,2}(\mathbf{x}) = \mathcal{C}_{\hat{\gamma}}(\hat{\pi}_1(\mathbf{x}), \hat{\pi}_2(\mathbf{x})), \quad (3.32)$$

where

$$\hat{\pi}_j(\mathbf{x}) = \frac{\partial_1 \mathcal{C}_{\hat{\theta}_j}(1-p, \widehat{F}_{n1}(x_1), \widehat{F}_{n2}(x_2)) - \partial_1 \mathcal{C}_{\hat{\theta}_j}(1-p, \widehat{F}_{n1}(x_1), \widehat{F}_{n2}(x_2-1))}{\partial_1 \mathcal{C}_{\hat{\theta}_j}(1, \widehat{F}_{n1}(x_1), \widehat{F}_{n2}(x_2)) - \partial_1 \mathcal{C}_{\hat{\theta}_j}(1, \widehat{F}_{n1}(x_1), \widehat{F}_{n2}(x_2-1))}, \quad j = 1, 2.$$

To this end, we generate our data according to the following scenario which is defined by alternating combinations of the following parameters $(n, \tau = \tau_j, p_j)$ for a total of $3 \times 2 \times 3 = 18$ scenarios. However, for the copula \mathcal{C}_{γ} and $\mathcal{C}_{\theta_j}^{cm}$ modeling the outcomes dependence and (Y_j, X_1, X_2) respectively. For each scenario, we generate our data as follows :

Algorithm 2 Simulation Algorithm

Require: $n, p_1, p_2, \boldsymbol{\theta}_j, j = 1, 2,$ and γ

1: **for** $i = 1, \dots, n$ **do**

2: Generate X_1 from a parametric continuous distribution $F_1(\cdot)$, and generate the discrete variable $X_2 \sim F_2(\cdot)$, where :

$$F_2^{-1}(t) = \inf\{v \in \{1, 2, 3\} : F_2(v) > t\}.$$

3: For $j = 1, 2$, set :

$$\pi_j(\mathbf{x}) = \mathcal{C}_{\boldsymbol{\theta}_j}^{cm}(1 - p_j \mid \mathbf{F}(\mathbf{x})),$$

where $\mathcal{C}_{\boldsymbol{\theta}_j}^{cm}(\cdot)$ is a specific copula in the presence of mixed variables. The marginal probabilities of success, $p_j = P(Y_j = 1)$, for $j = 1, 2$, are population-level parameters and thus do not vary per subject.

For all scenarios in our simulation study, we set $p_1 = 0.3$ and $p_2 = 0.4$.

4: Generate $U_{i,1}, U_{i,2}$ from a copula $\mathcal{C}_\gamma(\cdot)$.

5: For $j = 1, 2$, set :

$$Y_j = \mathbb{1}(U_j \geq \pi_j(\mathbf{x})).$$

6: **end for**

7: Use the generated sample $(Y_i, X_{i,1}, X_{i,2}), i = 1, \dots, n$ to estimate $\boldsymbol{\theta}_j$ and γ .

8: Evaluate the estimator $\hat{\pi}_j(x_1, x_2)$ for (x_1, x_2) and $j = 1, 2$ belonging to the grid defined by :

$$K = E \times \{1, 2, 3\}, \quad \text{where } E = \{i/K, i = 1, \dots, K\},$$

where the sets E and K define the evaluation grid for $\hat{\mathbb{M}}_{(0,0)}^{1,2}(\mathbf{x})$. Specifically, E represents a grid over the continuous covariate X_1 , and $K = E \times \{1, 2, 3\}$ extends this grid to include the discrete levels of X_2 .

For a fixed pair $(x_1, x_2) \in K$, the theoretical value $\mathbb{M}_{(0,0)}^{1,2}(x_1, x_2)$ is first computed. Subsequently, the estimator $\hat{\mathbb{M}}_{(0,0)}^{1,2}(x_1, x_2)$ is evaluated based on B independent random samples of size n , generated following the scenario described in algorithm 2. The corresponding estimates are denoted by $\hat{\mathbb{M}}_{(0,0)}^{1,2(b)}(x_1, x_2)$, where $b = 1, \dots, B$.

To assess the proposed estimator's performance, we employ the empirical integrated mean squared error

(IMSE), defined as :

$$\widehat{\text{IMSE}} = \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{|K|} \sum_{(x_1, x_2) \in K} \left(\widehat{\mathbb{M}}_{(0,0)}^{1,2(b)}(x_1, x_2) - \mathbb{M}_{(0,0)}^{1,2}(x_1, x_2) \right)^2 \right),$$

where $|K|$ denotes the cardinality of the evaluation grid K . The grid K is constructed as $K = E \times \{1, 2, 3\}$, with $E = \{i/K : i = 1, \dots, K\}$ representing a uniformly spaced grid over the continuous covariate X_1 .

Notice that the empirical Integrated Mean Squared Error (IMSE) can be decomposed into two distinct components, the squared empirical bias (IBIAS) and the empirical variance (IVAR), given by :

$$\widehat{\text{IMSE}} = \text{IBIAS}^2 + \text{IVAR}. \quad (3.33)$$

Proof. We recall the Monte-Carlo definition of the Integrated Mean Squared Error (IMSE). For each evaluation point $\mathbf{x} = (x_1, x_2) \in K$ and each replicate $b \in \{1, \dots, B\}$, let

$$\widehat{\mathbb{M}}^{(b)}(\mathbf{x}) \quad \text{and} \quad \mathbb{M}(\mathbf{x})$$

denote, respectively, the estimator and the true value of $\mathbb{M}_{(0,0)}^{1,2}(\mathbf{x})$. Setting

$$\overline{\mathbb{M}}(\mathbf{x}) := \frac{1}{B} \sum_{b=1}^B \widehat{\mathbb{M}}^{(b)}(\mathbf{x}),$$

recall Equation (3.33)

$$\widehat{\text{IMSE}} := \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{|K|} \sum_{\mathbf{x} \in K} \left(\widehat{\mathbb{M}}^{(b)}(\mathbf{x}) - \mathbb{M}(\mathbf{x}) \right)^2 \right).$$

Thus for a fixed grid point \mathbf{x} , one can write

$$\frac{1}{B} \sum_{b=1}^B \left(\widehat{\mathbb{M}}^{(b)}(\mathbf{x}) - \mathbb{M}(\mathbf{x}) \right)^2 = \left(\overline{\mathbb{M}}(\mathbf{x}) - \mathbb{M}(\mathbf{x}) \right)^2 + \frac{1}{B} \sum_{b=1}^B \left(\widehat{\mathbb{M}}^{(b)}(\mathbf{x}) - \overline{\mathbb{M}}(\mathbf{x}) \right)^2, \quad (3.34)$$

subtract and add $\overline{\mathbb{M}}(\mathbf{x})$ inside the square, expand the square, and note that the cross-term vanishes because

$$\sum_{b=1}^B \left(\widehat{\mathbb{M}}^{(b)} - \overline{\mathbb{M}} \right) = 0.$$

We next average this equality given by (3.34) over all points $\mathbf{x} \in K$. Because the IMSE itself is the mean of the left-hand side over the grid, the same decomposition persists after the averaging step. More precisely,

$$\text{IMSE} = \frac{1}{|K|} \sum_{\mathbf{x} \in K} \left(\overline{\mathbb{M}}(\mathbf{x}) - \mathbb{M}(\mathbf{x}) \right)^2 + \frac{1}{|K|} \sum_{\mathbf{x} \in K} \frac{1}{B} \sum_{b=1}^B \left(\widehat{\mathbb{M}}^{(b)}(\mathbf{x}) - \overline{\mathbb{M}}(\mathbf{x}) \right)^2.$$

The first sum is precisely the squared empirical bias

$$\text{IBIAS}^2 := \frac{1}{|K|} \sum_{\mathbf{x} \in K} \left(\overline{\mathbb{M}}(\mathbf{x}) - \mathbb{M}(\mathbf{x}) \right)^2,$$

while the second sum is simply the empirical variance,

$$IVAR := \frac{1}{|K|} \sum_{\mathbf{x} \in K} \frac{1}{B} \sum_{b=1}^B (\widehat{\mathbb{M}}^{(b)}(\mathbf{x}) - \overline{\mathbb{M}}(\mathbf{x}))^2.$$

Collecting these definitions gives the desired decomposition

$$\boxed{IMSE = IBIAS^2 + IVAR}.$$

which completes the proof \square

In the IMSE formulation given in (3.33), IBIAS measures the systematic deviation of the estimator $\widehat{\mathbb{M}}_{(0,0)}^{1,2}(x_1, x_2)$ from its theoretical counterpart across the evaluation grid K , whereas IVAR quantifies the estimator's variability over the B independent resamples. This decomposition therefore provides a comprehensive assessment of both the accuracy (bias) and precision (variance) of the proposed estimation method.

3.6.1 Results

Table 3.1 presents the Integrated Mean Squared Error (IMSE) statistic multiplied by 100 based on $B = 1000$ replications for various copula models under different dependence levels ($\tau = \tau_j = 0.3, 0.5, 0.7$) and sample sizes $n = 100, 300, 500$. The results demonstrate that the proposed estimator consistently outperforms competing methods across all settings.

Under a Clayton dependence, the proposed method achieves substantially lower IMSE values (e.g., 1.297, 1.124, and 1.096 for $n = \{100, 300, 500\}$ at $\tau = 0.3$, compared to VGLM (Yee, 2015) and C-logit (Genest *et al.*, 2013), which report significantly higher errors. In the case of the Frank copula, the proposed estimator also exhibits superior performance, with IMSE values increasing moderately as dependence strengthens from approximately 1.53 at $\tau = 0.3$ to around 4.78 and 4.82 at $\tau = 0.7$ remaining far lower than those of alternative methods, which yield errors exceeding 20 in some scenarios. The performance of the proposed approach under asymmetric dependence structures, such as those characterized by the Gumbel copula, is particularly notable. In addition, the proposed method consistently achieves low IMSE values across varying sample sizes ($n = 100, 300, 500$) and dependence levels quantified by Kendall's $\tau = \tau_j = 0.3, 0.5, 0.7$. In contrast, alternative methods, including VGLM and C-logit, produce significantly higher IMSE values, highlighting the superior efficiency of our estimator in capturing asymmetric dependencies. Under the Gaussian copula scenario, the proposed approach similarly outperforms competing methods across both correlation settings specified in (3.35), as detailed in Table 3.2.

Table 3.1 – Algorithm 2 : Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, for $d = 2$ and $p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
<i>Proposed method</i>	1.297	1.124	1.096	0.415	0.287	0.250	0.758	0.762	0.723
<i>VGLM</i>	5.099	4.536	4.996	11.427	10.487	10.341	18.911	19.496	19.474
<i>C-logit</i>	7.391	5.651	4.353	3.592	2.121	1.684	6.050	5.640	4.895
Frank									
<i>Proposed method</i>	1.531	1.364	1.355	3.455	3.281	3.230	4.782	4.818	4.510
<i>VGLM</i>	2.216	3.373	5.801	13.109	12.761	12.602	21.550	21.377	20.931
<i>C-logit</i>	6.168	5.735	5.023	7.973	9.520	9.631	10.693	10.731	10.651
Gumbel									
<i>Proposed method</i>	0.198	0.075	0.043	0.227	0.082	0.042	0.197	0.179	0.150
<i>VGLM</i>	2.489	2.103	2.092	3.847	3.455	3.394	6.841	6.745	6.861
<i>C-logit</i>	4.476	4.500	3.148	5.417	4.773	4.580	6.015	5.742	5.829

Table 3.2 – Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, using the correlation matrices defined in equation (3.35).

	$n = 100$	$n = 300$	$n = 500$
Gaussian			
<i>Proposed method</i>	1.487	1.302	1.271
<i>VGLM</i>	3.348	3.097	3.090
<i>C-logit</i>	4.496	3.856	3.607

$$\Sigma_1 = \begin{bmatrix} 1.00 & 0.36 & 0.40 \\ 0.36 & 1.00 & 0.30 \\ 0.40 & 0.30 & 1.00 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \quad (3.35)$$

Furthermore, we examined additional scenarios, including cases involving imbalanced data structures with a rare event probability $p_1 = 0.05$, while maintaining $p_2 = 0.4$ as specified in Algorithm 2. Moreover, we extended our investigation to scenarios involving three-dimensional responses ($d = 3$), setting $p_3 = 0.2$, and similarly assessed the performance under imbalanced conditions with $p_1 = 0.05$ for $d = 3$. The results

of these additional scenarios are differed in Appendices A.2.1 and A.2.2.

Overall, these results highlight the effectiveness of the proposed methods for evaluating $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$ in the presence of mixed covariates (e.g, $d = 2, p = 2$, and $d = 3, p = 2$). The proposed estimators consistently yield lower Integrated Mean Squared Error (IMSE) values across various copula families and simulation scenarios, accurately capturing complex dependence structures between binary outcomes and mixed covariates. Consequently, the proposed methods provide more precise and efficient estimates of predictive probabilities compared to conventional approaches such as VGLM and C-logit.

3.7 Conclusion

This paper introduces a methodology to model the dependence structure among binary outcomes in the presence of mixed covariates. Extending the framework developed in Chapter 2, each marginal outcome is represented through a copula-based link function. Specifically, the method proposed by Aldahmani *et al.* (2024) is employed to adapt the predictive probabilities $\pi_j(\mathbf{x})$, $j = 1, \dots, d$, accommodating mixed covariates for assessing the quantity $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$. The estimation procedure associated with this approach is thoroughly detailed and validated by an extensive simulation study conducted under various scenarios using different copula families. The simulation results demonstrate that the proposed model consistently outperforms traditional regression methods, underscoring its robustness and efficacy in handling mixed-covariate data.

This methodology provides a valuable tool for dependence modeling with binary outcomes, applicable to diverse fields such as medicine, finance, and other disciplines. A promising avenue for future research is to extend this approach to models featuring multivariate discrete responses with multiple response categories, while continuing to incorporate mixed covariates.

CHAPITRE 4

COPULAS BASED INFERENCE FOR ASYMMETRIC BINARY REGRESSION

Dans le cadre des travaux antérieurs présentés dans les chapitres 2 et 3, un modèle théorique novateur a été développé en s'appuyant sur les copules afin de modéliser de façon flexible les dépendances complexes et non linéaires entre les variables binaires dans un contexte de régression. Ce modèle propose d'estimer les marges des variables binaires à l'aide d'une fonction de lien fondée sur les copules, conformément à l'approche récemment développée par Mesfioui *et al.* (2023). Toutefois, la probabilité de prédiction ainsi obtenue ne reflète qu'une tendance moyenne entre la variable réponse et les covariables, pouvant ainsi s'écarter de la réalité observée dans les applications pratiques, où les données ne se limitent généralement pas à une simple tendance moyenne.

Dans ce quatrième chapitre, nous introduisons une méthodologie innovante visant l'estimation du prédicteur dans le cadre de la régression binaire, particulièrement adaptée aux données déséquilibrées (c.-à-d., présentant une distribution asymétrique). Afin de traiter efficacement cette problématique, nous recourons à des fonctions de perte robustes de type Huber (1964), et expectile (Newey et Powell, 1987). De plus, la distribution conditionnelle est modélisée en s'appuyant sur une formulation basée sur les copules, selon la perspective récemment développée par Mesfioui *et al.* (2023). L'approche basée sur les copules permet ainsi de relâcher les hypothèses d'isotonicité et d'unicité du point de croisement, généralement imposées par les modèles classiques tels que la régression logistique, ou encore dans les travaux comme ceux de Manski et Thompson (1989), où l'estimation du meilleur prédicteur repose explicitement sur ces contraintes. Par ailleurs, cette méthodologie offre la possibilité de modéliser efficacement des relations non linéaires et complexes entre les variables étudiées. Dans ce contexte, nous proposons une solution théorique au problème d'optimisation associé, accompagnée d'un estimateur semi-paramétrique. Par la suite, les propriétés asymptotiques de cet estimateur sont rigoureusement établies, démontrant notamment sa consistance ainsi que sa normalité asymptotique sous certaines conditions précises. Ces résultats attestent de l'efficacité de la méthodologie proposée en termes de prédiction. Une application empirique réalisée sur des données réelles de détection de fraudes par carte de crédit souligne clairement la supériorité de cette méthode comparativement à d'autres approches concurrentes telles que la régression logistique, les machines à vecteurs de support (SVM), les SVM asymétriques ainsi que les forêts aléatoires. Cette supériorité se manifeste notamment au niveau des critères d'évaluation tels que la précision, la sensibilité et le score

F1 . En particulier, cette approche se démarque par sa capacité accrue à gérer efficacement des données déséquilibrées, grâce à sa flexibilité intrinsèque et à son aptitude à capturer finement des dépendances complexes et asymétriques. Ce travail a été soumis à la revue “Journal of Statistical Theory and Practice”. Les retours, globalement favorables, ont conduit à une révision, sur laquelle nous travaillons activement afin de la soumettre de nouveau le mois prochain.

4.1 Abstract

In this paper, we propose a novel inference approach tailored for binary regression, specifically addressing challenges associated with imbalanced data. Our methodology determines the optimal predictor by employing Huber-type and expectile loss functions within a copula-based framework. This approach models the predictive probability of success using copulas, following the method investigated by Mesfioui *et al.* (2023), thus providing greater flexibility than traditional techniques such as logistic and probit regressions by overcoming linear assumptions and monotonicity constraints. In contrast to the approach proposed by Manski et Thompson (1989), which relies on the single-crossing condition, our method relaxes this restriction and effectively captures complex, nonlinear dependencies between the covariates and the response variable through the use of copulas. Furthermore, we introduce an estimator for the proposed optimal predictor based on sample data and rigorously establish its asymptotic properties. To validate the effectiveness of our proposed methodology, we carry out a comprehensive simulation study and illustrate its practical applicability using real-world datasets, particularly highlighting scenarios involving imbalanced data.

4.2 Introduction

In binary regression, the optimal predictor under the quadratic loss function is the conditional expectation $\mathbb{E}\{Y \mid \mathbf{X} = \mathbf{x}\}$, obtained by minimizing the conditional expectation of the squared loss. Here, $\mathbf{x} \in \mathbb{R}^d$ denotes the vector of observed covariates. However, relying exclusively on the quadratic loss may be insufficient to capture certain features of the data, particularly in the presence of asymmetry in the underlying distribution. Such asymmetry in binary regression typically arises when one category is underrepresented in the sample, reflecting the outcome distribution at the population level. This issue is common in machine learning applications, for instance, in the detection of fraudulent credit card transactions, where fraudulent cases are rare compared to legitimate ones. In such contexts, models that rely solely on modeling the mean tendency tend to produce high classification error rates and fail to adequately address class imbalance. Several strategies have been proposed in the literature to address the persistent challenge of class imba-

lance in binary regression and classification tasks. An early direction involved modifying the likelihood in logistic regression through inverse-frequency weighting, as examined by King et Zeng (2001). A second major direction in the literature focused on resampling strategies, with the Synthetic Minority Oversampling Technique (SMOTE) representing a foundational contribution (Chawla *et al.*, 2002). The SMOTE approach generates artificial instances of the minority class by interpolating between nearest neighbors, thereby mitigating imbalance and reducing bias toward the majority class. To enhance this approach, Akbani *et al.* (2004) combined SMOTE with a cost-sensitive Support Vector Machine (SVM), reporting significant gains in predictive performance, albeit at the cost of increased sensitivity to oversampling parameters and a heightened risk of overfitting due to noisy synthetic points. In contrast, Han *et al.* (2005) introduced a weighted SVM that assigns penalties proportional to class imbalance, significantly improving predictive accuracy for rare classes. Subsequent work refined synthetic sample generation by targeting more informative regions of the feature space. Borderline-SMOTE (Huang et Du, 2005) emphasizes instances near class boundaries, thereby improving class separation while limiting the influence of outliers. In a similar vein, He *et al.* (2008) proposed ADASYN, an adaptive mechanism that generates additional synthetic observations in regions where minority instances are more difficult to classify, aligning sample generation with local learning complexity. Building on this idea, SMOTE-IPF (Sáez *et al.*, 2015) introduced an ensemble-based noise filtering procedure, the Iterative Partitioning Filter (IPF), which iteratively removes noisy and borderline instances after each round of oversampling, progressively yielding a dataset that is both cleaner and more representative of the true minority distribution. More recently, SVM-SMOTE (Suh *et al.*, 2017) extended the principle of boundary-focused oversampling by leveraging SVMs to identify support vectors that define the decision boundary, and generating synthetic points along these vectors. This ensures that oversampling strengthens the minority class representation precisely where the risk of misclassification is most pronounced. Despite these advances, most methods remain constrained by heuristic resampling procedures, sensitivity to hyperparameter choices, restrictive linearity assumptions, and the absence of a rigorous inferential foundation. To address these limitations, our approach integrates loss functions from the literature that explicitly account for distributional asymmetries. For instance, Huber (1964) introduced robust loss functions adapted to M-estimators, which provide resilience against outliers while accounting for asymmetric distributions. Similarly, expectiles (Pakes et Pollard, 1989) extend the concept of the mean in a distribution, similarly to quantiles (Koenker et Bassett, 1978) generalize the notion of the median. In binary regression, the predictive probability induced by these loss functions can be expressed in terms of the predictive probability of success, as demonstrated by Manski et Thompson (1989); Dawber *et al.* (2022). These approaches typically assume an isotonic response structure and often rely on the weaker single-crossing model to ensure the

uniqueness of the optimal predictor under asymmetric loss (Manski et Thompson, 1989). It is also worth noting that these methods generally presume a linear relationship between the outcome and the covariates. Recently, many authors have reformulated the mean regression through the expectation of a multivariate function evaluated on its univariate marginals, using a function named a copula (Sklar, 1959). Copulas have gained significant popularity in dependence modeling due to their flexibility in capturing joint dependence independently from the marginal behavior of each variable. This distinctive feature allows copulas to effectively represent complex dependencies and nonlinear relationships. For example, Noh *et al.* (2015) proposed an extension of mean regression based on copulas, initially introduced by Noh *et al.* (2013), to quantile regression using copula-based estimation specifically adapted for continuous response variable. This method utilizes the quantile loss function introduced by Koenker et Bassett (1978) to capture various aspects of the conditional distribution of $Y|\mathbf{X} = \mathbf{x}$. Furthermore, the association between a univariate continuous outcome and the covariates is modeled through a copula-based framework, offering considerable flexibility in capturing complex dependency structures within the data. The estimation is conducted using a semi-parametric approach, which mitigates the risk of model misspecification bias while effectively controlling variance. This methodology facilitates more balanced and robust parameter estimation, enhancing the reliability of the inferred relationships. Recently, Mesfioui *et al.* (2023) proposed a novel copula-based link function for binary regression that does not rely on common assumptions such as the isotonic response or the more general single-crossing condition, which are typically required for consistent estimation. Moreover, this framework includes the classical logistic and probit models as special cases when using the Frank and Gaussian copulas, respectively. It is important to note that the aforementioned approach assumes the covariates to be random variables, unlike traditional approaches that consider them as non-random or fixed quantities.

In this work, we propose a unified framework to find an optimal predictor denoted by $\lambda_0(\mathbf{x})$, which can be formulated through the following optimization problem :

$$\lambda_0(\mathbf{x}) = \underset{\lambda \in [0,1]}{\operatorname{argmin}} \mathbb{E} [\rho_\tau(Y - \lambda) | \mathbf{X} = \mathbf{x}]. \quad (4.1)$$

Here, $\lambda_0(\mathbf{x})$ represents the optimal prediction of $Y \in \{0, 1\}$, while $\rho_\tau(\cdot)$ denotes an asymmetric loss function. In this framework, we utilize both Huber-type and expectile loss functions. Furthermore, the predictive probability of success is modeled using copulas, as examined by Mesfioui *et al.* (2023). The detailed steps for deriving the optimal predictor $\lambda_0(\mathbf{x})$ are provided in Section 4.3.1.

The structure of this work is organized as follows. Section 4.3 examines the assessment and uniqueness of the optimal predictor $\lambda_0(\mathbf{x})$ based on the Huber-type and expectile loss functions. Additionally, it provides

an expression for $\lambda_0(\mathbf{x})$ using copulas and introduces its corresponding estimator. Section 4.4 is dedicated to discussing the asymptotic properties of the proposed estimator of $\lambda_0(\mathbf{x})$. Section 4.5 focuses on illustrating $\lambda_0(\mathbf{x})$ and its estimator $\hat{\lambda}_n(\mathbf{x})$ based on Huber-type and expectile loss functions, as well as conducting simulations to evaluate the performance of the proposed method. Finally, Section 4.6 presents the results of applying the proposed method to real data, alongside comparisons with competitive methods such as SVM, random forest, and logistic regression.

4.3 Model and Inference

4.3.1 Model Formulation

Let $Y \in \{0, 1\}$ be a binary response variable, and let $\mathbf{X} = (X_1, \dots, X_d)$ represent a continuous random vector of predictor variables. Let F_0, F_1, \dots, F_d be the distribution functions of Y, X_1, \dots, X_d , respectively, and denote $\mathbf{F}(\mathbf{x}) = (F_1(x_1), \dots, F_d(x_d))$ where $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. In regression analysis, the primary objective is to minimize the conditional expectation based on a specific loss function in order to find the optimal predictor. The selection of this loss function depends on the specific goals of the analysis. In our study, we employ asymmetric loss functions, which are particularly well-suited for regression problems where the response variable exhibits an asymmetric distribution. For instance, in binary regression, these asymmetric loss functions address the challenges posed by imbalanced data distributions by ensuring that the regression model appropriately penalizes errors in accordance with the inherent asymmetry of the data. Consequently, our problem reduces to solving the following optimization problem as given in (4.1), where $\rho_\tau(\cdot)$ belongs to a set of asymmetric loss functions expressed as follows :

$$\rho_\tau(Y - \lambda) = \begin{cases} \tau\rho(1 - \lambda), & \text{if } Y = 1, \\ (1 - \tau)\rho(-\lambda), & \text{if } Y = 0, \end{cases} \quad (4.2)$$

where $\lambda \in [0, 1]$ and $\rho(u)$ is a convex function that measures deviance, ensuring both continuity and differentiability of ρ_τ around small deviations. In this work, we utilize the expectile loss function and a smoothed quantile loss function, adapted for M-estimators based on the Huber-type loss function (Huber, 1964). This approach is particularly advantageous for handling skewed distributions and influential or outlier observations, as validated by Bianco et Yohai (1996). It is important to note that $\rho_\tau(u)$, under the earlier mentioned loss functions (i.e, Huber type and expectile), admits a minimum at zero, convex and of class $C^1(\mathbb{R})$ and differentiable almost everywhere in the interval $[0, 1]$. The properties of these loss function are cited in the Section 1.3. To determine the optimal predictor, one can derive the conditional expectation defined in (4.1)

and find the roots of its derivative with respect to λ . Therefore, our optimization problem reduces to find the root of the following equation :

$$\begin{aligned} f(\lambda) &= \mathbb{E} \left[\psi_\tau(Y - \lambda) | \mathbf{X} = \mathbf{x} \right] \\ &= \psi_\tau(-\lambda) \mathbb{P}(Y = 0 | \mathbf{X} = \mathbf{x}) + \psi_\tau(1 - \lambda) \mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x}) = 0, \end{aligned} \quad (4.3)$$

where

$$\psi_\tau(Y - \lambda) = \frac{\partial \rho_\tau(Y - \lambda)}{\partial \lambda} = \begin{cases} -\tau \cdot \rho'(1 - \lambda), & \text{if } Y = 1, \\ -(1 - \tau) \cdot \rho'(-\lambda), & \text{if } Y = 0. \end{cases} \quad (4.4)$$

Note that the root of (4.3) must be unique, which is essential for obtaining a reliable predictor and accurately interpreting our results, as stated in the following Proposition.

Proposition 4.1 Consider the asymmetric loss function $\rho_\tau(Y - \lambda)$ defined in (4.2), and the function $f(\lambda)$ in (4.3). Then the equation $f(\lambda) = 0$ admits a unique solution denoted by $\lambda_0(\mathbf{x}) \in [0, 1]$.

Proof. To establish our proof, we employ the Intermediate Value Theorem (I.V.T). Based on the construction of $\rho_\tau(\cdot)$, the function f satisfies the necessary conditions for applying the I.V.T. Specifically, f must be continuous, which is ensured by its construction as presented in equation (4.3). Additionally, it is required to verify that f changes sign at the endpoints of the interval $[0, 1]$. Based on (4.4), one can write that :

$$f(0) = \psi_\tau(0)(1 - p_1) + \psi_\tau(1)p_1,$$

and

$$f(1) = \psi_\tau(-1)(1 - p_1) + \psi_\tau(0)p_1,$$

where $p_1 = \mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x})$. Furthermore, considering the expression of the influence function $\psi_\tau(Y - \lambda)$ given in (4.4) and the convexity of ρ , which exhibits asymmetry around λ and attains its minimum within the interval $[0, 1]$, the following results can be established.

$$\psi_\tau(Y - \lambda) = \frac{\partial \rho_\tau(Y - \lambda)}{\partial \lambda} = \begin{cases} \psi_\tau(0) = -(1 - \tau)\rho'(0) = 0 & \text{if } Y - \lambda = 0, \\ \psi_\tau(1) = -\tau\rho'(1) < 0 & \text{if } Y - \lambda = 1, \\ \psi_\tau(-1) = -(1 - \tau)\rho'(-1) > 0 & \text{if } Y - \lambda = -1. \end{cases} \quad (4.5)$$

Based on (4.5), one can determine the sign of f .

If $\lambda = 1 \implies f(1) > 0$, and for $\lambda = 0 \implies f(0) \leq 0$, then $f(0) \times f(1) \leq 0$.

Since f is continuous on $[0, 1]$ and $f(0) \times f(1) \leq 0$, by the Intermediate Value Theorem, it follows that (4.3) admits at least one root $\lambda_0(\mathbf{x}) \in [0, 1]$ such that $f(\lambda_0(\mathbf{x})) = 0$.

To prove the uniqueness of the root of $f(\lambda) = 0$, one can verify that $f(\lambda)$ is strictly increasing on $[0, 1]$. The derivative of f is expressed as follows :

$$\begin{aligned} f'(\lambda) &= \frac{d}{d\lambda} [\psi_\tau(-\lambda)(1 - p_1) + \psi_\tau(1 - \lambda)p_1] \\ &= -\psi'_\tau(-\lambda)(1 - p_1) - \psi'_\tau(1 - \lambda)p_1, \end{aligned}$$

where

$$\psi'_\tau(Y - \lambda) = \begin{cases} -\tau\rho''(1 - \lambda), & \text{if } Y = 1, \\ -(1 - \tau)\rho''(-\lambda), & \text{if } Y = 0. \end{cases}$$

Since, for $Y \in \{0, 1\}$, the considered Huber-type and expectile loss functions satisfy $\rho''(Y - \lambda) > 0$, it follows directly that $-\psi'_\tau(Y - \lambda) > 0$. Consequently, we have $f'(\lambda) > 0$ for all $\lambda \in [0, 1]$, indicating that $f(\lambda)$ is strictly increasing on the interval $[0, 1]$. Given that $f(\lambda)$ is continuous, strictly increasing, and changes sign exactly once over the interval $[0, 1]$, the equation $f(\lambda) = 0$ admits a unique solution in $[0, 1]$.

□

Under the conditions outlined in Proposition 4.1, Equation (4.3) has a unique solution, denoted by $\lambda_0(\mathbf{x}) \in [0, 1]$ which represents the best predictor of Y . This solution adjusts the conditional probability of success, $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$, by balancing the contributions of observations where $Y = 0$ and $Y = 1$. As a result, it effectively addresses data imbalances by enabling the tuning of the parameter τ in the asymmetric loss function $\rho_\tau(u)$, as detailed in the following result.

Result 4.2 Based on Equation (4.3), the parameter τ , which serves as the asymmetry coefficient in the loss function, balances the contributions of observations $Y = 0$ and $Y = 1$ in the presence of imbalanced data. In other words, the optimal solution $\lambda_0(\mathbf{x})$ satisfies the following equation :

$$-\psi_\tau(-\lambda_0(\mathbf{x})) \cdot P(Y = 0 \mid \mathbf{X} = \mathbf{x}) = \psi_\tau(1 - \lambda_0(\mathbf{x})) \cdot P(Y = 1 \mid \mathbf{X} = \mathbf{x}), \quad (4.6)$$

where $\psi_\tau(y - \lambda) = -|\tau - \mathbf{1}_{(y-\lambda < 0)}|\rho'(y - \lambda)$, with $\psi_\tau(-\lambda_0(\mathbf{x})) < 0$ and $\psi_\tau(1 - \lambda_0(\mathbf{x})) > 0$ due to the convexity of $\rho_\tau(\cdot)$.

By developing (4.6), one can obtain the following representation :

$$\frac{\tau}{1 - \tau} = -\frac{\frac{P(Y=0|\mathbf{X}=\mathbf{x})}{P(Y=1|\mathbf{X}=\mathbf{x})}}{\frac{\rho'(1-\lambda_0(\mathbf{x}))}{\rho'(-\lambda_0(\mathbf{x}))}} = -\frac{1}{\text{Odds}(Y = 1 \mid \mathbf{X} = \mathbf{x})} \times \frac{\rho'(-\lambda_0(\mathbf{x}))}{\rho'(1 - \lambda_0(\mathbf{x}))}, \quad (4.7)$$

with $\frac{1}{\text{Odds}(Y=1|\mathbf{X}=\mathbf{x})} = \frac{\mathbb{P}(Y=0|\mathbf{X}=\mathbf{x})}{\mathbb{P}(Y=1|\mathbf{X}=\mathbf{x})}$.

Equation (4.7) in Result 4.2 highlights the role of the parameter τ as a balancing mechanism between the classes $Y = 0$ and $Y = 1$ in the context of imbalanced data. Specifically, τ adjusts the class weights by incorporating the inverse odds ratio, which measures the degree of imbalance between the minority and majority classes. This ensures that the loss function appropriately accounts for the relative frequencies of each class. Moreover, τ refines the symmetry of penalization for specific types of errors, as reflected in the derivative ratio $\frac{\rho'(-\lambda_0(\mathbf{x}))}{\rho'(1-\lambda_0(\mathbf{x}))}$. This illustrates how the loss function differentiates the penalization of errors across classes, providing a more nuanced approach to handle imbalanced data.

4.3.2 Illustration of $\lambda_0(\mathbf{x})$ based on some common loss functions

This section presents examples of $\rho_\tau(\cdot)$ from the class of asymmetric loss functions, with particular focus on the Huber loss function (Huber, 1964) adapted for quantile regression. This adaptation offers improved performance compared to the classical quantile loss function. As a second example, we consider the expectile loss function. Both classes satisfy the conditions outlined in Proposition 4.1, guaranteeing the uniqueness of the optimal predictor $\lambda_0(\mathbf{x})$.

4.3.2.1 Huber-type loss functions

- **Huber loss function** : The Huber function, introduced by Huber (1964), is expressed as follows :

$$\rho_\tau(Y - \lambda) = \begin{cases} \rho(1 - \lambda) \cdot |\tau|, & \text{if } Y = 1, \\ \rho(-\lambda) \cdot |\tau - 1|, & \text{if } Y = 0, \end{cases} \quad (4.8)$$

where

$$\rho(u) = \begin{cases} \frac{1}{2k} u^2, & \text{if } |u| \leq k, \\ |u| - \frac{k}{2}, & \text{otherwise.} \end{cases}$$

The Huber loss function provided in (4.8) preserves the asymmetric properties adapted to quantile regression (Koenker et Bassett, 1978) while introducing differentiability around zero. This behavior is illustrated in Figure 4.1. Furthermore, Equation (4.3) based on Huber loss function verifies the condition cited in Proposition 4.1, admits a unique solution $\lambda_0(\mathbf{x})$ within the interval $[0, 1]$.

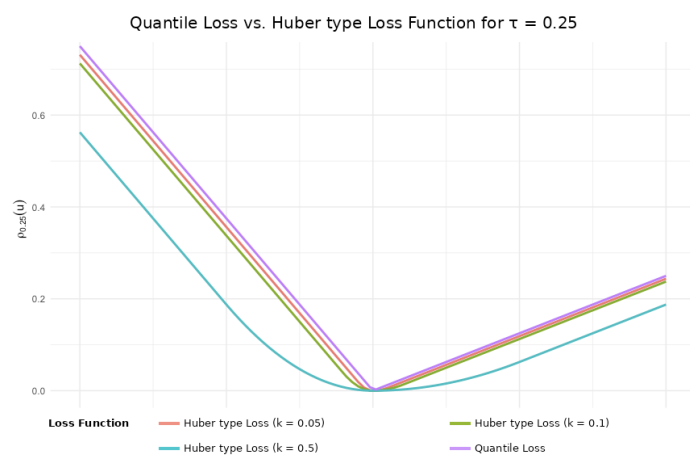


Figure 4.1 – Comparison of quantile and Huber-type loss defined in (4.8) functions for $\tau = 0.25$

- **Pseudo Huber loss** : The second example to be considered from the family of Huber type loss functions is the pseudo-Huber function, which is defined as follows :

$$\rho_{\tau}(Y - \lambda) \begin{cases} \rho(1 - \lambda)|\tau|, & \text{if } Y = 1, \\ \rho(-\lambda)|\tau - 1|, & \text{if } Y = 0. \end{cases} \quad (4.9)$$

where

$$\rho(u) = \sqrt{1 + u^2} - 1.$$

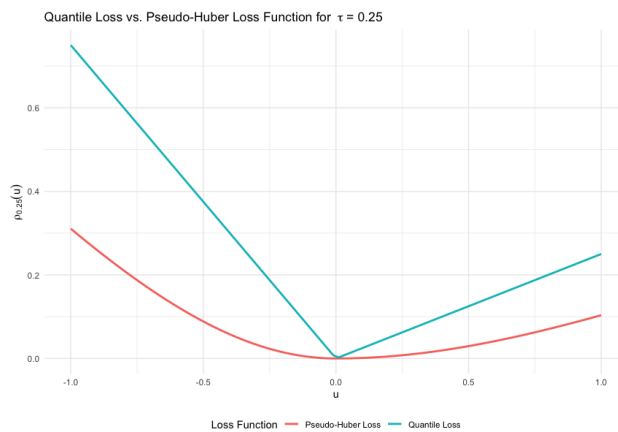


Figure 4.2 – Comparison of quantile and pseudo Huber-type loss functions for $\tau = 0.25$

The influence function ψ_τ is defined as follows :

$$\psi_\tau(Y - \lambda) = \begin{cases} \frac{-(1-\lambda) \cdot |\tau|}{\sqrt{1+(1-\lambda)^2}} & \text{if } Y = 1, \\ \frac{\lambda \cdot |\tau-1|}{\sqrt{1+\lambda^2}}, & \text{if } Y = 0. \end{cases} \quad (4.10)$$

Therefore, based on proposition 4.1, the function $f(\lambda)$ admits a unique solution.

- **Extension of the Huber Loss Function :** Another examples belonging to the Huber type loss function, is an extension of the first example defined in (4.8), where combines quadratic and linear components by selecting the minimum of $\frac{u^2}{2}$ and $|u| - \frac{1}{2}$. This approach ensures a quadratic behavior for small errors, providing smoothness and differentiability near zero, while for larger deviations takes a linear form. The expression of this function is giving as follows :

$$\rho_\tau(u) = |\tau - \mathbf{1}_{(Y < \lambda)}| \min \left(\frac{u^2}{2}, |u| - \frac{1}{2} \right), \quad (4.11)$$

where its influence function $\psi(y - \lambda)$ is expressed as follows :

$$\psi_\tau(Y - \lambda) = \begin{cases} -(Y - \lambda) \cdot |\tau - \mathbf{1}_{(Y < \lambda)}|, & \text{if } \frac{(Y-\lambda)^2}{2} \leq |Y - \lambda| - \frac{1}{2}, \\ -\text{sign}(Y - \lambda) \cdot |\tau - \mathbf{1}_{(Y < \lambda)}|, & \text{if } \frac{(Y-\lambda)^2}{2} > |Y - \lambda| - \frac{1}{2}. \end{cases} \quad (4.12)$$

Therefore by the convexity of the influence function in (4.12), the function $f(\lambda)$ admits a unique solution $\lambda_0(\mathbf{x})$.

Remark 4.3 Note that when the loss function ρ_τ is Huber type loss function, an explicit expression for $\lambda_0(\mathbf{x})$ cannot be derived. As a result, numerical methods are necessary to compute the solution.

4.3.2.2 Expectile loss function

Expectiles are similar to quantiles (Koenker et Bassett, 1978), but they generalize the mean of a distribution in the same way that quantiles generalize the median, offering a certain level of control. First introduced by Newey et Powell (1987), where its expression is given as follows :

$$\rho_\tau(Y - \lambda) = \begin{cases} |\tau - 1| \lambda^2 & \text{if } Y = 0, \\ |\tau| (1 - \lambda)^2 & \text{if } Y = 1; \end{cases}$$

and the influence function is expressed as follows :

$$\psi_\tau(Y - \lambda) = \begin{cases} 2\lambda(1 - \tau), & \text{if } Y = 0, \\ -2(1 - \lambda)\tau, & \text{if } Y = 1. \end{cases} \quad (4.13)$$

According to Proposition 4.1, Equation (4.3) based on the expectile loss function admits a unique root $\lambda_0(\mathbf{x}) \in [0, 1]$, where its expression is given by :

$$\lambda_0^{\text{ex}}(\mathbf{x}) = \frac{\tau \mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})}{\tau \mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x}) + (1 - \tau)(1 - \mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x}))}. \quad (4.14)$$

Equation (4.14) provides an adjusted representation of the conditional probability of success, $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$ (i.e., $\lambda_0^{\text{ex}}(\mathbf{x})$). The parameter τ governs the skewness of the conditional distribution, allowing the adjusted probability to capture different aspects of the conditional distribution effectively.

Specifically, when $\tau = 0.5$, the solution $\lambda_0^{\text{ex}}(\mathbf{x})$ equals the conditional probability $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$, leaving the original probability unchanged. In this case, the transformation reflects the median tendency. For $\tau < 0.5$, $\lambda_0^{\text{ex}}(\mathbf{x})$ decreases the adjusted probability of success compared to the initial probability, particularly when $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$ is low. This adjustment is useful in scenarios where less weight is assigned to rare success events. Conversely, for $\tau > 0.5$, $\lambda_0^{\text{ex}}(\mathbf{x})$ increases the adjusted probability of success, especially when $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$ is already high. This facilitates modeling situations where greater emphasis is placed on frequent success events.

By tuning the parameter τ , the influence of $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$ and $1 - \mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$ on the final estimate can be adjusted. This flexibility is particularly advantageous in contexts with asymmetric data distributions, such as cases with imbalanced responses.

Thus far, we have presented a unified framework to derive the best predictor $\lambda_0(\mathbf{x})$, based on Equation (4.3), which is expressed as a function of $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$. While many methods exist for estimating this probability, our approach leverages copulas to model the conditional probability of success due to their capacity to capture complex dependencies. The following section introduces the copula-based representation of $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$.

4.3.3 Copula-based representation of $\mathbb{P}(Y = y \mid \mathbf{X} = \mathbf{x})$

Accurately estimating the probability $\mathbb{P}(Y = y \mid \mathbf{X} = \mathbf{x})$ is essential for determining the optimal solution $\lambda_0(\cdot)$, which is derived from Equation (4.3). Various methods are available for this estimation, including widely used approaches such as logistic and probit regression. However, these models are inherently limited by their reliance on linearity assumptions, which may fail to capture complex dependencies in the data. Additionally, they are sensitive to multicollinearity, leading to inflated standard errors and potential inconsistencies in estimation.

To overcome these limitations, Mesfioui *et al.* (2023) introduced a copula-based link function for binary regression, providing greater flexibility in modeling the dependence structure between the outcome Y and

the explanatory variables. Unlike logistic and probit models, the copula-based approach expresses the probability of success, $\mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x})$, in terms of a conditional copula. This formulation allows for the modeling of a wide range of dependencies, including complex and nonlinear relationships.

The expression for the conditional probability of success using copulas was derived in Mesfioui *et al.* (2023) as follows :

$$\mathbb{P}(Y \leq y | \mathbf{X} = \mathbf{x}) = F(y | \mathbf{X} = \mathbf{x}) = C_{\theta}^c(F_0(y) | \mathbf{F}(\mathbf{x})), \quad (4.15)$$

where $C_{\theta}^c(u | \mathbf{v}) := \frac{\partial^d C_{\theta}(u, \mathbf{v})}{\partial v_1 \dots \partial v_d} \Big/ \frac{\partial^d C_{\theta}(1, \mathbf{v})}{\partial v_1 \dots \partial v_d}$ denotes the conditional distribution of the copula proposed by Bouyé et Salmon (2009).

Note that the copula C_{θ} has a unique representation on $Ran(F_0) \times Ran(\mathbf{F})$ as demonstrated in Genest et Nešlehová (2007), where $Ran(F)$ stands for the range of F . According to the representation of $\mathbb{P}(Y \leq y | \mathbf{X} = \mathbf{x})$ in equation (4.15), one can express the following probability $\mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x})$ as follows :

$$\mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x}) = 1 - C_{\theta}^c(1 - p | \mathbf{F}(\mathbf{x})), \quad (4.16)$$

where $p = \mathbb{P}(Y = 1)$.

Proposition 4.4 *Substituting the conditional probability of success given in (4.16) into (4.3), one can get :*

$$f^c(\lambda_0(\mathbf{x})) = \psi_{\tau}(-\lambda_0(\mathbf{x})) (1 - \pi(\mathbf{x})) + \psi_{\tau}(1 - \lambda_0(\mathbf{x})) \pi(\mathbf{x}) = 0, \quad (4.17)$$

where $\pi(\mathbf{x}) = 1 - C_{\theta}^c(1 - p | \mathbf{F}(\mathbf{x}))$ and $\lambda_0(\mathbf{x})$ is the root of Equation (4.17).

Proposition 4.4 provides an alternative representation of Equation (4.3), offering a copula-based solution. To illustrate this, we consider the case where the influence function $\psi_{\tau}(\cdot)$ takes the form of the expectile, as defined in (4.13). In this scenario, the theoretical solution $\lambda_0(\mathbf{x})$ has the same structure as equation (4.14), with $\mathbb{P}(Y = 1 | \mathbf{X} = \mathbf{x})$ replaced by the conditional distribution of copulas, as defined in (4.16).

Remark 4.5 *Of note, if the dependence between $\{Y, X_1\}$ is modeled by a Frank copula, the predictive probability of success $\mathbb{P}(Y = 1 | X_1 = x_1)$ can be expressed as follows :*

$$\begin{aligned} \mathbb{P}(Y = 1 | X_1 = x_1) &= \pi(x_1) \\ &= 1 - C_{\theta}^c\left(1 - p | F_1(x_1)\right) \\ &= \frac{\exp\left(\theta F_1(x_1) + \log(\exp[\varphi_{\theta}(1 - p)] - 1)\right)}{1 + \exp\left(\theta F_1(x_1) + \log(\exp[\varphi_{\theta}(1 - p)] - 1)\right)}, \end{aligned} \quad (4.18)$$

here φ_θ represents the generator of the Frank copula. Combining the result found in equation (4.18) and (4.14), one can express $\lambda_0(x_1)$ based on Frank copula as follows :

$$\begin{aligned} \lambda_0(x_1) &= \frac{\tau\pi(x_1)}{(1-\tau)(1-\pi(x_1)) + \tau\pi(x_1)} \\ &= \frac{\exp\left(\theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1) + \log\frac{\tau}{1-\tau}\right)}{1 + \exp\left(\theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1) + \log\frac{\tau}{1-\tau}\right)}. \end{aligned} \quad (4.19)$$

The complete proof of the expression provided in (4.19) is deferred to Appendix A.3.1. Furthermore, this derivation introduces a novel expression for $\lambda_0(x_1)$ based on copulas, which bears a close resemblance to the logistic regression form. This expression is evaluated at the margin of the covariate, with the slope coefficient determined by the Frank copula parameter θ . The intercept is defined as follows :

$$\beta_0 = \log(\exp[\varphi_\theta(1-p)] - 1) + \log\left(\frac{\tau}{1-\tau}\right).$$

Therefore, controlling the intercept through the parameter τ , maintaining our adjusted probability $\lambda_0(x_1)$ based on expectile given in (4.19) for different values of τ non-crossing, we achieve non crossing condition as assumed in (Manski et Thompson, 1989), in order to obtain the best predictor for binary response.

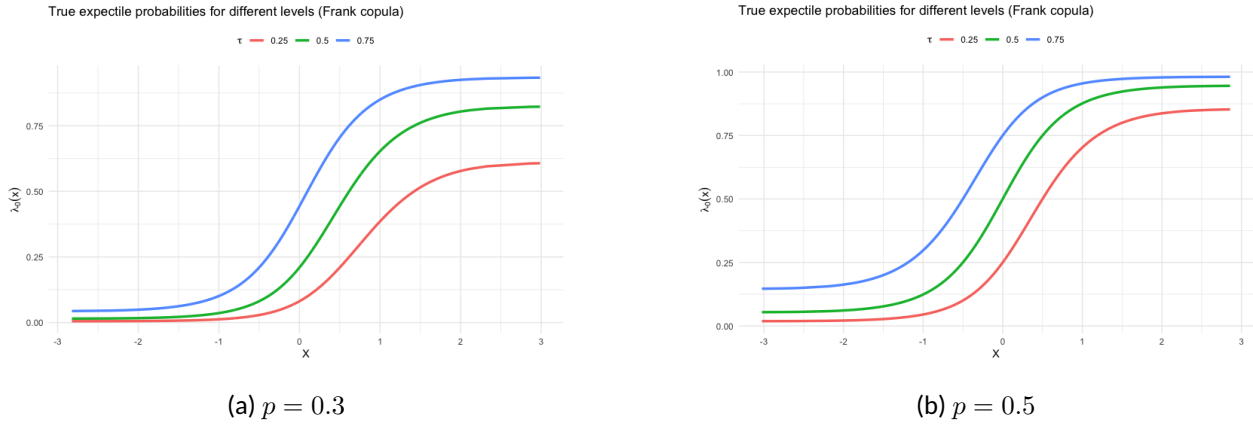


Figure 4.3 – Graphical representation of $\lambda_0(\cdot)$ across expectile levels τ based on equation (4.17), with the Kendall parameter fixed at 0.5

To illustrate, the fact that our solution meets the single crossing condition, Figure 4.3 displays the theoretical solution $\lambda_0(x_1)$ giving in (4.19) for different expectile levels $\tau = 0.25, 0.5$, and 0.75 . This visualization highlights different aspects of the conditional probability of success, with each curve representing an expectile, a generalization of the mean that asymmetrically adjusts the deviations from the mean tendency

(i.e., $\mathbb{P}(Y = 1|X_1 = x_1)$), which improves the modelling of asymmetric or heavy-tailed distributions. Also, each curve of $\lambda_0(x_1)$ for a specific τ adheres to the single-crossing principle, showing that the estimation curves intersect only once. This ensures the monotonicity of the optimal predictor giving in (4.19) in each expectile level.

Up to now, we derived the theoretical solution $\lambda_0(x_1)$, which represents the root of equation (4.17) and describes several aspects of $\mathbb{P}(Y = 1|X_1 = x_1)$. The next subsection presents a semi-parametric estimator of $\lambda_0(\cdot)$ based on a given sample.

4.3.4 Semi-parametric estimator of $\lambda_0(\cdot)$

This section presents a semi-parametric estimator for the theoretical solution $\lambda_0(\cdot)$ given in Proposition 4.4, based on an observed sample. Before exploring the details, consider independent copies $(Y_1, \mathbf{X}_1), (Y_2, \mathbf{X}_2), \dots, (Y_n, \mathbf{X}_n)$ of (Y, \mathbf{X}) where $Y_i \in \{0, 1\}$ and $\mathbf{X}_i = (X_{i1}, \dots, X_{id}) \in \mathbb{R}^d$.

Proposition 4.6 *The sample version of $\lambda_0(\cdot)$, denoted by $\hat{\lambda}_n(\cdot)$ which represents the root of the following equation :*

$$f_n^c(\hat{\lambda}_n(\mathbf{x})) = \psi_\tau(1 - \hat{\lambda}_n(\mathbf{x})) \hat{\pi}_n(\mathbf{x}) + \psi_\tau(-\hat{\lambda}_n(\mathbf{x})) (1 - \hat{\pi}_n(\mathbf{x})) = 0, \quad (4.20)$$

where $\hat{p} = n^{-1} \sum_{i=1}^n Y_i$, $\hat{\mathbf{F}}(\mathbf{x}) = (\hat{F}_{n1}(x_1), \dots, \hat{F}_{np}(x_p))$, $\hat{F}_{nk}(x_k) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(X_{ik} \leq x_k)$ and $\hat{\pi}_n(\mathbf{x}) = 1 - C_\theta^c(1 - \hat{p} | \hat{\mathbf{F}}(\mathbf{x}))$ and $\hat{\theta}$ is the pseudo-likelihood estimator of θ as investigated in Mesfioui et al. (2023).

Proof. Based on equation (4.17), one can write $f^c(\lambda)$ as follows :

$$f^c(\lambda) = \psi_\tau(1 - \lambda)\pi(\mathbf{x}) + \psi_\tau(-\lambda)(1 - \pi(\mathbf{x})) = 0.$$

The estimator $\hat{\lambda}_n(\mathbf{x})$ of $\lambda_0(\mathbf{x})$, will be the solution of the following equation :

$$f_n^c(\hat{\lambda}_n(\mathbf{x})) = \psi_\tau(1 - \hat{\lambda}_n(\mathbf{x})) \hat{\pi}_n(\mathbf{x}) + \psi_\tau(-\hat{\lambda}_n(\mathbf{x})) (1 - \hat{\pi}_n(\mathbf{x})),$$

this completes the proof. \square

Example 4.7 Building on the example presented in Remark 4.5, the solution to equation (4.20), as established in Proposition 4.6, provides an explicit expression for $\hat{\lambda}_n(x_1)$ given by :

$$\begin{aligned}\hat{\lambda}_n(x_1) &= \frac{\tau \hat{\pi}(x_1)}{(1 - \tau)(1 - \hat{\pi}(x_1)) + \tau \hat{\pi}(x_1)} \\ &= \frac{\exp\left(\hat{\theta} \hat{F}_{n1}(x_1) + \log\left(\exp[\varphi_{\hat{\theta}}(1 - \hat{p})] - 1\right) + \log\left(\frac{\tau}{1 - \tau}\right)\right)}{1 + \exp\left(\hat{\theta} \hat{F}_{n1}(x_1) + \log\left(\exp[\varphi_{\hat{\theta}}(1 - \hat{p})] - 1\right) + \log\left(\frac{\tau}{1 - \tau}\right)\right)}.\end{aligned}\quad (4.21)$$

Equation (4.21) provides a plug-in version of the theoretical solution given in Equation (4.19).

4.4 Asymptotic properties

Before presenting the asymptotic properties of the proposed estimator, $\hat{\lambda}_n(\mathbf{x})$, we introduce several assumptions to establish its i.i.d. asymptotic representation. These properties are derived from the behavior of $\hat{\pi}(\mathbf{x})$, as studied in Mesfioui *et al.* (2023). This section also includes two key theorems : Theorem 4.8 establishes the consistency of the proposed estimator, which is essential for demonstrating its asymptotic properties. Theorem 4.9 provides the asymptotic representation of the difference between the theoretical solution and its estimator, (i.e., $\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x})$), showing that it converges to a normal distribution as n tends to infinity.

4.4.1 Assumption

1. $\rho_\tau(\cdot)$ belongs to the asymmetric loss function defined in (4.2).

2. for a compact set $\mathbb{V} = [0, 1]$ we have :

- (i) $\mathbb{E}\left[\sup_{\lambda \in \mathbb{V}} \{\rho_\tau^2(Y - \lambda)\}\right] < \infty$,
- (ii) $\mathbb{E}\left[\left|\rho_\tau(Y - \lambda)\right| \middle| \mathbf{X} = \mathbf{x}\right] < \infty$.

Note that conditions provided in Assumption 2 are sufficient requirement for applying the uniform law of large numbers.

3. The hypotheses **B** presented bellow (Mesfioui *et al.*, 2023) hold for a semi parametric estimator of θ to guaranty the following representation

$$\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathcal{Z}_i(\mathbf{x}) + o_p\left(n^{-1/2}\right)$$

$$\begin{aligned} \mathcal{Z}_i(\mathbf{x}) &= (\mathbb{1}(Y_i = 1) - p) \partial_u \mathcal{C}_\theta^c(1 - p | \mathbf{F}(\mathbf{x})) - \zeta_i^\top \Delta_\theta \mathcal{C}_\theta^c(1 - p | \mathbf{F}(\mathbf{x})) \\ &\quad - (\mathbb{1}(\mathbf{X}_i \leq \mathbf{x}) - \mathbb{P}(\mathbf{X} \leq \mathbf{x})) \Delta_{\mathbf{v}} \mathcal{C}_\theta^c(1 - p | \mathbf{F}(\mathbf{x})), \end{aligned} \quad (4.22)$$

where $\partial_u \mathcal{C}_\theta^c(u | \mathbf{v}) = \frac{\partial_u \mathcal{C}_\theta^c(u | \mathbf{v})}{\partial u}$, $\Delta_\theta \mathcal{C}_\theta^c(u | \mathbf{v}) = \left(\frac{\partial \mathcal{C}_\theta^c(u | \mathbf{v}; \theta)}{\partial \theta_1}, \dots, \frac{\partial \mathcal{C}_\theta^c(u | \mathbf{v})}{\partial \theta_r} \right)^\top$,
 $\Delta_{\mathbf{v}} \mathcal{C}_\theta^c(u | \mathbf{v}) = \left(\frac{\partial \mathcal{C}_\theta^c(u | \mathbf{v})}{\partial v_1}, \dots, \frac{\partial \mathcal{C}_\theta^c(u | \mathbf{v})}{\partial v_d} \right)^\top$, $\mathbb{P}(\mathbf{X} \leq \mathbf{x}) = (\mathbb{P}(X_1 \leq x_1), \dots, \mathbb{P}(X_d \leq x_d))$, and
 $\mathbb{1}(\mathbf{X}_i \leq \mathbf{x}) = (\mathbb{1}(X_{i1} \leq x_1), \dots, \mathbb{1}(X_{ip} \leq x_p))$.

Here, ζ_i denotes a random vector with zero mean and finite second moment, i.e., $\mathbb{E}\|\zeta_i\|^2 < \infty$, as defined in Lemma 1 of Mesfioui *et al.* (2023).

— **Hypothesis B**

(B.1) $\mathcal{C}_\theta^c(\cdot)$ is twice continuously differentiable in a neighborhood of the true parameter θ and it is bounded away from zero.

(B.2) The information matrix $I(\theta) = \mathbb{E} \left(-\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta^\top} \right)$ exists and is non-singular, where $\mathcal{L}(\cdot)$ is defined in (4.25).

(B.3) $\frac{\partial^2 \log \mathcal{C}_\theta^c(u | \mathbf{v})}{\partial u \partial \theta_t}$ and $\frac{\partial^2 \log \mathcal{C}_\theta^c(u | \mathbf{v})}{\partial v_k \partial \theta_t}$ are continuous on $\Theta \times (0, 1) \times (0, 1)^d$, for $k = 1, \dots, d$, and $\Theta \subset \mathbb{R}^r$.

Theorem 4.8 (Consistency of $\hat{\lambda}_n(\mathbf{x})$) *Under Assumptions 1 through 3, the estimator $\hat{\lambda}_n(\mathbf{x})$ is consistent for the true parameter $\lambda_0(\mathbf{x})$. Specifically, for any fixed $\mathbf{x} \in \mathbb{R}^d$, the following convergence in probability holds :*

$$\hat{\lambda}_n(\mathbf{x}) \xrightarrow{P} \lambda_0(\mathbf{x}) \quad \text{as } n \rightarrow \infty.$$

Proof. Under Equation (4.1), one can write $\lambda_0(\mathbf{x})$, and $\hat{\lambda}_n(\mathbf{x})$ based on copulas as follows :

$$\lambda_0(\mathbf{x}) = \operatorname{argmin}_{\lambda \in \mathbb{V}} \rho_\tau(1 - \lambda) \pi(\mathbf{x}) + \rho_\tau(-\lambda) (1 - \pi(\mathbf{x})), \quad (4.23)$$

and

$$\hat{\lambda}_n(\mathbf{x}) = \operatorname{argmin}_{\lambda \in \mathbb{V}} \rho_\tau(1 - \lambda) \hat{\pi}_n(\mathbf{x}) + \rho_\tau(-\lambda) (1 - \hat{\pi}_n(\mathbf{x})). \quad (4.24)$$

Fix \mathbf{x} and define, for $\lambda \in \mathbb{V} = [0, 1]$,

$$Q(\lambda; \mathbf{x}) := \rho_\tau(1 - \lambda) \pi(\mathbf{x}) + \rho_\tau(0 - \lambda) [1 - \pi(\mathbf{x})], \quad Q_n(\lambda; \mathbf{x}) := \rho_\tau(1 - \lambda) \hat{\pi}_n(\mathbf{x}) + \rho_\tau(0 - \lambda) [1 - \hat{\pi}_n(\mathbf{x})].$$

Let

$$\lambda_0(\mathbf{x}) = \arg \min_{\lambda \in \mathbb{V}} Q(\lambda; \mathbf{x}), \quad \hat{\lambda}_n(\mathbf{x}) = \arg \min_{\lambda \in \mathbb{V}} Q_n(\lambda; \mathbf{x}).$$

For any $\lambda \in \mathbb{V}$,

$$\begin{aligned} Q_n(\lambda; \mathbf{x}) - Q(\lambda; \mathbf{x}) &= \rho_\tau(1 - \lambda)(\hat{\pi}_n(x) - \pi(\mathbf{x})) + \rho_\tau(0 - \lambda)(1 - \hat{\pi}_n(\mathbf{x}) - (1 - \pi(\mathbf{x}))) \\ &= (\hat{\pi}_n(\mathbf{x}) - \pi(x)) (\rho_\tau(1 - \lambda) - \rho_\tau(0 - \lambda)), \end{aligned}$$

hence

$$|Q_n(\lambda; \mathbf{x}) - Q(\lambda; \mathbf{x})| = |\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x})| |\rho_\tau(1 - \lambda) - \rho_\tau(0 - \lambda)|.$$

Taking the supremum over $\lambda \in \mathbb{V}$ yields

$$\sup_{\lambda \in \mathbb{V}} |Q_n(\lambda; \mathbf{x}) - Q(\lambda; \mathbf{x})| \leq |\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x})| \sup_{\lambda \in \mathbb{V}} |\rho_\tau(1 - \lambda) - \rho_\tau(0 - \lambda)|.$$

Assume that $\lambda \mapsto \rho_\tau(y - \lambda)$ is continuous on \mathbb{V} for $y \in \{0, 1\}$. Since \mathbb{V} is compact, the function $\lambda \mapsto |\rho_\tau(1 - \lambda) - \rho_\tau(0 - \lambda)|$ attains its bounds on \mathbb{V} , so

$$M := \sup_{\lambda \in \mathbb{V}} |\rho_\tau(1 - \lambda) - \rho_\tau(0 - \lambda)| < \infty.$$

Therefore,

$$\sup_{\lambda \in \mathbb{V}} |Q_n(\lambda; \mathbf{x}) - Q(\lambda; \mathbf{x})| \leq M |\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x})|.$$

Under our assumptions, $\hat{\pi}_n(\mathbf{x}) \xrightarrow{P} \pi(x)$, which implies

$$\sup_{\lambda \in \mathbb{V}} |Q_n(\lambda; \mathbf{x}) - Q(\lambda; \mathbf{x})| \xrightarrow{P} 0.$$

Finally, combining this uniform convergence with the uniqueness of the minimizer $\lambda_0(\mathbf{x})$ (Proposition 4.1), one can deduce that :

$$\hat{\lambda}_n(\mathbf{x}) \xrightarrow{\mathbb{P}} \lambda_0(\mathbf{x}).$$

This complete the proof. \square

After demonstrating the consistency of our estimator $\lambda_n(\mathbf{x})$, we now introduce some useful notation to facilitate the proof of its asymptotic normality. To this end, we define the following two functions :

$$\begin{aligned} f^c(\lambda) &= \psi_\tau(1 - \lambda) \pi(\mathbf{x}) + \psi_\tau(-\lambda) (1 - \pi(\mathbf{x})), \\ f_n^c(\lambda) &= \psi_\tau(1 - \lambda) \hat{\pi}_n(\mathbf{x}) + \psi_\tau(-\lambda) (1 - \hat{\pi}_n(\mathbf{x})). \end{aligned}$$

Theorem 4.9 (Asymptotic representation) Under Assumptions 1 to 3. One has that :

$$\sqrt{n} \left(\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x}) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{K}_i(\mathbf{x}) + o_p(n^{-\frac{1}{2}}),$$

which implies that :

$$\sqrt{n} \left(\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x}) \right) \sim \mathcal{N} \left(0, \sigma^2(\mathbf{x}) \right),$$

where $\sigma^2(\mathbf{x}) = \mathbb{E} \left\{ [\mathcal{K}_i(\mathbf{x})]^2 \right\}$, $\mathcal{K}_i(\mathbf{x}) = [(f^c(\lambda_0(\mathbf{x})))']^{-1} A(\lambda_0(\mathbf{x})) \mathcal{Z}_i(\mathbf{x})$,

$A(\lambda) = \psi_\tau(-\lambda) - \psi_\tau(1-\lambda)$, and $\mathcal{Z}_i(\cdot)$ is given in (4.22), with $[(f^c(\lambda_0(\mathbf{x})))']^{-1} > 0$ due to the convexity of $\rho_\tau(\cdot)$.

Proof. By definition, $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\mathbf{x})$ are the roots to equations (4.17) and (4.20) respectively.

Applying the mean value theorem between $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\mathbf{x})$, one can write :

$$0 = \sqrt{n} f^c(\lambda_0(\mathbf{x})) = \sqrt{n} f^c(\hat{\lambda}_n(\mathbf{x})) - (f^c(\tilde{\lambda}(\mathbf{x})))' \sqrt{n} \left(\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x}) \right),$$

where $\tilde{\lambda}(\mathbf{x})$ lies between $\hat{\lambda}_n(\mathbf{x})$ and $\lambda_0(\mathbf{x})$, and $(f^c(\cdot))'$ is a continuous function in $\lambda(\mathbf{x})$. Hence, by Theorem (4.8), one can write that :

$$\begin{aligned} \sqrt{n} \left(\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x}) \right) &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} \sqrt{n} f^c \left(\hat{\lambda}_n(\mathbf{x}) \right) + o_p(1) \text{ since } \hat{\lambda}_n \text{ is the root of } f_n^c(\cdot) \\ &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} \sqrt{n} \left(f^c \left(\hat{\lambda}_n(\mathbf{x}) \right) - f_n^c \left(\hat{\lambda}_n(\mathbf{x}) \right) \right) + o_p(1) \\ &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} \sqrt{n} \left[\psi_\tau \left(1 - \hat{\lambda}_n(\mathbf{x}) \right) (\pi(\mathbf{x}) - \hat{\pi}_n(\mathbf{x})) \right. \\ &\quad \left. + \psi_\tau \left(0 - \hat{\lambda}_n(\mathbf{x}) \right) ((1 - \pi(\mathbf{x})) - (1 - \hat{\pi}_n(\mathbf{x}))) \right] + o_p(1) \\ &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} \left[\psi_\tau(0 - \hat{\lambda}_n(\mathbf{x})) - \psi_\tau(1 - \hat{\lambda}_n(\mathbf{x})) \right] \\ &\quad \times \sqrt{n} (\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x})) + o_p(1) \\ &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} \left[A(\hat{\lambda}_n(\mathbf{x})) \right] \times \sqrt{n} (\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x})) + o_p(1), \end{aligned}$$

where

$$A(\lambda(\mathbf{x})) = \psi_\tau(-\lambda(\mathbf{x})) - \psi_\tau(1-\lambda(\mathbf{x})).$$

Since $A(\cdot)$ is continuous, it follows that $A(\hat{\lambda}_n(\mathbf{x})) = A(\lambda_0(\mathbf{x})) + o_p(1)$, one can write :

$$\begin{aligned} \sqrt{n} \left(\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x}) \right) &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} A(\lambda_0(\mathbf{x})) \sqrt{n} [\hat{\pi}_n(\mathbf{x}) - \pi(\mathbf{x})] + o_p(1) \\ &= [(f^c(\lambda_0(\mathbf{x})))']^{-1} A(\lambda_0(\mathbf{x})) \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{Z}_i(\mathbf{x}) + o_p(1), \text{ (By assumption 3)} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathcal{K}_i(\mathbf{x}) + o_p(1), \text{ with } \mathcal{K}_i(\mathbf{x}) = [(f^c(\lambda_0(\mathbf{x})))']^{-1} A(\lambda_0(\mathbf{x})) \mathcal{Z}_i(\mathbf{x}), \end{aligned}$$

which completes the proof \square

Theorem 4.9 states that the random quantity $\sqrt{n} \left(\hat{\lambda}_n(\mathbf{x}) - \lambda_0(\mathbf{x}) \right) \sim \mathcal{N}(0, \sigma^2(\mathbf{x}))$.

In practice, the variance $\sigma^2(\mathbf{x})$ can be approximated using the sample variance of $\mathcal{K}_i(\mathbf{x})$:

$$\hat{\sigma}_n^2(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \left(\hat{\mathcal{K}}_i(\mathbf{x}) - \frac{1}{n} \sum_{i=1}^n \hat{\mathcal{K}}_i(\mathbf{x}) \right)^2,$$

where $\hat{\mathcal{K}}_i(\mathbf{x})$ takes the same form as $\mathcal{K}_i(\mathbf{x})$, but is calculated using $\hat{\pi}_n(\mathbf{x})$ and $\hat{\boldsymbol{\theta}}$ instead of $\pi(\mathbf{x})$ and $\boldsymbol{\theta}$.

In our context we employ a bootstrap approach to estimate the variance $\sigma_n^2(\mathbf{x})$. The complete bootstrap procedure is described in Algorithm 5 of Appendix A.3.3.1. The consistency of the bootstrap method can be shown using arguments similar to those assumed in Rémillard *et al.* (2017).

4.5 Illustration and simulation of the optimal predictor based on copulas

This section presents several illustrations of the optimal predictor based on copulas, $\lambda_0(\mathbf{x})$, together with its estimator, $\hat{\lambda}_n(\mathbf{x})$, obtained using the Huber-type and expectile loss functions introduced in Section 4.3.2. The aim of these illustrations is to verify that the predictor satisfies both the single-crossing and isotonicity conditions, and to visually demonstrate that, when evaluated at different τ levels, the resulting predictors remain non-intersecting.

4.5.1 Data generation process

The data were generated following Algorithm 3. Each variable's marginal distribution was specified using a parametric copula family, with the number of covariates fixed at $d = 2$. The subsequent steps of the procedure are described in the next algorithm.

Algorithm 3 Data Generation Process (DGP)

- 1: Generate $U, V_1, V_2 \sim \mathcal{C}_\theta$.
- 2: Set $Y_i = \mathbb{1}(U_i \geq 1 - p)$ and $X_{i,j} = F_j^{-1}(V_{i,j}), i = 1, \dots, n, j = 1, 2$.
- 3: We estimate p and $F_j(x_j)$ by $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n Y_i$, and $\hat{F}_{nj}(x_j) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{1}(X_{ij} \leq x_j)$, for $j = 1, 2$, respectively.
- 4: Estimate $\hat{\theta} = \underset{\theta \in \mathbb{R}^l}{\operatorname{argmax}} \mathcal{L}(\theta)$, where

$$\mathcal{L}(\theta) = \sum_{i=1}^n \left\{ Y_i \log \left(1 - \mathcal{C}_\theta \left(1 - \hat{p}_n | \hat{\mathbf{F}}(\mathbf{x}) \right) \right) + (1 - Y_i) \log \left(\mathcal{C}_\theta \left(1 - \hat{p}_n | \hat{\mathbf{F}}(\mathbf{x}) \right) \right) \right\}, \quad (4.25)$$

with $\hat{\mathbf{F}}(\mathbf{x}) = \left(\hat{F}_{n1}(x_1), \hat{F}_{n2}(x_2) \right)$, and $\mathbf{x} = (x_1, x_2)$.

- 5: Based on $\hat{\theta}$ and $\hat{\mathbf{F}}(\mathbf{x})$, compute $\hat{\pi}_n(\mathbf{x})$.
-

After generating the data according to the DGP described in Algorithm 3, the next step is to find the roots of equations (4.17) and (4.20), denoted respectively by $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\mathbf{x})$. The process for obtaining these values is outlined in the following algorithm :

Algorithm 4 Evaluation $\lambda_0(\mathbf{x})$, and $\hat{\lambda}_n(\mathbf{x})$

1: Initialization :

- Compute $\hat{\pi}_n(\mathbf{x}) = 1 - \mathcal{C}_\theta(1 - \hat{p}_n | \hat{\mathbf{F}}(\mathbf{x}))$ where \mathbf{x} is value taken from the support of the covariates \mathbf{X} as defined in Algorithm 3.
- Choose an asymmetric loss function $\rho_\tau(u)$ from the set of loss function based on the Huber function, adapted for quantile regression or expectile.

2: Optimization Procedure :

- Solve equation (4.17) to obtain $\lambda_0(\mathbf{x})$.
 - Solve equation (4.20) to obtain $\hat{\lambda}_n(\mathbf{x})$.
-

4.5.2 Solution based on Huber-type loss function

Figure 4.4 presents the fitted adjusted predictive probability, $\lambda_0(\mathbf{x})$, alongside its estimator, $\hat{\lambda}_n(\mathbf{x})$, for various quantile levels, using a pseudo Huber-type loss function. The dependence between outcomes and covariates is modeled using a Joe copula. Notably, the curves for the adjusted probability and its estimator remain closely aligned across different covariate values, demonstrating the estimator's consistency, as

detailed in Section 4.4. Moreover, the curves for $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\mathbf{x})$ exhibit monotonicity and do not intersect as covariate values increase, regardless of the quantile levels ($\tau = 0.25$, $\tau = 0.5$, and $\tau = 0.75$). These quantile levels capture distinct features of the conditional distribution $\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x})$. For instance, $\tau = 0.5$ corresponds to the median, reflecting the central tendency of the conditional probability. In contrast, $\tau = 0.75$ emphasizes the upper tail of the distribution, highlighting scenarios with a higher likelihood of $Y = 1$. Conversely, lower quantile levels, such as $\tau = 0.25$, focus on the lower tail, assigning important weight to the complement probability of success, as illustrated by the red curve in Figure 4.4. The representation of $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\mathbf{x})$ in the presence of covariates based on the pseudo Huber loss is shown in Figure 4.5. The same conclusion applies to the representation of $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\mathbf{x})$ based on the Huber-type loss function, as presented in Figures A.3 and 4.6 in Appendix A.3.2.

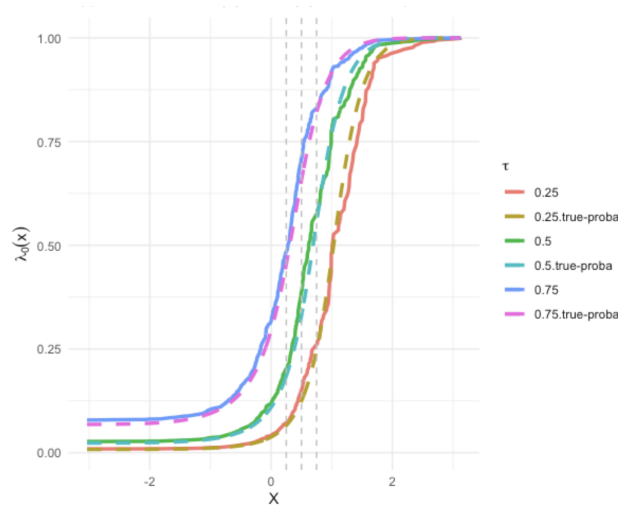
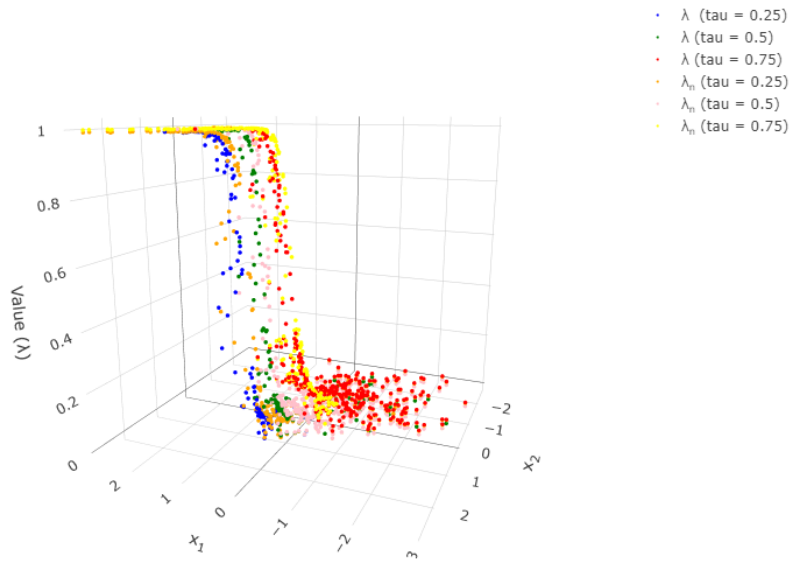
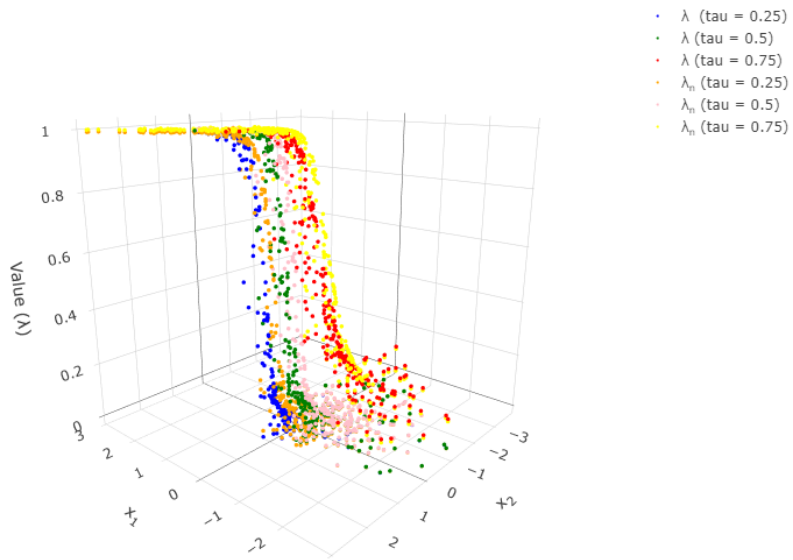


Figure 4.4 - Representation of $\lambda_0(\mathbf{x})$, and $\hat{\lambda}_n(\mathbf{x})$ on different quantile level based on pseudo Huber loss function with $p = 0.3$.

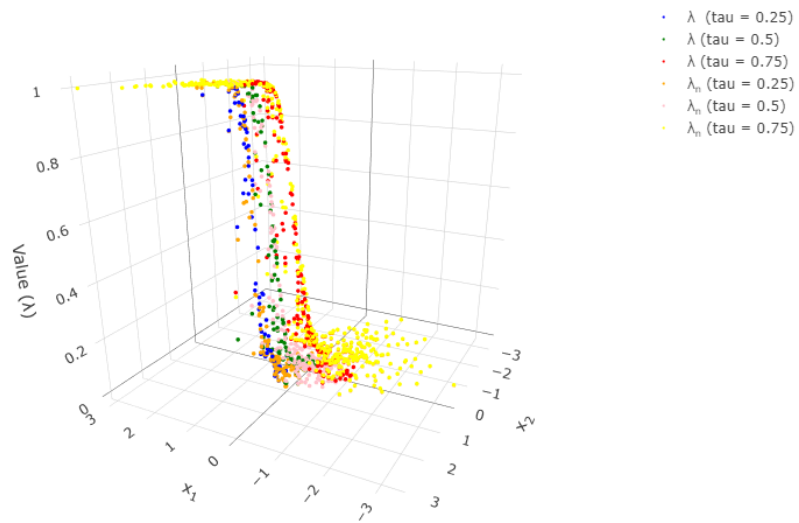


(a) $p = 0.3$

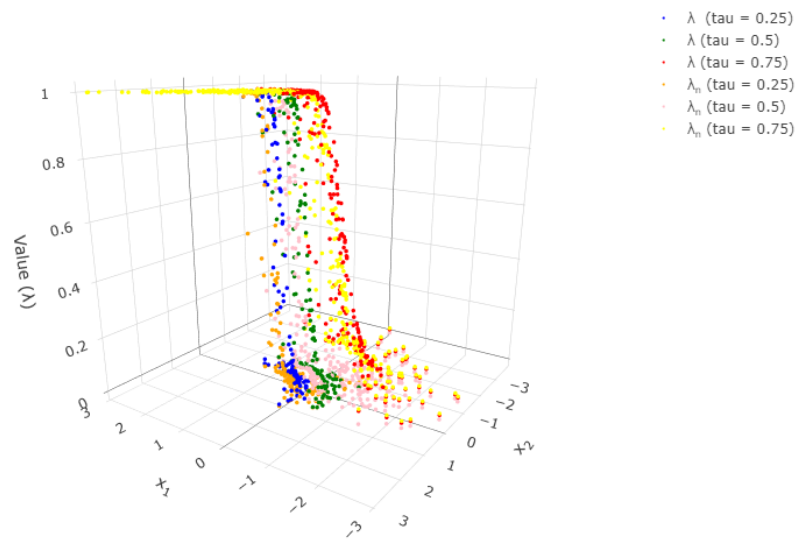


(b) $p = 0.6$

Figure 4.5 - Representation of $\lambda_0(\mathbf{x})$, and $\hat{\lambda}_n(\mathbf{x})$ on different quantile level based on pseudo Huber loss function, where $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$



(a) $p = 0.3$



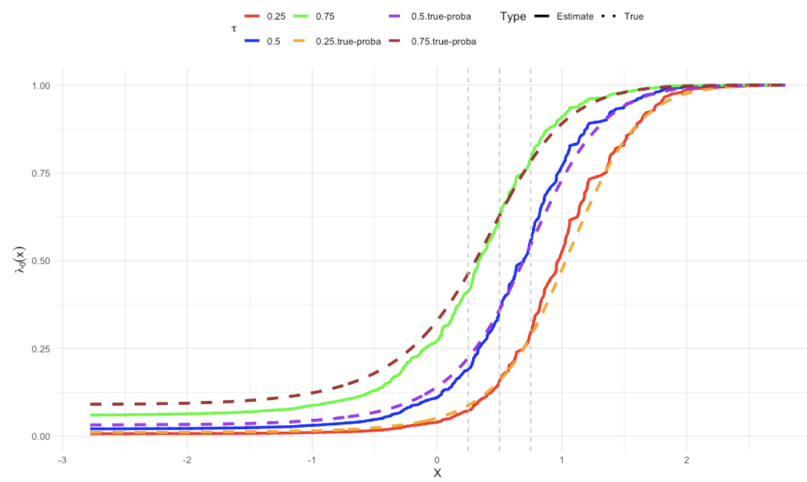
(b) $p = 0.6$

Figure 4.6 - Representation of $\lambda_0(\mathbf{x})$ and $\hat{\lambda}_n(\cdot)$ based on Huber-type loss function defined in (4.8) with $k = 0.05$ and $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$.

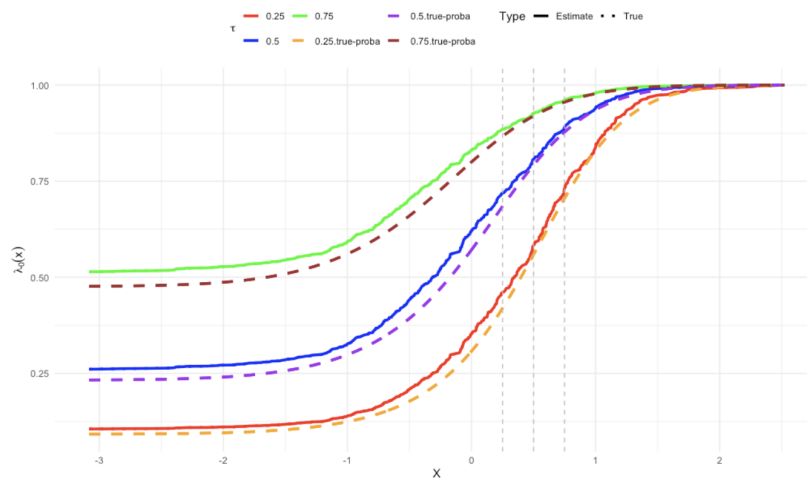
4.5.3 Optimal predictor based on expectile loss

Figure 4.7 illustrates the adjusted probability of success $\lambda_0(\mathbf{x})$ (i.e., the root of equation (4.17)) and its empirical version $\hat{\lambda}_n(\mathbf{x})$ (i.e., the root of equation (4.20)), based on the expectile loss function and Joe copula across different expectile levels ($\tau = 0.25$, $\tau = 0.5$, and $\tau = 0.75$). The solid lines represent $\hat{\lambda}_n(\mathbf{x})$, while the dashed lines indicate $\lambda_0(\mathbf{x})$.

The close alignment between the estimated and true curves demonstrating the consistency of our proposed estimator of $\lambda_0(\mathbf{x})$, as shown in Section 4.4. Moreover, for higher expectile levels ($\tau = 0.75$), our predictor captures the upper tail of the conditional distribution $\mathbb{P}(Y = 1 \mid X = x)$ by emphasizing larger values of x and giving more weight to the class $Y = 1$; its curve shifts rightward from the median tendency ($\tau = 0.5$). Similarly, smaller expectile levels ($\tau = 0.25$) capture the lower tail of the distribution. Furthermore, the method respects the one-crossing and isotonicity assumptions, ensuring monotonicity across different τ levels without the need to explicitly impose these conditions within the model.

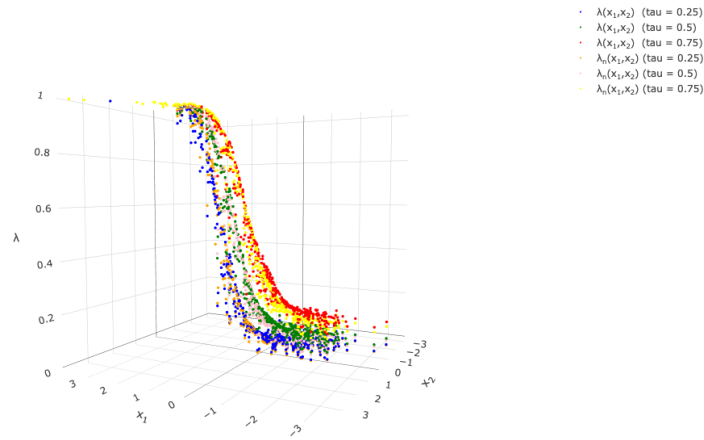


(a) $p = 0.3$

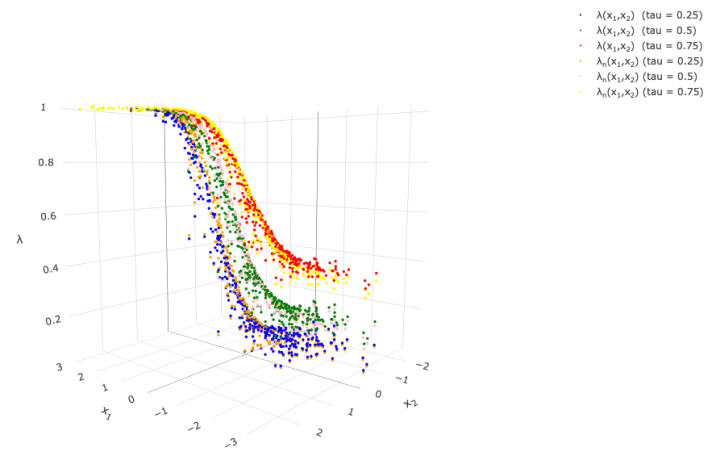


(b) $p = 0.6$

Figure 4.7 – The representations of $\lambda_0(\cdot)$ and $\hat{\lambda}_n(\cdot)$ are derived based on the expectile loss function and Joe copula, evaluated at different expectile levels τ , where $X_1 \sim \mathcal{N}(0, 1)$.



(a) $p = 0.3$



(b) $p = 0.6$

Figure 4.8 – The representations of $\lambda_0(\cdot)$ and $\hat{\lambda}_n(\cdot)$ are derived based on the expectile loss function and Joe copula, evaluated at different expectile levels τ , with $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$

4.5.4 Simulation study

Monte Carlo simulations were conducted to evaluate and compare our proposed method, outlined in Section 4.3.4, with competing approaches such as random forests (Breiman, 2001), logistic regression (Cox, 1958), support vector machines (SVM) (Cortes et Vapnik, 1995), and asymmetric SVM (Huang et Du, 2005). The asymmetric SVM modifies the standard SVM by adjusting its objective function to assign greater weight

to errors in the minority class, thereby balancing errors across classes. Details of the algorithm implementation are provided in Appendix A.3.4. Synthetic data were generated using Algorithm 3 with various parameter combinations ($n = 1000, \tau^c$), where τ^c represents Kendall's tau for a specified copula, characterizing the dependence structure between the outcome and explanatory variables. By selecting τ^c values of 0.5 and 0.7, four distinct scenarios were created, each corresponding to a different dependence structure (Joe, Gumbel, Frank, and Gaussian copula). Additionally, an imbalanced dataset was simulated by setting the class proportion to $p = 0.1$, ensuring underrepresentation of the $Y = 1$ class.

To assess $\hat{\lambda}_n(\mathbf{x})$, we first estimate the copula parameter θ by maximizing the log likelihood defined in (4.25) in order to assess $\hat{\pi}(\mathbf{x})$. Next, we define a grid of τ values, specifically $\tau \in \{0.001, 0.0015, 0.002, \dots, 0.95\}$. We then perform a cross-validation procedure to evaluate $\hat{\lambda}_n(\mathbf{x})$ at each value within this grid. For a chosen τ from the grid, $\hat{\lambda}_n(\mathbf{x})$ is computed using the *uniroot* function in R. Finally, we select the τ that maximizes the F1-score, which evaluates the method's ability to balance between the two categories ($Y = 0$ and $Y = 1$), thereby identifying the optimal τ .

For the competing methods, including Support Vector Machines (SVM) and its asymmetric variant (asymmetric SVM), we used the *svm* function from the **e1071** package. Additionally, for the random forest method, we employed the **randomForest** function available in the **randomForest** package. All the aforementioned packages were implemented in the R software.

Noting the similarity of the results obtained with the different Huber-type loss functions, we present only those derived from the pseudo-Huber loss function. The simulation results, reported in Tables A.27 to A.40 of Appendix A.3.6, demonstrate that the proposed method, based on Huber-type and expectile loss functions, achieves superior performance across a variety of dependence structures and scenarios. For both settings ($d = 1$, one covariate, and $d = 2$, two covariates), the proposed approach consistently outperforms benchmark methods, including Support Vector Machines (SVM), Random Forests, and Logistic Regression. The improvement is particularly pronounced in terms of standard classification metrics, namely precision, sensitivity, specificity, accuracy, and the F1-score. Notably, under the Joe dependence structure with higher Kendall's tau ($\tau^c = 0.7$), the proposed approach achieves higher accuracy and a better balance between precision and sensitivity compared to other methods. Even for lower Kendall's tau values ($\tau^c = 0.5$), the proposed method maintains robust performance, particularly in imbalanced settings where high precision is crucial. Additionally, the proposed method demonstrates consistently strong performance under other dependence structures, such as Gumbel, Frank, and Gaussian copulas, as detailed in Appendix A.3.6. Across all dependence scenarios, the proposed predictors, $\hat{\lambda}_n(\mathbf{x})$, leveraging Huber-type and expectile loss functions, show their adaptability and reliability, achieving higher accuracy and a better trade-off between precision

and sensitivity than traditional methods.

To further investigate the capacity of the proposed method to accommodate data that may deviate from the underlying theoretical assumptions, we present an empirical application. The results of this analysis are reported in the following section.

4.6 Real data illustration

4.6.1 Data description

The Credit Card Fraud Detection dataset comprises 284,807 transactions made by European cardholders over two days of September in 2013, of which only 492 (i.e., approximately 0.172%) fraudulent transactions were observed. This highly imbalanced dataset poses a significant challenge for traditional classification models, as they tend to prioritize non-fraudulent transactions, thereby compromising the accuracy of fraud detection. The dataset includes a target variable, `Class`, indicating whether a transaction is fraudulent or not, along with the transaction amount and 28 anonymized variables (`V1` to `V28`) derived through Principal Component Analysis (PCA). The proposed method was initially evaluated using the `V17` variable alone and subsequently by incorporating the `V14` variable.

The predictive performance of the proposed method was evaluated in comparison with competitive approaches, including Support Vector Machine (SVM) and its asymmetric version, Logistic Regression, and Random Forest, using key metrics such as Precision, Sensitivity, Specificity, Accuracy, and F1-score. To ensure a robust and fair evaluation, the dataset was split into separate training and testing sets.

4.6.2 Results

Before presenting the results, we trained our proposed method alongside competitive methods using a training set comprising 80% of the raw data. During this training phase, we estimated the family of copulas that best fit our data. Specifically, we focused on the `Class` and `V17` variables in the first application. Subsequently, we incorporated the `V14` variable as a second covariate to extend the analysis. To estimate the copula parameters for assessing the adjusted probability estimator $\hat{\lambda}_n(\mathbf{x})$, we employed the **bicop** and **rvine** functions available in the R library **rvinecopulib** (Nagler et Vatter, 2018). The **rvinecopulib** library is particularly well-suited for modeling mixed variables, which enhances its adaptability to the characteristics of our dataset. Specifically, the **bicop** function is employed when analyzing two variables, while the **rvine** function is used for datasets comprising more than two variables. The **rvine** function performs vine

decomposition (Joe, 1996; Bedford et Cooke, 2001, 2002) and selects the optimal decomposition among R-vine, C-vine, and D-vine structures based on the Akaike Information Criterion (AIC) (Akaike, 1973). For the competitive Support Vector Machine (SVM) (Cortes et Vapnik, 1995), and its asymmetric version of SVM (asymmetric SVM), we utilized the `svm` function implemented in the **e1071** package in **R**.

Regarding the selection of the parameter τ , we relied on a cross-validation process to determine the optimal τ that maximizes precision, accuracy, and the F1-score. To this end, we established a grid values of τ , specifically $\tau = \{0.01, 0.05, 0.1, \dots, 0.9\}$. For each value of τ , we evaluated the best predictor $\hat{\lambda}_n(\mathbf{x})$ and subsequently applied classification rules to predict the `Class` value. The dependence structure between the `Class` and `V17` variables is modeled with a t-Student copula with a correlation coefficient $\sigma = -0.1151$ and degree of freedom $\nu = 2$, which was selected by the **bicop** function.

Table A.31 displays the accuracy metrics for each method evaluated on the test set. Our proposed methods, based on expectile and quantile Huber-type loss functions, as well as the SVM-Asymmetric method, outperform the other competitive methods across multiple metrics in identifying the minority class within a highly imbalanced dataset.

The proposed method based on the expectile loss function achieved a precision of 0.7143, meaning that approximately 71.43% of the transactions it identified as fraudulent were correct. The sensitivity (recall) was 0.7143, indicating it correctly identified about 71.43% of actual fraudulent transactions, with a specificity of 0.9995, the method correctly recognized non-fraudulent transactions, minimizing false positives. Additionally, it recorded a high accuracy of 0.9990 and an F1-score of 0.7255, demonstrating a balanced performance between precision and recall.

The proposed method based on Huber-type loss function showed promising results, achieving a precision of 70.41%, which means that the proposed method correctly identifying 70.41% of actual fraudulent transactions and a sensitivity of 69%. The specificity was 99.95%, and the accuracy was 99.86%. The Huber-type method achieved an F1-score of 0.6970, indicating a good balance between precision and recall. The standard SVM method achieved a precision of 79.55%, sensitivity of 35.71%, specificity of 99.98%, accuracy of 99.77%, and an F1-score of 49.64%. While its high precision highlights its ability to correctly identify positive cases, the low sensitivity reveals a significant limitation in detecting fraudulent transactions, as many were missed, while the SVM-Asymmetric method demonstrated improved performance in addressing class imbalance. It achieved a precision of 68.37% and a sensitivity of 67.00%, along with a specificity of 99.95% and an accuracy of 99.89%. The F1-score of 67.68% reflects its greater effectiveness in balancing the trade-off between precision and recall.

In contrast, Logistic Regression showed a precision of 59.18% and a sensitivity of 55.24%, with a specificity

of 99.93% and an accuracy of 99.85%. The F1-score was 57.14%, suggesting moderate performance in detecting fraudulent transactions.

The Random Forest method had a lower precision of 35.71% but a higher sensitivity of 81.40%, indicating it correctly identified approximately 81.40% of actual fraudulent transactions, although with more false positives. The specificity was 99.89%, and the accuracy was 99.62%. The F1-score was 51.18%, reflecting an imbalance between precision and recall.

The results show that a copula-based predictor trained with asymmetric loss functions (expectile and Huber-type) handles pronounced class imbalance effectively. Among all candidates, the expectile-copula model attains the best F1-score 72.55%, indicating the strongest trade-off between precision and recall (sensitivity). It is therefore well suited to applications in which rare-event detection is essential. Including the covariate V_{14} further improves performance : our estimator outperforms competing approaches (Table A.41, Appendix A.3.7). The vine decomposition and parameter estimates for the conditional copula linking V_{17} and V_{14} are reported in Table A.42 (Appendix A.3.7). Copula families were selected using the Akaike information criterion (AIC).

Table 4.1 – Accuracy metrics for each method

Method	Precision	Sensitivity	Specificity	Accuracy	F1-score
Expectile, $\tau = 0.1$	0.7143	0.7370	0.9995	0.9990	0.7255
Huber-type, $\tau = 0.2$	0.7041	0.6900	0.9995	0.9986	0.6970
SVM-Asymmetric	0.6837	0.6700	0.9994	0.9989	0.6768
Logistic Regression	0.5918	0.5524	0.9993	0.9985	0.5714
Random Forest	0.3700	0.8300	0.9989	0.9962	0.5118
SVM	0.3571	0.8140	0.9990	0.9989	0.4964

4.7 Discussion and extensions

In this paper, we present a unified inference process specifically designed for binary regression in the presence of anomalies, such as imbalanced data. This approach integrates Huber-type loss functions and expectile loss functions, both of which are particularly useful for addressing asymmetric regression problems, along with copulas to derive the optimal predictor through a straightforward equation. These predictors overcome several limitations of traditional methods, such as isotonic modeling and the single-crossing pro-

property assumed in (Manski et Thompson, 1989) for estimating the best predictor. Furthermore, under the expectile loss function, explicit formulations for the optimal predictor can be derived, enabling the development of a semi-parametric estimator that ensures both consistency and efficiency. To demonstrate the practical utility of our approach, we apply it to a credit card fraud detection dataset. This application allows us to estimate the probability of fraudulent transactions and classify each transaction, effectively addressing the challenges posed by imbalanced data. The effectiveness of the proposed method is evaluated through comparative analyses with established techniques, including Support Vector Machines (SVM) and its asymmetric version, which is designed for imbalanced data, Random Forest, and Logistic Regression. Building on these results, we are exploring an extension of this methodology to accommodate multi-categorical responses, which will be addressed in future research.

CONCLUSION

Cette thèse contribue aux travaux récents sur la régression basée sur les copules, en mettant l'accent sur les modèles de régression à réponse binaire multiple. Ainsi que sur la régression binaire univariée, dans un cadre plus général où la modélisation de la tendance moyenne ne s'avère pas adéquate pour traiter efficacement des cas où les données de réponse binaire sont fortement déséquilibrées (imbalanced). En effet, certaines catégories peuvent être sous-représentées dans l'échantillon sous étude par rapport à d'autres, ce qui requiert des approches statistiques spécifiques pour analyser un tel type de données. L'objectif principal de cette recherche est de proposer une méthodologie novatrice pour modéliser efficacement la structure de dépendance entre des variables réponse binaires ainsi que leur association avec un ensemble de covariables.

Dans le chapitre 2, nous avons développé une méthodologie théorique visant à modéliser la dépendance entre plusieurs variables binaires. À cet effet, nous avons introduit une fonction de lien basée sur les copules pour caractériser les distributions marginales conditionnelles associées aux variables réponses binaires. Ces marges nous ont permis de décrire précisément la structure de dépendance entre les différentes réponses binaires. Nous avons également démontré que, pour certaines familles spécifiques de copules, notre approche généralise des méthodes existantes, confirmant ainsi sa validité et sa cohérence avec la littérature existante. De plus, la probabilité prédictive jointe, qui intègre explicitement la structure de dépendance, possède une forme analytique claire facilitant à la fois son interprétation et sa mise en œuvre pratique. Nous avons également développé les propriétés asymptotiques de notre estimateur de probabilité prédictive, ce qui nous a permis d'évaluer sa précision d'estimation. Une étude de simulation a été menée pour tester l'efficacité de notre estimateur dans deux contextes distincts : le premier repose sur la généralisation de données à partir de notre modèle théorique, tandis que le second génère les données à partir du modèle C-logit (Genest *et al.*, 2013). Cette deuxième approche vise à évaluer les performances de l'approche proposée en cas de mauvaise spécification du modèle générateur des données. Les résultats obtenus montrent que notre estimateur offre de bonnes performances dans les deux scénarios, surpassant ou égalant les méthodes C-logit et VGLM en termes de précision et de robustesse.

Dans le chapitre 3, nous avons généralisé la méthode introduite au chapitre 2. Plus précisément, nous avons adapté la probabilité de prédiction, $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$, afin d'y intégrer des covariables mixtes, regroupant à la fois des variables continues et discrètes. Contrairement au modèle initial présenté au chapitre 2, qui ne consi-

dérait que des covariables continues, cette extension permet désormais une analyse plus complète et plus réaliste. Nous avons ensuite proposé une expression mathématique explicite de la probabilité $\mathbb{M}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$ basée sur les copules en présence de covariables mixtes. Des exemples concrets ont également été présentés afin d'illustrer cette probabilité, en utilisant des familles de copules largement utilisées, telles que les copules archimédiennes et elliptiques. Cette extension constitue ainsi une contribution significative au domaine de l'analyse de régression, en fournissant aux chercheurs un nouvel outil adapté à l'étude de données en présence de variables explicatives mixtes.

Dans le chapitre 4, nous avons introduit un nouveau modèle et une nouvelle méthode d'inférence pour déterminer le prédicteur optimal dans un contexte de régression binaire en présence de données non-balancées. Dans la pratique, il est courant de rencontrer des données où la variable cible contient une catégorie sous-représentée par rapport à la catégorie opposée. Cette anomalie peut biaiser les estimations des paramètres des méthodes traditionnelles, qui se concentrent généralement sur la modélisation de la tendance moyenne. Dans notre approche, nous avons utilisé des fonctions de perte adaptées aux régressions asymétriques, telles que celles de type Huber et l'expectile. Par ailleurs, le lien entre la réponse binaire et les covariables continues est modélisé à l'aide d'une copule, permettant ainsi de capturer des dépendances complexes. Notre approche repose sur la résolution d'une équation d'estimation pour déterminer le prédicteur optimal, basée sur la fonction de perte asymétrique Huber-type ou expectile.

Cette solution se traduit simplement par un réajustement de la probabilité de succès, comme nous l'avons démontré dans le cas de l'expectile. En outre, nous avons montré que le paramètre d'asymétrie dans la fonction de perte joue un rôle clé en équilibrant l'influence des observations associées à $Y = 0$ et $Y = 1$. Cette approche constitue ainsi une alternative adaptée aux situations où les données présentent des déséquilibres significatifs ou des structures de dépendance complexes.

Par la suite, nous avons proposé un estimateur pour notre prédicteur et étudié ses propriétés asymptotiques. Afin de tester l'efficacité de notre approche, nous avons mené une étude de simulation ainsi qu'une illustration sur des données réelles, où nous avons comparé les performances prédictives de notre méthode avec celles de méthodes concurrentes, telles que SVM, SVM asymétrique, la régression logistique et les forêts aléatoires. Les résultats montrent que notre méthode offre de bonnes performances en termes de prédiction, évaluées à l'aide des métriques telles que la précision, le score F1, la sensibilité et la spécificité.

Les travaux réalisés dans le cadre de cette thèse ouvrent plusieurs perspectives prometteuses dans le domaine de la modélisation statistique des dépendances. En particulier, les développements présentés dans les chapitres 2, 3 et 4 mettent en évidence une voie particulièrement intéressante, consistant à étendre le cadre classique de la régression à des réponses discrètes comportant plusieurs modalités, qu'elles soient de nature univariée ou multivariée, tout en intégrant simultanément des covariables de types mixtes. Par ailleurs, l'approche proposée au chapitre 3 ouvre la voie à une extension novatrice de l'étude des structures de dépendance dans un vecteur de réponses mixtes, associé à un vecteur de covariables mixtes.

Bien que la méthodologie introduite dans cette thèse propose un cadre unifié et flexible, elle présente néanmoins deux limitations complémentaires. D'abord, lorsque l'estimation de la fonction de régression ($\pi_j(\mathbf{x})$, $j = 1, \dots, d$ définie dans l'équation (2.18)) repose sur une copule paramétrique, une mauvaise spécification de la famille de copules peut induire un biais dans l'estimation de la probabilité, c'est-à-dire $\hat{\pi}_j(\mathbf{x})$, $j = 1, \dots, d$. Dans un tel cas, l'estimateur ne converge pas vers la véritable probabilité $\pi(\mathbf{x})$, mais vers une probabilité limite $\pi^*(\mathbf{x})$ déterminée par la famille de copules choisie. Cette fonction limite représente la meilleure approximation possible au sens de la vraisemblance dans la famille spécifiée, mais elle peut s'écarter substantiellement de la fonction cible. Cette divergence s'explique notamment par le fait que l'estimation impose généralement une relation conditionnelle monotone entre la réponse et les covariables. Ainsi, des formes non monotones de la fonction de régression ne peuvent pas être correctement reproduites. Ce phénomène persiste même après sélection du modèle à l'aide d'un critère tel que l'AIC, et ce, y compris dans le cas où l'estimateur est construit via les vignes à partir des familles usuelles (Gaussienne, t-Student, Clayton, Gumbel, Frank, Joe) Tel que démontré par Dette *et al.* (2014).

Ensuite, le cadre théorique développé dans cette thèse repose sur l'hypothèse de simplification, selon laquelle la structure de dépendance entre les variables réponses binaires demeure constante quelle que soit la valeur prise par les covariables. Cependant, Mesfioui et Quessy (2008) ont démontré une propriété importante concernant la classe des copules archimédiennes : celle-ci jouit d'une stabilité remarquable vis-à-vis du conditionnement. Plus précisément, lorsqu'un vecteur aléatoire constitué de réponses et de covariables est décrit par une copule archimédienne, la copule conditionnelle des réponses, obtenue en fixant un ensemble spécifique de covariables, demeure elle-même archimédienne. Toutefois, cette copule conditionnelle est caractérisée par un générateur mis à jour à partir de la distribution conditionnelle des réponses. Autrement dit, lorsque la dépendance conjointe entre les réponses et les covariables relève de la classe archimédienne, cette propriété est préservée après conditionnement. Il devient ainsi possible de modéliser directement les probabilités conditionnelles des réponses sans recourir à l'hypothèse, souvent jugée restrictive, de simplification. Cette observation permet de contourner la limite de l'hypothèse de simplification

dans le cas des familles de copules Archimédiennes dont le générateur de copule dépend explicitement des covariables, tout en maintenant la tractabilité analytique propre aux copules archimédiennes. Néanmoins, dans le cas lorsque la dépendance est décrite par une copule de Clayton, le paramètre de la copule conditionnelle reste indépendant de la partie du conditionnement (covariables), ce qui limite la variabilité de la dépendance en fonction des covariables.

Une seconde voie de recherche, s'inscrivant dans la continuité du Chapitre 4, vise à développer une procédure d'automatisation de la sélection du paramètre d'asymétrie τ associé aux fonctions de perte de type expectile et Huber, dans le contexte de la régression binaire fondée sur les copules, en tenant compte de la présence de données présentant un déséquilibre marqué entre les classes. Cette avenue se base sur une approche à deux étapes interdépendantes. Premièrement, pour une valeur fixée de τ , le prédicteur optimal $\lambda_\tau(\cdot)$ est estimé selon la procédure décrite au Chapitre 4. Deuxièmement, ce prédicteur est intégré dans l'expression d'un risque quadratique pondéré, spécifiquement conçu pour rééquilibrer les erreurs de classification en fonction de la rareté de la classe minoritaire, défini par

$$\mathcal{R}(\tau) = \mathbb{E}_{Y, \mathbf{X}} \left\{ \omega(\mathbf{X}) (Y - \lambda_\tau(\mathbf{X}))^2 \right\},$$

où $\omega(\mathbf{X}) = \frac{1 - \pi(\mathbf{X})}{\pi(\mathbf{X})}$. Les fonctions $\pi(\cdot)$ et $\lambda_\tau(\cdot)$ sont, respectivement, définies dans la Proposition 4.4. Le paramètre τ est ensuite déterminé en minimisant l'estimateur empirique de ce risque. La procédure, entièrement non supervisée, ne requiert aucune étape de validation croisée, ce qui en accroît l'efficacité computationnelle.

ANNEXE A
APPENDICES

A.1 Chapter 2 Appendices

A.1.1 Proof for Lemma 2.2

Proof. The Proof of the lemma is established by induction. For $d = 1$, the following holds :

$$\mathbb{M}_1^1(\mathbf{x}) = 1 - \mathbb{M}_0^1(\mathbf{x}).$$

For $d = 2$, the following holds :

$$\mathbb{M}_{(1,0)}^{1,2}(\mathbf{x}) = \mathbb{M}_0^1(\mathbf{x}) - \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}),$$

and

$$\mathbb{M}_{(0,1)}^{1,2}(\mathbf{x}) = \mathbb{M}_0^2(\mathbf{x}) - \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}),$$

with

$$\mathbb{M}_{(1,1)}^{1,2}(\mathbf{x}) = 1 - \mathbb{M}_0^1(\mathbf{x}) - \mathbb{M}_0^2(\mathbf{x}) + \mathbb{M}_{(0,0)}^{1,2}(\mathbf{x}).$$

Assume the statement holds for d outcomes, and it will be shown that it holds for $d + 1$ outcomes. Consider a vector of $d + 1$ binary outcomes $(0, \dots, 1, \dots, 1, \dots, 0)$ with r binary outcomes equal to 1 and $d + 1 - r$ equal to 0. Assume that Y_i is among the r indices where $y_i = 1$. The following holds :

$$\mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=1, y_{i+1}=1, \dots, y_{d+1}=0}(\mathbf{x}) = \mathbb{P}(y_1 = 0, \dots, y_{i-1} = 0, y_i = 1, y_{i+1} = 1, \dots, y_{d+1} = 0 | \mathbf{x}),$$

which can be rewritten as :

$$\begin{aligned} \mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=1, y_{i+1}=1, \dots, y_{d+1}=0}^{1, \dots, d}(\mathbf{x}) &= \mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_{i+1}=1, \dots, y_{d+1}=0}^{(1, \dots, d+1)-i}(\mathbf{x}) \\ &\quad - \mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_{d+1}=0}^{(i)}(\mathbf{x}). \end{aligned} \quad (\text{A.1})$$

Here, $\mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_{i+1}=1, \dots, y_{d+1}=0}^{(1, \dots, d+1)-i}(\mathbf{x})$ represents the probability of d variables equal to 0, excluding the i^{th} index of the outcomes vector \mathbf{y} , and $\mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_{d+1}=0}^{(i)}(\mathbf{x})$ is the same initial probability without excluding the i^{th} , but with $y_i = 0$. The induction assumption can then be applied to $\mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_{d+1}=0}^{(i)}(\mathbf{x})$.

With exactly $r - 1$ binary results equal to 1 and $d - r + 1$ equal to 0, choose an index j between $(1, \dots, r - 1)$.

The following holds :

$$\begin{aligned} \mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_j=1, \dots, y_{d+1}=0}^{(i)}(\mathbf{x}) &= \mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_{d+1}=0}^{(1, \dots, d+1)-j}(\mathbf{x}) \\ &\quad - \mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_j=0, \dots, y_{d+1}=0}^{(j)}(\mathbf{x}). \end{aligned}$$

The same argument applies for $\mathbb{M}_{y_1=0, \dots, y_{i-1}=0, y_i=0, y_{i+1}=1, \dots, y_j=0, \dots, y_{d+1}=0}^{(j)}(\mathbf{x})$.

□

A.1.2 Proof of the Theorem 2.7

Proof. First, sitting $\boldsymbol{\pi}(\mathbf{x}) = (\pi_1(\mathbf{x}), \dots, \pi_d(\mathbf{x}))$, $\widehat{\boldsymbol{\pi}}(\mathbf{x}) = (\widehat{\pi}_1(\mathbf{x}), \dots, \widehat{\pi}_d(\mathbf{x}))$ and $\widetilde{\boldsymbol{\pi}}(\mathbf{x}) = (\widetilde{\pi}_1(\mathbf{x}), \dots, \widetilde{\pi}_d(\mathbf{x}))$.

Next, applying the mean value Theorem around $\pi_j(\mathbf{x}) = \mathcal{C}_{\theta_j}^c(1 - p_j \mid \mathbf{F}(\mathbf{x}))$,

$j = 1, \dots, d$ and γ to $\widehat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \mathcal{C}_{\widetilde{\gamma}}(\widehat{\boldsymbol{\pi}}(\mathbf{x}))$ with $\mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x}))$, one has

$$\widehat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \sum_{j=1}^d \partial_j \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x})) (\widehat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})) + (\widehat{\gamma}_n - \gamma)^\top \Delta_{\gamma} \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x})), \quad (\text{A.2})$$

where $\partial_j \mathcal{C}_{\gamma} = \partial \mathcal{C}_{\gamma} / \partial \pi_j$, $\Delta_{\gamma} \mathcal{C}_{\gamma}(\cdot) = \left(\frac{\partial \mathcal{C}_{\gamma}(\cdot)}{\partial \gamma_1}, \dots, \frac{\partial \mathcal{C}_{\gamma}(\cdot)}{\partial \gamma_n} \right)^\top$, $j = 1, \dots, d$, with

$$\widetilde{\pi}_j(\mathbf{x}) = \pi_j(\mathbf{x}) + \tau_j (\widehat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})),$$

and

$$\widetilde{\gamma} = \gamma + \tau_{(d+1)} (\widehat{\gamma}_n - \gamma),$$

for a constant quantities $\tau_j \in [0, 1]$, $j = 1, \dots, d + 1$. Then, one can write :

$$\begin{aligned} \widehat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) &= \sum_{j=1}^d \partial_j \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) (\widehat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})) + (\widehat{\gamma}_n - \gamma)^\top \Delta_{\gamma} \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) \\ &\quad - \sum_{j=1}^d (\partial_j \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) - \partial_j \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x}))) (\widehat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})) \\ &\quad + (\widehat{\gamma}_n - \gamma)^\top (\Delta_{\gamma} \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x})) - \Delta_{\gamma} \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x}))) \\ &= \sum_{j=1}^d \partial_j \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) (\widehat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})) + (\widehat{\gamma}_n - \gamma)^\top \Delta_{\gamma} \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) \\ &\quad + \sum_{j=1}^{d+1} \mathbb{O}_n^j, \end{aligned} \quad (\text{A.3})$$

with

$$\mathbb{O}_n^j = (\partial_j \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) - \partial_j \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x}))) (\widehat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})), \quad j = 1, \dots, d,$$

and

$$\mathbb{O}_n^{d+1} = (\widehat{\gamma}_n - \gamma)^\top (\Delta_{\gamma} \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x})) - \Delta_{\gamma} \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x}))).$$

Under assumption **A2**, one can write that

$$\partial_j \mathcal{C}_{\gamma}(\boldsymbol{\pi}(\mathbf{x})) - \partial_j \mathcal{C}_{\widetilde{\gamma}}(\widetilde{\boldsymbol{\pi}}(\mathbf{x})) = o_p(1), \quad \forall j = 1, \dots, d,$$

$$\Delta_\gamma \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x})) - \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) = o_p(\mathbf{1}),$$

and based on assumption **A1** and **A4** imply the root- n rates $\hat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x}) = O_p(n^{-1/2})$ for $j = 1, \dots, d$ and $\hat{\gamma}_n - \gamma = O_p(n^{-1/2})$. Hence $\hat{\boldsymbol{\pi}}(\mathbf{x}) \rightarrow_p \boldsymbol{\pi}(\mathbf{x})$ and $\hat{\gamma}_n \rightarrow_p \gamma$, and by the continuity in **A2**, $\partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) - \partial_j \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x})) = o_p(1)$ and $\Delta_\gamma \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x})) - \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) = o_p(1)$. Therefore,

$$O_n^j = (\partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) - \partial_j \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x}))) (\hat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x})) = o_p(1) O_p(n^{-1/2}) = o_p(n^{-1/2}),$$

and similarly

$$O_n^{d+1} = (\hat{\gamma}_n - \gamma)^\top (\Delta_\gamma \mathcal{C}_{\tilde{\gamma}}(\tilde{\boldsymbol{\pi}}(\mathbf{x})) - \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x}))) = O_p(n^{-1/2}) o_p(1) = o_p(n^{-1/2}).$$

Moreover, from assumption **A1** (i.e Theorem 1 of (Mesfioui et al., 2023)), one has

$$\begin{aligned} \hat{\pi}_j(\mathbf{x}) - \pi_j(\mathbf{x}) &= \sum_{i=1}^n (p_j - \mathbb{1}(Y_{ij} = 1)) \partial_u \mathcal{C}_{\boldsymbol{\theta}_j}^c(1 - p_j | \mathbf{F}(\mathbf{x})) + \boldsymbol{\zeta}_{ij}^\top \Delta_{\boldsymbol{\theta}_j} \mathcal{C}_{\boldsymbol{\theta}_j}^c(1 - p_j | \mathbf{F}(\mathbf{x})) \\ &\quad + (\mathbb{1}(\mathbf{X}_i \leq \mathbf{x}) - \mathbb{P}(\mathbf{X} \leq \mathbf{x}))^\top \Delta_{\mathbf{v}} \mathcal{C}_{\boldsymbol{\theta}_j}^c(1 - p_j | \hat{\mathbf{F}}_n(\mathbf{x})) \\ &= \frac{1}{n} \sum_{i=1}^n \mathcal{Z}_i(\mathbf{x}) + o_p(n^{-\frac{1}{2}}), \end{aligned} \tag{A.4}$$

and under assumption **A4** one can write that

$$(\hat{\gamma}_n - \gamma)^\top \Delta_\gamma \mathcal{C}_\gamma(\pi_1(\mathbf{x}), \dots, \pi_d(\mathbf{x})) = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma)^\top \Delta_\gamma \mathcal{C}_\gamma(\pi_1(\mathbf{x}), \dots, \pi_d(\mathbf{x})) + o_p(n^{-\frac{1}{2}}). \tag{A.5}$$

From the representation in (A.4) and (A.5), one can write equation (A.3) as follows :

$$\begin{aligned} \hat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) &= \sum_{j=1}^d \partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) \left(\frac{1}{n} \sum_{i=1}^n \mathcal{Z}_{ij}(\mathbf{x}) \right) \\ &\quad + \left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma)^\top \right) \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) + o_p(n^{-1/2}). \end{aligned}$$

Interchanging the sums leads to

$$\begin{aligned} \hat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^d \partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) \mathcal{Z}_{ij}(\mathbf{x}) + \boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma)^\top \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) \right) \\ &\quad + o_p(n^{-1/2}). \end{aligned} \tag{A.6}$$

Now define :

$$\mathcal{Z}_i(\mathbf{x}) = \sum_{j=1}^d \partial_j \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})) \mathcal{Z}_{ij}(\mathbf{x}) + \boldsymbol{\nu}_i(\boldsymbol{\pi}(\mathbf{x}); \gamma)^\top \Delta_\gamma \mathcal{C}_\gamma(\boldsymbol{\pi}(\mathbf{x})).$$

Therefore, one has

$$\hat{\mathbb{M}}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) - \mathbb{M}_{(0, \dots, 0)}^{1, \dots, d}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathcal{Z}_i(\mathbf{x}) + o_p(n^{-1/2}).$$

This completes the proof. \square

A.1.3 Evaluation of the variance of $\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d}(\mathbf{x})$

Algorithm A.1

Steps to estimate the variance of $\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d}(\cdot)$:

1. Generate $X_{i,k}^{(b)}$ from a $\widehat{F}_{n,k}(\cdot)$
2. Generate $(U_{i1}^{(b)}, \dots, U_{id}^{(b)})$ from copula $\mathcal{C}_{\widehat{\gamma}}(\cdot)$
3. Compute $Y_{ij}^b = \mathbb{1}(U_{i,j}^b \geq (1 - p_j))$
4. Based on $(Y_{i1}^{(b)}, \dots, Y_{id}^{(b)}, X_{i1}^{(b)}, \dots, X_{ip}^{(b)})$ compute $\widehat{p}_{nj}^{(b)}$ and $\widehat{\mathbf{F}}_n^{(b)}$ the empirical margins of the covariates. Then compute $\mathcal{C}_{\widehat{\theta}_j}^c(1 - p_j \mid \widehat{\mathbf{F}}_n(\mathbf{x}_i^{(b)}))$ for $j = 1, \dots, d, k = 1, \dots, p$ where $\widehat{\mathbf{F}}_n(\mathbf{x}_i^{(b)}) = (\widehat{F}_{n1}(x_{ik}^{(b)}), \dots, \widehat{F}_{np}(x_{ik}^{(b)}))$
5. Compute our estimator

$$\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,1(b)}(\mathbf{x}) = \mathcal{C}_{\widehat{\gamma}}\left(\mathcal{C}_{\widehat{\theta}_1}^c(1 - p_{n1}^{(b)} \mid \widehat{\mathbf{F}}_n^{(b)}(\mathbf{x})), \dots, \mathcal{C}_{\widehat{\theta}_d}^c(1 - p_{nd}^{(b)} \mid \widehat{\mathbf{F}}_n^{(b)}(\mathbf{x}))\right),$$

note the bootstrap estimator by $\widehat{\pi}_{boot}(\mathbf{x}) = B^{-1} \sum_{b=1}^B \widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d(b)}(\mathbf{x})$

6. Estimate the variance of the estimator by

$$\widehat{\sigma}_{boot}^2 = B^{-1} \sum_{b=1}^B \left(\widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d(b)}(\mathbf{x}) - B^{-1} \sum_{b=1}^B \widehat{\mathbb{M}}_{(0,\dots,0)}^{1,\dots,d(b)}(\mathbf{x}) \right)^2$$

The following figure displays the plots of the Confidence Interval (CI) for the proposed estimator based on the plug-in technique described in section 2.4, in the presence of one covariate, where the CI is expressed as follows :

$$ICs(\mathbf{x}) = [\widehat{\pi}_{boot}(\mathbf{x}) - Z_{\alpha/2} * \widehat{\sigma}_{boot}; \widehat{\pi}_{boot}(\mathbf{x}) + Z_{\alpha/2} * \widehat{\sigma}_{boot}]$$

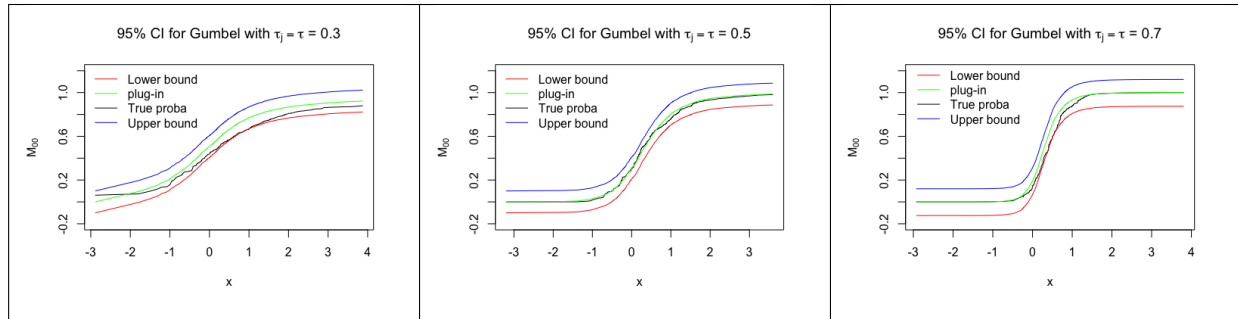


Figure A.1 – Setting 1 : Confidence interval of the estimator $\hat{M}_{0,0}^{1,2}(x)$ with $X \sim \mathcal{N}(0, 1)$ and sample size $n = 300$.

A.1.4 Parameters copula

Table A.1 – Setup of the data generation utilized in simulations.

Copulas	Parameters or Kendall's tau τ
Gaussian	ρ_γ corresponds $\tau = 0.3, 0.5, 0.7$ same for ρ_θ
Clayton	γ and θ_j corresponds $\tau = \tau_j = 0.3, 0.5, 0.7$
Gumbel	γ and θ_j corresponds $\tau = 0.3, 0.5, 0.7$
Joe	γ and θ_j corresponds $\tau = 0.3, 0.5, 0.7$

A.1.5 Plots of $\hat{M}_{y_1, y_2}^{1,2}$ versus the covariables, based on Western Collaborative Group Study data

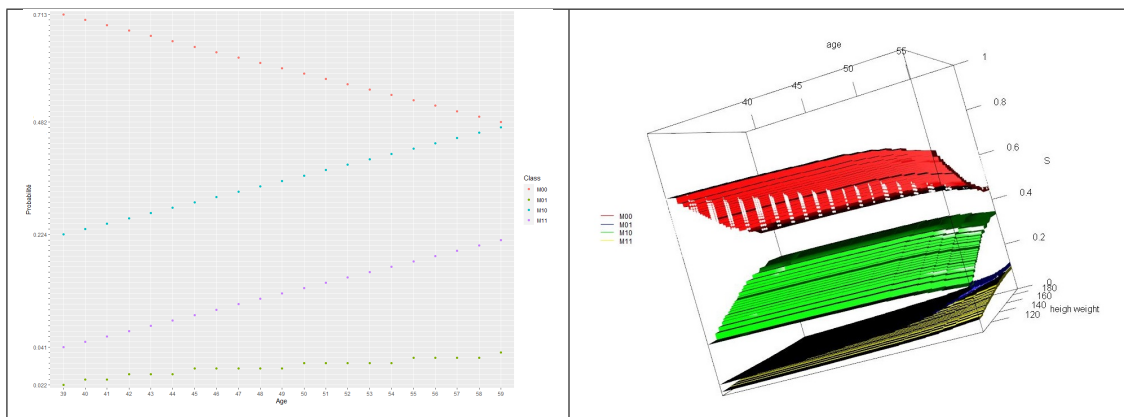


Figure A.2 – The left panel shows the probability $\hat{M}_{y_1, y_2}^{1,2}$ versus age. The right panel illustrates $\hat{M}_{y_1, y_2}^{1,2}$ versus age and BMI index. For $\{y_1, y_2\} \in \{0, 1\}^2$.

A.1.6 IMSE result of additional setting

Due to the similarity in the performance of the $\widehat{\text{IMSE}}$ across different scenarios, we report only selected examples of the $\widehat{\text{IMSE}}$ statistics from Settings 1 and 2 in this section.

A.1.6.1 Unbalanced data

Table A.2 – Setting 1 (Unbalanced data) : Results of the IMSE statistic ($\times 100$), based on K=1000 replications, for $d = 2, p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.886	0.420	0.113	0.301	0.159	0.133	0.900	0.602	0.247
<i>vine-based</i>	2.982	2.241	2.136	1.946	2.150	1.829	3.022	3.049	2.835
VGLM	4.123	3.757	3.937	9.710	9.832	9.937	13.789	13.696	13.042
<i>C-logit</i>	3.556	3.457	3.284	3.732	3.988	3.465	3.504	3.119	3.039
Joe									
plug-in	1.482	1.402	1.127	1.721	1.121	1.078	1.619	1.310	0.845
<i>vine-based</i>	2.331	3.882	2.690	4.537	4.285	4.231	5.948	5.318	5.174
VGLM	4.800	4.258	4.845	10.301	10.533	10.207	19.622	19.318	20.273
<i>C-logit</i>	5.662	5.207	5.098	6.464	6.631	6.433	8.815	8.835	8.676
Gumbel									
plug-in	1.082	0.833	0.158	1.236	1.049	0.716	1.661	1.202	0.986
<i>vine-based</i>	2.743	2.590	2.458	4.001	3.879	3.742	4.838	4.712	4.474
VGLM	5.256	5.291	5.074	10.413	10.353	10.053	13.440	13.651	13.621
<i>C-logit</i>	3.355	3.291	3.129	6.900	6.432	6.362	7.671	7.528	7.712
Gaussian									
plug-in	1.285	0.903	0.834	1.642	1.236	1.845	1.713	1.060	1.084
<i>vine-based</i>	2.968	2.715	2.811	2.548	2.953	2.923	2.138	2.404	2.517
VGLM	6.757	5.395	5.265	9.071	10.497	10.404	18.478	18.470	18.420
<i>C-logit</i>	3.597	3.790	3.643	3.948	3.466	3.652	4.664	4.114	3.755

Table A.3 – Setting 1 (Unbalanced data) : Results of the IMSE statistic ($\times 100$), based on K=1000 replications, for $d = 2, p = 1$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	1.188	0.553	0.152	0.405	0.210	0.178	1.203	0.792	0.326
<i>vine-based</i>	3.915	2.928	2.799	2.552	2.812	2.392	3.962	3.986	3.710
VGLM	5.387	4.902	5.143	12.723	12.882	13.017	18.062	17.942	17.085
C-logit	4.658	4.528	4.302	4.888	5.224	4.539	4.590	4.086	3.981
Joe									
plug-in	1.945	1.832	1.476	2.253	1.468	1.412	2.120	1.716	1.107
<i>vine-based</i>	3.055	5.084	3.524	5.943	5.613	5.542	7.792	6.966	6.778
VGLM	6.288	5.578	6.347	13.494	13.798	13.371	25.705	25.307	26.558
C-logit	7.417	6.821	6.678	8.468	8.687	8.427	11.548	11.574	11.366
Gumbel									
plug-in	1.418	1.091	0.207	1.619	1.374	0.938	2.176	1.575	1.292
<i>vine-based</i>	3.592	3.393	3.220	5.241	5.081	4.902	6.338	6.173	5.861
VGLM	6.885	6.931	6.647	13.641	13.562	13.169	17.606	17.883	17.844
C-logit	4.395	4.312	4.099	9.039	8.426	8.334	10.049	9.862	10.103
Gaussian									
plug-in	1.684	1.183	1.092	2.151	1.619	2.417	2.244	1.389	1.420
<i>vine-based</i>	3.888	3.556	3.682	3.338	3.869	3.829	2.801	3.149	3.297
VGLM	8.852	7.068	6.897	11.883	13.751	13.629	24.206	24.196	24.130
C-logit	4.713	4.965	4.772	5.172	4.540	4.784	6.110	5.389	4.919

Table A.4 – Setting 1 (Unbalanced data) : Results of the IMSE ($\times 100$), based on K=1000 replications, for $d = 3, p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	1.106	0.505	0.137	0.392	0.199	0.165	1.152	0.766	0.305
<i>vine-based</i>	3.795	2.957	2.772	2.457	2.664	2.272	3.853	3.877	3.658
VGLM	5.480	4.939	5.169	12.398	12.604	12.769	17.635	17.504	16.701
C-logit	4.743	4.588	4.378	4.948	5.298	4.635	4.661	4.166	4.040
Joe									
plug-in	1.868	1.758	1.428	2.178	1.415	1.358	2.055	1.663	1.058
<i>vine-based</i>	2.972	4.941	3.402	5.779	5.453	5.334	7.574	6.786	6.581
VGLM	6.222	5.538	6.279	13.667	13.969	13.531	26.088	25.495	26.788
C-logit	7.371	6.749	6.566	8.495	8.700	8.397	11.707	11.754	11.473
Gumbel									
plug-in	1.348	1.050	0.199	1.534	1.280	0.854	2.089	1.508	1.212
<i>vine-based</i>	3.473	3.283	3.097	5.110	4.955	4.780	6.149	6.000	5.698
VGLM	6.763	6.836	6.541	13.544	13.443	13.023	17.917	18.217	18.173
C-logit	4.320	4.218	4.026	8.980	8.378	8.276	9.985	9.779	9.982
Gaussian									
plug-in	1.632	1.134	1.054	2.075	1.558	2.325	2.156	1.310	1.348
<i>vine-based</i>	3.783	3.461	3.594	3.245	3.776	3.748	2.723	3.062	3.207
VGLM	8.662	6.933	6.765	11.978	13.867	13.740	24.465	24.453	24.387
C-logit	4.630	4.871	4.683	5.096	4.472	4.707	6.075	5.359	4.892

Table A.5 – Setting 1 (Unbalanced data) : Results of the IMSE statistic ($\times 100$), based on K=1000 replications, for $d = 3$, $p = 3$, with $p_1 = 0.05$, $p_2 = 0.3$, and $p_3 = 0.4$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.790	0.361	0.098	0.280	0.142	0.118	0.823	0.547	0.218
<i>vine-based</i>	2.711	2.112	1.980	1.755	1.903	1.623	2.752	2.769	2.613
VGLM	3.914	3.528	3.692	8.856	9.003	9.121	12.596	12.503	11.929
C-logit	3.388	3.277	3.127	3.534	3.784	3.311	3.329	2.976	2.886
Joe									
plug-in	1.334	1.256	1.020	1.556	1.011	0.970	1.468	1.188	0.756
<i>vine-based</i>	2.123	3.529	2.430	4.128	3.895	3.810	5.410	4.847	4.701
VGLM	4.444	3.956	4.485	9.762	9.978	9.665	18.634	18.211	19.134
C-logit	5.265	4.821	4.690	6.068	6.214	5.998	8.362	8.396	8.195
Gumbel									
plug-in	0.963	0.750	0.142	1.096	0.914	0.610	1.492	1.077	0.866
<i>vine-based</i>	2.481	2.345	2.212	3.650	3.539	3.414	4.392	4.286	4.070
VGLM	4.831	4.883	4.672	9.674	9.602	9.302	12.798	13.012	12.981
C-logit	3.086	3.013	2.876	6.414	5.984	5.912	7.132	6.985	7.130
Gaussian									
plug-in	1.166	0.810	0.753	1.482	1.113	1.661	1.540	0.936	0.963
<i>vine-based</i>	2.702	2.472	2.567	2.318	2.697	2.677	1.945	2.187	2.291
VGLM	6.187	4.952	4.832	8.556	9.905	9.814	17.475	17.461	17.419
C-logit	3.307	3.479	3.345	3.640	3.194	3.362	4.339	3.828	3.494

A.1.6.2 Case $d = 3, p = 2, 3$

Table A.6 – Setting 1 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications, for $d = 3, p = 3$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.930	0.441	0.119	0.316	0.167	0.140	0.945	0.632	0.259
vine-based	3.131	2.353	2.243	2.043	2.258	1.920	3.173	3.201	2.977
VGLM	4.329	3.945	4.134	10.196	10.324	10.434	14.478	14.381	13.694
C-logit	3.734	3.630	3.448	3.919	4.187	3.638	3.679	3.275	3.191
Joe									
plug-in	1.556	1.472	1.183	1.807	1.177	1.132	1.700	1.376	0.887
vine-based	2.448	4.076	2.825	4.764	4.499	4.443	6.245	5.584	5.433
VGLM	5.040	4.471	5.087	10.816	11.060	10.717	20.603	20.284	21.287
C-logit	5.945	5.467	5.353	6.787	6.963	6.755	9.256	9.277	9.110
Gumbel									
plug-in	1.136	0.875	0.166	1.298	1.101	0.752	1.744	1.262	1.035
vine-based	2.880	2.720	2.581	4.201	4.073	3.929	5.080	4.948	4.698
VGLM	5.519	5.556	5.328	10.934	10.871	10.556	14.112	14.334	14.302
C-logit	3.523	3.456	3.285	7.245	6.754	6.680	8.055	7.904	8.098
Gaussian									
plug-in	1.349	0.948	0.876	1.724	1.298	1.937	1.799	1.113	1.138
vine-based	3.116	2.851	2.952	2.675	3.101	3.069	2.245	2.524	2.643
VGLM	7.095	5.665	5.528	9.525	11.022	10.924	19.402	19.394	19.341
C-logit	3.777	3.980	3.825	4.145	3.639	3.835	4.897	4.320	3.943

Table A.7 – Setting 2 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications where $d = 3, p = 3$.

	$\tau = 0.3$			$\tau = 0.5$			$\tau = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
vine-based	2.307	2.148	1.998	1.992	1.950	1.926	4.634	4.889	4.957
VGLM	3.949	4.201	4.089	9.915	9.469	9.438	7.478	12.100	8.394
C-logit	2.220	2.216	2.102	6.447	5.872	6.044	6.754	10.284	9.465
Joe									
vine-based	2.932	2.905	3.903	3.297	2.920	3.284	2.899	3.096	2.624
VGLM	2.381	3.288	2.108	3.374	3.381	5.181	6.680	6.725	6.775
C-logit	4.513	5.063	4.519	3.854	3.481	3.290	5.211	5.108	4.764
Gumbel									
vine-based	2.590	2.602	2.564	3.704	4.410	3.582	6.601	6.728	6.029
VGLM	2.243	2.190	2.226	5.371	5.606	5.503	11.437	12.149	11.491
C-logit	3.415	1.239	1.243	2.599	2.477	2.473	6.217	6.302	6.159
Gaussian									
vine-based	1.224	1.239	1.069	1.528	1.366	1.351	3.053	2.752	2.116
VGLM	1.569	1.797	1.787	2.330	2.166	2.188	3.828	3.592	3.311
C-logit	2.851	2.802	2.805	2.568	2.518	2.493	3.361	3.722	4.506

Table A.8 – Setting 1 : Results of the IMSE statistic ($\times 100$), based on $K = 1000$ replications where $d = 3, p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	1.116	0.529	0.143	0.379	0.200	0.168	1.134	0.758	0.311
vine-based	3.757	2.824	2.692	2.452	2.710	2.304	3.808	3.841	3.572
VGLM	4.329	3.945	4.134	10.196	10.324	10.434	14.478	14.381	13.694
C-logit	3.734	3.630	3.448	3.919	4.187	3.638	3.679	3.275	3.191
Joe									
plug-in	1.867	1.766	1.420	2.168	1.412	1.358	2.040	1.651	1.064
vine-based	2.938	4.891	3.390	5.717	5.399	5.332	7.494	6.701	6.520
VGLM	5.040	4.471	5.087	10.816	11.060	10.717	20.603	20.284	21.287
C-logit	5.945	5.467	5.353	6.787	6.963	6.755	9.256	9.277	9.110
Gumbel									
plug-in	1.363	1.050	0.199	1.558	1.321	0.902	2.093	1.514	1.242
vine-based	3.456	3.264	3.097	5.041	4.888	4.715	6.096	5.938	5.638
VGLM	5.519	5.556	5.328	10.934	10.871	10.556	14.112	14.334	14.302
C-logit	3.523	3.456	3.285	7.245	6.754	6.680	8.055	7.904	8.098
Gaussian									
plug-in	1.619	1.138	1.051	2.069	1.558	2.324	2.159	1.336	1.366
vine-based	3.739	3.421	3.542	3.210	3.721	3.683	2.694	3.029	3.172
VGLM	7.095	5.665	5.528	9.525	11.022	10.924	19.402	19.394	19.341
C-logit	3.777	3.980	3.825	4.145	3.639	3.835	4.897	4.320	3.943

Table A.9 – Setting 2 : Results of the IMSE statistic ($\times 100$), based on $K=1000$ replications where $d = 3, p = 2$.

	$\tau = 0.3$			$\tau = 0.5$			$\tau = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
vine-based	2.415	2.052	1.875	2.143	1.872	1.845	4.825	5.142	5.218
VGLM	4.128	4.315	4.203	10.127	9.582	9.551	7.692	12.415	8.607
C-logit	2.335	2.328	2.215	6.558	5.985	6.157	6.867	10.397	9.578
Joe									
vine-based	3.128	3.102	4.105	3.402	3.028	3.392	3.005	3.204	2.731
VGLM	2.488	3.395	2.215	3.481	3.488	5.288	6.787	6.832	6.882
C-logit	4.620	5.170	4.626	3.961	3.588	3.397	5.318	5.215	4.871
Gumbel									
vine-based	2.695	2.708	2.671	3.811	4.517	3.689	6.708	6.835	6.136
VGLM	2.350	2.297	2.333	5.478	5.713	5.610	11.544	12.256	11.598
C-logit	3.522	1.346	1.350	2.706	2.584	2.580	6.324	6.409	6.266
Gaussian									
vine-based	1.331	1.346	1.176	1.635	1.473	1.458	3.160	2.859	2.223
VGLM	1.676	1.904	1.894	2.437	2.273	2.295	3.935	3.699	3.418
C-logit	2.958	2.909	2.912	2.675	2.625	2.600	3.468	3.829	4.613

A.1.6.3 Scenario of mixed Kendall's tau

Table A.10 – Setting 1 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications case $d = 3, p = 3$.

	$\tau = 0.3, \tau_j = 0.5$			$\tau = 0.5, \tau_j = 0.3$			$\tau = 0.5, \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.75	0.35	0.09	0.25	0.12	0.10	0.80	0.50	0.20
<i>vine-based</i>	2.10	1.85	1.70	1.60	1.75	1.55	2.50	2.60	2.30
VGLM	5.20	5.00	4.80	11.0	11.5	11.2	15.0	15.5	14.8
<i>C-logit</i>	4.50	4.30	4.00	5.00	5.20	4.80	4.80	4.50	4.20
Joe									
plug-in	1.30	1.20	1.00	1.50	1.00	0.90	1.40	1.10	0.70
<i>vine-based</i>	2.00	3.20	2.40	3.80	3.50	3.40	4.50	4.20	4.00
VGLM	6.00	5.50	6.20	12.5	12.8	12.0	22.0	21.5	23.0
<i>C-logit</i>	6.50	6.00	5.80	7.20	7.50	7.00	9.50	9.60	9.30
Gumbel									
plug-in	0.95	0.70	0.12	1.10	0.90	0.60	1.40	1.00	0.80
<i>vine-based</i>	2.30	2.10	2.00	3.50	3.30	3.10	4.00	3.90	3.70
VGLM	6.50	6.60	6.30	12.0	11.8	11.5	16.0	16.2	15.8
<i>C-logit</i>	4.00	3.80	3.60	7.50	7.00	6.90	8.20	8.00	8.30
Gaussian									
plug-in	1.10	0.75	0.70	1.40	1.10	1.60	1.50	0.90	0.95
<i>vine-based</i>	2.50	2.30	2.40	2.20	2.60	2.55	1.90	2.10	2.20
VGLM	8.00	6.20	6.00	10.5	12.0	11.8	20.0	20.1	19.8
<i>C-logit</i>	4.50	4.70	4.50	4.80	4.20	4.40	5.50	4.80	4.30

Table A.11 – Setting 1 : Results of the IMSE statistic ($\times 100$), based on K=1000 replications case $d = 3, p = 3$.

	$\tau = 0.3, \tau_j = 0.7$			$\tau = 0.7, \tau_j = 0.3$			$\tau = 0.7, \tau_j = 0.5$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.70	0.33	0.08	0.23	0.11	0.09	0.75	0.47	0.18
<i>vine-based</i>	2.00	1.76	1.62	1.52	1.66	1.47	2.38	2.47	2.19
VGLM	5.46	5.25	5.04	11.55	12.08	11.76	15.75	16.28	15.54
C-logit	4.73	4.52	4.20	5.25	5.46	5.04	5.04	4.73	4.41
Joe									
plug-in	1.23	1.14	0.95	1.43	0.95	0.86	1.33	1.05	0.67
<i>vine-based</i>	1.90	3.04	2.28	3.61	3.33	3.23	4.28	3.99	3.80
VGLM	6.30	5.78	6.51	13.13	13.44	12.60	23.10	22.58	24.15
C-logit	6.83	6.30	6.09	7.56	7.88	7.35	9.98	10.08	9.77
Gumbel									
plug-in	0.90	0.67	0.11	1.05	0.86	0.57	1.33	0.95	0.76
<i>vine-based</i>	2.19	2.00	1.90	3.33	3.14	2.95	3.80	3.71	3.52
VGLM	6.83	6.93	6.62	12.60	12.39	12.08	16.80	17.01	16.59
C-logit	4.20	3.99	3.78	7.88	7.35	7.25	8.61	8.40	8.72
Gaussian									
plug-in	1.05	0.71	0.67	1.33	1.05	1.52	1.43	0.86	0.90
<i>vine-based</i>	2.38	2.19	2.28	2.09	2.47	2.42	1.81	2.00	2.09
VGLM	8.40	6.51	6.30	11.03	12.60	12.39	21.00	21.11	20.79
C-logit	4.73	4.94	4.73	5.04	4.41	4.62	5.78	5.04	4.52

Table A.12 – Setting 1 : Results of IMSE ($\times 100$), K=1000 replications case $d = 3, p = 2$.

	$\tau = 0.3, \tau_j = 0.5$			$\tau = 0.5, \tau_j = 0.3$			$\tau = 0.5, \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.82	0.38	0.11	0.29	0.14	0.12	0.88	0.55	0.23
<i>vine-based</i>	2.31	2.04	1.87	1.76	1.93	1.71	2.75	2.86	2.53
VGLM	5.72	5.50	5.28	12.10	12.65	12.32	16.50	17.05	16.28
C-logit	4.95	4.73	4.40	5.50	5.72	5.28	5.28	4.95	4.62
Joe									
plug-in	1.43	1.32	1.10	1.65	1.10	0.99	1.54	1.21	0.77
<i>vine-based</i>	2.20	3.52	2.64	4.18	3.85	3.74	4.95	4.62	4.40
VGLM	6.60	6.05	6.82	13.75	14.08	13.20	24.20	23.65	25.30
C-logit	7.15	6.60	6.38	7.92	8.25	7.70	10.45	10.56	10.22
Gumbel									
plug-in	1.05	0.77	0.13	1.21	0.99	0.66	1.54	1.10	0.88
<i>vine-based</i>	2.53	2.31	2.20	3.85	3.63	3.41	4.40	4.29	4.07
VGLM	7.51	7.62	7.28	13.20	12.97	12.64	18.48	18.71	18.25
C-logit	4.62	4.39	4.16	8.25	7.70	7.59	9.47	9.24	9.59
Gaussian									
plug-in	1.21	0.83	0.77	1.54	1.21	1.76	1.65	0.99	1.05
<i>vine-based</i>	2.75	2.53	2.64	2.42	2.86	2.81	2.09	2.31	2.42
VGLM	8.80	6.82	6.60	11.55	13.20	13.00	23.10	23.22	22.87
C-logit	5.23	5.46	5.23	5.57	4.87	5.10	6.36	5.54	4.97

Table A.13 – Setting 1 : Result of IMSE ($\times 100$) based on K=1000 replications where $d = 3, p = 2$.

	$\tau = 0.3, \tau_j = 0.7$			$\tau = 0.7, \tau_j = 0.3$			$\tau = 0.7, \tau_j = 0.5$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
plug-in	0.77	0.36	0.09	0.25	0.12	0.10	0.83	0.52	0.20
<i>vine-based</i>	2.20	1.94	1.78	1.67	1.83	1.62	2.62	2.72	2.41
VGLM	6.01	5.78	5.54	12.73	13.30	12.94	17.33	17.91	17.09
C-logit	5.20	4.97	4.62	5.78	6.01	5.55	5.55	5.21	4.86
Joe									
plug-in	1.35	1.25	1.05	1.57	1.05	0.95	1.46	1.15	0.74
<i>vine-based</i>	2.09	3.34	2.51	3.97	3.66	3.55	4.71	4.39	4.18
VGLM	6.93	6.35	7.15	14.44	14.78	13.86	25.41	24.83	26.57
C-logit	7.51	6.93	6.70	8.32	8.66	8.09	10.98	11.09	10.73
Gumbel									
plug-in	0.99	0.74	0.12	1.16	0.95	0.63	1.46	1.05	0.84
<i>vine-based</i>	2.41	2.20	2.09	3.66	3.45	3.24	4.18	4.08	3.87
VGLM	7.88	8.00	7.64	13.86	13.62	13.27	19.40	19.65	19.16
C-logit	4.85	4.61	4.37	8.66	8.09	7.97	9.94	9.70	10.07
Gaussian									
plug-in	1.16	0.78	0.74	1.46	1.15	1.67	1.55	0.95	0.99
<i>vine-based</i>	2.62	2.41	2.51	2.30	2.72	2.66	2.30	2.20	2.30
VGLM	9.24	7.16	6.93	12.13	13.86	13.65	24.26	24.39	24.01
C-logit	5.49	5.73	5.49	5.85	5.12	5.36	6.68	5.82	5.22

A.1.7 Results for bias, SD, SE and coverage probability based on different scenarios from Settings 1 and 2

Table A.14 – Results of Bias, empirical SD, average SE, and 95% coverage $n = 300$, where $d = 2$, $p = 2$, based on setting 1.

Copula	Method	$\tau = \tau_j = 0.3$				$\tau = \tau_j = 0.5$				$\tau = \tau_j = 0.7$			
		Bias×100	SD×100	SE×100	Coverage in %	Bias×100	SD×100	SE×100	Coverage in %	Bias×100	SD×100	SE×100	Coverage in %
Clayton	plug-in	0.18	1.12	1.10	94.7	0.21	1.30	1.25	94.0	0.45	1.37	1.15	93.9
	vine-based	0.48	1.25	1.23	92.7	0.67	1.47	1.43	92.0	1.75	1.98	1.85	91.5
	VGLM	0.95	2.10	2.05	91.2	1.56	2.06	2.35	88.2	1.95	2.87	3.00	88.4
	C-logit	1.14	1.98	1.97	92.8	1.67	1.76	1.63	90.3	1.72	2.10	2.06	90.6
Joe	plug-in	0.55	1.85	1.90	94.8	0.56	1.61	1.43	94.2	0.73	2.00	1.94	94.4
	vine-based	0.61	2.16	2.23	92.1	0.62	2.46	2.79	90.9	1.58	2.94	2.67	91.0
	VGLM	2.83	2.27	2.32	89.0	2.56	2.94	2.79	88.6	2.15	3.58	2.94	81.9
	C-logit	1.92	2.39	2.41	88.8	1.86	2.16	2.01	88.4	1.29	2.15	2.01	88.5
Gumbel	plug-in	0.67	1.53	1.66	94.6	0.96	1.72	1.66	94.4	1.05	1.83	1.89	94.0
	vine-based	1.25	1.26	1.55	92.5	1.29	1.66	1.79	91.1	1.69	1.94	2.06	90.0
	VGLM	1.64	2.17	2.49	90.8	1.70	2.63	2.75	88.0	1.23	3.14	3.36	87.5
	C-logit	0.42	1.56	1.72	94.1	1.10	1.77	1.88	92.0	1.25	1.44	2.16	87.9
Gaussian	plug-in	0.22	0.93	0.92	94.4	0.73	1.62	1.70	94.9	0.53	1.51	1.56	94.1
	vine-based	0.76	1.62	1.64	93.3	1.33	1.64	1.69	89.9	3.41	2.49	2.82	89.5
	VGLM	1.89	2.06	2.17	89.5	1.96	2.41	2.64	89.9	3.44	3.42	3.82	65.6
	C-logit	1.67	1.66	1.89	90.6	1.53	1.69	1.83	90.2	1.72	1.55	1.43	90.0

Table A.15 – Results of Bias, empirical SD, average SE, and 95% coverage probability for $n = 500$, where $d = 2$, $p = 2$, based on setting 1.

Copula	Method	$\tau = \tau_j = 0.3$				$\tau = \tau_j = 0.5$				$\tau = \tau_j = 0.7$			
		Bias×100	SD×100	SE×100	Coverage(%)	Bias×100	SD×100	SE×100	Coverage(%)	Bias×100	SD×100	SE×100	Coverage(%)
Clayton	plug-in	0.15	1.08	1.05	94.5	0.19	1.28	1.22	94.8	0.42	1.32	1.12	94.2
	vine-based	0.45	1.20	1.18	93.2	0.63	1.40	1.36	91.9	0.72	1.92	1.80	90.8
	VGLM	0.90	2.05	2.00	90.8	1.50	2.00	2.30	87.5	1.92	2.80	2.95	87.9
	C-logit	1.10	1.93	1.92	92.3	1.22	1.72	1.60	89.5	1.30	1.90	1.96	90.1
Joe	plug-in	0.50	1.80	1.85	94.2	0.52	1.58	1.40	94.6	0.68	1.95	1.90	94.9
	vine-based	0.58	2.10	2.18	89.5	0.76	2.48	2.52	90.2	2.72	0.90	1.00	89.4
	VGLM	2.78	2.22	2.28	88.5	2.51	2.90	2.75	85.8	3.52	2.80	2.63	81.3
	C-logit	3.20	2.35	2.37	87.5	2.31	2.80	2.73	87.5	2.95	1.90	1.97	87.9
Gumbel	plug-in	0.62	1.48	1.55	94.1	0.92	1.68	1.62	94.9	1.00	1.78	1.84	94.5
	vine-based	1.20	1.22	1.70	91.5	1.48	1.76	1.97	90.6	1.63	1.90	2.00	90.0
	VGLM	1.60	2.66	2.89	90.3	1.65	2.58	2.70	87.5	2.08	3.40	3.52	86.5
	C-logit	0.38	1.52	1.45	93.6	1.05	1.72	1.78	90.6	1.48	1.83	2.03	89.2
Gaussian	plug-in	0.20	0.90	0.89	94.0	0.70	1.58	1.65	94.3	0.50	1.47	1.52	94.6
	vine-based	0.72	1.78	1.63	90.6	0.87	1.33	1.37	90.3	0.77	1.12	1.20	91.1
	VGLM	1.85	2.02	2.13	89.0	1.28	1.72	1.60	89.5	3.40	2.78	2.94	85.1
	C-logit	1.63	1.65	1.85	90.1	1.10	1.65	1.80	89.7	1.22	1.50	1.40	89.5

Table A.16 – Results of Bias, empirical SD, average SE, and 95% coverage ($\times 100$) $n = 300$, where $d = 2$, $p = 2$, based on setting 2.

Copula	Method	$\tau = \tau_j = 0.3$				$\tau = \tau_j = 0.5$				$\tau = \tau_j = 0.7$			
		Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)
Clayton	vine-based	0.45	1.22	1.20	93.5	0.65	1.42	1.38	92.1	0.75	1.92	1.82	91.0
	VGLM	0.92	2.08	2.03	91.0	1.55	2.08	2.35	87.9	1.98	2.85	2.98	88.1
	C-logit	1.12	1.95	1.93	92.7	1.25	1.72	1.65	90.2	1.35	1.92	1.98	90.5
Joe	vine-based	0.58	2.12	2.20	90.3	0.78	2.48	2.53	90.8	2.75	0.92	1.02	90.0
	VGLM	2.80	2.24	2.30	89.1	2.53	2.92	2.76	86.4	3.55	2.82	2.65	82.0
	C-logit	3.22	2.36	2.38	88.2	2.33	2.82	2.75	88.1	2.98	1.93	1.99	88.5
Gumbel	vine-based	1.22	1.24	1.72	92.1	1.50	1.78	1.98	91.1	1.65	1.92	2.03	90.6
	VGLM	1.62	2.68	2.91	90.8	1.68	2.60	2.72	88.1	2.10	3.42	3.54	87.1
	C-logit	0.40	1.53	1.46	94.1	1.07	1.75	1.80	91.2	1.50	1.85	2.05	89.8
Gaussian	vine-based	0.74	1.80	1.65	91.1	0.89	1.35	1.39	90.9	0.79	1.14	1.21	91.7
	VGLM	1.87	2.04	2.15	89.5	1.31	1.74	1.62	90.0	3.42	2.80	2.96	85.7
	C-logit	1.65	1.66	1.87	90.6	1.13	1.67	1.81	90.3	1.25	1.53	1.41	90.1

Table A.17 – Results of Bias, empirical SD, average SE, and 95% coverage ($\times 100$) $n = 300$, with $d = 2$, $p = 2$ $p_1 = 0.05$, $p_2 = 0.4$ based on setting 1.

Copula	Method	$\tau = \tau_j = 0.3$				$\tau = \tau_j = 0.5$				$\tau = \tau_j = 0.7$			
		Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)
Clayton	plug-in	0.25	1.25	1.22	94.2	0.32	1.45	1.38	94.5	0.55	1.52	1.32	94.2
	vine-based	0.58	1.38	1.35	92.8	0.78	1.58	1.53	91.5	0.88	2.12	2.00	90.3
	VGLM	1.15	2.25	2.20	90.2	1.75	2.25	2.55	86.8	2.15	3.05	3.18	87.2
	C-logit	1.30	2.12	2.10	92.0	1.42	1.88	1.82	89.5	1.50	2.08	2.15	89.8
Joe	plug-in	0.65	2.00	2.05	94.2	0.68	1.75	1.58	94.5	0.85	2.15	2.10	94.8
	vine-based	0.72	2.30	2.38	89.6	0.95	2.65	2.72	90.1	2.95	1.10	1.20	89.3
	VGLM	3.00	2.42	2.48	88.3	2.70	3.10	2.94	85.7	3.75	3.00	2.83	81.2
	C-logit	3.40	2.55	2.57	87.5	2.50	3.00	2.93	87.4	3.15	2.10	2.17	87.8
Gumbel	plug-in	0.80	1.68	1.75	94.0	1.10	1.88	1.82	94.8	1.20	2.00	2.05	94.5
	vine-based	1.40	1.42	1.90	91.5	1.68	2.00	2.18	90.5	1.82	2.12	2.22	89.9
	VGLM	1.80	2.85	3.10	90.0	1.85	2.78	2.90	87.4	2.30	3.60	3.75	86.3
	C-logit	0.55	1.70	1.65	93.5	1.25	1.95	2.00	90.6	1.68	2.05	2.25	89.1
Gaussian	plug-in	0.30	1.10	1.08	94.0	0.85	1.78	1.85	94.2	0.65	1.65	1.70	94.5
	vine-based	0.90	2.00	1.85	90.5	1.05	1.55	1.60	90.2	0.95	1.32	1.40	91.0
	VGLM	2.05	2.25	2.35	88.8	1.50	1.95	1.82	89.3	3.60	3.00	3.15	85.0
	C-logit	1.85	1.85	2.05	90.0	1.30	1.85	2.00	89.7	1.42	1.70	1.60	89.5

Table A.18 – Results of Bias, empirical SD, average SE, and 95% coverage ($\times 100$) $n = 500$, where $d = 2, p = 2, p_1 = 0.05, p_2 = 0.4$ based on setting 1.

Copula	Method	$\tau = \tau_j = 0.3$				$\tau = \tau_j = 0.5$				$\tau = \tau_j = 0.7$			
		Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)
Clayton	plug-in	0.18	1.05	1.02	94.8	0.25	1.25	1.20	94.2	0.42	1.32	1.18	94.5
	vine-based	0.45	1.18	1.15	93.6	0.65	1.38	1.33	92.5	0.75	1.82	1.75	91.8
	VGLM	0.95	1.95	1.90	91.5	1.55	1.95	2.25	88.5	1.95	2.75	2.88	88.8
	C-logit	1.15	1.82	1.80	93.2	1.28	1.58	1.52	90.8	1.35	1.78	1.85	91.2
Joe	plug-in	0.52	1.75	1.80	94.0	0.55	1.50	1.38	94.8	0.72	1.85	1.80	94.1
	vine-based	0.60	2.05	2.12	91.0	0.82	2.35	2.42	91.5	2.75	0.90	1.00	90.8
	VGLM	2.75	2.15	2.20	89.8	2.50	2.80	2.64	87.2	3.55	2.70	2.58	83.0
	C-logit	3.15	2.25	2.27	89.0	2.35	2.70	2.63	89.0	2.95	1.85	1.92	89.5
Gumbel	plug-in	0.65	1.45	1.50	94.2	0.95	1.65	1.60	94.1	1.05	1.75	1.80	93.9
	vine-based	1.25	1.22	1.65	92.8	1.50	1.75	1.95	92.0	1.65	1.85	1.95	91.5
	VGLM	1.60	2.55	2.80	91.5	1.65	2.48	2.60	89.0	2.10	3.30	3.45	88.0
	C-logit	0.42	1.45	1.40	94.6	1.05	1.70	1.75	92.0	1.50	1.80	2.00	90.8
Gaussian	plug-in	0.22	0.85	0.83	95.1	0.70	1.50	1.55	95.0	0.50	1.40	1.45	94.8
	vine-based	0.75	1.65	1.55	92.0	0.90	1.30	1.35	91.8	0.80	1.10	1.18	92.5
	VGLM	1.80	1.95	2.05	90.2	1.35	1.70	1.60	90.8	3.35	2.70	2.85	86.8
	C-logit	1.65	1.60	1.80	91.5	1.15	1.65	1.80	91.2	1.25	1.50	1.40	91.0

Table A.19 – Results of Bias, empirical SD, average SE ($\times 100$) and 95% coverage, with $d = 3, p = 3, p_1 = 0.3, p_2 = 0.4, p_3 = 0.2$ based on setting 1, $n = 300$.

Copula	Method	$\tau = \tau_j = 0.3$				$\tau = \tau_j = 0.5$				$\tau = \tau_j = 0.7$			
		Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)	Bias $\times 100$	SD $\times 100$	SE $\times 100$	Coverage(%)
Clayton	plug-in	0.26	1.31	1.28	94.8	0.34	1.52	1.45	94.1	0.58	1.60	1.39	94.6
	vine-based	0.61	1.45	1.42	92.4	0.82	1.66	1.61	91.1	0.92	2.23	2.10	89.9
	VGLM	1.21	2.36	2.31	89.8	1.84	2.36	2.68	86.4	2.26	3.20	3.34	86.7
	C-logit	1.37	2.23	2.21	91.6	1.49	1.97	1.91	89.1	1.58	2.18	2.26	89.4
Joe	plug-in	0.68	2.10	2.15	92.8	0.71	1.84	1.66	91.1	0.89	2.26	2.21	91.4
	vine-based	0.76	2.42	2.50	89.2	1.00	2.78	2.86	89.7	3.10	1.16	1.26	88.9
	VGLM	3.15	2.54	2.60	87.9	2.84	3.26	3.09	85.3	3.94	3.15	2.97	80.8
	C-logit	3.57	2.68	2.70	87.1	2.63	3.15	3.08	87.0	3.31	2.21	2.28	87.4
Gumbel	plug-in	0.54	1.76	1.84	93.6	1.16	1.97	1.91	91.4	1.26	2.10	2.15	90.1
	vine-based	1.47	1.49	2.00	91.0	1.76	2.10	2.29	90.1	1.91	2.23	2.33	89.5
	VGLM	1.89	2.99	3.26	89.6	1.94	2.92	3.05	87.0	2.42	3.78	3.94	85.9
	C-logit	0.58	1.79	1.73	89.1	1.31	2.05	2.10	90.2	1.76	2.15	2.36	88.7
Gaussian	plug-in	0.32	1.16	1.13	92.2	0.89	1.87	1.94	93.8	0.68	1.73	1.79	92.1
	vine-based	0.95	2.10	1.94	90.1	1.10	1.63	1.68	89.8	1.00	1.39	1.47	90.6
	VGLM	2.15	2.36	2.47	88.4	1.58	2.05	1.91	88.9	3.78	3.15	3.31	84.6
	C-logit	1.94	1.94	2.15	89.6	1.37	1.94	2.10	89.3	1.49	1.79	1.68	89.1

A.1.8 Selection of pair-copula families for real-data application

Table A.20 – Specification of the copula family for each tree in the vine construction involving coronary heart disease Y_1 , Age and BMI.

Tree	Edge	Conditioned	Conditioning	Var_types	Family	Parameter	τ
1	1	(age, chd)	-	c, d	Survival Clayton	0.11	0.05
1	2	(BMI, age)	-	c, c	Frank	0.99	0.11
2	1	(BMI, chd)	age	c, d	Joe	1.05	0.03

Table A.21 – Specification of the copula family for each tree in the vine construction involving corneal arcus disease Y_2 , Age and BMI.

Tree	Edge	Conditioned	Conditioning	Var_types	Family	Parameter	τ
1	1	(age, arcus)	-	c, d	Gaussian	0.15	0.10
1	2	(BMI, age)	-	c, c	Survival Clayton	0.19	0.09
2	1	(BMI, arcus)	age	c, d	Clayton	0.014	0.007

Table A.22 – Copula selection between the conditional marginal distributions of corneal arcus and coronary heart disease based on the proposed method.

Pair	Copula family	Parameter 1	Parameter 2	τ
Arcus , Coronary heart disease	t	0.99	2	0.91

A.2 Chapter 3 Appendices

A.2.1 IMSE Results case $d=3, p=2$

Table A.23 – Algorithm 2 : Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, for $d = 3$ and $p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
<i>Proposed method</i>	1.842	1.573	1.425	0.873	0.642	0.518	1.215	1.084	0.967
VGLM	6.724	5.892	5.437	12.836	11.925	11.502	20.447	20.883	20.915
C-logit	8.127	6.892	5.643	4.927	3.215	2.784	7.385	6.972	6.218
Frank									
<i>Proposed method</i>	2.157	1.892	1.843	4.872	4.516	4.327	6.127	6.154	5.842
VGLM	3.672	4.218	6.247	14.556	13.218	13.047	23.897	22.724	22.378
C-logit	7.615	6.182	6.470	9.320	10.867	10.978	12.040	12.078	11.998
Gumbel									
<i>Proposed method</i>	0.543	0.428	0.392	0.672	0.435	0.391	0.642	0.526	0.497
VGLM	3.936	3.450	3.439	5.194	4.802	4.741	8.188	8.092	8.208
C-logit	5.823	5.847	4.495	6.764	6.120	5.927	7.362	7.089	7.176

Table A.24 – Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, using the correlation matrices defined in equation (A.7).

	$n = 100$	$n = 300$	$n = 500$
Gaussian			
<i>Proposed method</i>	1.629	1.415	1.328
VGLM	3.782	3.426	3.215
C-logit	4.937	4.128	3.842

$$\Sigma_1 = \begin{bmatrix} 1.00 & 0.36 & 0.40 \\ 0.36 & 1.00 & 0.30 \\ 0.40 & 0.30 & 1.00 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 1.0 & 0.3 \\ 0.3 & 0.3 & 1.0 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 1.00 & 0.45 & 0.28 \\ 0.45 & 1.00 & 0.32 \\ 0.28 & 0.32 & 1.00 \end{bmatrix} \quad (\text{A.7})$$

$$\Sigma = \begin{bmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 1.0 & 0.6 \\ 0.3 & 0.6 & 1.0 \end{bmatrix}$$

A.2.2 Result of IMSE based on imbalanced data scenario

Table A.25 – Algorithm 2 : Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, for $d = 2$ and $p = 2$.

	$\tau = \tau_j = 0.3$			$\tau = \tau_j = 0.5$			$\tau = \tau_j = 0.7$		
	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$	$n = 100$	$n = 300$	$n = 500$
Clayton									
<i>Proposed method</i>	1.642	1.428	1.392	0.728	0.592	0.515	1.152	1.067	1.028
<i>VGLM</i>	6.724	5.892	5.437	13.836	12.925	12.502	22.447	22.883	22.915
<i>C-logit</i>	8.127	6.892	5.643	4.927	3.215	2.784	7.385	6.972	6.218
Frank									
<i>Proposed method</i>	2.157	1.892	1.843	4.872	4.516	4.327	6.127	6.154	5.842
<i>VGLM</i>	3.672	4.218	6.247	15.556	14.218	14.047	24.897	23.724	23.378
<i>C-logit</i>	7.615	6.182	6.470	9.320	10.867	10.978	12.040	12.078	11.998
Gumbel									
<i>Proposed method</i>	0.543	0.428	0.392	0.672	0.435	0.391	0.642	0.526	0.497
<i>VGLM</i>	3.936	3.450	3.439	5.194	4.802	4.741	8.188	8.092	8.208
<i>C-logit</i>	5.823	5.847	4.495	6.764	6.120	5.927	7.362	7.089	7.176

Table A.26 – Results of the IMSE statistic ($\times 100$), based on $B = 1000$ replications, using the correlation matrices defined in equation (3.35) for $d = 2$ and $p = 2$.

	$n = 100$	$n = 300$	$n = 500$
Gaussian			
<i>Proposed method</i>	1.892	1.645	1.518
<i>VGLM</i>	4.127	3.782	3.625
<i>C-logit</i>	5.428	4.723	4.315

A.3 Chapter 4 Appendices

A.3.1 Proof of the Representation of $\lambda_0(x)$ given by (4.19)

Proof. From the Remark 4.5, one can get the expression for the conditional probability $\mathbb{P}(Y = 1 | X_1 = x_1)$ in the presence of one covariate, and the dependence between Y and X_1 is modeled by a Frank copula :

$$\pi(x) = \mathbb{P}(Y = 1 | X_1 = x_1) = \frac{\exp(\theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1))}{1 + \exp(\theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1))},$$

where :

- θ is the parameter of the Frank copula,
- $F_1(x_1)$ is the marginal distribution function of X_1 ,
- $p = \mathbb{P}(Y = 1)$,
- $\varphi_\theta(t)$ represents the Frank copula generator.

The objective is to derive the expression for $\lambda_0(x)$ given by Equation (4.19) :

$$\begin{aligned} \lambda_0(x_1) &= \frac{\tau \pi(x_1)}{(1-\tau)(1-\pi(x_1)) + \tau \pi(x_1)} \\ &= \frac{\exp\left(\theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1) + \ln\left(\frac{\tau}{1-\tau}\right)\right)}{1 + \exp\left(\theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1) + \ln\left(\frac{\tau}{1-\tau}\right)\right)}. \end{aligned}$$

Starting with the expression for $\lambda_0(x)$:

$$\lambda_0(x_1) = \frac{\tau \pi(x_1)}{(1-\tau)(1-\pi(x_1)) + \tau \pi(x_1)}.$$

From the expression of $\pi(x_1)$, let's define :

$$A = \theta F_1(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1).$$

Then :

$$\pi(x_1) = \frac{\exp(A)}{1 + \exp(A)},$$

and similarly :

$$1 - \pi(x_1) = \frac{1}{1 + \exp(A)}.$$

Substitute $\pi(x_1)$ and $1 - \pi(x_1)$ into $\lambda_0(x_1)$:

$$\begin{aligned}\lambda_0(x_1) &= \frac{\tau \cdot \frac{\exp(A)}{1 + \exp(A)}}{(1 - \tau) \cdot \frac{1}{1 + \exp(A)} + \tau \cdot \frac{\exp(A)}{1 + \exp(A)}} \\ &= \frac{\tau \exp(A)}{(1 - \tau) + \tau \exp(A)}.\end{aligned}$$

Simplify the numerator and denominator :

$$\lambda_0(x_1) = \frac{\tau \exp(A)}{(1 - \tau) + \tau \exp(A)}.$$

Note that one can express τ , and $1 - \tau$ respectively as follows $\tau = \exp(\ln \tau)$ and $1 - \tau = \exp(\ln(1 - \tau))$.

Rewriting $\lambda_0(x_1)$:

$$\begin{aligned}\lambda_0(x_1) &= \frac{\exp(A + \ln \tau)}{\exp(\ln(1 - \tau)) + \exp(A + \ln \tau)} \\ \lambda_0(x_1) &= \frac{\exp(A + \ln \tau)}{\exp(\ln(1 - \tau)) + \exp(A + \ln \tau)}.\end{aligned}$$

Factor out $\exp(\ln(1 - \tau))$ in the denominator :

$$\lambda_0(x_1) = \frac{\exp(A + \ln \tau)}{\exp(\ln(1 - \tau)) [1 + \exp(A + \ln \tau - \ln(1 - \tau))]}.$$

Since $\exp(\ln(1 - \tau)) = 1 - \tau$, we have :

$$\lambda_0(x_1) = \frac{\exp(A + \ln \tau)}{(1 - \tau) \left[1 + \exp \left(A + \ln \left(\frac{\tau}{1 - \tau} \right) \right) \right]}.$$

Multiply numerator and denominator by $1 - \tau$:

$$\begin{aligned}\lambda_0(x_1) &= \frac{\exp(A + \ln \tau)}{(1 - \tau) + \exp(A + \ln \tau)} \\ \lambda_0(x_1) &= \frac{\exp(A + \ln \tau)}{(1 - \tau) + \exp(A + \ln \tau)}.\end{aligned}$$

Divide numerator and denominator by $1 - \tau$:

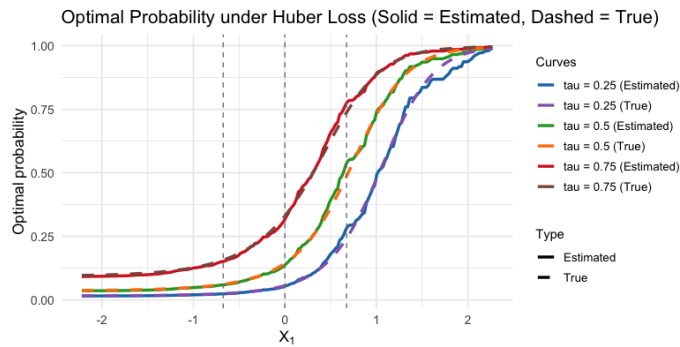
$$\lambda_0(x_1) = \frac{\frac{\exp(A + \ln \tau)}{1 - \tau}}{1 + \frac{\exp(A + \ln \tau)}{1 - \tau}} = \frac{\exp \left(A + \ln \left(\frac{\tau}{1 - \tau} \right) \right)}{1 + \exp \left(A + \ln \left(\frac{\tau}{1 - \tau} \right) \right)}.$$

Substitute A by its expression then one can get the result given by equation (4.19)

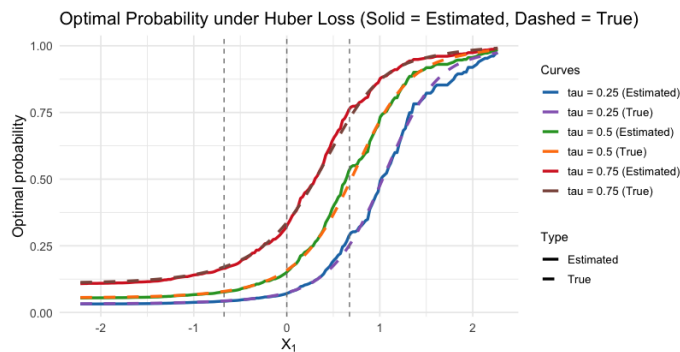
$$\lambda_0(x_1) = \frac{\exp\left(\theta F(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1) + \ln\left(\frac{\tau}{1-\tau}\right)\right)}{1 + \exp\left(\theta F(x_1) + \log(\exp[\varphi_\theta(1-p)] - 1) + \ln\left(\frac{\tau}{1-\tau}\right)\right)}.$$

□

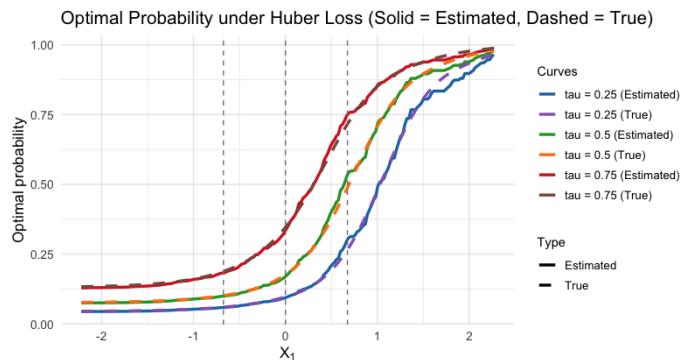
A.3.2 Graphic illustration of $\lambda_0(\cdot)$ and $\hat{\lambda}_n(\cdot)$ based on Huber-type loss function



(a) $p = 0.3$, and $k = 0.01$



(b) $p = 0.3$, and $k = 0.05$



(c) $p = 0.3$, and $k = 0.1$

Figure A.3 - Estimation of the probability of success on different quantile levels based on the Huber loss function defined in Eq. (4.8) for different values of k , dashed lined and solid lines represent $\lambda_0(x_1)$ and $\hat{\lambda}_n(x_1)$ respectively.

A.3.3 Algorithm

A.3.3.1 Bootstrap procedure to asses $\sigma^2(\mathbf{x})$

Algorithm 5 Based on the observations $(Y_i, X_{i,1}, \dots, X_{i,d}), i = 1, \dots, n$, we estimate the copula parameter θ by $\hat{\theta}$. For $b = 1, \dots, B$ and $i = 1, \dots, n$:

- 1: Generate $(U_i^{(b)}, V_{i,1}^{(b)}, \dots, V_{i,d}^{(b)})$ from the copula $\mathcal{C}_{\hat{\theta}}(\cdot)$.
- 2: Calculate $X_{i,j}^{(b)} = \hat{F}_{j,n}^{-1}(V_{i,j}^{(b)})$ for $j = 1, \dots, d$ and $Y_i^{(b)} = \hat{F}_{0n}^{-1}(U_i^{(b)})$, where \hat{F}_{0n} and $\hat{F}_{j,n}$ are the empirical distributions of Y_i and $X_{i,j}$, respectively, based on $i = 1, \dots, n$.
- 3: Based on the observations $(Y_i^{(b)}, X_{i,1}^{(b)}, \dots, X_{i,d}^{(b)})$, evaluate $\hat{\pi}_n^{(b)}$, the copula parameter $\hat{\theta}^{(b)}$ and the empirical distributions

$$\hat{\mathbf{F}}_n^{(b)}(\mathbf{x}) = (\hat{F}_{1n}^{(b)}(x_1), \dots, \hat{F}_{dn}^{(b)}(x_d)).$$

- 4: Calculate

$$\hat{\pi}_n^{(b)}(x) = \mathcal{C}_{\hat{\theta}^{(b)}}^c\{1 - \hat{p}_n^{(b)} \mid \hat{\mathbf{F}}_n^{(b)}(\mathbf{x})\}.$$

then evaluate $\hat{\lambda}_n^b(\mathbf{x})$ which represents the root of the following equation :

$$f_n^c(\hat{\lambda}_n(\mathbf{x})) = \psi_\tau(1 - \hat{\lambda}_n(\mathbf{x})) \hat{\pi}_n^{(b)}(\mathbf{x}) + \psi_\tau(-\hat{\lambda}_n(\mathbf{x})) (1 - \hat{\pi}_n^{(b)}(\mathbf{x})) = 0,$$

- 5: Estimate the variance $\sigma_n^2(\mathbf{x})$ by the empirical variance of the sample $\{\hat{\lambda}_n^{(b)}(\mathbf{x})\}_{b=1, \dots, B}$ given by

$$\sigma_n^2(\mathbf{x}) = B^{-1} \sum_{b=1}^B \left(\hat{\lambda}_n^{(b)}(\mathbf{x}) - B^{-1} \sum_{b=1}^B \hat{\lambda}_n^{(b)}(\mathbf{x}) \right)^2.$$

A.3.4 Monte Carlo Simulation Procedure

Simulation Parameters

- Number of observations : $n = 1000$
- Number of simulations : $N_{\text{sim}} = 1000$
- Probability of the rare event : $p = 0.1$
- Values of τ for the estimation of λ : $\tau \in \{0.2, 0.5, 0.7\}$
- Set $k = 0.01$ For quantile loss, and $\tau = 0.2, 0.5, 0.7$

$$f^c(\lambda(\mathbf{x})) = \psi(0 - \lambda(\mathbf{x})) \cdot (1 - \pi_n(\mathbf{x})) + \psi(1 - \lambda(\mathbf{x})) \cdot \pi_n(\mathbf{x}) = 0, \quad (\text{A.8})$$

where π_n is defined in Algorithm 6

Algorithm 6 Monte Carlo Simulation

1: **for** $s = 1$ **to** N_{sim} **do**

2: **Data Generation**

— **Copula Parameter** : Calculate the parameter θ of copula for the following kendalls tau $\tau^c = 0.2, 0.5, 0.75$:

$$\theta = \text{ITau}(C_\theta, \tau^c = 0.5)$$

— **Generate Random Variables** :

— The data is generated following the steps **1 to 3** in the Algorithm 3.

— **Dataset Construction** :

$$\mathcal{D} = \{(Y_i, X_{i1}, X_{i2}), \quad i = 1, \dots, n\}$$

3: **Data Splitting**

— Split data into $\mathcal{D}_{\text{train}}$ (80%) and $\mathcal{D}_{\text{test}}$ (20%).

4: **Parameter Estimation**

— **Estimate** \hat{p}_n :

$$\hat{p}_n = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} Y_i$$

— **Empirical Distribution Functions** :

$$\hat{F}_{nj}(x_j) = \frac{1}{n_{\text{train}} + 1} \sum_{i=1}^{n_{\text{train}}} \mathbb{1}(X_{ij} \leq x_j), \quad j = 1, 2$$

— **Estimate** $\hat{\theta}$: Solve :

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^l} \sum_{i=1}^{n_{\text{train}}} \left\{ Y_i \ln \left(1 - C_\theta^c \left(1 - \hat{p}_n \mid \hat{\mathbf{F}}_n(\mathbf{X}_i) \right) \right) + (1 - Y_i) \ln \left(C_\theta^c \left(1 - \hat{p}_n \mid \hat{\mathbf{F}}_n(\mathbf{X}_i) \right) \right) \right\}$$

where $\hat{\mathbf{F}}_n(\mathbf{X}_i) = (\hat{F}_{n1}(x_1), \hat{F}_{n2}(x_2))$, and C_θ^c is the conditional distribution of copulas.

5: **Predicted Probabilities**

— Compute $\hat{\pi}(\mathbf{x})$ on test data :

$$\hat{\pi}(\mathbf{x}) = 1 - C_\theta \left(1 - \hat{p}_n \mid \hat{\mathbf{F}}_n(\mathbf{x}) \right)$$

6: **Model Fitting and Evaluation**

— Fit models on $\mathcal{D}_{\text{train}}$ and predict on $\mathcal{D}_{\text{test}}$ using : Logistic Regression, Random Forest, SVM, etc.

— Calculate average performance metrics : accuracy, precision, recall, F1-score.

7: **end for**

A.3.5 Accuracy metrics

The accuracy metrics are calculated as follows :

— **Accuracy** :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

— **Precision :**

$$\text{Precision} = \frac{TP}{TP + FP},$$

— **Sensitivity :**

$$\text{Sensitivity} = \frac{TP}{TP + FN}.$$

— **Specificity :**

$$\text{Specificity} = \frac{TN}{TN + FP},$$

— **F1-score :**

$$F1 = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}.$$

A.3.6 Simulation Results

A.3.6.1 Case d=1

Table A.27 - Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.25$)	0.6090	0.7968	0.9138	0.9216	0.6904
Expectile ($\tau = 0.25$)	0.7729	0.8539	0.9704	0.9582	0.8056
SVM	0.4619	0.8828	0.7963	0.8779	0.6065
SVM-Asymmetric	0.4127	0.9723	0.9908	0.8624	0.7520
Random Forest	0.6682	0.8598	0.9497	0.9404	0.7473
Logistic Regression	0.3647	0.8544	0.8330	0.8348	0.5112

Table A.28 - Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.25$)	0.7621	0.8626	0.9712	0.9616	0.8091
Expectile ($\tau = 0.25$)	0.7729	0.8605	0.9715	0.9620	0.8140
SVM	0.3891	0.9870	0.8337	0.7512	0.5582
SVM-Asymmetric	0.8966	0.5939	0.9406	0.9543	0.7145
Random Forest	0.6009	0.8275	0.9511	0.9421	0.6962
Logistic Regression	0.5931	0.9466	0.9289	0.9305	0.7293

Table A.29 - Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.2$)	0.5051	0.7968	0.9138	0.9022	0.6183
Expectile ($\tau = 0.2$)	0.4746	0.8481	0.8965	0.8915	0.6086
SVM	0.3543	0.9587	0.8077	0.8225	0.5174
Asymmetric SVM	0.6751	0.3045	0.9843	0.9174	0.4197
Random Forest	0.4250	0.7173	0.8932	0.8756	0.5338
Logistic Regression	0.4002	0.9308	0.8460	0.8542	0.5597

Table A.30 – Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.2$)	0.5344	0.7365	0.9251	0.9054	0.6194
Expectile ($\tau = 0.3$)	0.5098	0.7884	0.9112	0.8987	0.6192
SVM	0.3370	0.9579	0.7821	0.8001	0.4986
SVM-Asymmetric	0.6995	0.2820	0.9859	0.9136	0.4020
Random Forest	0.4235	0.7107	0.8872	0.8693	0.5307
Logistic Regression	0.4041	0.9182	0.8423	0.8500	0.5612

Table A.31 – Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.2$)	0.6942	0.9089	0.9623	0.9424	0.7872
Expectile ($\tau = 0.2$)	0.6997	0.7582	0.9634	0.9429	0.7278
SVM	0.4667	0.9142	0.8820	0.8851	0.6179
SVM-Asymmetric	0.8964	0.4469	0.9941	0.9400	0.5964
Random Forest	0.5300	0.7707	0.9235	0.9079	0.6281
Logistic Regression	0.4667	0.9142	0.8820	0.8851	0.6179

Table A.32 - Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.2$)	0.8480	0.9174	0.9802	0.9742	0.8813
Expectile ($\tau = 0.2$)	0.8517	0.9119	0.9810	0.9743	0.8808
SVM	0.4099	1.0000	0.8356	0.8524	0.5815
SVM-Asymmetric	0.3759	1.0000	0.8123	0.8315	0.5464
Random Forest	0.7974	0.9153	0.9235	0.9666	0.8523
Logistic Regression	0.7452	0.9657	0.9603	0.8851	0.8412

Table A.33 - Mean of the accuracy metrics for each method based on Algorithm 3 (Gaussian dependence with $\rho_{Y,X} = 0.4$, $n = 1000$, and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.35$)	0.5465	0.5747	0.9441	0.9071	0.5602
Expectile ($\tau = 0.3$)	0.5109	0.6082	0.9331	0.9006	0.5553
SVM	0.3096	0.8827	0.7861	0.7952	0.4584
SVM-Asymmetric	0.0979	1.0000	0.0000	0.0979	0.1783
Random Forest	0.3594	0.6343	0.8764	0.8526	0.4588
Logistic Regression	0.3377	0.8460	0.8193	0.8216	0.4827

A.3.6.2 Case d=2

Table A.34 - Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.5$ $n = 1000$ and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.3$)	0.600	0.7120	0.9526	0.9316	0.6512
Expectile ($\tau = 0.25$)	0.5318	0.7195	0.9402	0.9232	0.6116
SVM	0.3445	0.8857	0.8127	0.8199	0.4961
SVM-Asymmetric	0.3677	0.8749	0.8324	0.8365	0.5178
Random Forest	0.3795	0.7786	0.8584	0.8506	0.5103
Logistic Regression	0.3896	0.8633	0.8499	0.8511	0.5369

Table A.35 - Mean of the accuracy metrics for each method based on Algorithm 3 (Gumbel dependence with $\tau^c = 0.7$, $n = 1000$ and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.3$)	0.8524	0.8025	0.9861	0.9698	0.8267
Expectile ($\tau = 0.45$)	0.8499	0.9112	0.9667	0.9750	0.8795
SVM	0.4774	0.9852	0.8785	0.8892	0.6431
SVM-Asymmetric	0.5136	0.9816	0.8951	0.9038	0.6744
Random Forest	0.6388	0.9175	0.9405	0.9382	0.7532
Logistic Regression	0.6445	0.9539	0.9398	0.9412	0.7693

Table A.36 – Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.5$, $n = 1000$, and $p = 0.1$)

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.3$)	0.6842	0.8527	0.9345	0.9238	0.7591
Expectile ($\tau = 0.4$)	0.6528	0.8324	0.9256	0.9152	0.7318
SVM	0.3682	0.9472	0.8124	0.8271	0.5303
Asymmetric SVM	0.6875	0.3182	0.9815	0.9228	0.4350
Random Forest	0.4386	0.7325	0.9014	0.8823	0.5487
Logistic Regression	0.4158	0.9254	0.8512	0.8597	0.5738

Table A.37 – Mean of the accuracy metrics for each method based on Algorithm 3 (Frank dependence with $\tau^c = 0.7$, $n = 1000$, and $p = 0.1$)

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.3$)	0.7125	0.8614	0.9423	0.9312	0.7799
Expectile ($\tau = 0.4$)	0.6837	0.8432	0.9335	0.9224	0.7551
SVM	0.3527	0.9638	0.7982	0.8124	0.5164
Asymmetric SVM	0.7124	0.2958	0.9872	0.9251	0.4180
Random Forest	0.4312	0.7254	0.8927	0.8742	0.5409
Logistic Regression	0.4125	0.9327	0.8475	0.8562	0.5720

Table A.38 - Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.5$, $n = 1000$ and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.25$)	0.7646	0.7562	0.9759	0.9550	0.7604
Expectile ($\tau = 0.35$)	0.7381	0.7695	0.9716	0.9538	0.7535
SVM	0.3780	0.9465	0.8294	0.8410	0.5402
SVM-Asymmetric	0.8814	0.5021	0.9925	0.9443	0.6398
Random Forest	0.5072	0.8688	0.9077	0.9039	0.6405
Logistic Regression	0.5025	0.9212	0.8996	0.9017	0.6503

Table A.39 - Mean of the accuracy metrics for each method based on Algorithm 3 (Joe dependence with $\tau^c = 0.7$, $n = 1000$ and $p = 0.1$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.25$)	0.9432	0.8531	0.9952	0.9828	0.8959
Expectile ($\tau = 0.25$)	0.9493	0.8638	0.9952	0.9860	0.9045
SVM	0.4787	0.9962	0.8763	0.8886	0.6467
SVM-Asymmetric	0.9736	0.7213	0.9977	0.9701	0.8287
Random Forest	0.7555	0.9641	0.9635	0.9638	0.8471
Logistic Regression	0.7578	0.9707	0.9641	0.9649	0.8511

Table A.40 – Mean of the accuracy metrics for each method based on Algorithm 3 (Gaussian dependence with the following correlation matrix defined in (A.9), $p = 0.1$ and $n = 1000$).

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Huber-type ($\tau = 0.3$)	0.6768	0.4475	0.9746	0.9208	0.5388
Expectile ($\tau = 0.35$)	0.5547	0.6353	0.9411	0.9096	0.5923
SVM	0.2299	0.7714	0.7067	0.7129	0.3542
SVM-Asymmetric	0.1010	1.0000	0.0000	0.1010	0.1835
Random Forest	0.2336	0.4568	0.8316	0.7930	0.3091
Logistic Regression	0.3694	0.8618	0.8334	0.8358	0.5171

$$\Sigma = \begin{bmatrix} 1 & 0.6 & 0.3 \\ 0.6 & 1 & 0.45 \\ 0.3 & 0.45 & 1 \end{bmatrix} \quad (\text{A.9})$$

A.3.7 Result of real data illustration case $d = 2$

A.3.8 Accuracy metric results

Table A.41 – Accuracy metrics for each method with $d = 2$.

Method	Precision	Sensitivity	Specificity	Accuracy	F1_Score
Expectile $\tau = 0.1$	0.7449	0.7228	0.9996	0.9991	0.7337
Huber-type $\tau = 0.2$	0.7755	0.6552	0.9996	0.9989	0.7103
SVM-Asymmetric	0.4898	0.8421	0.9991	0.9990	0.6194
Logistic Regression	0.5000	0.8978	0.9991	0.9990	0.6423
Random Forest	0.5408	0.8689	0.9992	0.9991	0.6667
SVM	0.5204	0.8500	0.9992	0.9990	0.6456

A.3.9 Configuration of conditional copula based on vine decomposition

Table A.42 – C-vine copula configuration.

Tree	Edge	Conditioned	Conditioning	Var_types	Family	Parameters	Tau
1	1	(V17, <i>Class</i>)	-	c, d	Gaussian	0.07	0.04
1	2	(V14, <i>Class</i>)	-	c, d	Gaussian	0.14	0.09
2	1	(V14, V17; <i>Class</i>)	<i>Class</i>	c,c; d	Frank	-0.67	-0.07

BIBLIOGRAPHIE

- Acar, E. F., Genest, C. et Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110, 74–90.
- Agresti, A. (2010). *Analysis of ordinal categorical data*. John Wiley & Sons.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Dans B. N. Petrov et F. Csaki (dir.). *Proceedings of the 2nd International Symposium on Information Theory*, 267–281., Budapest. Akademiai Kiado.
- Akbani, R., Kwek, S. et Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets. Dans *European conference on machine learning*, 39–50. Springer.
- Aldahmani, S., Kortbi, O. et Mesfioui, M. (2024). Copula-based regression with mixed covariates. *Mathematics*, 12(22), 3525.
- Bedford, T. et Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.
- Bedford, T. et Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4), 1031–1068.
- Bianco, A. M. et Yohai, V. J. (1996). Robust estimation in the logistic regression model. *The Annals of Statistics*, 24(3), 1994–2008.
- Bishop, Y. M., Fienberg, S. E., Holland, P. W., Light, R. J. et Mosteller, F. (1977). Book review : Discrete multivariate analysis : Theory and practice. *Applied psychological measurement*, 1(2), 297–306.
- Bouyé, E. et Salmon, M. (2009). Dynamic copula quantile regressions and tail area dynamic dependence in forex markets. *The European Journal of Finance*, 15(7-8), 721–750.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5–32.
- Briseño Sanchez, G., Klein, N., Klinkhammer, H. et Mayr, A. (2024). Boosting distributional copula regression for bivariate binary, discrete and mixed responses. *Statistical Methods in Medical Research*, p. 09622802241313294.
- Chawla, N. V., Bowyer, K. W., Hall, L. O. et Kegelmeyer, W. P. (2002). Smote : synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357.
- Chib, S. et Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 85(2), 347–361.
- Christopher, M. et Nasrabadi, N. M. (2006). Pattern recognition and machine learning. springer. vol. *Bishop*, p. 183.
- Cortes, C. et Vapnik, V. (1995). Support-vector network-. machine learning 20 : 273–297. *Portfolio Selection, Journal of Global Optimization*, 43(2-3).
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 20(2), 215–232.

- Czado, C. et Nagler, T. (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9(Volume 9, 2022), 453–477.
<http://dx.doi.org/https://doi.org/10.1146/annurev-statistics-040220-101153>.
 Récupéré de <https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-040220-101153>
- Dawber, J., Salvati, N., Fabrizi, E. et Tzavidis, N. (2022). Expectile regression for multi-category outcomes with application to small area estimation of labour force participation. *Journal of the Royal Statistical Society Series A : Statistics in Society*, 185(Supplement_2), S590–S619.
- De Leon, A. R. et Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*, 30(2), 175–185.
- Derumigny, A. et Fermanian, J.-D. (2017). About tests of the 'simplifying' assumption for conditional copulas. *Dependence Modeling*, 5(1), 154–197.
- Dette, H., Hecke, R. V. et Volgushev, S. (2014). Some comments on copula-based regression. *Journal of the American Statistical Association*, 109(507), 1319–1324.
<http://dx.doi.org/10.1080/01621459.2014.916577>. Récupéré de
<https://doi.org/10.1080/01621459.2014.916577>
- Dissmann, J., Brechmann, E. C., Czado, C. et Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59, 52–69.
- Fernández, A., Sorokin, A. et Thompson, P. D. (2007). Corneal arcus as coronary artery disease risk factor. *Atherosclerosis*, 193(2), 235–240.
- Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22, 700–725. Première formulation systématique de la méthode de scoring de Fisher.
- Genest, C. et Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4), 347–368.
- Genest, C., Ghoudi, K. et Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3), 543–552.
- Genest, C. et MacKay, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian journal of statistics*, 14(2), 145–159.
- Genest, C. et Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin : The Journal of the IAA*, 37(2), 475–515.
- Genest, C., Nikoloulopoulos, A. K., Rivest, L.-P. et Fortin, M. (2013). Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. *Brazilian Journal of Probability and Statistics*, 27(3), 265–284.
- Gijbels, I., Omelka, M., Pešta, M. et Veraverbeke, N. (2017). Score tests for covariate effects in conditional copulas. *Journal of Multivariate Analysis*, 159, 111–133.
- Han, H., Wang, W.-Y. et Mao, B.-H. (2005). Borderline-smote : a new over-sampling method in imbalanced data sets learning. Dans *International conference on intelligent computing*, 878–887. Springer.
- Hastie, T., Tibshirani, R. et Friedman, J. H. (2009). *The Elements of Statistical Learning : Data Mining, Inference, and Prediction* (2 éd.). New York : Springer.

- He, H., Bai, Y., Garcia, E. A. et Li, S. (2008). Adasyn : Adaptive synthetic sampling approach for imbalanced learning. Dans *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 1322–1328. Ieee.
- Huang, Y.-M. et Du, S.-X. (2005). Weighted support vector machine for classification with uneven training class sizes. Dans *2005 international conference on machine learning and cybernetics*, volume 7, 4365–4369. IEEE.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1), 73–101. <http://dx.doi.org/10.1214/aoms/1177733251>
- Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *Lecture notes-monograph series*, 120–141.
- Joe, H. (2014). *Dependence modeling with copulas*. CRC press.
- Jorgensen, B. (1997). *The theory of dispersion models*. CRC Press.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2), 81–93.
- King, G. et Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, 9(2), 137–163.
- Koenker, R. et Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- Kraus, D. et Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics & Data Analysis*, 110, 1–18.
- Manski, C. F. et Thompson, T. S. (1989). Estimation of best predictors of binary response. *Journal of Econometrics*, 40(1), 97–123.
- McCullagh, P. et Nelder, J. A. (1989). *Generalized Linear Models* (2 éd.). London : Chapman and Hall.
- McCulloch, C. E. et Searle, S. R. (2004). *Generalized, linear, and mixed models*. John Wiley & Sons.
- Mesfioui, M., Bouezmarni, T. et Belalia, M. (2023). Copula-based link functions in binary regression models. *Statistical Papers*, 1–29.
- Mesfioui, M. et Quesy, J.-F. (2008). Dependence structure of conditional archimedean copulas. *Journal of Multivariate Analysis*, 99(3), 372–385.
- Nagler, T. et Vatter, T. (2018). rvinecopulib : High performance algorithms for vine copula modeling. *R package version*, 2.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer.
- Newey, W. K. et Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica : Journal of the Econometric Society*, 819–847.
- Nikoloulopoulos, A. K. et Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9), 1555–1568.
- Noh, H., Ghouch, A. E. et Bouezmarni, T. (2013). Copula-based regression estimation and inference. *Journal of the American Statistical Association*, 108(502), 676–688.

- Noh, H., Ghouch, A. E. et Van Keilegom, I. (2015). Semiparametric conditional quantile estimation through copula-based multivariate models. *Journal of Business & Economic Statistics*, 33(2), 167–178.
- Omelka, M., Gijbels, I. et Veraverbeke, N. (2009). Improved kernel estimation of copulas : weak convergence and goodness-of-fit testing. *Annals of Statistics*, (37), 3023–3058.
- Pakes, A. et Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica : Journal of the Econometric Society*, 1027–1057.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3), 1065–1076.
- Peter, X.-K. S. et Song, K. (2007). *Correlated data analysis : modeling, analytics, and applications*. Springer.
- Rémillard, B., Nasri, B. et Bouezmarni, T. (2017). On copula-based conditional quantile estimators. *Statistics & Probability Letters*, 128, 14–20.
- Rosenman, R. H., Brand, R. J., Jenkins, C. D., Friedman, M., Straus, R. et Wurm, M. (1975). Coronary heart disease in the western collaborative group study : Final follow-up experience of 8 1/2 years. *Jama*, 233(8), 872–877.
- Sáez, J. A., Luengo, J., Stefanowski, J. et Herrera, F. (2015). Smote–ipf : Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. *Information Sciences*, 291, 184–203.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. Dans *Publications de l'Institut de Statistique de l'Université de Paris*, volume 8, 229–231.
- Song, P. X.-K., Li, M. et Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics*, 65(1), 60–68.
- Stöber, J. (2013). *Regular vine copulas with the simplifying assumption, time-variation, and mixed discrete and continuous margins*. (Thèse de doctorat). Technische Universität München.
- Stöber, J., Hong, H. G., Czado, C. et Ghosh, P. (2015). Comorbidity of chronic diseases in the elderly : Patterns identified by a copula design for mixed responses. *Computational Statistics & Data Analysis*, 88, 28–39.
- Stoeber, J., Joe, H. et Czado, C. (2013). Simplified pair copula constructions—limitations and extensions. *Journal of Multivariate Analysis*, 119, 101–118.
- Suh, Y., Yu, J., Mo, J., Song, L. et Kim, C. (2017). A comparison of oversampling methods on imbalanced topic classification of korean news articles. *Journal of Cognitive Science*, 18(4).
- Xue-Kun Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2), 305–320.
- Yamaguchi, Y. et Maruo, K. (2019). Bivariate beta-binomial model using gaussian copula for bivariate meta-analysis of two binary outcomes with low incidence. *Japanese Journal of Statistics and Data Science*, 2(2), 347–373.
- Yee, T. W. (2015). *Vector generalized linear and additive models : with an implementation in R*, volume 10. Springer.
- Zhao, L. P. et Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, 77(3), 642–648.