



Thèse

présentée pour obtenir le grade de **Docteur**

de l'**Université de Reims Champagne-Ardenne**

Spécialité : **Mathématiques**

Option : **Statistique**

Alexandre DEPIRE

Modélisation Markovienne - Modèles de régression de copules et valeurs extrêmes - Application aux systèmes d'aide à la conduite

soutenue le 18/02/2008 devant le jury composé de

Président	Professeur Armelle GUILLOU	Université de Strasbourg Louis Pasteur
Rapporteur	Professeur Belkacem ABDOUS	Université Laval - Québec (Canada)
Rapporteur	Professeur Armelle GUILLOU	Université de Strasbourg Louis Pasteur
Examineurs	Professeur Michel BRONIATOWSKI	Université Pierre et Marie Curie
	Professeur Georges DELAUNAY	Université Reims Champagne-Ardenne
	Sylvain LASSARRE	Institut National de Recherche sur les Transports et leur Sécurité
	Farida SAAD	Institut National de Recherche sur les Transports et leur Sécurité
Directeur de thèse	Professeur Djamel LOUANI	Université de Reims Champagne-Ardenne

Remerciements

En tout premier lieu, je tiens à remercier très vivement mon directeur de thèse le Professeur Michel Broniatowski pour son soutien constant, ses conseils pertinents et ses encouragements qui m'ont permis de mener à bien ce travail.

Je suis reconnaissant au Professeur Djamal Louani de l'Université de Reims Champagne-Ardenne pour avoir accepté de prendre la direction de cette thèse durant cette soutenance.

Je remercie très chaleureusement Madame le Professeur Armelle Guillou et Monsieur le Professeur Belkacem Abdous pour avoir accepté d'être rapporteurs de ma thèse. Je leur suis très reconnaissant du temps qu'ils ont consacré pour évaluer ce travail.

Je remercie l'Institut National sur les Transports et leur Sécurité (INRETS) pour le financement de mes travaux qui sont tant appliqués que théoriques. Je remercie Sylvain Lassarre et Farida Saad d'avoir accepté l'encadrement de cette thèse et pour tous leurs conseils et les efforts dont ils ont fait preuve.

Je remercie Vincent Spenlehauer de m'avoir accueilli au sein du département *Groupe d'Analyse du Risque Routier et de sa Gouvernance* (GARIG).

Je voudrais remercier très chaleureusement Pierre-Alain Hoyau, Cyril Bouvier, Alain Carton, Michel Gueguen, Gérard Leblus, Pierre-Olivier Flavigny, Frédéric Jeanniard et Edgard Mazouth pour leur aide et leur sympathie quotidienne. Je tiens à m'excuser pour tous ceux, plus nombreux, que j'ai oubliés, qu'ils aient toute ma reconnaissance.

Je tiens à remercier tous les membres du L.S.T.A. et notamment le directeur Paul Dehevels pour m'avoir admis en DEA, et les membres du Groupe de travail *Entropie-Divergence* qui, de près ou de loin, ont participé à la réalisation de ce travail.

Je souhaite enfin et surtout remercier toute ma famille pour le soutien qu'elle m'a apporté tout au long de l'élaboration de cette thèse.

Alexandre DEPIRE

*À ma mère et mon père, et à toute ma famille
modeste témoignage de mon infinie tendresse.*

*À ma compagne Belinda, pour son soutien permanent.
À ma fille Anaïs, mon ange*

Notations

Nous adopterons dans l'ensemble de ce texte les notations suivantes

\mathbb{R}	l'ensemble des réels complété par $-\infty$ et $+\infty$
F	une loi de probabilité
$\text{Im } F$	Image par l'application F
$\text{Dom } F$	Domaine de F
$\mathcal{H}(H)$	désigne l'entropie d'une loi H
f	une densité de probabilité associée à une loi F
\mathbf{Z}	un vecteur dans \mathbb{R}^d , $\mathbf{Z} = (Z_1, \dots, Z_d)$
$X \perp\!\!\!\perp Y$	X est indépendant de Y
r_{X_i}	désigne le rang de X_i dans l'échantillon considéré
$\mathbf{1}_{\{A\}}(x)$	désigne la fonction caractéristique de l'ensemble A , prenant 1 si x appartient à A et 0 sinon
$\mathbb{E}_L X$	désigne l'espérance mathématique de X sous la loi L , qui ne sera pas notée s'il n'y a pas ambiguïté
$\text{Cov}(X, Y)$	désigne la covariance entre les variables aléatoires X et Y
\triangleq	désigne une définition
f^{\leftarrow}	désigne l'inverse généralisé de la fonction f
f^{-1}	désigne l'inverse usuel de la fonction f s'il existe.

Conventions typographiques

Concernant les termes mathématiques issus de la langue anglaise, nous avons décidé de ne pas les traduire si leur traduction n'apportait aucune information quant à leur compréhension. De plus leur traduction n'étant pas unique, afin d'éviter un quelconque contre-sens, ils seront le plus souvent tels quels. Nous adoptons la convention typographique de les écrire avec le présent style.

Table des matières

Introduction générale	xix
1 Introduction	1
1.1 Intelligent Transport System (ITS)	1
1.2 Adaptive Cruise Control	4
1.2.1 Histoire - Fonctionnement	4
1.2.2 Études réalisées	6
2 Stratégies de conduite	9
2.1 Données	9
2.2 Réflexion sur l'analyse de l'impact d'un système d'aide en situation réelle	10
2.3 Stratégie de conduite	11
2.4 Résultats empiriques	12
2.4.1 Résultats descriptifs	12
2.4.2 Variation de vitesse - Variation de TIV	14
2.5 Conclusion	18
3 Automates cellulaires	21
3.1 Historique	21
3.2 Cadre des Automates Cellulaires	22
3.2.1 Modèle à file unique	23
3.2.2 Modèle à plusieurs files	24
3.2.3 Distance de sécurité	25
3.2.4 Impact de la distance	27
3.3 Résultats	27
3.4 Conclusion - Extensions envisageables	33
4 Modélisation markovienne cachée - Validation de l'importance du contexte	35
5 Modélisation markovienne avec exogènes	39
5.1 Introduction	39
5.2 Les Modèles Markoviens Cachés (HMM) - Réseau Bayésien Dynamique (DBN)	40
5.3 Les Modèles Markoviens Cachés avec Exogènes (IOHMM)	41
5.3.1 Différence avec HMM	43

5.3.2	Avantage de la modélisation IOHMM	43
5.4	Questions - Problèmes théoriques	44
5.5	Cadre - Hypothèses	45
5.6	Modèle considéré	46
5.7	Étude pratique	46
5.7.1	Cadre	47
5.7.2	Comparaison	49
5.7.3	Interprétation	55
5.7.4	Choix de la variable de contexte	56
5.8	Algorithme de Viterbi	57
5.9	Notion de seuil d'influence	59
5.10	Optique de comparaison	60
5.11	Conclusion	61
6	État de l'art sur les copules et la fonction de linkage - Quelques avancées	63
6.1	Copules	63
6.2	Simulation d'un échantillon bivarié de copule donnée	67
6.3	Intérêt des copules – Dépendance	68
6.4	Dépendance multidimensionnelle	70
6.4.1	Mesures de dépendance	70
6.4.2	Mesures d'association multivariées	72
6.5	La fonction de linkage	74
6.5.1	★ - Lien entre la copule et la fonction de linkage dans le cadre bivarié	77
6.5.2	★ - Lien entre la fonction de linkage et la loi jointe multidimensionnelle	78
6.5.3	★ - Lien entre la loi jointe et la n -copule - Lien entre la n -copule et la fonction de linkage	80
6.6	★ - Une nouvelle copule C^*	81
6.7	★ - Famille de fonctions de linkage	83
6.8	Démonstrations	83
7	Lien entre la transformation du linkage et l'information mutuelle	89
7.1	Lien entre l'Information mutuelle et la Copule	89
7.2	Information mutuelle dans le cas multidimensionnel	90
7.3	Fonction de linkage	91
7.4	Remarques	93
7.5	Preuves	93
8	Bivariate Cox models and dependence analysis	101
8.1	Some model of dependence for duration data	101
8.2	Cox models	101
8.2.1	Univariate Cox models	101
8.2.2	Bivariate Cox models	103
8.2.3	Model based on $\lambda_{Y X=x}$	104
8.2.4	Statistical estimation	105
8.3	Some useful facts in bivariate dependence	106

8.3.1	Max-id distribution - Min-id distribution - Positive Quadrant Dependence (PQD)	106
8.3.2	Survival Copulas	109
8.3.3	Copulas - Max-infinitely divisible Distribution	109
8.3.4	Extreme-value Copulas	110
8.4	Approximation of a PQD copula by Extreme-Value Copula	113
8.5	Coherence	116
8.6	Covariates and estimation	120
8.7	Application - Regression of Spearman's rho on covariates	123
8.8	A case study	124
8.9	Detailed proofs	126
8.9.1	Approximation of a PQD copula by Extreme-Value Copula	126
8.9.2	Study of the estimator	128
9	Estimation of a dependence function	133
9.1	Notations	133
9.2	The form of the estimate and its limit properties	134
9.3	Asymptotic Law of the process $t \rightarrow \sqrt{n} [A_n(s) - A(s)]$	136
9.4	Comparison with other estimators	138
9.5	Discussion	141
	Graphiques	i
	Input-Output Hidden Markov Model	v
.1	Main Results	v
	Package R - Modèles de Cox	xiii
.2	Programmes principaux	xiii
.2.1	Simulation d'un échantillon suivant le modèle de Cox - Simulation.R	xiv
.2.2	Estimation des coefficients marginaux du modèle de Cox - Estimation.R	xxv
.2.3	Prevision du rho spearman - Prevision spearman.R	xxix
.3	Librairies C	xxxvii
.3.1	Copule empirique - copuleempirique.c	xxxvii
.3.2	Estimation des paramètres marginaux du modèle de Cox - implémentation de la méthode d'Efron au cas d'égalité	xxxix
.3.3	Inversion d'une fonction - inv.c	xlii
.3.4	Copule de survie empirique - scopuleempirique.c	xliii
	Package R - Estimation de la fonction de dépendance d'une copule bivariable de valeurs extrêmes	xlvi
.4	Estimation de la fonction de dépendance dans le cadre non paramétrique - estimation A test.R	xlvi

Table des figures

1.1	Famille d'aides à la conduite	2
1.2	Peloton de voitures ayant le système AHS	4
1.3	Principe du régulateur intelligent de vitesse (ACC)	5
1.4	Performance en fonction de la charge de travail	6
1.5	Différentes définitions du terme <i>Évaluation</i>	7
2.1	Schéma explicatif pour la variable <i>Contrainte</i>	10
2.2	Distribution des temps intervéhiculaires	12
2.3	Densité de probabilité de l'interdistance	13
2.4	Distribution des interdistances en situation de suivi	17
3.1	Impact du coefficient β	28
3.2	Diagramme temps-espace - coefficient de sécurité, $\alpha = 1$	29
3.3	Diagramme temps-espace - coefficient de sécurité, $\alpha = 0,3$	29
3.4	Apparition du bouchon - $\alpha = 0.8$ - phénomène de recul	30
4.1	Variable <i>Contrainte</i>	36
5.1	Schématization d'un modèle markovien caché	41
5.2	Schématization d'un modèle markovien d'ordre 1 caché avec prise en compte d'exogène	42
5.3	Vitesse du conducteur 2	47
5.4	Vitesse du conducteur 3 - bornes des modélisation HMM et IOHMM sur l'intervalle $[1 : 300]$	50
5.5	Comparaison simplifiée HMM - IOHMM	51
5.6	Comparaison HMM (6) et IOHMM (2x3)	52
5.7	Chaînes cachées (1)	54
5.8	Chaînes cachées (2)	54
5.9	Chaînes cachées - modèle HMM et IOHMM	55
8.1	Parametrization de I^2	113
8.2	Oscillation	116
8.3	Illustration of formula (8.16)	121
8.4	Two estimators of B^z	122
8.5	Forecasted dependence functions	123
8.6	Forecasted Spearman's rho	124

8.7	The LAVIA system	125
8.8	Data X, Y	126
8.9	Parametrisation of $[0; 1]^2$	129
9.1	Gumbel marginals - Gumbel-Hougaard copula ($\theta = 1.5$)	139
9.2	Non-extreme value margins - Gumbel-Hougaard copula ($\theta = 1.5$)	140
9.3	Gumbel-Hougaard copula with $\theta = 4$	141
9.4	Galambos copula with $\theta = 4 - (A(t) = 1 - [t^{-\theta} + (1 - t)^{-\theta}]^{-1/\theta})$ $C(u, v) =$ $uv \exp \left\{ [(-\log u)^{(-\theta)} + (-\log v)^{(-\theta)}]^{(-1/\theta)} \right\}$	141
5	Allure des vitesses pratiquées par les conducteurs	ii
6	Densité des vitesses de chaque conducteurs	iii
7	Distribution des temps intervéhiculaires (s)	iv

Liste des tableaux

2.1	Variation des indicateurs	14
2.2	Test d'égalité de la moyenne ($m = 0$)	15
2.3	Variance des variations relatives de vitesse et d'interdistance	16
2.4	Temps intervéhiculaire moyen en phase de suivi	18
3.1	Paramètres estimés des modèles d'automates cellulaires	31
4.1	Pourcentage d'occurrence maximal entre la chaîne cachée et la variable <i>Contexte</i>	36
4.2	Paramètres du modèle markovien obtenu (les valeurs ont été arrondies à l'unité près)	37
5.1	Comparaison de modèles (en terme de loi d'émission) pour le conducteur 2	48
5.2	Comparaison HMM et IOHMM	50
5.3	Paramètres du modèle HMM (conducteur 2)	52
5.4	Paramètres du modèle IOHMM (conducteur 2)	53
5.5	Comparaison des chaînes cachées (HMM et IOHMM)	53
5.6	Répartition du temps dans chaque phase	56
5.7	Comparaison HMM et IOHMM (variable d'input : file)	57
5.8	Seuil d'influence	59
8.1	Test of proportional hazard rates	126
9.1	$\sup \left \hat{A}_*(s) - A(s) \right $ - case of extreme-value margins	139
9.2	$\sup \left \hat{A}_*(s) - A(s) \right $ - case of non-extreme value margins	140

Introduction générale

L'Institut National de Recherche sur les Transports et leur Sécurité (INRETS) créé en 1985 de la fusion de l'Institut de Recherche sur les Transports (IRT) et de l'Observatoire National sur la Sécurité Routière (ONSER) est un Établissement Public à caractère Scientifique et Technologique (EPST), placé sous la tutelle des ministères chargés de la recherche et des transports.

Cet institut a pour mission

- Rassembler et développer les connaissances scientifiques par une approche globale du système des transports.
- Assurer des missions d'expertise et de conseil.
- Contribuer à l'effort général de l'enseignement supérieur.
- Valoriser les résultats de recherche dans le cadre de partenariats économiques.
- Diffuser les connaissances acquises auprès d'un large public.

J'ai effectué ma thèse au sein du Département Évaluation et Recherche en Accidentologie (DERA) devenu l'unité de recherche nommée Groupe d'Analyse du Risque routier et de sa Gouvernance (GARIG) en 2005, encadré par Sylvain LASSARRE et Farida SAAD. Ce travail a été mené sous la direction du Professeur Michel BRONIATOWSKI initialement à l'Université de Reims Champagne-Ardenne puis au sein du Laboratoire de Statistiques Théorique et Appliquées (LSTA) de l'Université Pierre et Marie Curie (Paris VI) et je tiens à remercier le Professeur Djamel LOUANI de l'Université de Reims Champagne-Ardenne pour son aide administrative et scientifique.

Cette thèse s'organise en deux grandes parties.

La première partie s'articule en quatre chapitres et porte sur le système d'aide à la conduite ACC (Automated Cruise Control). Tout d'abord, une analyse portant sur les stratégies est présentée. Nous étudions les stratégies en terme de vitesse et de temps inter-véhiculaire et nous confirmons l'hypothèse de *F. Saad* que les conducteurs privilégient un maintien de leur vitesse au détriment d'une réduction du temps inter-véhiculaire, (Saad et Villame, 1996).

Le second chapitre aborde une approche assez récente dans les domaines de la modélisation routière; il s'agit des automates cellulaires. Initialement utilisés dans la Physique, cette notion, comme nous le montrons, présente de nombreux avantages dans la modélisation

du processus vitesse. Une extension de ces modèles est également exposée.

Le troisième chapitre propose l'utilisation concrète de la modélisation markovienne cachée dans le cadre de la validation d'une hypothèse dans le domaine de la psychologie de la conduite, issue de Saad et Villame (1999). Cette hypothèse basée sur la contextualisation de la tâche de conduite nécessite l'utilisation de l'algorithme de Viterbi (exposé au chapitre 4) qui permet d'obtenir la réalisation la plus probable de la chaîne cachée du modèle de Markov proposé.

Le dernier chapitre de cette partie porte sur une extension de la modélisation markovienne usuelle. La modélisation IOHMM (Input-Output Hidden Markov Model) introduite par Bengio et Frasconi (1995) est très récente. Dans notre cadre, nous comparons les apports de cette méthode, avec comme variable d'entrée (input) la contextualisation évoquée au chapitre 4, avec une modélisation cachée classique. D'autre part, un cadre théorique est brièvement présenté et une piste pour démontrer les aspects théoriques est donnée en annexe, voir annexe 9.5.

L'objet central de cette thèse est l'étude et la mise en place d'outils pour l'évaluation d'un système d'aide à la conduite par le biais d'une comparaison avec/sans et cela sur des données réelles non contrôlées. L'étude des effets d'un système dans des contextes non comparables, nous a amené à considérer la notion de scène. La conduite peut être vue comme une succession d'action ayant un début et une fin ; ces actions pouvant être brutalement interrompue par la survenue d'une autre, tel un dépassement succédant à une conduite en situation de suivi. Ceci définit la notion de scènes. Il est apparu que la scène peut se définir comme la modification de la dépendance entre plusieurs indicateurs, telle la vitesse et l'interdistance par exemple. Par conséquent, nous nous sommes intéressés à la notion de dépendance dont l'objet central est la copule.

La seconde partie a pour objet d'étude la copule. Cette notion apparue au début du siècle par le biais de Fréchet puis reprise en 1959 par Sklar s'est vu grandement popularisée dans les domaines universitaires sous la poussée de champs d'application tels que la Finance ou l'Assurance. Cette partie s'articule sur quatre chapitres.

Le premier chapitre constitue un état de l'art sur les copules et les mesures de dépendance, mais présente également une notion très récente, la fonction de linkage. Cette notion est fortement liée à celle de la copule comme nous le montrerons et peut être considérée comme une extension de cette dernière. Des résultats classiques mais d'autres novateurs y sont exposés.

Le second chapitre, s'inscrivant dans la suite du premier, aborde une mesure de dépendance multivariée, l'information mutuelle. Nous démontrons une formule de décomposition de l'information mutuelle où apparaît exclusivement les fonctions de linkage des vecteurs aléatoires considérés.

Le troisième chapitre de cette seconde partie expose une extension du modèle de Cox au cas bivarié. Ce développement s'appuie sur la classe des copules de valeurs extrêmes. De multiples notions telle que les ordres stochastiques en multivarié ont été abordées et exhibées comme étant naturelles dans le problème initial du modèle de Cox. Des résultats

d'approximation et de convergence sont prouvés. Pour conclure, une application de l'extension du modèle de Cox bivarié est présentée sur les données du système LAVIA (limiteur de vitesse s'adaptant à la vitesse autorisée).

Le quatrième chapitre de cette partie porte l'estimation de la fonction de dépendance d'une copule bivariée. Cet objet est crucial dans le domaine des valeurs extrêmes bivariées. L'extension du modèle de Cox au cas bivarié a permis d'obtenir un estimateur extrêmement simple de cette fonction. Ce nouvel estimateur présente de multiples avantages, le premier et le plus important est qu'il permet de s'affranchir de la connaissances des marges du phénomène bivarié observé. Cette hypothèse était cruciale dans les estimateurs précédents proposés par Tiago de Oliveira, Peter Hall ou encore Paul Deheuvels par exemple. Le second avantage est sa simplicité puisqu'il s'agit d'un estimateur plug-in calculé à partir de la copule empirique, objet introduit par Deheuvels en 1979. Cette simplicité a permis d'obtenir la convergence uniforme ainsi que la loi asymptotique de notre estimateur. Une comparaison de la variance limite a même permis de conclure que ce nouvel estimateur est meilleur, sous cet angle.

En annexe, sont donnés les codes informatiques du package R développés dans le cadre des chapitres 8 et 9 portant sur le modèle de Cox et l'estimation non paramétrique de la fonction de dépendance dans le domaine des valeurs extrêmes bivariées.

Chapitre 1

Introduction

Le comportement du conducteur a été étudié depuis le début des années 1950, mais durant les années 1990 l'intérêt pour ce sujet a considérablement augmenté. Tous les aspects de la conduite sont étudiés : la modélisation de la vitesse, du changement de file, des modèles psychologiques, avec les concepts de but et de mode de réalisation. Avec la compréhension accrue de la conduite, est apparue la notion de charge de travail du conducteur. Son étude a abouti à l'idée naturelle de diminuer celle-ci, pour plusieurs raisons ; on peut citer l'amélioration du confort, la réduction du nombre et/ou de la gravité des accidents. La notion d'assistance à la conduite trouve son origine dans les programmes européens EUREKA, PROMETHEUS et CARMINAT qui se sont déroulés de 1987 à 1994. Cette notion est née de la constatation que, dans un grand nombre de situations accidentogènes, un gain d'une seconde voire d'une fraction de seconde dans la juste compréhension d'une situation par une information adéquate ou bien l'amorce d'une réaction par un automatisme approprié, pouvait faire la différence entre accident et non-accident ou encore engendrer une réduction de la gravité de l'accident. Naturellement, cela a mené au développement de l'automatisation de la conduite automobile. La notion de systèmes intelligents de transports – sur le plan international, on utilise le sigle « ITS » pour *Intelligent Transport System* et en Europe, on parle d'« ADASE » pour *Advanced Driver Assistance Systems in Europe* – est alors apparue, concept sous lequel est regroupé tout système d'aide à la conduite informatisé.

Pour éviter toute erreur quant à la traduction des termes spécifiques au domaine des transports, nous conserverons les termes anglais dans l'ensemble de nos propos.

1.1 Intelligent Transport System (ITS)

On retient communément comme définition des « ITS » l'ensemble des applications (matériels, logiciels, services, effets d'organisation) découlant des progrès des technologies suivantes :

- informatique ;
 - télécommunications ; l'ensemble informatique et télécommunications a récemment pris une dimension nouvelle ; on parle de « nouvelles technologies de l'information et de la communication » NTIC ;
 - services satellitaires (GPS) ;
-

- automatismes de tous genres appliqués aux véhicules automobiles : la gestion des moteurs est le point d'entrée de l'électronique le plus ancien.

Pour bien situer et comprendre ce domaine de recherche, il convient de préciser quelques points :

- deux voies distinctes d'automatisation sont empruntées dans le domaine des ITS :
agrégation de technologies ;
création de grands centres de gestion du trafic.
- différence importante entre USA-Japon et Europe ;
- l'automatisation absolue et centralisée (à l'image du système AHS américain, fig. 1.2) n'est pas la voie désirée par tous.

Les différences entre la diffusion des systèmes d'aide à la conduite entre l'Europe et les USA/Japon apparaissent comme majeures. On peut constater le faible taux de pénétration de tels systèmes en Europe alors qu'ils sont déjà très utilisés ailleurs, pour les systèmes les plus basiques (Cruise control, parfois appelé passive cruise control¹), ou en passe d'être introduits pour les systèmes plus automatisés (ACC ou freinage automatique), Orselli et Chanaron (1999). Nous allons présenter à présent l'optique européenne, les systèmes de la famille ADASE.

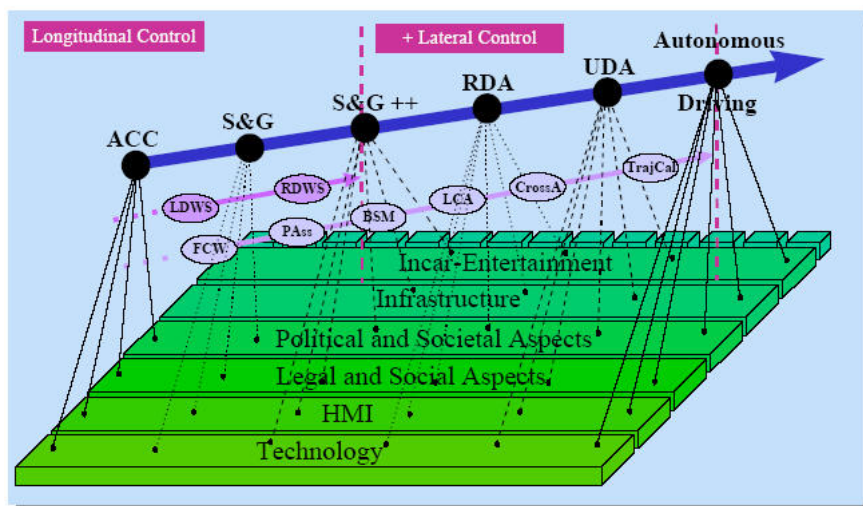


FIG. 1.1 – Famille d'aides à la conduite²

¹ITS American News, Vol. 13, No. 8, Août 2003.

²Source : http://www.aida.utwente.nl/Onderzoek/onderzoekpublicaties/arem_soeteman.pdf

- LDWS : Lane Departure Warning System
- RDWS : Road Departure Warning System
- FCW : Front Collision Warning
- PAss : Parking Assistant
- BSM : Blind Spot Monitoring
- LCA : Lane Changing/Merging Assistance
- CrossA : Crossing Assistance
- TrajCal : Trajectory Calculation

La Figure 1.1 montre l'évolution envisagée par les spécialistes du domaine en matière d'aides à la conduite, Carsten (2003). La ligne bleue montre la direction principale envisagée, et les produits dérivés sont indiqués en rose dessous. On peut voir deux grandes familles se dessiner : les systèmes d'aide longitudinale, les systèmes d'aide latérale, avec un futur commun, la conduite automatisée. Dans la catégorie des aides latérales, les systèmes RDA (Rural Drive Assistance) et UDA (Urban Drive Assistance) sont des projets aidant le conducteur à contrôler la direction de son véhicule. Il s'agit des concepts envisagés pour le futur. Les contrôles longitudinaux sont plus anciens ; le premier est l'ACC. Ce système est très adapté à des conduites sur autoroutes. Il gère la vitesse et l'interdistance. Mais dans des conditions de trafic dense, il peut s'avérer dangereux, dû en particulier aux nombreuses insertions qui réduisent brutalement l'interdistance. En outre, le système se désactive automatiquement si la vitesse passe en dessous de 30 km/h, ce qui rend inutilisable son utilisation dans le milieu urbain ou en situation de forte congestion comme les bouchons. Pour pallier ce problème, ce dispositif a été couplé avec le système Stop&Go. Ce dernier est basé sur une technique de *lissage* du trafic sur voie rapide permettant d'anticiper et de coordonner localement, véhicule par véhicule, les mouvements d'arrêt et de redémarrage. Les objectifs de gain de capacité à la congestion sont très importants ; en outre il permet des démarrages plus rapides.

Le contrôle latéral est en parallèle développé et une évolution mixant les deux approches est envisagée, sous le nom S&G++. Ce dernier couple à la fois l'ACC+S&Go et un contrôle latéral (*Lane Keeping*). L'introduction des premiers systèmes (ACC, S&G) est effective dans les modèles haut de gamme de Mercedes-Benz (Système Distronic et Tempomat), mais aussi chez BMW, Prestl et al. (2000). Les prévisions pour la mise en place des systèmes de seconde génération (S&G++) dans la voiture de *monsieur tout le monde* tablent sur 2010.

En parallèle de telles recherches, certains étudient directement l'objectif final affiché de la conduite autonome. Cette optique se distingue de la précédente ; on quitte le principe d'agrégation de systèmes. Cet axe est principalement occupé par les USA et le Japon, en particulier les premiers avec leur projet d'AHS (Automated Highway System) ; le Japon a lui eu deux types d'actions, d'une part la création d'un programme AHS propre et d'autre part une participation au projet américain.

Une telle automatisation n'a pas été accueillie avec un fort élan, pour preuve les arrêts récents de ces projets (en réalité, l'arrêt du projet américain a entraîné l'arrêt de son homologue japonais) en 1997. De multiples raisons sont également évoquées ; on peut citer l'absence de gain en terme de sécurité (Orselli et Chanaron (1999)), mais aussi des luttes politiques. On pourra se reporter à Ygnace et Benouar (1998) pour de plus amples informations. La volonté est en effet plutôt la réalisation de systèmes informatiques d'aides à la conduite que la construction de réseaux dédiés (voir figure 1.2).

On peut consulter le site de IVsource (site d'information relative au concept de véhicule intelligent)³ pour obtenir le planning envisagé pour le déploiement de tels systèmes.

³IVsource : http://www.ivsource.net/archivep/2002/feb/020203_AHSeditorial.html



FIG. 1.2 – Peloton de voitures ayant le système AHS

1.2 Adaptive Cruise Control

Après avoir présenté les différents systèmes d'aide à la conduite, nous allons examiner de près l'Adaptive Cruise Control. Ce système est en soi une révolution ; il s'agit de substituer en partie le conducteur dans les phases de conduite en file. Révolution car c'est la première fois où l'action de conduite est gérée par un système informatisé. Au conducteur, incombe la tâche ensuite, de diriger le véhicule latéralement (gestion de la direction), de surveiller le fonctionnement du système et éventuellement de reprendre les commandes en cas d'urgence. Il s'agit d'un premier pas d'une évolution dont les plans ont déjà été tracés comme nous l'avons vu précédemment. On rencontre plusieurs appellations pour ce système :

- ACC : Adaptive Cruise Control ;
- ICC : Intelligent Cruise Control ;
- AICC : Autonomous Intelligent Cruise Control.

On utilise l'adjectif d'autonome pour faire référence à l'absence de dispositifs dans les véhicules autres ; le dispositif n'a besoin d'être implanté que dans son propre véhicule, il est donc autonome. Dans la suite, on utilisera l'unique acronyme ACC pour l'ensemble de ces sigles.

1.2.1 Histoire - Fonctionnement

Amélioration du système de Ralph Teetor, dit passif, l'ACC permet de réguler automatiquement la vitesse tout en maintenant une interdistance minimale, choisie également par le conducteur. Cette interdistance est en terme de secondes et non de mètres ; on introduit alors la notion de temps intervéhiculaire (TIV). En l'absence de véhicule précédent, le système maintient une vitesse préalablement choisie par le conducteur. Avant de pour-

suivre, précisons deux notions qui nous seront utiles dans la suite. Le temps intervéhiculaire (TIV) est défini comme suit :

$$TIV_t = \frac{D_t}{V_t}$$

où D_t désigne l'interdistance avec le véhicule précédent, V_t désigne sa vitesse absolue ; le temps intervéhiculaire s'exprime en secondes. On parle aussi de temps avant collision (TTC ou Time-to-collision), cet indicateur est défini par :

$$TTC_t = \frac{D_t}{\Delta V_t}$$

où ΔV_t désigne la vitesse relative entre les deux véhicules, $\Delta V_t = V_t - V_t^{precedent}$. Le régulateur intelligent de vitesse gère par le biais d'un radar (voir Figure 3) l'interdistance.

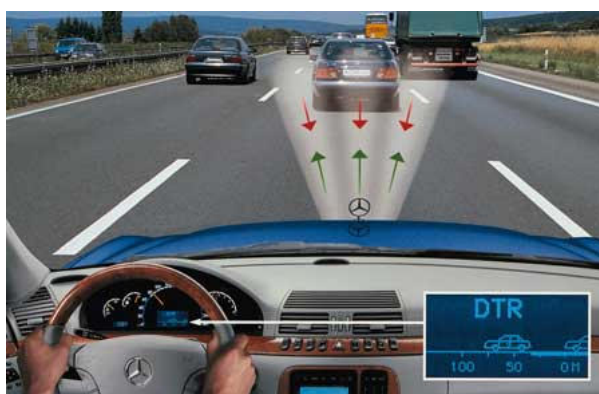


FIG. 1.3 – Principe du régulateur intelligent de vitesse (ACC)⁴

Le système contrôle la puissance du moteur, par le biais de l'arrivée de carburant ; dans le cas où la régulation n'est pas suffisante, il peut agir sur les freins du véhicule. On note qu'il existe plusieurs types de régulateurs intelligents de vitesse, différence se tenant principalement au niveau de la technologie choisie (gestion de la commande d'accélération ou de l'arrivée de carburant par exemple). Le conducteur peut reprendre à tout moment le contrôle du véhicule par pression sur la pédale d'accélérateur ou de frein.

L'introduction d'un tel système est motivée par des gains envisageables et cela dans des domaines divers, ainsi on peut s'attendre à :

- une réduction du nombre d'accidents par le maintien d'une interdistance minimale ;
- une diminution de la fatigue (Kopf (2002)) ;
- l'augmentation du rendement en diminuant les changements de régimes cause d'une plus grande consommation ;
- l'augmentation de la performance du conducteur.

L'objectif des systèmes d'aide à la conduite de manière générale est indirectement d'augmenter la performance du conducteur. En effet, dans des situations ne nécessitant pas une très grande attention (conduite en situation libre), le système ne revêt qu'un aspect de

⁴Source : <http://www.auto-innovations.com>

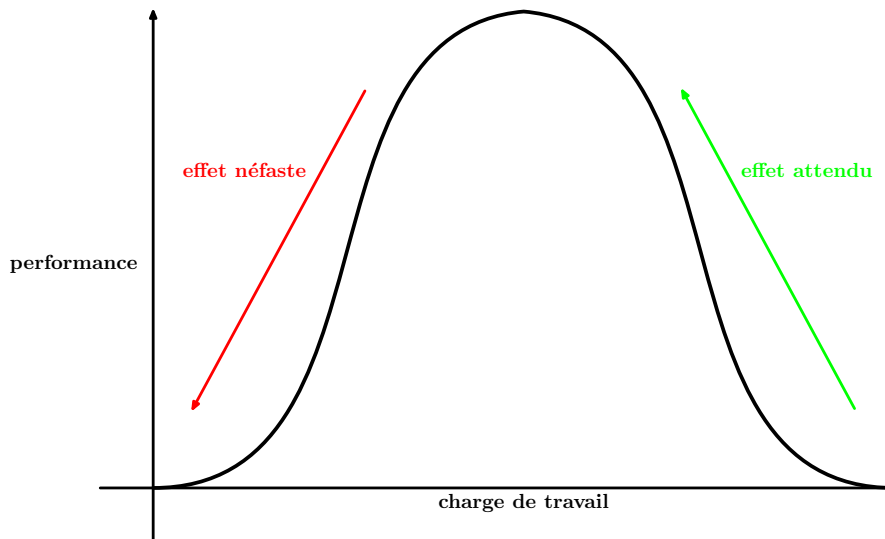


FIG. 1.4 – Performance en fonction de la charge de travail

confort en maintenant la vitesse désirée. On retrouve l'aspect premier du système, le cruise control. Par contre, dans des situations nécessitant une plus grande vigilance de la part du conducteur (conduite en file), le système permet de réduire la charge et donc permet de maintenir une performance correcte. On entend ici par charge, la charge mentale du conducteur. Ce fait est illustré par la flèche verte sur la Figure 4. On doit aussi signaler que la mise en place d'un tel système peut possiblement engendrer des effets néfastes, illustrés par la flèche rouge. Le conducteur étant assisté dans sa tâche sa vigilance tend à diminuer. Cet effet est simplement évoqué ici, mais n'est pas traité.

1.2.2 Études réalisées

De nombreuses recherches ont pour objet d'étude l'ACC, on peut distinguer plusieurs types de recherches, liés aux objectifs envisagés :

- acceptabilité du système de la part du conducteur ;
- amélioration de la sécurité routière, par une diminution du nombre d'accidents ou de leur gravité ;
- homogénéisation des vitesses pratiquées ;
- normalisation des distances intervéhiculaires avec plausiblement une réduction des distances très faible ;
- interaction avec les véhicules ne disposant pas de l'ACC ;
- diminution des émissions de polluants par une réduction des variations de vitesse ;
- augmentation de la capacité des voies de circulation par une homogénéisation des vitesses et création d'un régime dit synchronisé du trafic.

De multiples études ont été menées sur ce système, en simulateur ou situation réelle avec des objectifs divers, parmi ceux cités ci-dessus. On peut signaler que sur le plan de

la sécurité, peu de résultats existent. Une étude menée par l'Université du Michigan a abouti à des résultats peu convaincants; les auteurs précisent que si gain il y a, il est faible en regard des risques nouveaux créés (Fancher et Ervin (1998)). (Carsten (2000)) pose le problème en s'interrogeant sur notre capacité à savoir si de tels systèmes sont sûrs et ce même auteur (Carsten (2003)) cite une diminution de 20% des accidents sur autoroutes dus à des collisions frontales, soit environ 0,8% des accidents, au sens là, d'accidents corporels; mais il ne s'agit que de prévisions, aucun résultat empirique n'est venu confirmer ce point. En outre, des expériences menées en simulateur montrent que l'introduction de l'ACC peut engendrer des accidents (Nilsson (1995)) et augmenter les situations de risques; pour preuve on a constaté une diminution significative des TTC (Time to Collision) dans une expérience sur le simulateur de HUSAT (Richardson et al. (1996)). En outre l'introduction d'un tel système peut réduire l'attention du conducteur; il est moins sollicité et alors la charge mentale réclamée étant trop faible, la performance décroît, point pouvant avoir de multiples conséquences, voir (Endsley (1996)).

Un point reste encore à analyser, il s'agit de la modification du comportement de l'automobiliste. Cette question est centrale mais difficile. L'OECD (Organisation for Economic Cooperation and Development) définit le concept de « behavioural adaptation » par les termes suivants :

Les adaptations comportementales sont des comportements qui apparaissent suite à la modification du système utilisateur-véhicule-route et qui n'ont pas été prévus par les initiateurs de la modification. OECD (1990)

Une expérience sur le simulateur de Groningen montre que les conducteurs rapides (voir West et al. (1992)) ont une tendance accrue à *naviguer* latéralement avec l'ACC dans des conditions de trafic peu dense (Hoedemaeker et Brookhuis (1999)). En outre, les conducteurs lents sont enclins à diminuer leur SDLP (Standard Deviation Lateral Position). On voit là un exemple d'une modification de comportement. La modification de comportement peut aussi s'opérer en terme de stratégie de conduite que nous allons aborder dans la partie suivante.

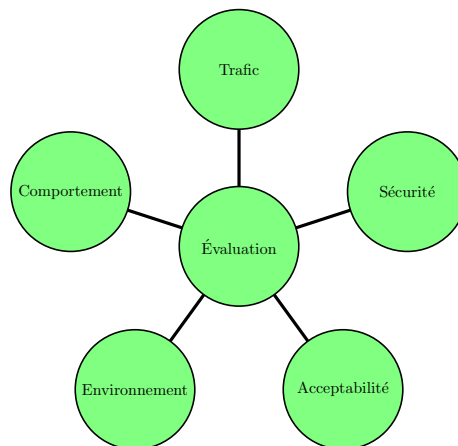


FIG. 1.5 – Différentes définitions du terme *Évaluation*

Chapitre 2

Stratégies de conduite

2.1 Données

Les données sont issues d'une expérience mise en œuvre par Farida Saad (INRETS - Laboratoire de Psychologie de la Conduite) et Thérèse Villame (RENAULT) en 1994. Neuf conducteurs ont participé (5 hommes et 4 femmes) à cette recherche. Ils sont tous expérimentés (plus de 5 ans de permis, plus de 20 000 km annuels et plus de 200 000 km au total) et familiers de la conduite sur autoroute. Après une phase de familiarisation avec le système, les conducteurs ont tous réalisé un même parcours autoroutier, situé aux environs de Paris. Nous n'utiliserons ici que les données issues des premiers 90 km réalisés sans l'emploi du système ACC. La seconde partie n'étant pas à notre disposition pour l'instant. Les données recueillies sont de plusieurs types : des mesures cinématiques, des mesures du contexte et des verbalisations. Ces dernières ne seront pas exploitées ici. Une mesure est effectuée chaque seconde. Pour les données cinématiques, on donne ici un descriptif de la variable ainsi que son code

- la vitesse du véhicule, Vveh ;
- la vitesse du véhicule cible, Vcib ;
- les manœuvres du sujet, ManS ;
- les manœuvres du véhicule cible, ManA ;
- l'action sur le frein, Frein ;
- la file empruntée, File ;
- le dépassement d'autres usagers par le sujet, SD ;
- le dépassement du sujet par d'autres usagers, DS ;
- l'interdistance, DistCib.

On dispose également d'indicateurs de trafic :

- la densité du trafic devant le sujet, TAV ;
 - la densité du trafic à l'arrière, sur la file du sujet, TAF ;
 - la densité du trafic à l'arrière, sur les voies adjacentes du sujet, TAA ;
 - le niveau de contrainte, Contrainte ;
 - le type de conduite, TyCdte ;
 - le nombre de voies totales, NbVoie ;
 - le nombre de voies praticables, Prat.
-

On va décrire que les quelques variables qui seront utilisées dans cette étude. Les variables telles que la vitesse, l'interdistance se comprennent, mais nous allons utiliser la variable *Contrainte*, qu'il convient d'expliquer. Cette variable peut prendre 3 valeurs distinctes :

- 1 : il n'y a aucune voiture devant le véhicule sur sa propre file, il peut y avoir d'autres usagers sur les voies adjacentes ;
- 2 : un véhicule se trouve sur sa propre voie, mais au moins une voie devant lui est libre ;
- 3 : toutes les voies devant le conducteur sont occupées.

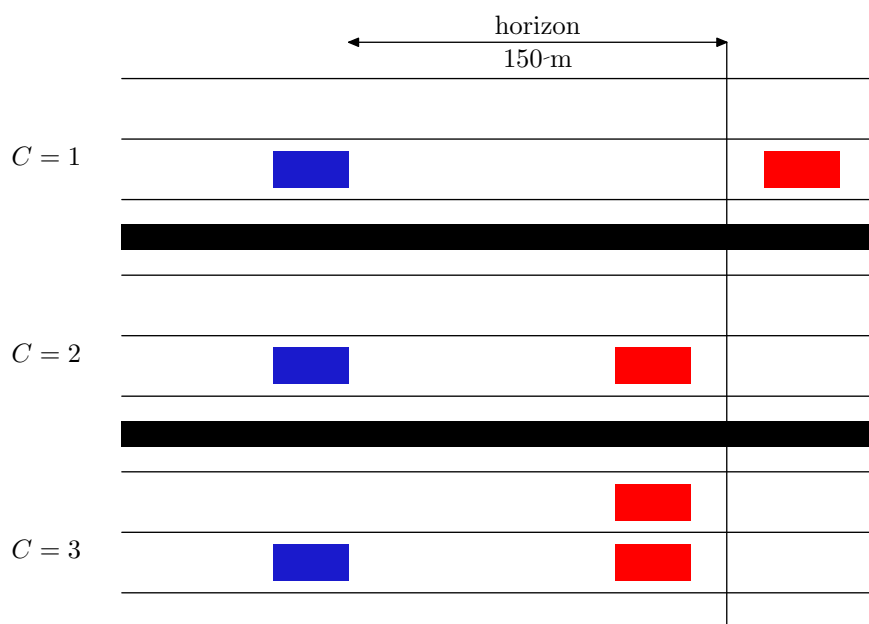


FIG. 2.1 – Schéma explicatif pour la variable *Contrainte*

2.2 Réflexion sur l'analyse de l'impact d'un système d'aide en situation réelle

La richesse de cette étude repose sur la multitude d'indicateurs recueillies mais aussi sur le fait qu'il s'agit d'une expérience réalisée dans un contexte réel, par opposition à des expériences réalisées en simulateur. Ce point est crucial, hormis la remise en cause de la véracité de résultats obtenus lors d'expérimentations sur simulateur ; on dispose d'un cadre précis pour une évaluation, en fait on dispose au sens littéral de données non contrôlées. De manière générale, l'approche retenue pour évaluer un système est l'approche avec/sans. Il s'agit de l'approche la plus naturelle, mais dans le cadre des aides à la conduite, cela revêt une connotation précise. Il est important de s'attarder sur ce point. Dans l'optique de l'évaluation d'un système, on réalise des mesures relatives au système, en l'occurrence vitesse, interdistance, activité du frein, etc, et cela avec et sans le système afin de comparer les différentes caractéristiques de ces mesures (moyenne, variance, loi de

probabilité, etc). Cela repose sur une hypothèse cruciale afin de pouvoir conclure. En effet, une différence de caractéristiques entre les deux essais incombe naturellement au système mis en place ; cela suppose que tout changement n'est dû qu'au système et à rien d'autre. Ce mécanisme hypothético-déductif est applicable dans le cadre d'un simulateur où l'on peut reproduire la configuration à l'identique avec et sans le système, et par conséquent en tirer une conclusion sur l'ACC. Dans le cadre d'expériences dans le monde réel, il en va tout autrement, il convient avant de conclure de s'assurer que la différence n'incombe pas à une autre grandeur, qui aurait évolué entre les deux phases, grandeurs non contrôlables dans le monde réel. Une grande partie de la difficulté de l'analyse de l'impact d'un système d'aide à la conduite provient de cette considération. En pratique, on essaie de réduire au maximum cet écueil en réalisant les expériences dans des conditions les plus voisines possibles. Malgré tout, quelques grandeurs semblent échapper à cette précaution ; il s'agit en particulier du trafic, et de manière générale du contexte local.

2.3 Stratégie de conduite

Le système de régulation automatique de vitesse de croisière a pour objectif d'atteindre et de maintenir une vitesse de croisière et le cas échéant de conserver une interdistance minimale avec le véhicule précédent. Clairement, le système est réellement actif en situation de conduite en file. Lors d'un changement de configuration de l'horizon de conduite (insertion ou fuite d'un véhicule), le système provoque un changement d'allures soit en diminuant la vitesse (cas d'insertion) soit en augmentant cette dernière, si on se trouve en deçà de la vitesse de croisière choisie, en cas de fuite du véhicule précédent. Cela signifie que le système a une stratégie ; on parle de stratégie de conduite.

On peut donner des stratégies de conduite sommaire, on peut désirer :

- maintenir sa vitesse ;
- maintenir son interdistance ;
- maintenir son temps intervéhiculaire constant.

Plusieurs écoles s'opposent sur cette question. Le maintien de la vitesse au prix d'interdistances faibles, si ces dernières ont lieu pendant un laps de temps court est une vision rencontrée, voir Monseur et Malaterre (1970). On parle de maintien d'une allure préférentielle, ou *vitesse de base*. C'est une motivation très fortement exprimée par les conducteurs et fréquemment évoquée dans les entretiens comme justification des choix opérés au cours de la conduite (choix de la file de circulation, changement de file), Saad et Schnetzler (1994).

Le maintien du temps intervéhiculaire est une stratégie plus récente ; elle est baptisée CTH (Constant Time Headway). Cette stratégie établit que la vitesse désirée doit être proportionnelle à l'interdistance. Cette idée est intéressante car on a observé que la conduite naturelle tend à diminuer la vitesse lors d'une diminution de l'interdistance. Certes, ceci apparaît comme évident, mais c'est un point clé dans les simulations, en particulier dans les règles régissant le comportement du véhicule. Plus récemment, cette vision est remise en cause par la stratégie nommée VTH (Variable Time Headway), voir Zou et Levinson (2000). On doit noter que certaines recherches présente le système ACC comme un système ayant pour *stratégie* le maintien du temps intervéhiculaire. Cette optique ne

semble pas réellement adaptée. En effet, la technologie tient compte de la vitesse relative également, ainsi lors de l'insertion d'un véhicule plus rapide (hormis dans le cas d'une situation critique), le système ne va pas réduire la vitesse car il déterminera l'absence de risque malgré une modification du temps intervéhiculaire. Par conséquent, il semble que cette approche soit mise en défaut par cet exemple.

A chaque instant, le conducteur peut décider soit de modifier sa vitesse afin de maintenir une interdistance, soit de maintenir son allure en faisant varier son interdistance. Une interrogation naturelle suit, c'est l'étude de la stratégie des conducteurs. Nous allons tenter de répondre en nous appuyant sur nos données, à deux questions :

- y a-t-il une stratégie unique et prédominante adoptée par les conducteurs ?
- si oui, laquelle ?

2.4 Résultats empiriques

2.4.1 Résultats descriptifs

Dans cette partie, nous allons étudier simultanément les variations du temps intervéhiculaire et de la vitesse. En amont, on se place dans des situations de suivi, afin d'avoir la possibilité de calculer le TIV.

Nous allons donner quelques exemples graphiques caractérisant les conducteurs.

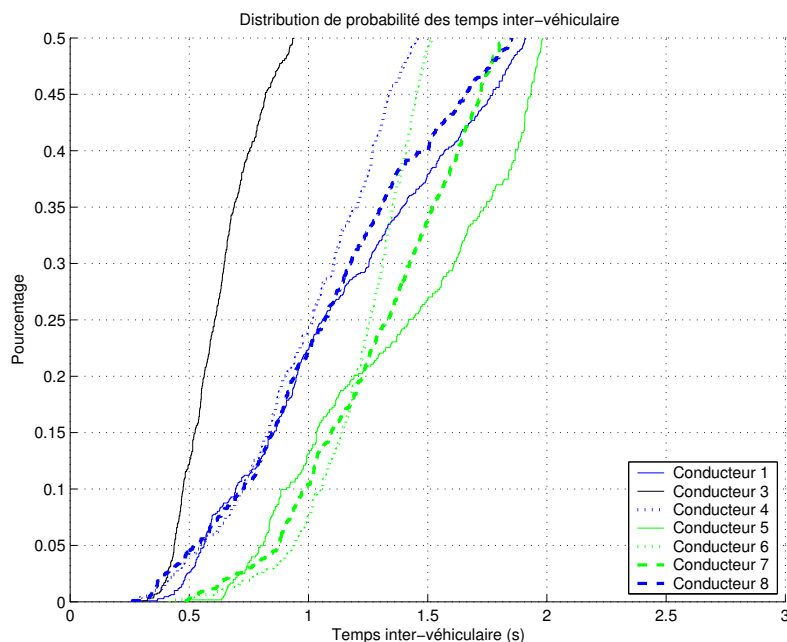


FIG. 2.2 – Distribution des temps intervéhiculaires

On peut voir sur la Figure 2.2 (on ne présente que la partie la plus intéressante de la loi, le graphique *non tronqué* est en annexe p.iv) que deux groupes de conducteurs

apparaissent sur le plan des temps intervéhiculaires pratiqués. Cette figure présente la loi des temps inter-véhiculaire (TIV) lors de phase de suivi, i.e. lorsque le radar identifie un véhicule à moins de 150 m.

Deux groupes de conducteurs se distinguent (couleur verte et bleue); les conducteurs 1, 4 et 8 présentent des temps intervéhiculaires petits plus fréquents que les autres conducteurs; on peut dire qu'ils passent presque 25% de leur temps en situation très critique (25% du temps passé en situation de conduite en file), car pratiquant un temps intervéhiculaire inférieur à 1 s. A contrario, les conducteurs 5, 6 et 7 ne passent qu'environ 10% de leur temps dans cette même situation.

Aron et al. (1999) montrent que, dans cette même configuration de conduite en autoroute interurbaine, 11% des cas rencontrés sont des phases où le temps intervéhiculaire est de l'ordre ou inférieur à 1 s. On recoupe donc cette information avec nos données, et on peut ainsi qualifier nos conducteurs 5, 6 et 7 de conducteurs *moyens* en référence à cette dernière étude; mais on peut en déduire un premier trait sur le style de conduite des usagers 1, 4 et 8 : ils sont risquophiles. Nous reviendrons sur cette dernière notion dans le chapitre 3.

En outre, pour les valeurs plus élevées de temps intervéhiculaires, les conducteurs ne se différencient plus comme l'illustre la loi de répartition en annexe. On remarque que le conducteur 3 est atypique : il présente un profil très spécifique, avec presque plus de la moitié de son temps passé en situation critique. Ce point est appuyé par la figure 2.3. Nous avons tracé la densité des interdistances, en ramenant toute interdistance *infinie* (i.e. supérieure à l'horizon du radar) à 150 mètres.

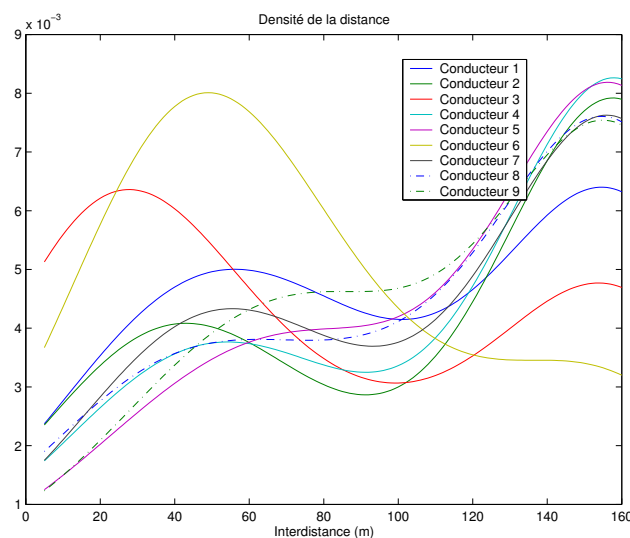


FIG. 2.3 – Densité de probabilité de l'interdistance

Sur un plan général, on distingue une répartition ayant deux modes, le principal autour d'une interdistance de 150 m., traduisant la conduite libre et ainsi la propension des usagers à adopter cette configuration où il peut atteindre sa vitesse préférentielle, et un second autour de la valeur de 50 m., où les interactions avec les autres ne lui permettent

plus d'adopter sa vitesse désirée. Nous reviendrons sur ce chiffre de 50 m. dans le chapitre 5 et sur la notion de seuil d'influence.

Pour revenir au conducteur 3, on constate que l'utilisateur présente le plus souvent les plus faibles interdistances. On remarque même qu'il adopte ou subit plus souvent des situations où l'interdistance est de l'ordre de 30 mètres que des situations où l'interdistance traduit une conduite libre.

2.4.2 Variation de vitesse - Variation de TIV

Nous avons évoqué précédemment la notion de stratégie. Deux stratégies apparaissent : la première est basée sur une vitesse constante et la seconde est basée sur un temps intervéhiculaire constant. Nous allons confronter ces deux visions sur nos données afin de déterminer les stratégies éventuelles adoptées par nos conducteurs.

Avant de répondre à notre question, il convient de décrire précisément dans quel cadre nous allons opérer. L'utilisation de l'indicateur TIV oblige à ne considérer que la situation de suivi, qui d'ailleurs est la seule configuration où l'interrogation a raisonnablement un sens. Ensuite, l'insertion ou la *fuite* engendre des variations absolues ou relatives considérables pour les grandeurs que sont l'interdistance ou le temps intervéhiculaire. En effet, à une vitesse moyenne de 130 km/h par exemple, le passage d'une situation où le véhicule précédent est à environ 150 mètres à la situation d'un véhicule à 20 mètres, résultante d'une insertion, engendre :

	Phase 1	Phase 2	Variation	
			Absolue	Relative
Vitesse	130 km/h 36,11 m/s	130 km/h		
Interdistance	150 m	20 m	-130 m	} -87%
TIV	4,15 s	0,55 s	-3,6 s	

TAB. 2.1 – Variation des indicateurs

Tout d'abord, la stratégie s'appuyant sur le maintien d'une interdistance fixe ne semble avoir aucun écho dans la littérature ; les spécialistes s'accordent plus sur une stratégie (*policy*) nommée CTH (Constant Time Headway). Zou et Levinson (2000) opposent et étudient les deux notions, CTH et VTH (Variable Time Headway) dans l'optique d'une meilleure simulation pour mieux appréhender l'impact de l'ACC sur le trafic.

En absolue, l'interdistance est plus sujette à des variations qui peuvent être importantes, on peut envisager des variations assez fortes, chose non observable physiquement sur les deux autres grandeurs. L'interdistance peut varier considérablement du fait des interactions avec le contexte, l'insertion ou la *fuite* d'un véhicule devant l'utilisateur entraîne une variation importante de l'interdistance. Par conséquent, on se focalise uniquement sur la vitesse et le temps intervéhiculaire.

Le questionnement relatif au maintien de la vitesse ou du temps intervéhiculaire, revient donc à regarder si on a $V(t) = V(t + 1)$ et/ou $TIV(t) = TIV(t + 1)$, ces égalités traduisent le maintien de la vitesse ou du temps inter-véhiculaire. Ceci revient à tester si $\Delta V = V(t + 1) - V(t) = 0$ et $\Delta TIV = 0$. On parvient aux résultats suivants :

Conducteur	Test			
	ΔV	ΔTIV	$\Delta V2$	$\Delta TIV2$
1	Non rejet	Non rejet	Non rejet	Rejet
3	Non rejet	Non rejet	Rejet	Rejet
4	Non rejet	Non rejet	Non rejet	Rejet
5	Non rejet	Non rejet	Rejet	Rejet
6	Non rejet	Non rejet	Non rejet	Rejet
7	Non rejet	Non rejet	Non rejet	Rejet
8	Non rejet	Non rejet	Non rejet	Rejet

TAB. 2.2 – Test d'égalité de la moyenne ($m = 0$)

On a réalisé un test d'égalité de moyenne nulle sur les échantillons ΔV_t et ΔTIV_t et les résultats sont présentés dans le tableau 2.2.

On peut voir que dans tous les cas, on ne peut rejeter l'hypothèse que la moyenne est nulle, *i.e.* que l'usager maintient sa vitesse et son temps intervéhiculaire constants. Cette conclusion est trop hâtive car les calculs sont effectués sur l'historique entier de la conduite, donc mélangeant des phases de conduite en file et de conduite libre. Naturellement, la conduite libre induit un temps intervéhiculaire infini (en pratique il est de l'ordre de 3000 s. cela étant dû à des choix de codage), par conséquent la variance est très importante. La conséquence est l'obtention d'un intervalle de confiance large qui amène à accepter l'hypothèse, puisque contenant la valeur 0.

Ainsi, il nous faut nous focaliser uniquement sur les phases où de tels phénomènes n'ont pas lieu, ce qui nous oblige à isoler certaines fenêtres temporelles adéquates pour nos calculs. On ne conserve que les données liées à une interdistance inférieure à 150 mètres (limite physique du radar), on se trouve ainsi dans des situations où un véhicule se trouve sur la même file et à une distance « finie ». Il s'agit des quantités $\Delta V2$ et $\Delta TIV2$. Les résultats sont affichés dans le tableau 2.2.

Tout d'abord, on peut voir que la majorité des cas, l'hypothèse de nullité de la moyenne (et l'hypothèse du maintien de la vitesse) ne peut être rejetée. On note que pour les conducteurs 3 et 5, le maintien de la vitesse ne semble pas pouvoir être retenu. Nous ne pouvons pas réellement apporter de réponse à ce fait. Certes le fait que le conducteur 3 soit dans la plupart du temps dans des situations de suivi, comme peut le traduire la figure 2.3, pourrait peut-être nous éclairer, mais nous n'avons plus étudié ce cas de figure.

La chose pertinente à remarquer est que pour tous les conducteurs, l'hypothèse de maintien de temps intervéhiculaire est rejetée (tableau 2.2). On voit là un changement radical sur la réponse au test.

Par conséquent, dans des phases de suivi (au sens large), on montre que la vitesse peut être considérée comme constante contrairement au temps intervéhiculaire. On en déduit

qu'il en est de même de l'interdistance¹.

Un autre angle d'études peut être envisagé, il s'agit de l'analyse des phases de suivi pur. Nous appelons situation de suivi pur, une situation de suivi ne comportant aucun changement de file de la part de l'utilisateur, aucune insertion ou fuite de la part d'un autre usager, sur sa voie ou en partance de celle-ci.

Pour cela, nous avons regardé pour chaque conducteur la présence de séquence de type suivi pur. Dans une première étape, on s'intéressait au caractère constant d'indicateurs, et pour cela nous avons regardé $\Delta V2$ et $\Delta TIV2$. À présent, on veut comparer les fluctuations de ces mêmes indicateurs et en déduire, lequel varie le moins. Pour cela, il est nécessaire d'étudier les variations relatives :

$$\tilde{\Delta}V2(t+1) = \frac{\Delta V2(t+1)}{V(t)} = \frac{V(t+1) - V(t)}{V(t)}$$

et

$$\tilde{\Delta}TIV2(t+1) = \frac{\Delta TIV2(t+1)}{TIV(t)} = \frac{TIV(t+1) - TIV(t)}{TIV(t)}.$$

La variance est la statistique qui convient pour déterminer lequel de ces deux indicateurs varient le plus. On présente les résultats dans le tableau 2.3.

Nous n'avons considéré que des séquences *suffisamment longues* ; en moyenne il s'agit de séquences de 140 secondes.

Conducteur	Variance	
	$\tilde{\Delta}V2$	$\tilde{\Delta}TIV2$
1	0,0003	0,007
3	0,0003	0,002
	0,0002	0,0145
4	0,0003	0,0023
6	0,0007	0,0014
	0,0002	0,0029
	0,0001	0,0006
7	0,0006	0,0082

TAB. 2.3 – Variance des variations relatives de vitesse et d'interdistance

Tous les conducteurs ne sont pas présents dans le tableau, car parfois il a été difficile d'exhiber des périodes satisfaisant aux critères et suffisamment longues pour pouvoir faire des calculs fiables.

De manière générale, on constate que la variation absolue d'interdistance est moins dispersée que celle de la vitesse, en d'autres termes la variance de $\Delta TIV2$ est inférieure à la variance de $\Delta V2$. Mais le bon critère est la variance de la variation relative. Là les résultats montrent dans tous les cas la variance de $\tilde{\Delta}V2$ est inférieure à celle de $\tilde{\Delta}TIV2$. Il existe en moyenne un rapport de l'ordre de 17 entre les variances.

¹En effet, par définition TIV est le ratio entre l'interdistance et la vitesse.

Ceci traduit le fait que la vitesse varie moins que le temps intervéhiculaire. Ceci conforte le premier résultat concernant le caractère constant de la vitesse.

En conclusion, on peut dire que la majorité des conducteurs adopte une stratégie basée sur un maintien de la vitesse. La maintien du temps intervéhiculaire est à rejeter ainsi que celle basée sur l'interdistance. Cette constatation amène à remettre en cause certains algorithmes et simulations basés sur l'hypothèse de *constant time headway*. Brackstone et McDonald (2003) ont soulevé ce problème dans l'optique des simulations utilisées pour mesurer l'impact de l'introduction de système intelligent de transport (ITS). Ils remettent en cause plusieurs hypothèses communément admises dans les modèles de suivi (interdistance inférieure à 150 mètres pour nous). En outre, ils remettent en cause l'hypothèse liée, de maintien d'une interdistance de sécurité. Nous avons vu que dans au moins 10% des cas, les usagers ont un temps intervéhiculaire inférieur à 1 seconde, ce chiffre est considéré comme illustrant une situation critique. En outre, dans les phases de suivi, on peut constater que l'interdistance est assez faible au regard de la vitesse moyenne pratiquée, comme le montre la figure 2.4.

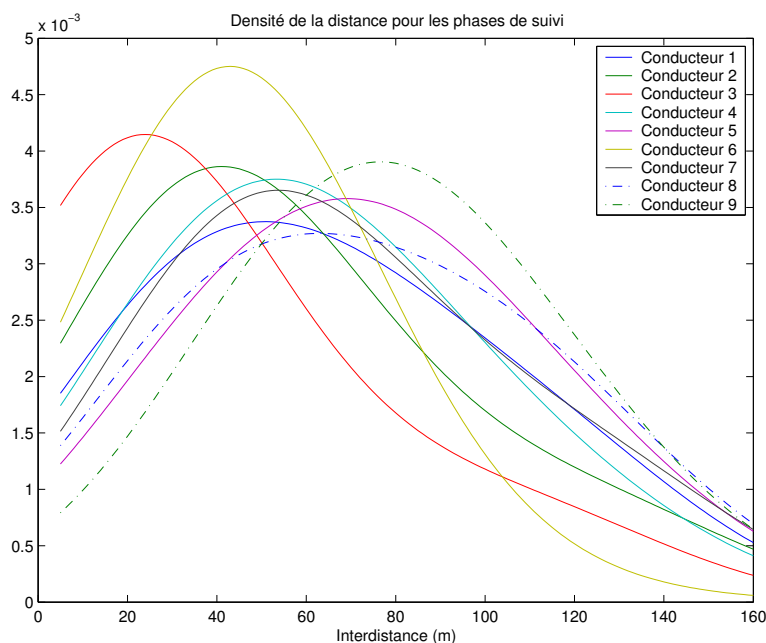


FIG. 2.4 – Distribution des interdistances en situation de suivi

On constate une répartition semblable à celle précédente, mais on peut mieux apprécier le mode (valeur la plus fréquente) des interdistances en situation de suivi, elle oscille selon les conducteurs entre 20 mètres et 80 mètres. Le tableau 2.4 montre que le temps intervéhiculaire est très faible et qu'ainsi l'hypothèse de maintien d'une distance de sécurité peut être remise en cause. Les résultats sur le conducteur 9 ne figure pas ici, pour des raisons relatives à la mesure de sa vitesse.

Dans ce tableau, nous présentons les moyennes calculées en période de suivi (interdistance inférieure à 150 mètres) et lors de dépassement. On constate dans tous les cas (sauf pour le conducteur 3 que nous avons déjà signalé), que le TIV lors du dépassement

Conducteur	Temps inter-véhiculaire	
	Suivi	Lors de dépassement
1	1,91 s	1,49 s
3	1,34 s	1,79 s
4	1,62 s	1,42 s
5	2,16 s	1,55 s
6	1,74 s	1,41 s
7	1,99 s	1,64 s
8	1,93 s	1,42 s

TAB. 2.4 – Temps intervéhiculaire moyen en phase de suivi

est plus faible que celui pratiqué dans la phase de suivi. Ceci traduit le fait que lors du dépassement, les usagers augmentent leur vitesse, pour pouvoir dépasser, et simultanément ils diminuent leur interdistance, impliquant une diminution du TIV. Ceci est tout à fait logique.

Pour tous ces résultats, il convient de signaler que les conducteurs ne sont pas des conducteurs *ordinaires*, dans le sens où il s'agit de conducteurs expérimentés et précisément de *commerciaux*. Par conséquent, les chiffres et conclusions sur la notion de distance de sécurité sont à prendre au conditionnel. Malgré tout, cette hypothèse semble être de plus en plus remise en cause, et notre étude n'est qu'un argument de plus.

2.5 Conclusion

Toutes les études menées sur nos données nous amène à deux conclusions.

Tout d'abord, l'hypothèse de maintien de la vitesse peut être de manière générale retenue et cela pour la majorité des conducteurs. Ensuite, supposer qu'il existe une stratégie adoptée par les usagers basées sur le maintien de l'interdistance ou du temps intervéhiculaire semble peu vraisemblable.

Tout ceci provient naturellement du fait de la variabilité plus grande et intrinsèque de l'interdistance et par conséquent du temps intervéhiculaire. En effet, sur autoroutes, les changement de file des autres usagers (insertion, rabattement) provoquent des variations importantes de l'interdistance. Parfois, le conducteur ne cherche pas à rétablir une interdistance précise, et parfois accepte d'avoir une interdistance faible si ceci a lieu pendant un laps de temps très court. Nous avons vu ceci sur la loi de répartition des temps intervéhiculaire.

Par conséquent, on peut considérer que la majeure partie des usagers considérés optent dans la majorité des cas pour une stratégie basée sur le maintien de la vitesse.

Un tel résultat est favorable dans l'optique de la mise en place de l'Adaptive Cruise Control. En effet, ce système est un régulateur de vitesse, ainsi son objectif est le maintien d'une vitesse préférentielle et dans les cas critiques l'obtention de marges *raisonnables*.

Pour bien comprendre, il faut préciser que dans la majorité des systèmes d'aides à la conduite de type ACC, le paramètre vitesse relative est considéré. Ainsi dans une situation

d'insertion où la vitesse relative est positive, traduisant le fait que le véhicule qui s'insère est plus rapide que vous, le système n'a pas d'effets. On voit qu'il s'agit bien d'une stratégie de maintien de la vitesse et non pas de maintien du temps intervéhiculaire. En effet, si cela était le cas, le système aurait réduit la vitesse afin d'augmenter l'interdistance avec pour objectif le retour à un TIV précis.

Chapitre 3

Automates cellulaires

Une manière de simuler le comportement d'un conducteur et l'évolution dynamique du véhicule est de dissocier les phases dans le processus complexe qu'est la tâche de l'automobiliste. Les modèles psychologiques, contrairement aux approches physiques, ont opté pour cette approche. Nous avons vu apparaître les notions de motivation, d'intention et de modes de réalisation, Saad et Schnetzler (1994). Cette vision est idéale pour comprendre les modèles basés sur les automates cellulaires.

La simulation du trafic peut se faire à plusieurs échelles, microscopique et macroscopique. Le niveau microscopique s'intéresse à la modélisation de l'individu unique dans un flot donné et la modélisation macroscopique s'intéresse à un objet plus global, un flot de véhicules par exemple par le biais des caractéristiques comme le débit q ou la concentration k . Souvent, ces deux approches s'excluent mutuellement ; il est difficile de modéliser un flot entier avec une approche microscopique du fait de la difficulté à modéliser les interactions entre véhicules ; déduire de l'approche macroscopique une vision microscopique semble impossible.

Néanmoins, une vision intermédiaire, nommée mésoscopique, semble apparaître sous la forme, par exemple, de modèles basés sur les automates cellulaires. Le point fondamental distinguant ces modèles CA (cellular automaton) et de ceux microscopiques usuels est le caractère discret dans le temps et l'espace ; cette restriction permet à l'opposé de modéliser via des règles simples les diverses phases d'un trafic (aspect macroscopique).

3.1 Historique

Dans les dernières années, la modélisation du trafic s'est appuyée sur des approches de physique statistique, en particulier via des automates cellulaires ; Nagel et Schreckenberg (1992) en furent les précurseurs. Les véhicules sont considérés comme des particules circulant et engendrant un flux, le trafic ; ainsi l'étude de l'écoulement d'un véhicule ou d'un peloton de véhicules permet de prévoir l'évolution global du flux.

Ces modèles se distinguent des précédents par le caractère discret ; en effet on discrétise le temps en tranche de t secondes et l'espace, ici la route, en tranches de L mètres ; le plus souvent cette longueur est prise comme étant la longueur d'un véhicule. Outre cela, il faut quelques paramètres décrivant le cadre :

- la vitesse maximale V_{max} ;
-

- le débit q (nombre de véhicules passant en un point fixé par unité de temps) ;
- la densité spatiale ρ (nombre de véhicules par unité de longueur de la section).

Cette approche a beaucoup d'avantages :

- la simulation de toutes les phases du trafic et pas simplement de la poursuite ;
- moins de difficultés dans sa compréhension car reposant sur des règles d'évolution et non sur des équations à résoudre ;
- la vitesse de simulation de tels modèles est élevée car reposant sur des règles simples et explicites.

3.2 Cadre des Automates Cellulaires

Le succès de cette approche est multiple comme nous venons de l'évoquer, mais un point supplémentaire est à noter, c'est le parallélisme important entre cette approche et la vision du psychologue de l'utilisateur. Ici, je simplifie volontairement la vision du psychologue de la conduite afin de pouvoir effectuer un lien avec la modélisation basée sur les automates cellulaires, mais à mes yeux, cette simplification est licite. Les deux approches sont régies par le biais de règles basées sur la notion de préférence ou d'objectif mais également d'interactions avec les autres, empêchant la réalisation de son projet. Les règles traduisent le lien entre l'intention et un mode de réalisation. En outre, on trouve la notion de vitesse préférentielle, absente dans beaucoup de modèles physiques, qui est souvent la clé de voûte des modèles psychologiques.

Dans un premier temps, la route est discrétisée spatialement en cellules, ou sites, de taille L mètres. On considère le plus souvent une valeur de 7,5 mètres pour L , soit environ la longueur du véhicule plus l'interdistance entre véhicule dans une situation de congestion maximale (bouchon) (Larraga (2002)).

Ensuite, la discrétisation au niveau du temps est communément prise à $\Delta t = 1$ s., qui correspond au temps de réaction de l'ensemble conducteur-véhicule.

Chaque site¹ est soit occupé par un véhicule, soit vide. Chaque véhicule possède une vitesse entière comprise entre 0 et V_{max} . Cette valeur est fonction des pas de temps et d'espace choisis, comme nous le verrons dans la partie suivante, cf. section 3.3.

Une fois la discrétisation opérée, il faut être mis en garde sur les simplifications d'écriture adoptées. En effet, toute vitesse écrite dans la suite s'exprime en m/s et un abus d'écriture est alors employé. Dans les règles régissant les automates cellulaires, on s'autorise à écrire $V + D$ où V représente une vitesse et D une distance mais en fait ces quantités sont sans unité. Bien évidemment, il faut comprendre que ceci est écrit à la place de $V \cdot \Delta L \cdot \Delta t + D \cdot \Delta L$. Comme le pas de temps est pris égal à 1 seconde, l'usage est que le terme Δt disparaît ; ainsi il ne reste que des termes s'exprimant en terme de *nombres de cellules*. Cette mise en garde étant faite, on peut présenter les différents modèles.

¹Ceci n'est valable que pour les véhicules légers. Des versions du modèle existent pour des mélanges entre véhicules légers et poids lourds, où les poids lourds occupent deux cellules.

3.2.1 Modèle à file unique

On va exposer les règles régissant les modèles basés sur les automates cellulaires, et en précisant à chaque étape les paramètres nécessaires en vue d'une simulation ou d'une estimation sur des données réelles, on parle alors plutôt de calibrage

Contrairement aux modèles précédents, où l'accélération était la grandeur d'intérêt, ce type de modèle s'appuie sur la vitesse individuelle ; il s'agit d'un modèle microscopique. Pour cela, on se donne des règles simples, dans lesquelles réside l'avantage de ces modèles, permettant d'obtenir à chaque instant t la vitesse du véhicule i .

On va énumérer les règles minimales, dans ce genre de modèle : elles sont au nombre de 4 et s'appuient sur une hypothèse psychologique du comportement des conducteurs :

leur propension à maintenir leur vitesse préférentielle.

On retrouve là la conclusion apportée par la partie précédente sur la notion de stratégie. Ici, cette vitesse préférentielle sera la vitesse maximale V_{max} . En notant V_{t+1}^i la vitesse du véhicule i à l'instant $t + 1$, les règles suivantes exposent la méthode pour obtenir la valeur de la vitesse (notée $V_{t+1}^i(3)$, car obtenue après trois étapes). La première règle est naturelle et s'écrit :

Règle 1 *Propension à atteindre sa vitesse préférentielle*

$$V_{t+1}^i(1) = \max(V_t^i + 1, V_{max})$$

Cette règle est à confronter à la réalité physique, qui est la régulation due aux autres véhicules, ceci est l'objet de la seconde règle qui traduit l'absence de collisions. Si on note X_t^i la position du véhicule i au temps t avec la convention que le véhicule de tête est noté $i = 0$ et le véhicule en queue est noté $i = n$, alors on peut définir la distance entre deux véhicules successifs par la quantité :

$$d_t(i - 1, i) = X_t^{i-1} - X_t^i$$

La seconde règle s'énonce alors :

Règle 2 *Incidence du véhicule précédent*

$$V_{t+1}^i(2) = \min(V_{t+1}^i(1), d_t(i - 1, i))$$

en d'autres termes, cette expression signifie que la vitesse adoptée, $V_{t+1}^i(1)$ ne doit pas engendrer une collision.

La troisième règle permet d'introduire un caractère aléatoire à cette construction, cette étape illustre la tendance de certains conducteurs à ralentir pour aucune raison a priori. Si on considère un nombre $p \in [0; 1]$, alors cette règle est :

Règle 3 *Décélération aléatoire*

$$V_{t+1}^i(3) = \begin{cases} \max(V_{t+1}^i(2) - 1, 0) & \text{avec la probabilité } p \\ V_{t+1}^i(2) & \text{avec la probabilité } 1 - p \end{cases}$$

Cette règle s'avère nécessaire pour simuler, ou prendre en compte, la formation spontanée de ralentissements, menant même aux bouchons, qui est un phénomène macroscopique. En effet, elle regroupe les trois comportements suivants :

- fluctuations en régime stationnaire : un véhicule à la vitesse V_{max} peut ralentir aléatoirement, avec une probabilité p . Ceci reflète par exemple une distraction du conducteur, ou le dénivelé variable de la route, qui font que même dans un trafic faible les véhicules n'ont pas une vitesse constante ;
- accélération retardée : au lieu d'accélérer linéairement de 0 à V_{max} , un véhicule peut prendre du retard au démarrage ;
- surcompensation au freinage : lorsqu'un conducteur voit un véhicule ralentir devant lui, il a tendance à *surcompenser* et à ralentir plus que nécessaire.

On a ici une différence avec les modèles de type « voiture suiveuse », plus classique.

Ces règles permettent de définir la vitesse de chaque véhicule i à l'instant $t + 1$, la dernière étape est le déplacement proprement dit s'énonçant :

Règle 4 *Déplacement*

$$X_{t+1}^i = X_t^i + V_{t+1}^i(3)$$

Le modèle énoncé est minimal et les paramètres à préciser sont la vitesse maximale V_{max} et la probabilité p de décélération.

Des extensions d'un tel modèle existent, on va citer les possibles améliorations :

- extension à un réseau à plusieurs files ;
- amélioration par l'insertion d'un paramètre de sécurité (voir paragraphe 3.2.3).

3.2.2 *Modèle à plusieurs files*

Dans cette partie, on tente d'étendre le modèle précédent à une configuration plus réaliste, i.e. la conduite sur un réseau à plusieurs files (Knospe et al. (2002)). La différence réside, bien évidemment, dans la capacité de dépassement. On conserve les règles initiales et on ajoute celle relative au processus de dépassement. Dans sa version simple, si on se situe dans le cas où $V_{t+1}^i > d_t(i - 1, i)$, on regarde la possibilité d'un changement de file. Ce dernier est effectué si les conditions suivantes sont vérifiées :

- $d_{\text{autre file}}^{\text{devant}} > V_{t+1}^i(1)$, reflétant l'intérêt de changer de file ;
- $d_{\text{autre file}}^{\text{derrière}} > V_{max}$ afin de ne pas gêner un autre usager.

Ces règles n'incluent pas de paramètres subjectifs liés au conducteur, les quantités $d_{\text{autre file}}^{\text{devant}}$ et $d_{\text{autre file}}^{\text{derrière}}$ sont plutôt comme des valeurs physiques et ainsi le processus de changement de file est totalement automatique. On peut modifier cela, en considérant deux paramètres d'horizon propres à chaque conducteur, h_{devant}^i et $h_{\text{derrière}}^i$ et en substituant dans les règles précédentes, les quantités $V_{t+1}^i(1)$ et V_{max} par respectivement h_{devant} et $h_{\text{derrière}}$. Mais cette voie souffre de l'estimation, ou calibrage, de ces quantités. En outre, cette extension est intéressante mais un écueil existe sur le plan de l'estimation : il s'agit de la disponibilité des données. En effet, en pratique, peu, pour ne pas dire aucune, étude ne dispose d'informations permettant l'estimation de tels modèles, les quantités relatives aux distances avant ou arrière sur l'autre file sont rarement disponibles pour des raisons techniques ; le plus souvent, on ne dispose que d'un radar aligné sur la voie pratiquée.

3.2.3 Distance de sécurité

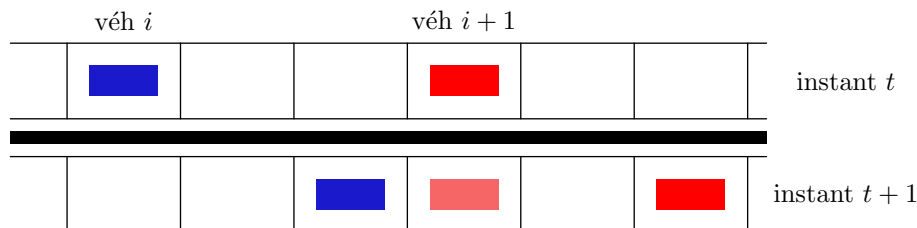
Cette partie expose une idée nouvelle dans ces modèles, l'ajout de paramètres liés à la notion de risque.

De nombreuses améliorations sont venues enrichir le modèle de base, mais le plus souvent il s'agit d'améliorations en vue d'obtenir des résultats macroscopiques plus proche de la réalité, comme la prise en compte des dépassements.

Peu d'améliorations d'ordre microscopique ont été apportées, et de plus aucune sur l'angle sécuritaire. Nous présentons ici une modification faisant partie de cette dernière classe et tentant de palier ce manque. Il s'agit de l'insertion d'un paramètre relatif à la sécurité traduisant deux notions présentes dans la théorie du trafic dès les premières modélisations, mais des modélisations rarement vues sous l'angle de la sécurité. Il s'agit des modèles de Greenshields (1934) et de Greenberg (1959).

Pour une même vitesse relative et une même interdistance, un conducteur régi par le modèle de *Greenshield* possédera une accélération plus faible qu'un conducteur régi par le modèle de *Greenberg*. En d'autres termes, un conducteur de type *Greenberg* est plus risquophile qu'un conducteur de type *Greenshield*.

Cet aspect n'est pas présent dans le modèle initial de Nagel-Schreckenberg. La modification d'une règle, en l'occurrence la règle 2, va permettre de prendre en compte cette notion. La règle illustre une contrainte physique trop forte dans le sens où au mieux la position de véhicule i sera celle occupée par le véhicule $i - 1$ à l'instant précédent.



Or durant ce laps de temps, ce dernier a avancé d'une vitesse V_t^{i-1} , par conséquent on

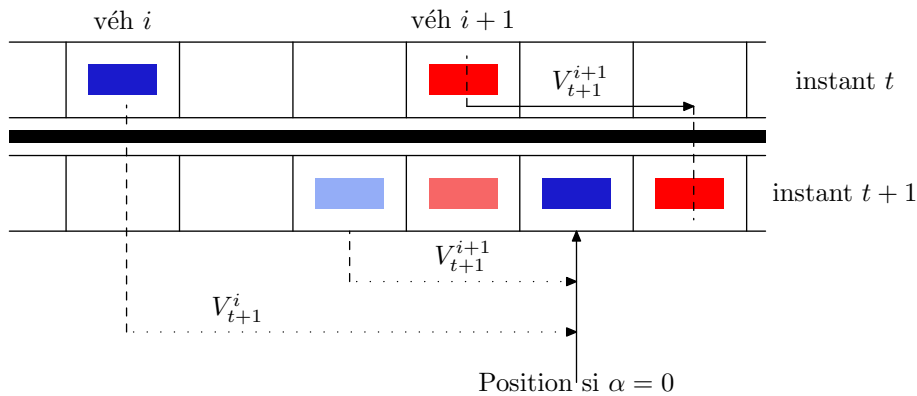
pourrait modifier cette règle en tenant compte de cette vitesse. Pour cela, on considère un paramètre α et alors on obtient :

Règle 5 *Incorporation de la vitesse du véhicule précédent (Règle 2 bis)*

$$V_{t+1}(2) = \min(V_{t+1}^i(1), d_t(i-1, i) + (1 - \alpha)V_t^{i-1})$$

Pour bien comprendre cette nouvelle règle et en quoi elle reflète un souci de sécurité de la part des conducteurs, on va prendre des valeurs remarquables. Si $\alpha = 0$, alors la vitesse V_t^i est maximale et dans ce cas les deux véhicules i et $i - 1$ seront *collés* à l'instant $t + 1$. Ce cas de figure illustre un cas extrême, l'absence de distance de sécurité ; à l'inverse si on considère $\alpha = 1$ alors on retrouve l'expression antérieure où la vitesse du conducteur précédent n'intervient pas. Là, la sécurité est *absolue* dans le sens où l'interdistance sera au minimum² de V_t^{i-1} .

En outre, cette règle permet d'obtenir des vitesses plus importantes qu'avec le modèle initial. Les deux notions de conducteur risquophobe et risquophile apparaissent ici. La valeur $\alpha = 0$ traduit un comportement risquophile, l'usager conduit au plus près du véhicule précédent, à l'opposé la valeur $\alpha = 1$ illustre un comportement risquophobe, ou extrêmement prudent, car le choc est quasiment nul.



On voit sur le schéma précédent, les évolutions entre deux pas de temps du véhicule selon la valeur du paramètre α . Nous présentons une configuration à l'instant t avec un véhicule (en rouge) animé d'une vitesse donnée et un véhicule (en bleu) régi par le modèle présenté, donc celui d'intérêt. À l'instant $t + 1$, le véhicule de tête a changé de position et en parallèle le véhicule de test aussi. Selon la valeur du paramètre α , on voit que ce véhicule suit et même *colle* le véhicule (en rouge), traduisant l'aspect risquophile du conducteur considéré. À contrario, si $\alpha = 1$, alors la future position du véhicule (en bleu) est juste en deçà de son ancienne position (représenté par un carré rouge atténué).

Bien évidemment, les valeurs 0 et 1 sont des cas limites, il est intéressant de voir en pratique si des valeurs précises peuvent être obtenues et si on peut scinder un groupe de conducteurs en risquophile et risquophobe.

²Il s'agit d'une distance minimum car avec la règle 3, elle peut s'accroître.

3.2.4 Impact de la distance

Dans cette partie, nous allons examiner la seconde règle et voir que nous sommes contraints de la modifier pour être réaliste.

Tout d'abord, nous rappelons que l'interdistance est comprise entre 0 et 150 mètres ; du fait des discrétisations considérées, la vitesse maximale est de l'ordre de 40 m/s, ce qui correspond à 40 unités d'espace par unité de temps dans notre choix de discrétisation. Mais le point que nous voulons soulever ici n'a pas de rapport direct avec le choix de la discrétisation. En effet, une rapide étude montre que toute interdistance supérieure à la vitesse maximale n'intervient pas dans la régulation opérée par les automates cellulaires. Or on convient qu'en pratique, une interdistance supérieure à cette quantité conserve malgré tout un impact sur la vitesse adoptée. En d'autres termes, il semble clair qu'un usager ne pratique pas sa vitesse préférentielle dès que l'interdistance est supérieure à 50 m.

Par conséquent, il convient de modifier la règle 5 afin d'augmenter l'intervalle où l'interdistance a un impact. Nous proposons comme modification la règle suivante :

Règle 6 *Incorporation de la vitesse du véhicule précédent et modification de l'interdistance (Règle 2 ter)*

$$V_{t+1}^i(2) = \min(V_{t+1}^i(1), \beta d_t(i-1, i) + (1 - \alpha)V_t^{i-1})$$

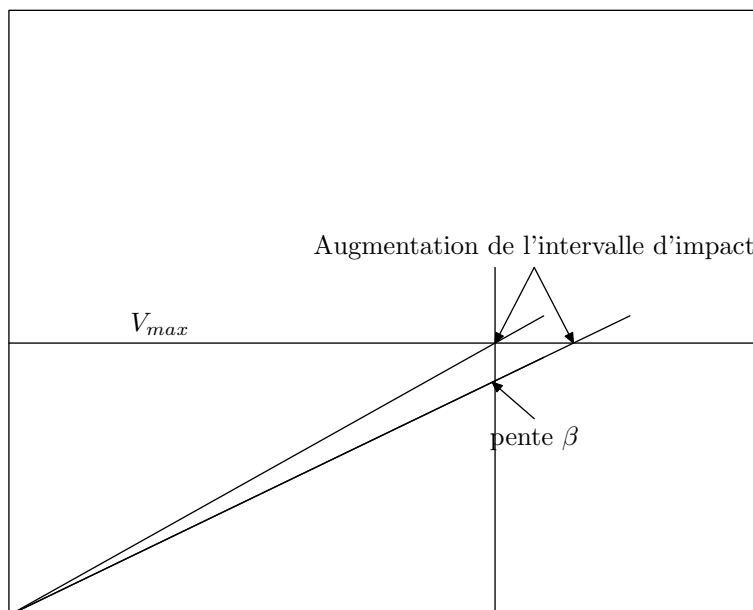
Si on trace le lien entre vitesse et interdistance dans le cas où $\alpha = 1$, on voit l'interprétation du coefficient β , voir figure 3.1. Dans la partie suivante, nous allons estimer la valeur de la pente afin d'obtenir la meilleure adéquation sur les données.

3.3 Résultats

Dans cette partie, nous allons présenter les résultats de calibrage obtenus sur les données de nos conducteurs. Tout d'abord, cette partie présente une différence notable par rapport aux approches basées sur les automates cellulaires, classiquement ils sont employés dans le domaine de la simulation avec comme objectif la calibration des paramètres pour avoir une adéquation correcte au niveau du diagramme fondamental liant le débit d'un réseau à la vitesse moyenne pratiquée sur ce dernier.

Ici, on se focalise sur l'estimation de ces modèles sur nos données.

Afin d'avoir un modèle, nous avons considéré un pas de temps de 1 seconde et un pas spatial de 1 mètre. Ceci diffère du cadre classique exposé ci-dessus. Nous allons voir les raisons de ce choix. Comme dit précédemment, le choix des pas temporel et spatial influe sur le *pas de vitesse*. En effet, une fois choisis les pas, dans la première règle le chiffre 1 acquiert immédiatement une réalité physique. La règle 1 expose la mise à jour de la vitesse, en l'incrémentant si cela est possible. Mais que signifie réellement cette règle ? Si la vitesse maximale n'est pas atteinte, alors la vitesse peut augmenter d'une unité mais

FIG. 3.1 – Impact du coefficient β .

que vaut cette unité? Pour répondre à cette interrogation, il suffit de mener un petit calcul.

Le pas de temps étant d'une seconde, la vitesse passe de V à $V + 1$, autrement dit le véhicule avance d'une cellule, soit donc de L mètres. Dans le cas usuel, L a pour valeur 7,5 mètres, ce qui nous amène à considérer un modèle où la vitesse peut subir d'une seconde à l'autre une augmentation de :

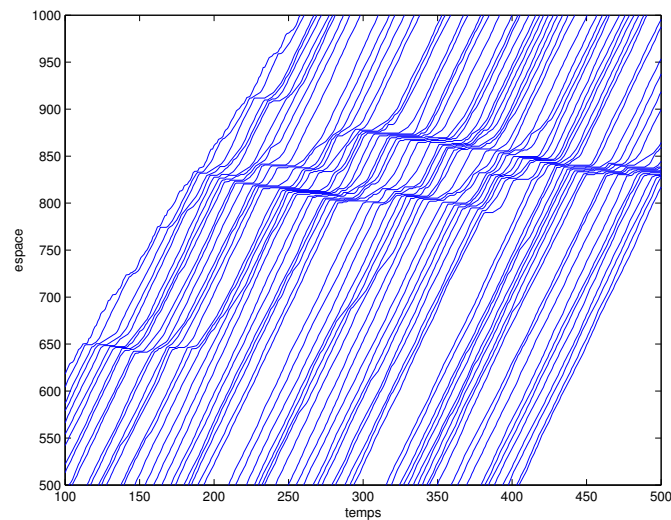
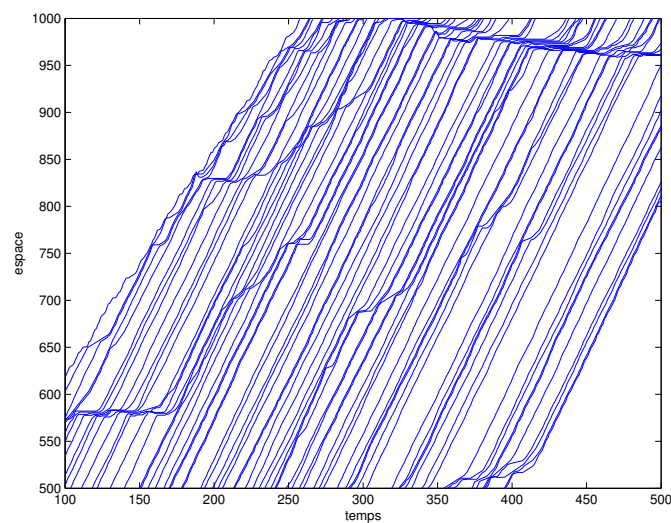
$$\Delta V = \frac{L}{\Delta t} = \frac{7,5 \text{ m}}{1 \text{ s}} = 7,5 \text{ m s}^{-1} \quad \text{soit } 27 \text{ km h}^{-1}$$

Cette augmentation semble bien peu réaliste, par conséquent nous avons considéré la configuration suivante. Un véhicule peut occuper plusieurs cellules, nécessairement contiguës, et chaque cellule est prise de longueur égale à 1 mètre. Cette modification entraîne que l'augmentation de la vitesse est d'une seconde à l'autre de :

$$\Delta V = \frac{L}{\Delta t} = \frac{1 \text{ m}}{1 \text{ s}} = 1 \text{ m s}^{-1} \quad \text{soit } 3,6 \text{ km h}^{-1}$$

On parvient à des augmentations beaucoup plus réalistes et plus adéquates pour une estimation.

Nous allons présenter deux types de résultats, un diagramme temps-espace illustrant la capacité de tels modèles à simuler les bouchons et les délais au démarrage observés. Pour cela, nous allons considérer un flot de véhicules dont le véhicule est animé d'un mouvement pseudo-aléatoire autour de la vitesse maximale. Les autres véhicules sont simulés à l'aide d'automates cellulaires. Deux valeurs sont testées, $\alpha = 1$ et $\alpha = 0,3$. Pour pouvoir comparer les deux diagrammes suivants, nous avons conservé la loi d'évolution du véhicule de tête. De plus, on peut voir la nécessité d'avoir la règle 3 (décélération aléatoire) pour avoir l'apparition de bouchons de manière spontanée.

FIG. 3.2 – Diagramme temps-espace - coefficient de sécurité, $\alpha = 1$ FIG. 3.3 – Diagramme temps-espace - coefficient de sécurité, $\alpha = 0,3$

En outre, les diagrammes présentent un phénomène bien connu en théorie du trafic, il s'agit du *recul* du bouchon. En effet, en pratique, on constate que la zone spatiale congestionnée se déplace dans le sens opposé à l'arrivée des véhicules. Cette constatation est parfois un critère de réalisme d'une simulation. Dans la figure suivante, on a noté deux types de ralentissement. Dans les deux cas, on note bien le recul, illustré par la pente négative de la droite tracée dans le diagramme temps-espace. De plus, on peut voir que la vitesse est plus grande si la quantité de véhicules impliquée est importante.

Après avoir présenté des résultats macroscopiques, nous allons donner quelques résultats d'estimation microscopique sur nos données.

L'objectif est d'estimer le paramètre de sécurité sur nos conducteurs et de montrer

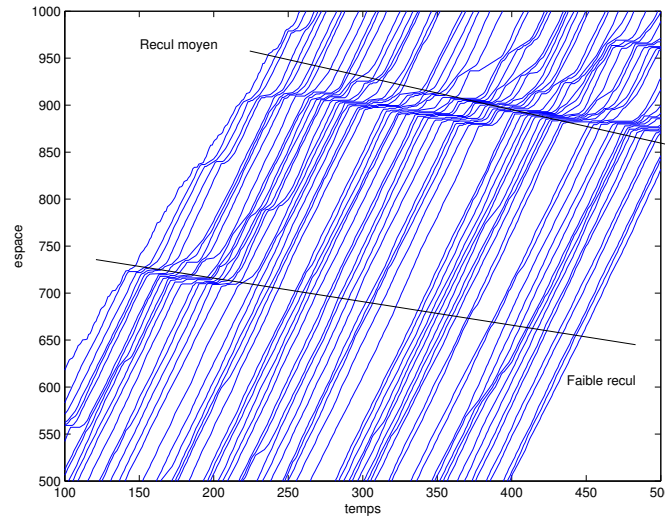


FIG. 3.4 – Apparition du bouchon - $\alpha = 0.8$ - phénomène de recul

que de tels modèles, très simples, permettent d'obtenir des ajustements très corrects en période de suivi, mais aussi mixte.

La méthode utilisée est un balayage simultané sur les paramètres α et β .

Les résultats sont donnés dans le tableau 3.1 où l'erreur relative exprimée en pourcentage est donnée. Pour certains conducteurs, nous avons considéré l'ensemble du parcours, soit 90 km, mais aussi des sous-périodes afin de montrer la puissance de l'adéquation. En effet, l'ensemble de la période est certes importante, mais son hétérogénéité, en termes de scènes de conduite (conduite libre, conduite en file, insertion de conducteurs, modification de l'infrastructure par la présence de voies d'insertion ou d'échappement) rend parfois ardue l'adéquation du modèle.

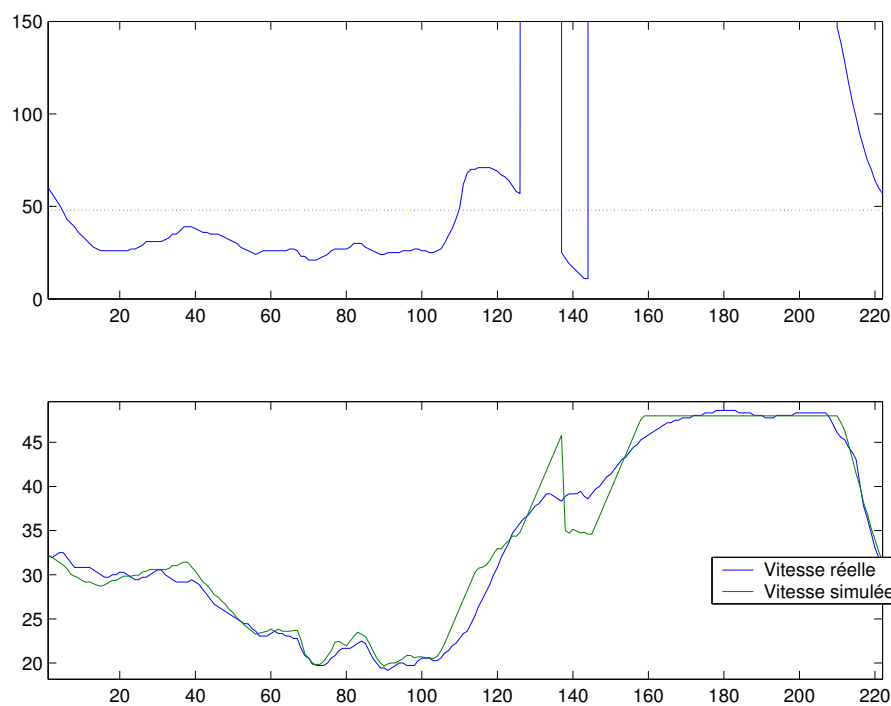
Plusieurs commentaires sont à tirer du tableau précédent. Tout d'abord, les conducteurs présentent tous ces coefficients α et β de l'ordre de 0, 1. La valeur du coefficient α estimé illustre le comportement risquophile de nos conducteurs, ceci est en adéquation avec le caractère *expérimenté* des sujets de l'étude et à mettre en parallèle avec la distribution de probabilité des TIV.

Ensuite, on constate que les erreurs relatives sont très correctes vu la simplicité des modèles en terme de structure mais aussi par le fait que contrairement à des modèles sur séries temporelles usuels, les modèles d'automates cellulaires présentés ici ne s'appuient pas sur la valeur de V_{t-1} pour prévoir V_t , il s'agit quasiment d'un prévision. En effet, dans le cadre des séries temporelles (par exemple ARMA), il faut distinguer deux approches. La première que nous appellerons estimation – ou prévision à un pas de temps – consiste à obtenir une approximation de la valeur de V_{t+1} à partir de la valeur connue de V_t ; la seconde consiste à obtenir une approximation de V_{t+1} en s'appuyant sur l'estimation faite au pas précédent pour V_t , on parle alors de prévision à n pas de temps. En d'autres termes, dans l'estimation à un pas de temps, on s'appuie sur la valeur exacte de V_t , et dans la prévision à n pas de temps, on utilise que l'approximation de V_t réalisée à partir de la connaissance de V_{t-1} . Ici, nous sommes dans le cas d'une prévision à n pas

Conducteur	Période analysée	Coefficient		Erreur relative
		α	β	
1	globale	0,1	0,1	8,37%
	$1 < t < 118$	0	0	9,14%
	$1240 < t < 1400$	0	0	4,78%
3	globale	0,1	0,2	6,47%
	$1 < t < 300$	0	0	1,96%
4	globale	0,1	0,1	8,88%
	$t > 1550$	0,1	0,1	3,70%
5	globale	0,1	0,1	9,21%
6	globale	0,1	0,1	5,66%
	$1 < t < 600$	0,1	0,1	4,64%
7	globale	0,1	0,1	10,57%
	$1 < t < 200$	0	0	3,01%
8	globale	0,1	0,1	11,54%

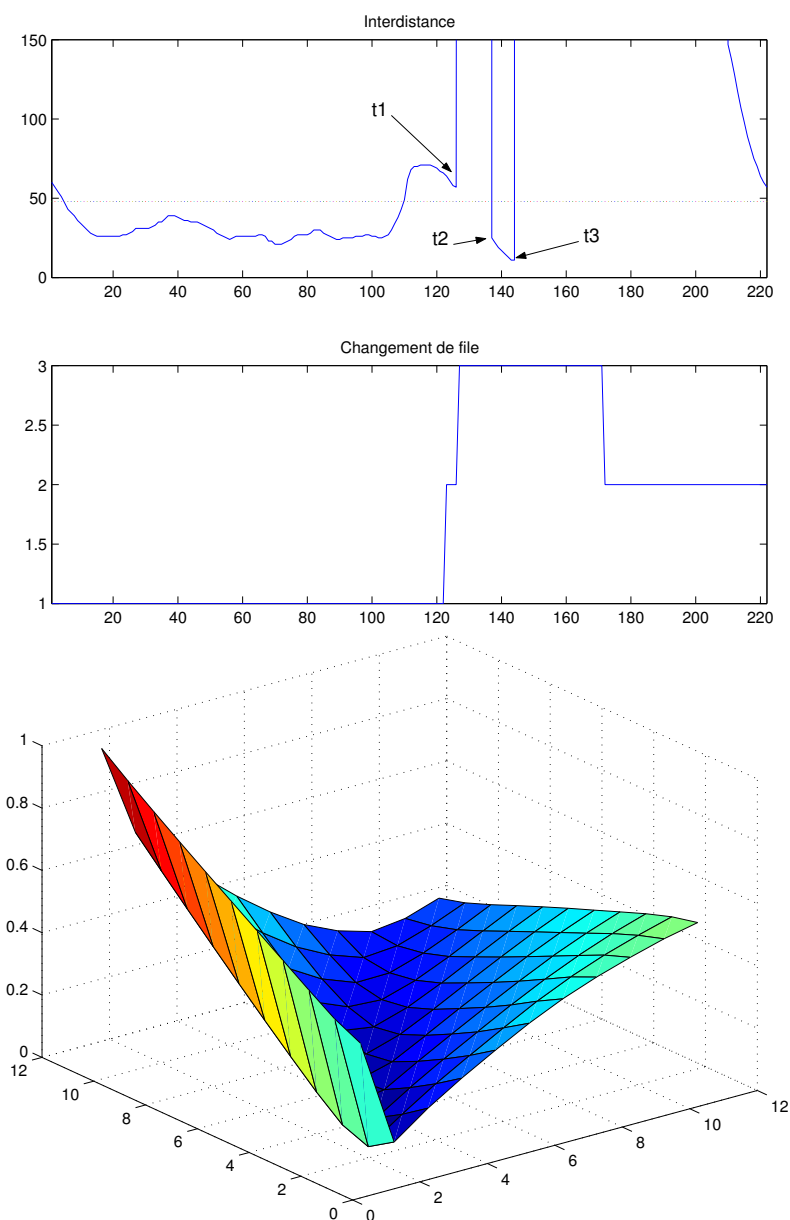
TAB. 3.1 – Paramètres estimés des modèles d’automates cellulaires

de temps. Cette approche est clairement sujette à des erreurs plus importantes, mais pourtant ici l’adéquation est très correcte dans un horizon de 200s. Nous concluons en exhibant quelques graphes montrant la qualité de l’estimation.



Sur les deux schémas précédents, on peut voir en haut l’évolution de l’interdistance pour le conducteur 4 pour un épisode de conduite ($1550 < t < 1770$ s.) et en bas l’évolution

simultanée de la vitesse réellement observée et celle de l'estimation réalisée. L'adéquation est très convenable, adéquation qui s'appuie sur la valeur initiale de la vitesse (à $t = 1550$ s) et sur l'interdistance observée à chaque instant. De plus, on voit que cette période test présente une phase de suivi puis un changement de file (t_1) pour la file de gauche où une insertion se produit (t_2) suivi de la disparition de ce même véhicule (t_3) et enfin un rabattement. On peut voir que l'adéquation ne souffre pas trop de cette succession de scènes distinctes.



La dernière illustration est la fonction erreur qui a été minimisée pour obtenir les valeurs des paramètres α et β .

3.4 Conclusion - Extensions envisageables

Nous avons présenté dans cette partie les automates cellulaires dans le domaine du trafic routier avec l'angle nouveau de l'estimation et pas uniquement de la simulation. Cette modélisation a de multiples avantages :

- une modélisation régie par des règles réalistes ;
- des règles simples ;
- la rapidité de l'algorithme ;
- une très bonne adéquation dans des situations raisonnablement hétérogènes et cela avec peu d'informations.

Les automates cellulaires semblent devoir prendre une grande place dans la modélisation du trafic au niveau mésoscopique dans le futur.

Sur le plan de l'évaluation, les automates cellulaires peuvent revêtir deux intérêts. En effet, une comparaison des estimations des deux paramètres α et β entrent dans l'évaluation comportementale du système. Une modification, probablement une augmentation, du paramètre α traduirait une disparition importante ou non du comportement risquophile de nos conducteurs, modification comportementale induit uniquement par la présence du système.

Pour conclure, des améliorations à nos modèles peuvent être apportées par l'insertion de manière plus fine du concept de vitesse préférentielle. Cela consisterait à distinguer la vitesse maximale et la vitesse préférentielle, en posant par exemple comme règle 1, une règle faisant intervenir une réalisation autour de la vitesse préférentielle.

Chapitre 4

Modélisation markovienne cachée - Validation de l'importance du contexte

Nous nous intéressons dans ce chapitre à la contextualisation de la tâche du conducteur. L'idée est de montrer que la variable *Contexte*, codée par F. Saad et son équipe, revêt une importance prépondérante dans l'analyse du processus vitesse. Cette variable prend les trois modalités suivantes :

- 1 - aucun véhicule devant dans sa file dans un horizon inférieur à 150 m
- 2 - un véhicule se situe devant lui mais au moins un file libre
- 3 - toutes les files devant lui sont occupées

Toutes ces informations sont à considérer uniquement dans un horizon de 150 m au plus, puisqu'il s'agit de la portée du radar embarqué.

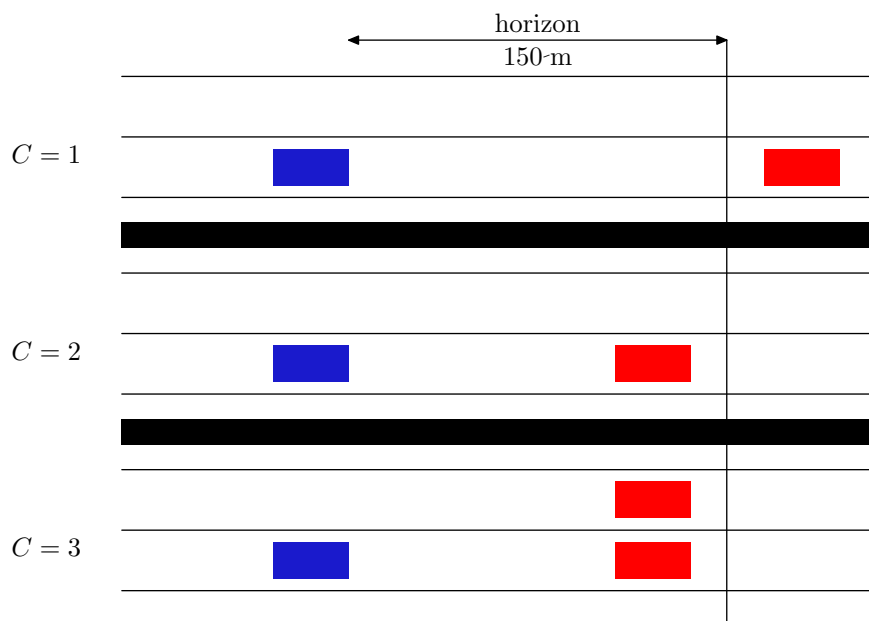
On rappelle ici le descriptif de la variable *Contrainte* pour une meilleure lisibilité du propos.

Afin de montrer l'importance de la variable *Contexte* dans la compréhension du processus vitesse, nous allons modéliser ce dernier à l'aide d'un modèle markovien caché dont on impose le nombre d'états, qui sera de 3.

L'objectif est de travailler sur le pourcentage de recouvrement entre la chaîne cachée que l'on exhibera et la variable *Contrainte*. Pour cela, nous nous appuyons sur l'algorithme de Viterbi qui permet d'obtenir la chaîne cachée la plus vraisemblable. Son algorithme est détaillé dans le chapitre 5.

La modélisation s'appuie sur un programme écrit en MatLab, permettant d'obtenir les matrices de transition, les paramètres de la loi d'émission telles que la moyenne et la variance. Dans le domaine des Transports, il est usuel d'utiliser une loi log-normale pour modéliser la vitesse (variable positive). Nous avons opté pour une loi d'émission gaussienne car étant sur autoroutes et travaillant sur des données où les points d'arrêt (vitesse nulle) ont été ôtés, la loi gaussienne peut parfaitement être utilisée. Ce programme permet d'obtenir la chaîne la plus vraisemblable.

Le point suivant porte sur la numérotation de cette chaîne, en d'autres termes l'étiquetage, ou plus exactement sur les permutations des étiquettes que le programme lui a conféré. En effet ce dernier comportant une partie aléatoire pour l'initialisation, il est concevable

FIG. 4.1 – Variable *Contrainte*

que la numérotation des états soit complètement factice. Pour être plus précis, ce même programme peut sur les mêmes données obtenir des résultats identiques à une permutation de la numérotation près. C'est justement ce point que nous allons exploiter. En effet, nous allons considérer le codage de la variable *Contexte* comme référence et via une boucle pour tester les 6 permutations d'étiquetage, nous calculons pour chaque permutation le pourcentage d'occurrence. On retient comme permutation adéquate, celle associée au pourcentage d'occurrence maximum.

Le tableau 4.1 ci-après montre les pourcentages d'occurrences maximum obtenus pour chaque conducteur.

Conducteur	Pourcentage maximum
1	50,13%
2	51,58%
3	53,14%
4	57,09%
5	46,44%
6	55,69%
7	55,97%
8	53,29%
9	63,58%

TAB. 4.1 – Pourcentage d'occurrence maximal entre la chaîne cachée et la variable *Contexte*

Conducteur	état 1 - μ_1	état 1 - σ_1	état 2 - μ_2	état 2 - σ_2	état 3 - μ_3	état 3 - σ_3
1	147 km/h	14	134 km/h	28	104 km/h	228
2	138 km/h	11	124 km/h	21	106 km/h	46
3	147 km/h	21	129 km/h	41	91 km/h	173
4	154 km/h	76	129 km/h	45	97 km/h	249
5	139 km/h	8	131 km/h	9	108 km/h	144
6	141 km/h	5	133 km/h	13	102 km/h	229
7	153 km/h	17	143 km/h	8	118 km/h	209
8	149 km/h	23	136 km/h	9	118 km/h	72
9	137 km/h	30	120 km/h	23	96 km/h	84

TAB. 4.2 – Paramètres du modèle markovien obtenu (les valeurs ont été arrondies à l'unité près)

Le tableau 4.2 donne les valeurs des paramètres moyenne et écart-type du modèle markovien ordonné suivant la permutation maximisant l'occurrence entre la chaîne cachée et la variable *Contexte* en prenant l'ordre de cette dernière comme base.

La permutation associée à l'occurrence maximale permet d'obtenir un résultat logique et naturel. En effet, on constate que cette permutation associe les instants correspondant à l'état 1, la moyenne sur la loi d'émission la plus élevée, aux instants pour lesquels la valeur de la variable *Contrainte* vaut 1 dans la majorité du temps. Ceci est naturel et rassurant puisque cette contrainte signifie qu'aucun véhicule n'est présent dans les 150 m précédent le véhicule et que par conséquent le conducteur peut adopter sa vitesse de croisière ou vitesse préférentielle.

De même, l'état intermédiaire (état 2) présentant une moyenne inférieure et une variance légèrement plus forte dans la majorité des cas est associé à la modalité 2 de la variable décrivant le contexte de conduite. Ceci est également logique, puisque cette modalité indique qu'un véhicule se trouve directement devant le conducteur, il est naturel que ce dernier diminue sa vitesse mais sans entrer dans un régime synchronisé de conduite ou dans un régime de conduite en file puisqu'au moins une file reste libre lui offrant la possibilité de dépasser, si la gêne engendrée est considérée comme trop forte.

Et finalement, le dernier état de la chaîne noté 3 est caractérisé d'une part par une moyenne de la loi d'émission très nettement inférieure, et d'autre part et surtout par une variance très supérieure aux valeurs des autres états.

Là encore, l'association est parfaitement logique. En effet, la modalité 3 de la variable *Contrainte* désigne un contexte de conduite où toutes les files en aval du véhicule sont occupées et cela dans un horizon de 150 m. Par conséquent, le conducteur est contraint de s'adapter au trafic, et cela engendre une diminution de sa vitesse et une augmentation de la variance.

Ce phénomène corrobore la conclusion du chapitre sur les stratégies de conduite où nous avons montré que la stratégie principale d'un conducteur est le maintien de la vitesse au détriment le plus souvent d'une interdistance. Bien évidemment, dans les cas de congestion ou de manière générale de contrainte forte, le conducteur doit diminuer sa vitesse et sa tendance à adopter une vitesse préférentielle étant contrariée on aboutit à une aug-

mentation de la variance.

En effet, si la stratégie était basée sur le maintien de l'interdistance alors la vitesse pourrait diminuer mais sans que la variance n'augmente, en restant comparable aux autres valeurs caractérisant la conduite de la personne, mais ceci n'a pas lieu.

Pour conclure, nous venons de montrer que la permutation associée à la plus grande similarité entre la chaîne cachée du modèle markovien du processus de la vitesse et la variable recueillie par *F. Saad* aboutit directement à une association naturelle et logique.

Un second point doit être abordé à présent, il agit du pouvoir explicatif de cette variable *Contrainte*. Un outil pour répondre à cela peut être également ce pourcentage d'occurrence. En effet, le tableau 4.1 présente des pourcentages tous, exceptés un, supérieur à 50%. Ce chiffre est à comparer au pourcentage d'occurrence entre deux chaînes indépendantes à 3 états, c'est-à-dire 33%. Nous avons donc un gain de plus de 50%, et même de près de 90% avec le conducteur 9.

Ce résultat confirme l'hypothèse de *Farida Saad* selon laquelle le contexte de conduite joue un rôle prépondérant dans le processus de conduite.

Chapitre 5

Modélisation markovienne avec exogènes

5.1 Introduction

Les chaînes de Markov cachées (HMM) sont des modèles statistiques qui ont eu un succès important dans les dernières décennies. Ils sont utilisés pour l'étude de données séquentielles et cela dans de multiples domaines. On peut citer la reconnaissance vocale comme premier champ d'application, mais plus récemment, la reconnaissance d'écriture, de séquences en biologie moléculaire, ou encore détection d'erreurs dans le domaine industriel.

Les modèles Input/Output HMM (IOHMM) sont des extensions de ces derniers. Ils ont été étudiés pour la première fois par Bengio et Frasconi (1995) et Bengio et Frasconi (1996). Parfois appelés modèles markoviens cachés conditionnels, ils constituent un développement par la possibilité de faire dépendre la loi d'émission et de transition de variables exogènes, nommés *input*. Dans ce cadre, les observations modélisées par les lois d'émission sont appelées *output*, et le modèle représente la loi conditionnelle de l'*output* étant donné une séquence cachée dépendante de l'entrée, dit *input*.

Nous allons présenter rapidement les modèles markoviens cachés, puis présenter les modèles IOHMM. Nous ferons un lien avec les réseaux bayésiens dynamiques.

Ensuite, nous expliciterons les hypothèses sous-jacentes de tels modèles et leurs traductions dans le cadre qui nous intéresse de la modélisation de la vitesse.

Après avoir explicité la procédure d'estimation de tels modèles, nous présenterons les résultats obtenus et une comparaison avec les modélisations usuelles basées sur les HMM. Enfin, nous expliquerons la technique d'estimation des valeurs des états cachés par l'algorithme de Viterbi, pour finir par l'introduction et l'estimation du concept de seuil d'influence.

5.2 Les Modèles Markoviens Cachés (HMM) - Réseau Bayésien Dynamique (DBN)

Les modèles de Markov cachés ont été introduits par Baum et Petrie (1966). D'abord utilisés en reconnaissance de la parole à partir des années 80 (Rabiner (1989)), ils ont été appliqués ensuite à la reconnaissance de textes manuscrits et à la bioinformatique.

Leur succès est dû au développement de la modélisation de processus non stationnaire. On entend par processus stationnaire (au second ordre), un processus $(X_t)_t$ vérifiant les hypothèses suivantes :

- son espérance mathématique, $\mathbb{E}[X_t]$, est indépendante du temps t ;
- sa covariance, $\text{Cov}(X_t, X_{t+h})$, ne dépend que de h , et par conséquent sa variance est indépendante du temps.

Un processus non stationnaire est un processus ne vérifiant pas l'une de ces propriétés. En pratique, de nombreux phénomènes ne satisfont pas de tels hypothèses. Le processus de vitesse en est un cas particulier ; par exemple l'hypothèse d'une espérance constante et indépendante du temps n'est pas tenable. Une analyse de cette quantité fait apparaître la notion de régime, de régime multiple et de régime indexé par le trafic local.

Revenons sur le modèle markovien caché. Il représente un objet donné par deux suites de variables aléatoires : l'une dite cachée et l'autre observable. Dans la suite, on désignera la suite cachée par X et celle observée par Y . Y est l'*output*.

Un modèle markovien caché (HMM) est donc un processus doublement stochastique au sens où il est constitué d'un processus stochastique X sous-jacent qui n'est pas observable directement ; le dernier n'est perceptible que par l'observation de Y , suite observée. En d'autres termes, nous observons les réalisations d'un processus Y que nous supposons être dépendant d'un certain processus X (voir figure 5.1).

La dépendance entre deux instants s'interprètent via le processus caché $(X_t)_{t \geq 1}$.

Cette représentation sous forme géométrique est empruntée au domaine des réseaux bayésiens dynamiques (RBD), ou en anglais *dynamic bayesian network* (DBN). Une généralisation naturelle des modèles de Markov cachés est à trouver dans les réseaux bayésiens, modèles hiérarchiques pour lesquels la chaîne de Markov sous-jacente est remplacée par un graphe markovien orienté. On parle alors de dynamiques. Chaque forme revêt une signification précise : une variable continue est symbolisée par un rond, une variable discrète par un carré. Toute dépendance est symbolisée par une flèche. Nous verrons qu'une telle représentation sera utile pour comprendre l'extension que sont les modèles IOHMM, et la difficulté précise qu'ils engendrent.

Le schéma 5.1 représente le principe de base des modèles HMM.

On observe les données Y_t ; ces dernières sont les réalisations d'un processus $(Y_t)_t$ tel que les réalisations présentent une dépendance, symbolisé par la flèche dans la partie haute du schéma. On s'intéresse à la modélisation de cette dépendance temporelle. L'hypothèse des modèles markoviens cachés est l'existence d'un processus $(X_t)_t$ expliquant cette dépendance. Le but est d'estimer la dépendance entre X_t et X_{t+1} (symbolisée par la flèche verte) et le lien entre X_t et Y_t (symbolisé par la flèche rouge). On voit que cette conception permet le passage d'un problème considéré comme compliqué entre Y_t et Y_{t+1} en deux problèmes plus simples, l'étude du lien entre X_t et X_{t+1} d'une part, et l'étude du

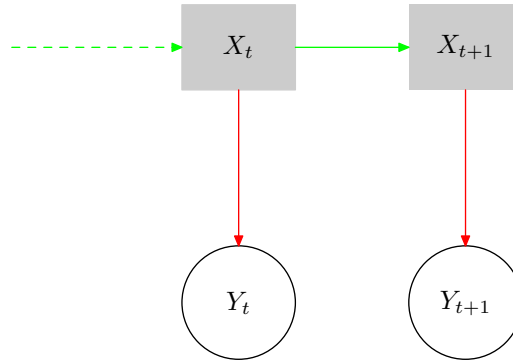


FIG. 5.1 – Schématisation d'un modèle markovien caché

lien entre X_t et Y_t . Il y a donc deux types de probabilités :

- une probabilité de changement d'état entre X_t et X_{t+1} ;
- une probabilité d'émission de symboles entre X_t et Y_t .

On parle de probabilité de changement d'état pour désigner la loi de X , et de probabilité d'émission pour la loi de Y_t sachant X_t .

La difficulté d'un tel raisonnement réside dans le caractère inobservable du processus $(X_t)_t$. L'adjectif markovien qualifie le lien entre X_t et son passé $(X_{t-1}, X_{t-2}, \dots)$. Le plus souvent, et ce sera le cas pour nous, on parle de markovien d'ordre 1. Ceci signifie que l'information apportée par tout le passé $(X_{t-1}, X_{t-2}, \dots)$ utile dans la prévision de X_t peut être résumée par la seule connaissance de X_{t-1} .

La formulation mathématique est la suivante, en notant $X_1^{t+1} = (X_1, \dots, X_{t+1})$:

$$\begin{aligned} \mathbb{P} [Y_{t+1} \leq y | Y_1^t = y_1^t, X_1^{t+1} = x_1^{t+1}] &= \mathbb{P} [Y_{t+1} \leq y | X_{t+1} = x_{t+1}] \\ \mathbb{P} [X_{t+1} = x | Y_1^t = y_1^t, X_1^t = x_1^t] &= \mathbb{P} [X_{t+1} = x | X_t = x_t] \end{aligned}$$

La première expression est la loi d'émission, et la seconde la loi de changement d'état. Ces expressions signifient que la connaissance de l'état de la chaîne résume toute la connaissance pour la valeur future à *un pas* de l'observation, et que le passé de la chaîne est résumé en la connaissance à l'instant t pour prévoir l'instant $t + 1$.

De nombreuses applications de tels modèles existent, en reconnaissance de parole, en médecine, Churchill et Lazareva (1999), Strat et Carrat (1999). Dans le domaine du transport routier, on peut citer les travaux sur la modélisation du changement de voie (Kuge et al. (2000), Bui (1996)).

5.3 Les Modèles Markoviens Cachés avec Exogènes (IOHMM)

Les modèles markoviens sont très puissants en terme de modélisation de phénomènes non stationnaires. Nous présentons une extension introduite par Bengio et Frasconi (1995).

Dans beaucoup d'études, on peut séparer les variables observées en deux familles. D'une part les variables, objet de l'étude, nommées variables dépendantes, qui sont fonction des variables de la seconde famille, dites exogènes ou de contrôle. On peut penser au

cadre de la régression, où on tente d'expliquer le niveau d'une caractéristique en fonction de certains paramètres.

Cet aspect, que nous n'avons pas encore présenté pour les modèles markoviens cachés (HMMs), peut être considéré à deux niveaux. En effet, nous avons présenté ce type de modèles comme un processus doublement stochastique. Par conséquent, l'introduction, ou plutôt la prise en compte, de termes exogènes peut se faire à chacun de ces deux niveaux.

Le premier niveau est une simple extension du modèle initial et n'en remet pas fondamentalement le cadre. Il s'agit de l'insertion au niveau de la loi d'émission. Ceci est représenté par la ligne noire dans la figure 5.2.

Plus précisément, si on observe un phénomène Y supposé modélisable par un modèle markovien caché et si de plus, une variable C influe sur cette mesure alors le modèle initial est modifié comme suit :

$$\begin{aligned} \mathbb{P} [Y_{t+1} \leq y | Y_1^t = y_1^t, X_1^{t+1} = x_1^{t+1}, C_1^{t+1} = c_1^{t+1}] &= \mathbb{P} [Y_{t+1} \leq y | X_{t+1} = x_{t+1}, C_{t+1} = c_{t+1}] \\ \mathbb{P} [X_{t+1} = x | Y_1^t = y_1^t, X_1^t = x_1^t, C_1^{t+1} = c_1^{t+1}] &= \mathbb{P} [X_{t+1} = x | X_t = x_t, C_{t+1} = c_{t+1}] \end{aligned}$$

Si on fait l'hypothèse d'une loi normale par exemple alors on parvient à l'expression suivante qui donne la loi de Y à l'instant t conditionnellement aux quantités X_t et C_t :

$$\mathbb{P} [Y_t \leq y | X_t, C_t] = \mathbb{P} [\mathcal{N}(\mu_{X_t} + f(C_t), \sigma_{X_t}^2) \leq y]$$

La première assertion est relative à la loi d'émission, loi de l'observation. La seconde décrit les transitions du processus caché.

Nous allons avoir recours à la schématisation des réseaux bayésiens dynamiques pour bien comprendre où figure la différence entre les modèles markoviens cachés classiques et ceux incorporant une variable d'entrée.

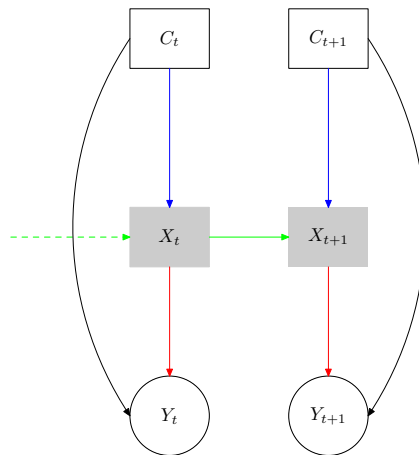


FIG. 5.2 – Schématisation d'un modèle markovien d'ordre 1 caché avec prise en compte d'exogène

Le second niveau (symbolisé par la flèche bleue) est beaucoup plus profond et remet en cause le cadre initial ; il s'agit de la modification du processus sous-jacent, celui du

processus caché, voir le paragraphe 5.3.1. C'est l'essence des modèles IOHMM signifiant Input Output Hidden Markov Models.

Bengio et *Frasconi* furent les premiers à considérer de tels modèles dans Bengio et Frasconi (1996). Cette extension est très intéressante car elle permet la prise en compte de connaissances *a priori* essentielles dans la compréhension de la dynamique d'un phénomène. On peut citer à titre d'exemple la première application qui en a découlé (application au domaine de la finance). Dans Bengio et al. (2001), les auteurs comparent les différents modèles usuels (mélange de lois, modèle markovien caché, modèle markovien caché avec exogènes au niveau de la loi d'émission et modèle markovien caché avec exogènes au niveau de la loi de transition), et montrent l'amélioration en terme d'ajustements avec leur nouveau modèle.

5.3.1 Différence avec HMM

La principale différence entre les deux modèles est le passage du cadre homogène à celui, plus compliqué, du cadre non homogène. Ce terme qualifie la loi du processus caché. Dans le cadre initial, la probabilité de transition est indépendante du temps. L'insertion du processus C rompt l'homogénéité; la loi de transition dépend d'une variable et par conséquent, la probabilité de passer d'un état i dans un état j est indiquée par la variable de contexte C .

Nous verrons la signification précise dans notre cadre ultérieurement.

5.3.2 Avantage de la modélisation IOHMM

L'étude de tels modèles peut s'avérer très utile dans le cadre de la comparaison et de l'évaluation. En effet, dans notre cas, s'appuyer sur les modèles HMM classiques, s'avère à nos yeux incorrects. Le principe est a priori dans un premier par exemple de comparer les paramètres avec et sans le système. Bien évidemment, une telle pratique ne revêt un sens que si l'on peut attribuer l'évolution au système que l'on veut tester, en l'occurrence ici, l'ACC. Or, s'appuyant sur des données non contrôlées (nous avons insisté sur ce point précédemment), les évolutions des paramètres (vitesse moyenne, variance, transition, etc) sont difficilement imputables au seul système. En effet, le contexte peut influencer grandement certaines mesures. C'est le problème central dans l'évaluation, il faut dans un premier temps, établir les variations et dans un second temps, imputer ou expliciter ces variations. La seconde partie est très difficile dans le cadre d'expériences en situations réelles. Pour pallier ce problème, il s'avère impératif de s'affranchir du contexte et pour cela, il faut le prendre en compte afin de rendre comparable les différentes caractéristiques. Ce propos est d'après nous la vision correcte à adopter, malheureusement comme nous le rappellerons dans la partie liée aux résultats, seuls des adéquations sur les données sans le système a été possible.

Dans cette optique, nous avons donc considéré les modèles IOHMM.

Sur le plan pratique, ces modèles ont déjà été l'objet d'études appliquées, mais sans investigation théorique. Il s'avère que sur le plan théorique, un obstacle majeur apparaît, nous y reviendrons dans la partie 5.5.

5.4 Questions - Problèmes théoriques

Une étape essentielle à toute modélisation est l'estimation des paramètres. Des méthodes, comme les moindres carrés et leurs variantes existent depuis de nombreuses années dans la boîte à outils du modélisateur.

Dans le domaine des modèles markoviens cachés, une approche spécifique fût élaborée par Baum et Petrie (1966). Cette technique est connue sous le nom d'algorithme EM. Il s'agit d'un algorithme itératif en deux phases. On calcule l'espérance à l'étape i (phase E : expectation) puis on détermine de nouveaux estimateurs en maximisant la vraisemblance (phase M : maximization), ces nouveaux estimateurs servant pour le calcul de l'espérance à l'étape $i + 1$.

Sous certaines hypothèses, cet algorithme converge vers les estimateurs du maximum de vraisemblance (estimateur MV), maxima de la vraisemblance *limite*. Le caractère fini de la vraisemblance *limite* ainsi que la consistance des estimateurs MV est due à Leroux (1992) dans le cas de lois de l'observation continues.

Nous venons de préciser et de dissocier les méthodes d'estimation, du problème plus ardu de la convergence de l'estimateur du maximum de vraisemblance.

Pour préciser ce point, nous rappelons le théorème suivant où l'espace Θ du paramètre θ est supposé métrique de métrique d . On note θ_0 la vraie valeur.

Théorème 1 (Vaart (1998), section 5.2 p.44) *Soit M_n une fonction aléatoire (resp. la vraisemblance empirique) et M une fonction fixe de θ (resp. la vraisemblance limite, si elle existe) telles que pour tout ε positif*

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{P} 0$$

$$\sup_{\theta: d(\theta, \theta_0) \geq \varepsilon} M(\theta) < M(\theta_0)$$

alors pour toute suite d'estimateurs $(\hat{\theta}_n)$ satisfaisant $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_P(1)$ converge en probabilité vers θ_0 .

Sur le plan calculatoire, et sous réserve de la validité apportée par la preuve de la convergence, l'algorithme EM peut s'appliquer au cadre plus large des modèles Input/Output HMM. Pour l'instant, nous nous placerons dans le cas d'un contexte discret ; dans ce cas, l'algorithme converge. Nous avons réalisé des simulations et des estimations consistantes ont été obtenues. Mais ces simulations nous ont permis de déceler des cas dans lesquels les estimations résultant de l'algorithme EM sont peu voisines des valeurs de notre modèle. Il s'avèrent que le contexte doit *a priori* satisfaire une propriété.

Sur le plan théorique, le problème de la convergence est encore ouvert. Aucun résultat n'existe dans la littérature et cela pour deux raisons à nos yeux : le domaine est assez récent et la difficulté est nette.

5.5 Cadre - Hypothèses

Soient Y_1, \dots, Y_T des réalisations d'un processus à valeurs dans \mathbb{R} . Simultanément on dispose des observations d'un contrôle, que l'on nommera contexte par la suite, C . (C_1, \dots, C_T) sont les réalisations d'un processus à temps discret et à valeurs dans un espace \mathcal{C} .

Considérons à présent un processus non observé X à temps discret et dont l'espace des états est noté \mathcal{X} .

Nous faisons les hypothèses suivantes.

A1 $(X_t)_{t \geq 1}$ est markovien conditionnellement à $(C_t)_{t \geq 1}$,

$$\begin{aligned} \mathbb{P} [X_t = x | X_1^{t-1} = x_1^{t-1}, C_1^T = c_1^T] &= \mathbb{P} [X_t = x | X_{t-1} = x_{t-1}, C_t = c_t] \\ &= T(x_{t-1}, x, c_t) \end{aligned}$$

A2 La suite Y_1, Y_2, \dots, Y_T sachant $(X_1, C_1), \dots, (X_T, C_T)$ est une suite indépendante de variables aléatoires.

La densité de Y_t conditionnellement à $(C_{1:t}, X_{1:t})$ est $g_\theta(Y_t; X_t, C_t)$, pour tout $t \leq T$.

A3 La suite C_1^∞ dépend de manière causale du processus Y , donc :

$$\mathcal{L}(C_t | C_1^{t-1}, Y_1^T, X_1^T) = \mathcal{L}(C_t | C_1^{t-1}, Y_1^t), \quad t \leq T$$

A4 pour tout t , $X_t \in \{x_1, \dots, x_K\}$ et $C_t \in \{c_1, \dots, c_M\}$.

A5 Θ est un ouvert d'un espace euclidien. Pour l'identifiabilité du modèle, on spécifie que si $\theta, \theta' \in \Theta^2$, $T_\theta(\cdot, \cdot, \cdot) = T_{\theta'}(\cdot, \cdot, \cdot)$ et $g_\theta(\cdot; \cdot, \cdot) = g_{\theta'}(\cdot; \cdot, \cdot)$ alors $\theta = \theta'$.

A6 on suppose que $\min_{x, x', c, \theta} T_\theta(x, x', c) > 0$.

A7 Pour tout $i = 1, \dots, K$, $\liminf_n n^{-1} \sum_{t=1}^T \mathbb{1}_{\{C_t = c_j\}} > 0$ avec probabilité 1.

A8 Soit G la loi de Y conditionnellement à $(X_{1:t}, C_{1:t})$, de densité g ; pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que,

$$S_\eta = \left\{ y : \inf_c \frac{\min_x g_\theta(y|x, c)}{\max_x g_\theta(y|x, c)} > \eta \right\},$$

on a $\lim_{\eta \rightarrow 0} \min_{x, c} G_\theta(S_\eta | x, c) = 1$.

Les hypothèses A1 et A2 sont les versions des hypothèses usuelles de la modélisation markovienne, adaptées à notre cadre. Connaissant la valeur du contexte en t , X_t ne dépend que de la connaissance apportée par X_{t-1} . La loi de Y_t , dite loi d'émission, est entièrement déterminée par la connaissance de (X_t, C_t) .

L'hypothèse A3 est cruciale, car elle permet de considérer l'étude du processus Y conditionnellement à C ; par conséquent, dans la suite C sera considéré comme non stochastique et fixé.

La difficulté de ce genre de modélisation réside dans son aspect non-homogène du fait de la présence de la variable de contrôle C modifiant la dynamique de la chaîne cachée et la rendant non-stationnaire.

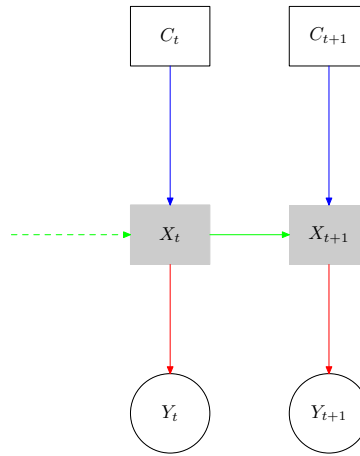
On peut trouver en annexe 9.5 une ligne de preuve de la convergence d'un estimateur θ et de sa loi limite.

Un cadre d'hypothèses a été avancé et nous nous placerons dans ce dernier pour la suite.

5.6 Modèle considéré

Nous avons exposé les modèles markoviens cachés avec exogènes généraux. Dans cette étude nous avons considéré une version simplifiée. On supposera que l'input ne joue de rôle qu'au niveau de la chaîne cachée et non pas au niveau des observations. Le modèle retenu est décrit par les équations suivantes et par le schéma suivant :

$$\begin{aligned}\mathbb{P} [Y_{t+1} \leq y | Y_1^t = y_1^t, X_1^{t+1} = x_1^{t+1}, C_1^{t+1} = c_1^{t+1}] &= \mathbb{P} [Y_{t+1} \leq y | X_{t+1} = x_{t+1}] \\ \mathbb{P} [X_{t+1} = x | Y_1^t = y_1^t, X_1^t = x_1^t, C_1^{t+1} = c_1^{t+1}] &= \mathbb{P} [X_{t+1} = x | X_t = x_t, C_{t+1} = c_{t+1}]\end{aligned}$$



L'introduction d'un tel modèle nécessite une justification, en particulier sur un point. Pour être clair, nous allons supposer que la chaîne cachée X est à deux états et que le contexte peut prendre trois valeurs. Le point sur lequel il faut s'attarder est la différence entre le modèle IOHMM avec C à valeurs dans C_1, C_2, C_3 et le modèle HMM avec une chaîne cachée à 6 états. Ce point est crucial pour justifier l'emploi d'un modèle IOHMM.

La différence entre les deux approches réside dans le lien entre C et X . Dans un cadre où le lien entre le contexte et la chaîne cachée, serait déterministe alors les deux modèles sont identiques, et il apparaît plus simple de considérer le modèle markovien caché à 6 états.

A l'opposé, dans un cadre où le lien entre l'input et la chaîne cachée peut être considéré comme aléatoire alors le modèle IOHMM ne peut être réduit au modèle classique.

Dans notre cas, le lien entre le contexte et le processus $(X_t)_t$ est situé au niveau de l'utilisateur. En effet, l'observation du contexte C_t va influencer l'utilisateur dans son choix d'allure (Y_t) et pour nous, dans son choix de la valeur de X_t . Il nous semble par conséquent justifier de considérer une telle modélisation et que celle-ci ne peut être réduite à un modèle markovien caché.

5.7 Étude pratique

Dans cette partie, nous allons présenter les résultats obtenus en terme de modélisation du processus vitesse. La vitesse est mesurée chaque seconde, ainsi que d'autres variables

liées au contexte. Le processus vitesse n'est pas stationnaire ; il présente une variance fonction du temps, comme le montre la figure 5.3. Nous rappelons que Y désigne l'observation, ici la vitesse ; C désigne le contexte et enfin X la chaîne cachée.

Pour des raisons relatives aux données, tous les conducteurs ne sont pas considérés ; on examine les usagers 2, 3, 5, 6, 7 et 8. Nous avons choisi de retenir la variable *Contrainte*, résumant la gêne subie par le conducteur.

On retrouve bien ici, le concept de variable de contrôle évoquée précédemment. De plus, de multiples études montrent qu'un même conducteur ne réagit pas de façon automatique en fonction du contexte. Il peut choisir de maintenir sa vitesse même si le niveau de contrainte est moyen ou fort. On retrouve là les considérations évoquées en terme de stratégies évoquées dans le chapitre Stratégies de conduite.

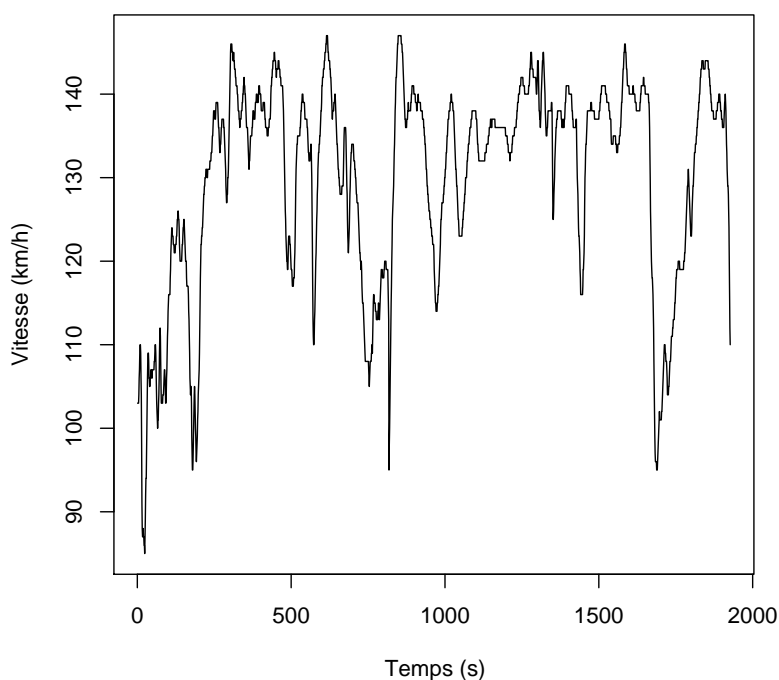


FIG. 5.3 – Vitesse du conducteur 2

5.7.1 Cadre

L'objet central est la vitesse pratiquée par les conducteurs. Cette dernière quantité étant positive par nature, la loi log-normale est choisie classiquement. Ainsi, en pratique, nous avons modélisé le logarithme de cette quantité qui suit par hypothèse une loi normale.

Dans tous les tableaux suivants, nous donnons les caractéristiques de la loi directement. En aparté, nous signalons qu'une modélisation basée sur une loi normale aboutit à des résultats très semblables ; en outre les estimations du processus caché dans les deux cas mène à des valeurs strictement identiques.

		Loi normale	Loi log-normale
État 1	Moyenne	138,06	138,09
	Variance	12,76	12,58
État 2	Moyenne	116,31	116,41
	Variance	101,62	110,58

Matrice de transition	0,99	0,01	0,99	0,01
	0,02	0,98	0,02	0,98
Différence entre les chaînes cachées (%) :			0%	

TAB. 5.1 – Comparaison de modèles (en terme de loi d'émission) pour le conducteur 2

Ensuite, nous considérons comme variable de contexte, la variable Contrainte. Cette dernière est prise en compte dans le modèle au niveau des probabilités de transition. Ainsi, alors que classiquement, les probabilités de passer d'un état à l'autre sont indépendantes du contexte, à présent, la connaissance de ce dernier influe sur l'état du régime à considérer. En d'autres termes, si l'état 1 signifie une conduite en haut régime (vitesse élevée) et l'état 2 une conduite en bas régime (vitesse plus faible), alors le fait que le contexte soit dans un état donné influe sur la valeur du régime *possible*. La connaissance de l'état du trafic influe sur la probabilité d'être à grande vitesse.

Dans la suite, nous optons pour le modèle suivant :

- la vitesse conditionnée par l'état de la chaîne cachée suit une loi log-normale de moyenne μ et de variance σ^2 :

$$V_t | X_t = x \rightsquigarrow \mathcal{N}(\mu_x, \sigma_x^2)$$

- les transitions s'écrivent :

$$\mathbb{P}[X_{t+1} = i | X_t = j, C_{t+1} = c] = p_{ijc}.$$

L'objectif, dans le cas d'une chaîne à deux états, sera d'estimer les quantités μ_1, μ_2, σ_1 et σ_2 , ainsi que les coefficients p_{ijc} . Les estimations sont obtenues par le biais de l'algorithme EM. Il s'agit de l'algorithme de Baum-Welch et de l'algorithme *forward-backward*. Nous avons adapté ces deux algorithmes à notre modèle.

Enfin, le choix de la variable d'input n'est pas aisée. Il s'agit d'introduire une variable qui influe sur la vitesse, non pas au niveau de la valeur ponctuelle de celle-ci mais plutôt au niveau de la valeur moyenne, le *régime*. Dans un premier temps, la contrainte a été retenue ; mais d'autres variables seront considérées, nous présenterons dans la partie 5.7.4 les résultats obtenus sur ce point.

5.7.2 Comparaison

Nous allons montrer les comparaisons entre les deux approches. Dans un premier temps, nous faisons la comparaison en terme d'erreur globale, en s'appuyant sur les bornes obtenues par les deux approches. Ensuite, nous regardons pour un conducteur précis, de plus près les deux estimations et voyons que l'approche tenant compte du contexte se révèle plus adéquate. Pour finir, nous étudions l'hypothèse de prise en compte du contexte par l'option choisie.

Afin de comparer les deux modélisations, on considère les critères d'erreurs habituels (somme des carrés des erreurs) mais aussi un critère heuristique basé sur la proportion d'observations inclus dans les bornes de confiance obtenues à chaque instant. La phase d'estimation nous permet d'obtenir une estimation pour la moyenne et la variance ; ces quantités permettent de construire des bornes de confiance par la formule suivante :

$$B_{inf}(Y_t) = \mu_{X_t} - \sigma_{X_t} \text{ et } B_{sup}(Y_t) = \mu_{X_t} + \sigma_{X_t}$$

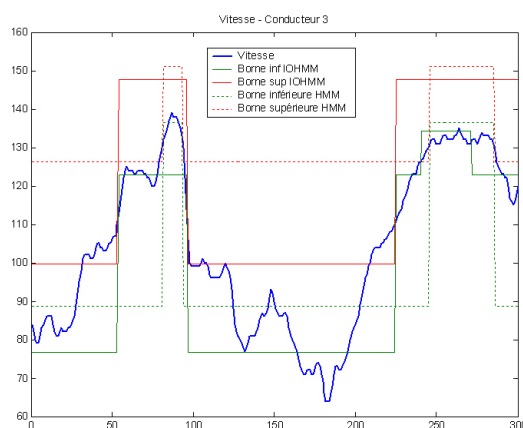
Nous avons choisi de considérer des bornes à un écart-type, car pour certaines phases, le choix de deux écarts-types aurait mené à des intervalles trop larges, ne conservant plus alors un réel intérêt. Ainsi nous avons comparé la proportion de temps où la vitesse observée se situait entre ces deux bornes.

La première constatation est le fait que les bornes obtenues par le modèle IOHMM sont plus resserrées ; on obtient des écarts-types plus faibles. Ensuite, le tableau suivant montre les gains conséquents obtenus en terme d'erreur. Ces derniers sont en moyenne de 25% ; on peut voir l'importance de la prise en compte de l'information donnée par le contexte pour une meilleure compréhension de la vitesse. En outre, en plus d'obtenir des bornes plus fines, l'adéquation du modèle est accrue ; dans la majorité des cas, on remarque une augmentation en terme de proportion d'observations dans les intervalles de confiance. En effet, on serait enclin à penser que plus les bornes de confiance sont larges, plus la proportion d'observations dans l'intervalle de confiance est grande ; or ici ce n'est pas le cas. C'est heureux puisqu'on a incorporé une variable supplémentaire, ce qui montre sa pertinence. En fait, on perçoit ainsi que la modélisation markovienne cachée sans variable de contrôle a plus de difficultés à saisir la dynamique du processus de vitesse et que cette dernière peut être mieux appréhendé par la connaissance du contexte de conduite. Ceci révèle une meilleure adéquation sur tous les plans, meilleure dynamique et meilleure estimation des caractéristiques. Il est à noter que pour le conducteur 6, aucune convergence n'a pu être obtenue. Nous reviendrons sur ce point dans la partie 5.7.4.

Sur le graphique précédent, on a tracé simultanément la vitesse ainsi que les bornes d'incertitude pour chaque modèle. Les bornes d'incertitude sont calculées à un écart-type et cela pour chaque modélisation. Les bornes du modèle markovien classique sont en pointillé, celles associées au modèle avec exogènes sont en train plein. On peut voir que presque toujours, la vitesse est comprise entre les deux bornes ; ici, nous souhaitons insister sur le *comportement* décrit par ces bornes. Pour simplifier et nous permettre de faire une constatation, on trace non plus les bornes mais directement les moyennes. Pour faciliter la lecture, on a tracé l'évolution simplifiée de la vitesse en utilisant un ajustement par un polynôme d'ordre 5 (voir figure 5.5).

Conducteur	HMM		IOHMM		
	Erreur	Proportion des observations entre $[B_{inf}; B_{sup}]$	Erreur	Proportion des observations entre $[B_{inf}; B_{sup}]$	Gain
2	$8,78E+04$	65,91%	$7,16E+04$	66,53%	18,53%
3	$2,84E+05$	62,43%	$1,57E+05$	66,65%	44,83%
5	$8,40E+04$	62,33%	$6,87E+04$	68,50%	18,21%
6	$2,26E+05$	70,78%		NC	
7	$1,77E+05$	65,37%	$1,23E+05$	63,49%	30,74%
8	$1,17E+05$	62,84%	$1,05E+05$	67,81%	10,87%

TAB. 5.2 – Comparaison HMM et IOHMM

FIG. 5.4 – Vitesse du conducteur 3 - bornes des modélisation HMM et IOHMM sur l'intervalle $[1 : 300]$.

Tout d'abord, en terme d'évolution, on voit clairement que le modèle HMM (courbe bleue) est loin d'être correct, en particulier pour la période $[100; 200]$, où l'estimation est très supérieure à la valeur. Par contre, dans l'ensemble le modèle IOHMM s'ajuste bien mieux en terme d'évolution. Tout ceci nous permet de constater que les valeurs estimées en terme de moyenne sont plus correctes.

Le point essentiel ici est lié à la structure même du modèle. En effet, hormis les valeurs moyennes, un tel modèle doit être adéquat en terme d'états. On peut voir les sauts (notés 1, 2, 3 et 4 sur la figure 5.5). Pour les sauts 1 et 2, il apparaît que le modèle HMM change d'état (point 2) trop *tardivement* par rapport à la valeur de la vitesse entre les points 1 et 2.

On peut donc noter que la prise en compte du contexte, ici par le biais de la variable *Contrainte*, permet d'avoir des changements d'états plus appropriés à la vitesse.

Pour conclure, nous allons comparer notre modèle avec un modèle HMM ayant un nombre d'état de 6 ; il s'agit du produit entre la dimension de l'espace d'état de la chaîne cachée et de l'espace des valeurs du contexte choisi ici.

Tout d'abord, un tel modèle n'est nullement justifié ici ; un test sur la dimension de l'espace d'état nous amène à rejeter un tel choix Rydén (1995) ; mais à titre de comparaison, nous conservons ce modèle.

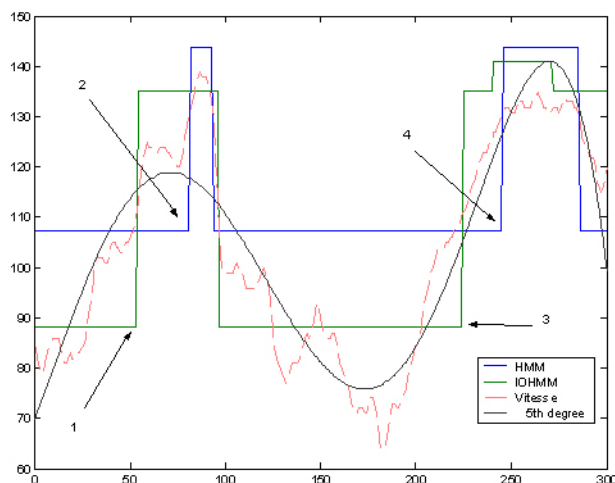


FIG. 5.5 – Comparaison simplifiée HMM - IOHMM

Ensuite, l'adéquation est bien évidemment excellente. Pour bien comprendre cela, il suffit de considérer un modèle ayant un espace d'état de taille la dimension de la série à modéliser ; l'adéquation est alors parfaite car il suffit de considérer la matrice de transition constituée de 1 aux coordonnées $(i, i + 1)$, traduisant le changement systématique d'un état i à l'état $i + 1$. Ensuite, pour chaque état, il suffit de prendre comme estimation de la moyenne l'observation (unique!). Ainsi la comparaison en terme d'erreur avec un tel modèle est vaine, mais nous nous intéressons à l'approche ponctuelle. Pour cela, on trace sur une période courte (300 secondes) les deux estimations, voir figure 5.6.

Comme nous l'avons signalé, la courbe associée au modèle HMM (6) constitue une excellente adéquation à la courbe de vitesse, mais on voit que les nouveaux états permettent de mieux décomposer l'observation lors de valeurs hautes ; une plus grande segmentation de l'estimation pour des valeurs supérieures à 120 km/h peut être en effet observée. Mais contrairement à l'observation faite sur la figure précédente, cette fois-ci les estimations sont très proches, autant en niveau qu'en saut. On peut même dire que les changements d'états sont identiques. Ces derniers sont signalés par les flèches.

Par conséquent, on peut voir que la prise en compte du contexte au niveau du processus stochastique sous-jacent permet de mieux saisir les changements d'états du processus vitesse observé.

Pour conclure, il faut rappeler que le modèle HMM (6) ne peut être retenu en pratique, et qu'entre les deux seul le modèle IOHMM (2×3) possède une justification comme reflétant le comportement du conducteur dans un contexte évolutif, et dont les réactions ne sont pas entièrement déterminées par ce dernier. Il s'agit d'une interaction non déterministe.

À présent, nous allons nous pencher sur les estimations numériques obtenues par les deux approches. Nous considérons les données du conducteur 2.

Pour finir, on regarde les estimations du processus caché sous-jacent. L'étape consis-

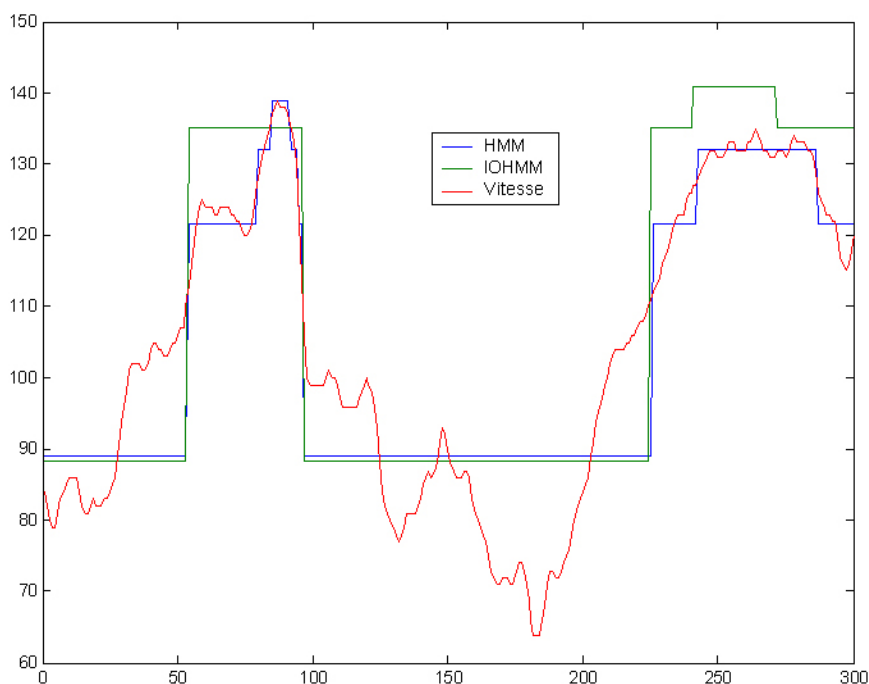


FIG. 5.6 – Comparaison HMM (6) et IOHMM (2x3)

HMM	État	
	$x = 1$	$x = 2$
Moyenne	138,06	116,31
Variance	12,76	101,62

$$P_{HMM} = \begin{pmatrix} 0,99 & 0,01 \\ 0,0165 & 0,9835 \end{pmatrix}$$

TAB. 5.3 – Paramètres du modèle HMM (conducteur 2)

tant à exhiber l'estimation proprement dite sera abordée dans la partie 5.8. Il s'agit de l'algorithme Viterbi.

On dispose de l'estimation de la chaîne cachée pour le modèle HMM et IOHMM. Nous pouvons comparer les deux modèles sur ce point car les processus cachés sont à valeurs dans un même espace, nous nous sommes placés ici dans le cas d'un processus caché à deux états. Dans le tableau 5.5, les chiffres représentent les nombres d'occurrences où les deux chaînes ne sont pas égales, exprimés en pourcentage. Si on suppose que les chaînes obtenues sont les suivantes : 1010 et 0010, alors la différence est de 50%.

Cette comparaison est utile pour comparer les modifications apportées par l'input au niveau de l'occurrence même de la chaîne.

		Contrainte					
		Contrainte = 1		Contrainte = 2		Contrainte = 3	
		État		État		État	
IOHMM	$x = 1$	$x = 2$	$x = 1$	$x = 2$	$x = 1$	$x = 2$	
Moyenne	138,23	121,07	137,75	113,93	136,55	110,58	
Variance	12,26	49,64	8,88	119,29	30,05	93,93	

$$P_{IOHMM}^{C=1} = \begin{pmatrix} 0,9923 & 0,0077 \\ 0,0263 & 0,9737 \end{pmatrix}, P_{IOHMM}^{C=2} = \begin{pmatrix} 0,9925 & 0,0075 \\ 0 & 1 \end{pmatrix}, P_{IOHMM}^{C=3} = \begin{pmatrix} 0,9821 & 0,0179 \\ 0,0073 & 0,9927 \end{pmatrix}$$

TAB. 5.4 – Paramètres du modèle IOHMM (conducteur 2)

Conducteur	Différence
2	2,54%
3	25,70%
5	2,64%
7	7,71%
8	12,15%

TAB. 5.5 – Comparaison des chaînes cachées (HMM et IOHMM)

Dans un premier temps, on constate que les chaînes sont assez semblables (en moyenne 10% de différences sont à constater). Par conséquent, on peut considérer que les chaînes ont une signification semblables, mais comme nous l'avons exposé dans le tableau 5.2, les gains sont assez considérables lors de l'adjonction du contexte. On note une très forte différence pour le conducteur 3, point supplémentaire corroborant l'aspect différent de cet usager par rapport aux autres.

Après avoir vu que les chaînes revêtent la même signification, il serait intéressant d'examiner si les différences ont lieu à des instants exacts ou si elles peuvent être liées à des phénomènes précis. Il s'avère, pour l'instant, qu'aucun lien précis ne peut être établi. Malgré tout, on constate que la majorité des écarts ont lieu lors des changements d'états (voir figure 5.7) et non par l'apparition de sauts fortuits (voir figure 5.8).

Dans la plupart des cas, la chaîne IOHMM anticipe les sauts à la hausse et retarde les sauts à la baisse comme l'illustre la figure 5.9.

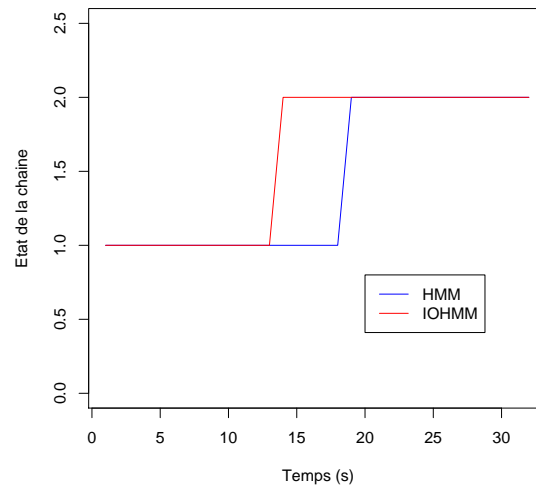


FIG. 5.7 – Chaînes cachées (1)

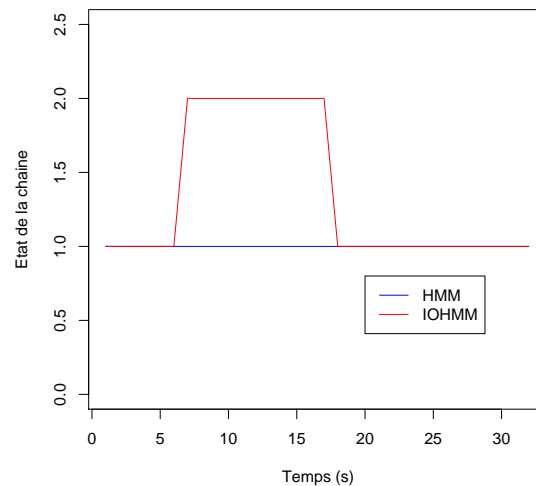


FIG. 5.8 – Chaînes cachées (2)

Sur les graphiques précédents, le tracé bleu (respectivement rose) désigne la chaîne pour le modèle HMM (respectivement IOHMM), en ordonnée est représenté les états de la chaîne.

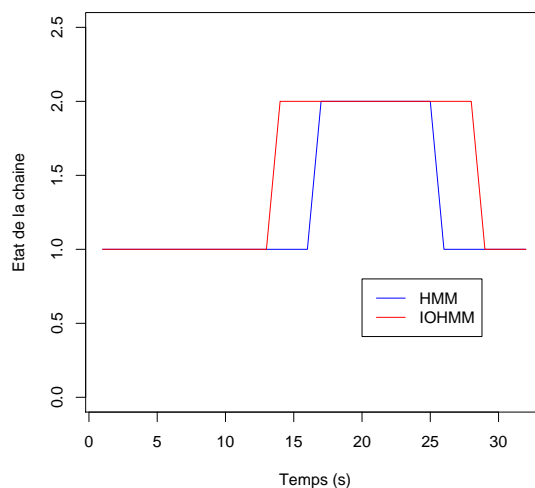


FIG. 5.9 – Chaînes cachées - modèle HMM et IOHMM

5.7.3 Interprétation

Après avoir comparé les résultats avec un modèle markovien usuel, nous allons commenter les valeurs obtenues et en donner une interprétation.

A la suite du tableau 5.2, on présente les estimations obtenues pour chaque conducteur en prenant toujours le niveau de contrainte comme variable d'input, en annexe 9.5.

Tout d'abord, les moyennes sont décroissantes en fonction de la contrainte et cela à état de la chaîne fixé. Ceci est naturel et traduit la gêne induite par les autres usagers. Dans la quasi-totalité des cas, on peut voir un des deux états comme un état représentant une phase libre tandis que l'autre est la phase contrainte.

Ensuite, toujours à contrainte fixée, les états associées à des moyennes élevées sont accolées à des variances faibles ; il s'agit là d'une illustration de la stratégie basée sur la vitesse. Lors de la conduite, si l'usager atteint sa vitesse préférentielle, qui peut être variable selon le contexte, alors il maintient cette dernière (variance faible) ; ceci est plus rare s'il se trouve contraint par les autres. Si on prend l'exemple du conducteur 2, on peut voir que pour chaque valeur de la contrainte, la variance de l'état 2 est supérieure à la variance de l'état 1.

Il serait intéressant de pouvoir caractériser chaque état pour chaque contrainte. Il semble, toujours pour l'usager 2, que l'état 1 dans le cas d'une contrainte faible (1) est la phase libre par excellence ; il parvient à sa vitesse moyenne la plus grande. A l'opposé, l'état 2 dans le cas d'une contrainte forte (3) est la phase où la pression par les autres usagers est maximale au sens où il pratique une vitesse moyenne faible, vitesse qui varie considérablement, des écarts de plus ou moins 10 km/h ($\sigma = 9,69$). Cette typologie est applicable à chaque conducteur ; la description des autres configurations (états \times contexte)

n'est pas aisée car non universelle ; malgré tout, on peut regarder la répartition en terme de temps passé dans chaque configuration.

		Contrainte		
		Faible	Moyenne	Forte
Conducteur 2	État 1	17,2%	5,7%	14,1%
	État 2	40,6%	13,8%	8,6%
Conducteur 3	État 1	15,0%	12,4%	28,3%
	État 2	23,3%	11,7%	9,3%
Conducteur 5	État 1	10,7%	5,5%	10,2%
	État 2	45,0%	19,6%	9,0%
Conducteur 7	État 1	12,1%	3,9%	16,3%
	État 2	40,0%	13,4%	14,2%
Conducteur 8	État 1	18,8%	4,7%	8,5%
	État 2	33,8%	22,0%	12,1%

TAB. 5.6 – Répartition du temps dans chaque phase

La première constatation est celle d'une répartition semblable pour tous les conducteurs. Seul le conducteur 3 se différencie par un plus faible pourcentage dans l'état 2 lors des faibles contraintes et par un plus fort pourcentage dans l'état 1 lors des phases de contrainte moyennes et fortes. Cela appuie la constatation faite sur sa faible interdistance pratiquée. En effet, dans les situations où la pression de l'utilisateur externe s'avère moyenne, il opte pour des vitesses très élevées contrairement aux autres usagers.

La majorité du temps est passée dans une configuration de contrainte faible mais pas pour autant dans celle d'une conduite à vitesse maximale (état 2 plutôt que 1)¹, dans le sens préférentiel. De manière générale, dans le cas de contrainte faible ou moyenne, la majorité du temps est passée dans l'état 2.

Cette propension à être dans l'état 2 pour une contrainte faible est assez étonnante. On rappelle que la contrainte faible signifie que personne n'est dans sa file dans les 150 m futurs ; ainsi on devrait plutôt observer un fort pourcentage passé dans l'état 1. Cette configuration peut être vue comme un état transitoire ; cette interprétation est appuyée par les estimations des moyennes obtenues pour l'état 1. Il s'agit de valeurs importantes, au delà des 130 km/h autorisé sur les autoroutes françaises. Cette phase peut être rapprochée des phases de dépassement - rabatement où la vitesse subit une forte augmentation.

5.7.4 Choix de la variable de contexte

Nous avons présenté ci-dessus les résultats de la modélisation du processus vitesse en tenant compte de la variable Contrainte. Bien évidemment, il est intéressant de comparer

¹Ici, tous les états 1 correspondent à des vitesses supérieures à celles de l'état 2, afin de pouvoir comparer les significations des états entre les usagers.

les résultats avec d'autres variables. Ce point, hormis une augmentation de l'adéquation, permet d'avoir une idée sur l'impact des différentes composantes du contexte (type de véhicules ou nombres de véhicules) et sur leur influence sur les changements de phase visibles sur le processus vitesse.

A titre d'exemple, nous avons considéré comme variable de contexte la file empruntée par le conducteur. Ci-dessous nous donnons les résultats obtenus.

Conducteur 2		
HMM	Erreur	$8,78E + 04$
	Proportion	65,91%
IOHMM variable input : file	Erreur	$8,62E + 04$
	Proportion	66,48%
	Gain	1,86%
	Différence entre les chaînes	

TAB. 5.7 – Comparaison HMM et IOHMM (variable d'input : file)

On voit que l'incorporation de la variable file est très peu informatif. Le gain est de 1,86%, de plus les estimations des processus cachés dans les deux modèles aboutissent à des processus quasi semblables, puisque les estimations ne diffèrent que dans 0,83% des cas.

Par cet exemple, il est visible que l'information de la file pratiquée n'apporte guère de connaissances sur la vitesse pratiquée tant en niveau (estimation des moyennes et erreur générale) qu'en dynamique (chaîne cachée quasi identique).

Pour conclure, un inconvénient d'une telle modélisation nous est apparu : il s'agit de la grande sensibilité à l'initialisation. On appelle initialisation les valeurs servant de point de départ à l'algorithme. Ici, il s'agit des valeurs des vitesses et des écarts-types, en insistant plus particulièrement sur les premières. Nous avons choisi dans tous les cas, une valeur initiale commune pour les moyennes indépendamment du contexte pour simplifier, et de même pour les variances.

En effet, nous avons constaté que pour certaines valeurs, pourtant proches, l'algorithme ne parvient pas à converger. Par exemple, si l'algorithme est initialisé avec $V \in [90110]$ pour le conducteur 3, l'algorithme stoppe car la vraisemblance décroît puis croît. Par contre, si les valeurs initiales sont $V \in [110,130]$, alors l'algorithme converge (bien évidemment, la question d'un minimum global ou local n'est pas réglée pour autant). Cette sensibilité semble être résolue dans la majorité des cas par un choix judicieux du point initial. Nous avons choisi de retenir comme valeurs de départ les estimations obtenues par le modèle markovien HMM.

5.8 Algorithme de Viterbi

Tout modèle markovien caché pose deux problèmes

- estimation de θ , paramètre du modèle (moyenne, variance, matrice de transition);

- obtention de la réalisation la plus probable de la chaîne conditionnellement à $Y_{1:T}$ et $C_{1:T}$.

L'algorithme de Baum-Welch s'intéresse au deuxième et l'algorithme de Viterbi au dernier.

Nous avons à plusieurs reprises évoqué l'estimation de la chaîne cachée ; il convient d'explicitier ceci. L'expérience menée permet de recueillir les valeurs de la vitesse, notée (Y_t) . Nous avons fait l'hypothèse que ce processus possède une dynamique régie par un processus caché (X_t) , et que connaissant ce dernier l'estimation des caractéristiques de la vitesse est simplifiée. En contrepartie de cette simplification, on introduit un processus dont aucune valeur n'est disponible.

Il existe un algorithme permettant d'obtenir la réalisation la plus probable de la chaîne cachée, c'est l'algorithme de Viterbi. Par cette méthode, on peut disposer d'une estimation de la chaîne cachée sous-jacente, et cela sans en observer une seule valeur.

L'algorithme de Viterbi est une solution récursive sous-optimale au problème de l'estimation de la séquence d'états dans un HMM. Cet algorithme est également utilisé dans le cadre plus général de la recherche d'un chemin sous-optimal dans un treillis. Nous avons considéré cet algorithme dans notre cas. La modification de l'algorithme se situe au niveau des probabilités de transition qui sont indicées par le contexte.

L'algorithme est itératif et se décompose en deux phases.

On cherche la réalisation de la chaîne la plus probable, notée $\tilde{x}_1^T = (\tilde{x}_1, \dots, \tilde{x}_T)$. Pour cela, on cherche la chaîne définie par :

$$\tilde{x}_1^T = \operatorname{argmax}_{x_1^T} \mathbb{P} [Y_1^T = y_1^T, X_1^T = x_1^T, C_1^T = c_1^T].$$

On note :

$$V(i, T) = \mathbb{P} [Y_1^T = y_1^T, X_1^{T-1} = x_1^{T-1}, X_T = i, C_1^T = c_1^T].$$

La première phase consiste à calculer cette quantité. Pour le calcul, on procède par itération en utilisant la structure du modèle. On écrit donc :

$$V(i, T) = \mathbb{P} [Y_y = y_t | X_t = i] \times \max_j \mathbb{P} [X_t = i | X_{t-1} = j, C_t = c] \times V(j, t-1)$$

En conditionnant par la chaîne, les observations deviennent indépendantes et on peut factoriser et faire apparaître la loi d'émission, la matrice de transition ainsi que la quantité désirée à $t-1$. Par conséquent, cette maximisation peut se faire pas à pas par itération. On initialise l'algorithme avec le calcul suivant :

$$V(i, 1) = \mathbb{P} [Y_1 = y_1 | X_1 = i] \times \mathbb{P} [X_1 = i].$$

On détermine i en maximisant la vraisemblance. En pratique, on peut choisir la valeur de la chaîne à l'état initial. Ceci n'a comme conséquence qu'un réétiquetage de la chaîne. Ensuite, on calcule le terme suivant :

$$V(i, t) = \mathbb{P} [Y_y = y_t | X_t = i] \times \max_j \mathbb{P} [X_t = i | X_{t-1} = j, C_t = c] \times V(j, t-1)$$

La différence par rapport à l'algorithme de Viterbi utilisé dans le cadre des modèles HMM est la présence de variable C_t . Une fois obtenue la quantité $V(i^*, T)$, maximisation

en i de $V(i, T)$, on procède à la seconde phase, la détermination à proprement parler de la chaîne. On procède cette fois-ci en partant de l'instant T et en finissant par la valeur de la chaîne à l'état initial. Avec les notations suivantes :

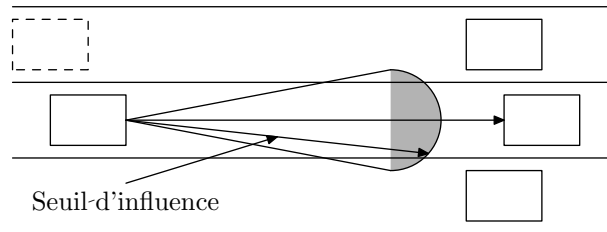
$$\Psi(i, t) = \max_j \mathbb{P} [X_t = i | X_{t-1} = j, C_t = c] \times V(j, t - 1).$$

On détermine la valeur \tilde{x}_t par :

$$\tilde{x}_t = \Psi(\tilde{x}_{t+1}, t + 1).$$

5.9 Notion de seuil d'influence

Dans cette partie on va présenter la notion de seuil d'influence. Il s'agit d'une interdistan-
tance en deçà de laquelle l'usager ne peut *ignorer* la présence du véhicule précédent, sur
la même voie. On peut interpréter ce concept comme une distance minimale au-dessus de
laquelle le conducteur peut se considérer comme en conduite libre.



Sur le schéma précédent, on a tracé la quantité que l'on appelle seuil d'influence. On considère que si un véhicule se situe dans la zone d'influence du véhicule alors le conducteur doit agir en conséquence, et dans la majorité des cas réduire sa vitesse. Dans le cadre proposé ici, cette notion prend une forme précise. En effet, nous avons modélisé la vitesse à l'aide d'un processus caché prenant deux états, et nous avons vu que les états sont associés à des vitesses distinctes en terme de moyenne. Par conséquent, le calcul du seuil revient à calculer la moyenne des interdistances lors des changements d'états. Pour mettre en relief cette valeur, nous avons calculé l'interdistance moyenne pratiquée lors des *dépassements*, plus exactement lors des changements de file avec la présence d'un véhicule à moins de 150 mètres. On présente quelques résultats intéressants dans le tableau 5.8.

Conducteur	Interdistance moyenne	Seuil d'influence
3	56,64 m	30,16 m
5	58,11 m	40,33 m
7	63,21 m	82,43 m
8	54,09 m	63,25 m

TAB. 5.8 – Seuil d'influence

La première constatation est que l'interdistance et le seuil d'influence ne sont pas nécessairement majorés l'un par l'autre. Pour les conducteurs 3 et 5, le seuil d'influence

est inférieur à l'interdistance moyenne pratiquée lors des déboîtements ; ceci est inversé pour les conducteurs 7 et 8. Pour ces derniers, les résultats sont assez logiques dans le sens où lors du rattrapage du véhicule précédent, l'utilisateur passe d'une situation quasi-libre à une situation où la présence du véhicule sur sa voie engendre une gêne traduite par une baisse de la vitesse ; on peut voir qu'il s'agit de valeurs de l'ordre de 70 mètres, ordre de grandeur acceptable. Ensuite, la gêne devient telle que, si le conducteur en a la possibilité, il entame un déboîtement ; pour ces mêmes conducteurs, cette distance est de l'ordre de 60 mètres. La différence entre le seuil et cette distance est faible ; ceci peut être l'illustration que ces conducteurs passent par une phase de transition courte (en terme de distance, mais pouvant être longue en temps si les vitesses sont voisines) avant de tenter de supprimer cette gêne.

A contrario, les conducteurs 3 et 5 évitent cette phase de transition. Les valeurs de seuil et d'interdistances tendent plus à montrer que ces conducteurs amorcent leur déboîtement avant même d'être gêné.

5.10 Optique de comparaison

Avant de conclure, nous allons présenter dans cette partie les possibilités apportées par ces modèles dans l'optique de l'évaluation du système ACC. Comme nous avons signalé précédemment, l'un des points cruciaux à étudier lors de l'introduction d'un tel système est son impact. Plusieurs axes d'évaluation ont été évoqués et en particulier celui de l'évaluation de son impact sur le conducteur. Les psychologues sont très intéressés par ceci, certes en terme d'acceptabilité mais aussi en terme de modification comportementale. Les modèles présentés ici peuvent être des outils permettant de répondre à certaines de leurs interrogations. On va évoquer quelques pistes envisageables, qui malheureusement non pu être plus étudiée par la non disponibilité des données de ces mêmes conducteurs à bord de véhicules équipés du système. Tout d'abord, on peut étudier la modification en terme de vitesse, dans le sens de moyenne ; pour cela les moyennes estimées de chaque état constituent de bons indicateurs. On pourrait constater des baisses (ou augmentations) de la vitesse moyenne de manière générale ou dans une situation caractérisée par l'input considéré. Là, la prise en compte du contexte s'avère cruciale car une comparaison directe des moyennes pratiquées avec et sans le système sera erronée car souffrirait de l'oubli d'une cause importante de la régulation de la vitesse. Pour bien saisir ce point, il suffit d'imaginer une phase sans le système dans un trafic faible et une phase avec le système dans un trafic dense ; alors certes une diminution de la vitesse moyenne serait mesurée, mais il est clair que son attribution au système serait une erreur. Ensuite, une modification des écarts-types traduirait une homogénéisation (ou son contraire) de la vitesse. En effet, si dans la première phase l'estimation mène à des écarts-types σ et que la seconde phase aboutit elle à des valeurs σ inférieurs aux précédents (significativement), alors l'introduction du système aurait comme conséquence une moins grande variabilité de la vitesse pratiquée par l'utilisateur. Ceci a de nombreuses conséquences, un plus grand confort pour le conducteur a priori mais aussi une augmentation du débit du réseau à plus grande échelle. À ces modifications, peuvent s'ajouter des modifications en terme de temps passé dans certaines phases (voir tableau 5.6). La comparaison est très intéressante

sur le plan du comportement de l'utilisateur, car la modification sera due au système, puisqu'on comparera des phases semblables en terme d'input (ici de contrainte). L'évaluation s'exprimera donc en terme : *l'utilisateur évite les situation où il peut être gêné plus qu'il ne le faisait sans le système*. Ce comportement a déjà pu être constaté et traduit souvent la volonté de l'utilisateur d'annihiler le système en optant pour des marges très importantes, court-circuitant le radar de l'*Adaptive Cruise Control*. On insiste ici sur le fait que la volonté n'est pas d'interpréter les modifications dues au système mais plutôt d'isoler les réelles modifications.

5.11 Conclusion

Nous avons présenté les modèles IOHMM en les comparant avec les modèles markoviens cachés couramment utilisés dans de multiples domaines. Nous avons vu que l'incorporation de l'information peut se faire à plusieurs niveaux. L'introduction d'une variable exogène au niveau même de la loi de l'observation n'engendre pas de problèmes et s'avère parfois en pratique intéressant, mais il existe des situations où l'introduction de l'input s'avère naturelle non pas au niveau de l'observation même mais plutôt au niveau du processus caché. Cette modification du modèle originel constitue une nouveauté dans le domaine de la modélisation. Une extension que nous envisageons est la possibilité d'introduire non pas une variable discrète mais une variable continue via une paramétrisation de la *matrice de transition* sous la forme suivante :

$$\mathbb{P}[X_{t+1} = i | X_t = j, C_{t+1} = c] = \exp(-(\alpha_{ij} + \beta_{ij}c)).$$

Cette avancée présente certes des avantages mais au prix d'un inconvénient majeur qui est l'existence intrinsèque d'une telle modélisation et sa validité statistique, base nécessaire à tout algorithme désirant approcher les solutions du problème. Ce point n'est pas le cœur du présent rapport. Nous avons vu en détail les différences sur le plan théorique mais aussi sur les données de vitesse de conducteurs entre les deux approches, et les indicateurs qui en découlent, seuil d'influence, temps passé. Ces derniers mais aussi les paramètres estimés du modèle peuvent servir de base dans les recherches sur l'évaluation de tels systèmes, s'appuyant sur des mesures réalisées en contexte réel non contrôlé.

Chapitre 6

État de l'art sur les copules et la fonction de linkage - Quelques avancées

Une partie des résultats présentées dans ce chapitre sont classiques et font partie de l'état de l'art sur la notion de copule et sur celle plus récente de la fonction de linkage. Nous notons par \star les énoncés originaux.

6.1 Copules

Cette notion date de 1959 et est due à *Sklar*. L'objectif est d'exprimer toute loi bivariée par l'intermédiaire d'une fonction ayant comme argument les marges.

Théorème 2 (Sklar(1959)) *Soit H une loi jointe de marge F et G univariée, alors il existe une copule C telle que :*

$$\forall (x, y) \in \bar{\mathbb{R}}, H(x, y) = C(F(x), G(y))$$

Réciproquement, si C est une copule et F, G des lois de probabilité alors $C(F(x), G(y))$ définit bien une loi jointe de marges F et G .

Si F, G sont continues alors on peut définir C de manière unique par :

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \tag{6.1}$$

Définition 1 (Copule) *Soit $C : \text{Dom}C \rightarrow [0; 1]$. C est une copule si :*

- le domaine de C est $[0; 1]^2$;
- elle satisfait la condition

$$\forall (x, y) \in [0; 1]^2, C(x, 0) = C(0, y) = 0.$$

On dit qu'elle vérifie une propriété nommée grounded.

- Conditions aux bords : $C(u, 1) = u$ et $C(1, v) = v$ pour tout $(u, v) \in [0; 1]^2$;

– 2-croissance : si pour tout $B = [x_1, x_2] \times [y_1, y_2] \subset \text{Dom } C$ alors

$$V_C(B) = C(x_2, y_2) - C(x_1, y_2) - C(x_2, y_1) + C(x_1, y_1) \geq 0.$$

Si on veut être précis, la notion de copule est l'extension de celle de *sous-copule*, où à la différence de la copule le domaine n'est pas nécessairement $[0; 1]^2$.

Avec cette précision, on peut démontrer le théorème 2 à l'aide de deux lemmes.

Lemme 1 *Soit H une loi jointe de marges F, G . Alors il existe une sous-copule C' telle que :*

- $\text{Dom } C' = \text{Im } F \times \text{Im } G$;
- $\forall (x, y) \in \bar{\mathbb{R}}, H(x, y) = C'(F(x), G(y))$.

Lemme 2 *Soit C' une sous-copule. Alors il existe une copule C telle que :*

$$\forall (u, v) \in \text{Dom } C', C(u, v) = C'(u, v)$$

En d'autres termes, toute sous-copule peut être prolongée en une copule. Ce prolongement n'est pas unique.

Démonstration du lemme 1. Tout d'abord, on fait la remarque suivante : si on considère une fonction *grounded* et 2-croissante de marges F, G alors

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |F(x_2) - F(x_1)| + |G(y_2) - G(y_1)|.$$

En effet, par l'inégalité triangulaire, on a :

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|$$

De plus, si $x_1 \leq x_2$, on a :

$$0 \leq H(x_2, y_2) - H(x_1, y_2) \leq F(x_2) - F(x_1).$$

Si $x_2 \leq x_1$, on a l'inégalité opposée d'où le résultat avec la valeur absolue.

On procède de même avec les points y_1 et y_2 . On obtient ainsi l'inégalité voulue.

Cette inégalité va nous servir pour prouver que le nuage de points dans $[0; 1]^3$ suivant est bien celui d'une fonction :

$$\{((F(x), G(y)), H(x, y)), (x, y) \in \bar{\mathbb{R}}\}.$$

En effet, si on considère les couples $(F(x_1), G(y_1))$ et $(F(x_2), G(y_2))$ tels qu'on ait $F(x_1) = F(x_2)$ et $G(y_1) = G(y_2)$ alors par l'inégalité précédente on prouve que $H(x_1, y_1) = H(x_2, y_2)$. Par conséquent, ce nuage de points dans $[0; 1]^3$ définit bien une application. De plus on prouve que l'application ainsi définie est *grounded* et 2-croissante d'après les propriétés de H . Il faut vérifier les conditions aux bords pour conclure qu'il s'agit d'une sous-copule. Si on considère un point u de $\text{Im } F$ alors il existe x tel que $F(x) = u$. Donc si on note C' l'application définie par ce nuage de points, on a :

$$C'(u, 1) = C'(F(x), G(\infty)) = H(x, \infty) = F(x) = u.$$

On procède de même pour $C'(1, v) = v$. Ainsi l'application C' ainsi définie est une sous-copule. ■

Démonstration du lemme 2. Voir Nelsen (1998). ■

On présente un exemple de cette « décomposition » .
Soit la loi H définie sur $\text{Dom } H = [-1; 1] \times [0; +\infty[$ par :

$$H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}$$

alors les marges sont :

$$F(x) = H(x, +\infty) = \frac{x+1}{2} \text{ et } G(y) = H(1, y) = 1 - e^{-y}$$

et par la formule (6.1), on obtient :

$$C(u, v) = \frac{uv}{u + v - uv}.$$

Cet exemple illustre le fait que plusieurs lois jointes peuvent avoir la même copule. En effet rien n'empêche à présent de considérer deux lois unidimensionnelles F', G' quelconques et de construire H' par $C(F', G')$. La classe de telles lois peut s'écrire :

$$\mathcal{C}_C = \{H / H(x, y) = C(F(x), G(y)), \text{ avec } F, G \text{ cdf}\}, \text{ avec } C \text{ fixée.}$$

Dans toute la suite, on se place sur $I = [0; 1]$.

Trois copules sont à distinguer des autres pour leur intérêt bien spécifique. La première notée Π illustre l'indépendance, il s'agit de $\Pi(u, v) = uv$. Les deux autres copules seront abordées dans le lemme 3. Tout d'abord, on a :

$$\forall (u, v) \in I^2, C(u, v) \leq C(u, 1) = u$$

car C est 2-croissante et de marge uniforme. On procède de même avec l'autre variable

$$\forall (u, v) \in I^2, C(u, v) \leq C(1, v) = v$$

donc on a :

$$\forall (u, v) \in I^2, C(u, v) \leq \min(u, v) \stackrel{\text{déf}}{=} M(u, v).$$

Pour la minoration, on s'appuie sur la 2-croissance d'une copule en prenant comme pavé $B = [u; 1] \times [v; 1]$. Ainsi

$$V_C(B) = C(1, 1) - C(u, 1) - C(1, v) + C(u, v) \geq 0.$$

Or $C(1, 1) = 1$, $C(u, 1) = u$ et $C(1, v) = v$ donc

$$C(u, v) \geq u + v - 1.$$

De plus comme une copule est **grounded** et 2-croissante, on en déduit qu'une copule est positive donc :

$$C(u, v) \geq \max(u + v - 1; 0) \stackrel{\text{déf}}{=} W(u, v)$$

Lemme 3 *Les applications M et W sont des copules. De plus, pour toute copule C , on a :*

$$\forall (u, v) \in [0; 1]^2, W(u, v) \leq C(u, v) \leq M(u, v).$$

Ceci permet d'obtenir les bornes, dites de Fréchet-Hoeffding, de l'ensemble $\mathbb{F}(F, G)$, ensemble des lois bivariées ayant comme marges F et G ,

$$\forall H \in \mathbb{F}(F, G), \max(F(x) + G(y) - 1; 0) \leq H(x, y) \leq \min(F(x); G(y)).$$

Des familles paramétriques de copules basées sur les trois copules Π, W et M ont été largement étudiées. On peut citer à titre d'illustration et on pourra se reporter à Nelsen (1998) et Joe (1997) pour une liste très complète.

– famille de Fréchet : $\forall (\alpha, \beta) \in [0; 1], \alpha + \beta \leq 1$

$$C_{\alpha, \beta}(u, v) = \alpha M(u, v) + (1 - \alpha - \beta) \Pi(u, v) + \beta W(u, v)$$

– famille de Mardia : $\forall \theta \in [-1; 1]$

$$C_{\theta}(u, v) = \frac{\theta^2(1 + \theta)}{2} M(u, v) + (1 - \theta^2) \Pi(u, v) + \frac{\theta^2(1 - \theta)}{2} W(u, v)$$

Un aspect intéressant des copules est le lien direct avec la notion de dépendance entre variables aléatoires. Le théorème suivant illustre ceci.

Théorème 3 *Soit X, Y deux variables aléatoires réelles continues de copule $C_{X, Y}$. Si α et β sont deux applications strictement croissante alors on a :*

$$C_{\alpha(X), \beta(Y)} = C_{X, Y}$$

On peut voir que comme la notion de dépendance, celle de copule est invariante par transformation croissante.

Démonstration. Soit $X \rightsquigarrow F_1, Y \rightsquigarrow G_1$, et $\alpha(X) \rightsquigarrow F_2, \beta(Y) \rightsquigarrow G_2$. On peut tout d'abord écrire que $F_2(x) = \mathbb{P}[\alpha(X) \leq x] = \mathbb{P}[X \leq \alpha^{-1}(x)] = F_1(\alpha^{-1}(x))$ et de même on obtient $G_2(y) = G_1(\beta^{-1}(y))$.

La seconde étape consiste à exploiter la notion de copule, en effet par définition,

$$\begin{aligned} H_2(x, y) &= C_{f(X), g(Y)}(F_2(x), G_2(y)) \\ &= \mathbb{P}[\alpha(X) \leq x, \beta(Y) \leq y] \\ &= \mathbb{P}[X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)] \\ &= C_{X, Y}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) \\ &= C_{X, Y}(F_2(x), G_2(y)) \end{aligned}$$

comme X et Y sont continues, l'inversion est possible pour tous les $(x, y) \in [0; 1]^2$, donc on a l'égalité $C_{X, Y} = C_{\alpha(X), \beta(Y)}$. ■

6.2 Simulation d'un échantillon bivarié de copule donnée

Ce paragraphe porte sur la génération d'un n -échantillon bivarié associée à une copule fixée.

Nous allons présenter un algorithme simple que l'on utilisera dans le package *R* développé pour l'extension du modèle de Cox afin de simuler des couples de variables aléatoires associée à une famille de copule spécifiée.

Soit C une copule bivariée. Notons c_u la dérivée première de C par rapport au premier argument

$$c_u(v) = \frac{\partial}{\partial x} C(x, y)|_{x=u, y=v}.$$

Nous notons $v \mapsto c_u^{\leftarrow}(v)$ l'inverse généralisé de l'application $v \mapsto c_u(v)$.

Lemme 4 (Nelsen (1998)) *Soient T_1, T_2 deux variables aléatoires uniformes indépendantes, nous construisons le couple (U, V) comme suit,*

$$\begin{aligned} U &= T_1 \\ V &= c_u^{\leftarrow}(T_2). \end{aligned}$$

alors la loi jointe de (U, V) est la copule C .

Démonstration. Tout d'abord, nous étudions les lois marginales. Bien évidemment, U suit une loi uniforme sur $[0; 1]$. Pour la loi de V , nous avons

$$\begin{aligned} \mathbb{P}[c_U^{\leftarrow}(T_2) \leq x] &= \mathbb{P}[T_2 \leq c_U(x)] \quad \text{Th 2.2.7 Nelsen (1998)} \\ &= \int_0^1 \mathbb{P}[T_2 \leq c_u(x)] \, d\mathbb{P}[U = u] \\ &= \int_0^1 c_u(x) \, du \\ &= [C(u, x)]_0^1 \\ &= C(1, x) - C(0, x) = x \end{aligned}$$

par conséquent, V est une variable aléatoire uniforme.

Étudions à présent la loi jointe du couple (U, V)

$$\begin{aligned} \mathbb{P}[U \leq x, V \leq y] &= \mathbb{P}[U \leq x, T_2 \leq c_U(y)] \\ &= \int_0^1 \mathbb{P}[U \leq x, T_2 \leq c_U(y) | U = u] \, d\mathbb{P}[U = u] \\ &= \int_0^1 \mathbb{P}[U \leq x, T_2 \leq c_u(y) | U = u] \, d\mathbb{P}[U = u] \\ &= \int_0^x \mathbb{P}[T_2 \leq c_u(y)] \, d\mathbb{P}[U = u] \\ &= \int_0^x c_u(y) \, d\mathbb{P}[U = u] \\ &= [C(u, y)]_0^x \\ &= C(x, y) - C(0, y) = C(x, y) \end{aligned}$$

■

Nous venons d'exposer et de démontrer un algorithme de simulation de données à copule donnée. Ce point est important car nous nous n'intéresserons ultérieurement au domaine des valeurs extrêmes, et en particulier à la génération d'échantillons et à l'estimation de la fonction de dépendance. D'autres algorithmes de génération existent mais sont plus restrictifs dans leur domaine de validité; voir par exemple Stephenson (2003). La méthode présentée ici couvre un domaine de validité plus large et permet de générer assez aisément des réalisations suivant une copule donnée donc en particulier une copule de valeurs extrêmes.

6.3 Intérêt des copules – Dépendance

La notion de copule est centrale dans l'étude des dépendances. Comme nous allons le voir, la majorité des indices usuels de dépendance s'exprime uniquement en fonction de la copule, traduisant l'indépendance d'un quelconque indice de dépendance par rapport aux lois des marges.

Lemme 5 *Soit X, Y deux v.a. continues alors on a l'équivalence suivante :*

$$X \perp\!\!\!\perp Y \Leftrightarrow C_{X,Y} = \Pi \text{ avec } \Pi(u, v) = uv.$$

Exemple 1 (Tau de Kendall) *Soit deux variables X, Y unidimensionnelles, on considère X_1, X_2 (respectivement Y_1, Y_2) deux variables indépendantes de même loi que X (respectivement de Y) alors :*

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

Proposition 1 *Soient X, Y deux variables aléatoires absolument continues. On a le résultat suivant :*

$$\tau_{X,Y} = 4 \iint_{I^2} C(u, v) \, dC(u, v) - 1.$$

Remarque 1 *Soient X, Y deux v.a. indépendantes, on sait qu'alors leur copule est nécessairement $\Pi(u, v) = uv$. Le calcul du tau de Kendall mène alors à :*

$$\tau_{X,Y} = 4 \iint_{I^2} uv \, d\Pi(u, v) - 1 = 4 \iint_{I^2} uv \, du \, dv - 1 = 0$$

car $d\Pi(u, v) = \frac{\partial^2 \Pi(u, v)}{\partial u \partial v} = 1 \, du \, dv$.

Exemple 2 Si on regarde la famille de copules Farlie-Gumbel-Morgenstern définie par :

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$$

un calcul aboutit à :

$$\tau_{X,Y} = \frac{2\theta}{9}$$

pour X, Y de copule C_θ , avec $\theta \in [-1; 1]$. Une conséquence de ce résultat est l'obtention du domaine de variation du tau de Kendall pour deux variables dont la copule est dans la famille de Farlie-Gumbel-Morgenstern, $\tau \in [-\frac{2}{9}; +\frac{2}{9}]$.

Exemple 3 Un autre angle de vue concernant la dépendance porte sur le lien entre les queues de lois de répartition.

Soit un couple de variables (X, Y) . Avec les notations usuelles, on s'intéresse à la dépendance des queues de lois de répartition, i.e. à la quantité :

$$\lim_{t \rightarrow +\infty} \mathbb{P}[X \geq t | Y \geq t]$$

or en introduisant la copule sous-jacente, on aboutit à :

$$\begin{aligned} \lim_{t \rightarrow +\infty} \mathbb{P}[X \geq t | Y \geq t] &= \lim_{t \rightarrow \infty} \mathbb{P}[F^{-1}(U) \geq t | G^{-1}(V) \geq t] \\ &= \lim_{t \rightarrow +\infty} \mathbb{P}[U \geq F(t) | V \geq G(t)] \\ &= \lim_{t \rightarrow 1} \mathbb{P}[U \geq t | V \geq t] \\ &= \lim_{t \rightarrow 1} \frac{\mathbb{P}[U \geq t, V \geq t]}{\mathbb{P}[V \geq t]} \\ &= \lim_{t \rightarrow 1} \frac{\bar{C}(t, t)}{1 - t} \text{ où } \bar{C}(t, t) = C(1 - t, 1 - t) \\ &= \lambda_U \end{aligned}$$

Par un calcul similaire, si on regarde la dépendance des queues de lois de répartition à gauche, ou inférieure, on étudie :

$$\lim_{t \rightarrow -\infty} \mathbb{P}[X \leq t | Y \leq t] = \lim_{t \rightarrow 0} \frac{C(t, t)}{t}$$

Après avoir présenté quelques notions de dépendance unidimensionnelles, on se penche sur la dépendance entre des vecteurs aléatoires. La n -copule peut être utilisée dans ce cadre mais celle-ci tient compte de la dépendance interne aux vecteurs. Une notion introduite par Li et al. (1996) est celle de *fonction de linkage*, elle permet de s'intéresser uniquement à la dépendance entre les vecteurs.

Nous allons voir ces notions de dépendance multidimensionnelle et le lien avec la n -copule, extension multivariée de la copule. Puis nous introduisons la notion de *fonction de linkage* plus récente et nous démontrons de multiples résultats de liens existants entre copule et linkage.

6.4 Dépendance multidimensionnelle

On va présenter l'extension multivariée de notions de dépendance unidimensionnelle et montrer que ces notions s'expriment uniquement en fonction de la fonction de linkage, montrant ainsi l'indépendance par rapport aux marges comme dans le cas unidimensionnel.

6.4.1 Mesures de dépendance

Plusieurs indicateurs ont été développés dans ce cadre, les notions de **Positive Upper Orthant Dependent (PUOD)** et **Positive Lower Orthant Dependent (PLOD)**, extension au cas n -varié de la dépendance positive par quadrant. Cette notion classique notée **Positive Quadrant Dependent (PQD)** s'écrit $\mathbb{P}[X > x, Y > y] \geq \mathbb{P}[X > x] \mathbb{P}[Y > y]$. Elle a été généralisée en **Setwise Positive Upper Dependent (SPUSD)** et **Setwise Positive Lower Dependent (SPLSD)** en particulier par Shaked et Shanthikumar (1994).

Pour bien comprendre, ce que l'on va faire en multidimensionnel, on va exposer le principe en unidimensionnel. La notion **Positive Quadrant Dependent (PQD)** exprime une mesure dépendance local en (x, y) par la quantité $\mathbb{P}[X > x, Y > y] - \mathbb{P}[X > x] \mathbb{P}[Y > y]$. Si on veut une mesure globale, il suffit d'intégrer sur \mathbb{R}^2 . On regarde donc :

$$\begin{aligned} D_{X,Y} &= \iint_{\mathbb{R}^2} \mathbb{P}[X > x, Y > y] - \mathbb{P}[X > x] \mathbb{P}[Y > y] \, dF(x) \, dG(y) \\ &= \iint_{\mathbb{R}^2} H(x, y) - F(x)G(y) \, dF(x) \, dG(y). \end{aligned}$$

Cette quantité est le rho de Spearman divisé par 12 (voir théorème 4); il s'agit d'une *mesure d'association*. On a fait le lien entre une mesure de dépendance et une mesure d'association ou de concordance.

Définition 2 (Mesure de dépendance) *Une mesure d'association δ entre deux v.a. continues X, Y de copule C est une mesure de dépendance si elle satisfait :*

- la mesure δ est définie pour tout couple de v.a. ;
- symétrie : $\delta_{X,Y} = \delta_{Y,X}$;
- $0 \leq \delta_{X,Y} \leq 1$;
- $\delta_{X,Y} = 0$ si et seulement si X et Y sont indépendants ;
- $\delta_{X,Y} = 1$ si et seulement si X est presque partout une fonction strictement monotone de Y et inversement ;
- si f et g sont presque sûrement monotone sur $Im X$ et $Im Y$ alors $\delta_{f(X),g(Y)} = \delta_{X,Y}$;
- si $\{(X_n, Y_n)\}$ est une suite de v.a. continues de copule C_n et si C_n converge simplement vers C alors $\lim_{n \rightarrow +\infty} \delta_{C_n} = \delta_C$.

Définition 3 (Rho de Spearman) *Soient $(X_1, Y_1), (X_2, Y_2)$ et (X_3, Y_3) trois vecteurs aléatoires indépendants de même loi H . Le rho de Spearman est défini par :*

$$\rho = 3 (\mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0])$$

Cette quantité est proportionnelle à la différence entre probabilité de concordance et de discordance des deux vecteurs (X_1, Y_1) et (X_2, Y_3) , i.e. entre deux vecteurs ayant même marges mais dont le premier a pour loi jointe H et l'autre le produit des marges (on pourrait utiliser tout aussi bien une formulation avec (X_3, Y_2)). Une autre formulation est la corrélation entre $F_X(X)$ et $G_Y(Y)$, donc :

$$\begin{aligned}\rho_{X,Y} &= \frac{\text{Cov}(F(X), G(Y))}{\sqrt{\text{Var}(F(X))\text{Var}(G(Y))}} \\ &= \frac{\iint F(x)G(y) \, dH(x, y) - \frac{1}{2} \times \frac{1}{2}}{\sqrt{\frac{1}{12} \frac{1}{12}}} \\ &= 12 \iint F(x)G(y) \, dH(x, y) - 3\end{aligned}$$

Théorème 4 Soit X, Y des v.a. dont la copule est C alors le rho de Spearman de X, Y est donné par :

$$\rho_{X,Y} = 12 \iint_{[0;1]^2} C(u, v) \, du \, dv - 3 = 12 \iint_{[0;1]^2} [C(u, v) - uv] \, du \, dv$$

On s'intéresse au cas de n variables unidimensionnelles.

Définition 4 (Positive Upper (Lower) Orthant Dependent (PUOD - PLOD))

Soit n variables aléatoires unidimensionnelles X_i , on dira qu'elles sont à dépendance positive par orthant supérieur (Positive Upper Orthant Dependent (PUOD)) si :

$$\mathbb{P}[X_1 > x_1, \dots, X_n > x_n] \geq \prod_{i=1}^n \mathbb{P}[X_i > x_i].$$

Elles sont à dépendance positive par orthant inférieur (Positive Lower Orthant Dependent (PLOD)) si :

$$\mathbb{P}[X_1 \leq x_1, \dots, X_n \leq x_n] \geq \prod_{i=1}^n \mathbb{P}[X_i \leq x_i].$$

On a la propriété dépendance par orthant supérieur négative (Negative Upper Orthant Dependence (NUOD)) en remplaçant le signe \geq par \leq dans les inégalités principales.

Définition 5 (Positive Dependent by Mixture (PDM)) La loi F est dépendante positive par mélange (Positive Dependent by Mixture (PDM)) si elle vérifie

$$F(x_1, x_2) = \int G(x_1, \alpha)G(x_2, \alpha) \, dM(\alpha)$$

où M est une distribution mélangeante et $G(\cdot, \alpha)$ est une famille de lois de répartition choisie de manière appropriée pour que la représentation soit valide.

Remarque 2 Pour $n = 2$, les propriétés PUOD et PLOD sont identiques car si on suppose que X, Y sont PUOD alors :

$$\begin{aligned} \mathbb{P}[X \leq x, Y \leq y] &= 1 - \mathbb{P}[X > x] - \mathbb{P}[Y > y] + \mathbb{P}[X > x, Y > y] \\ &\geq 1 - \mathbb{P}[X > x] - \mathbb{P}[Y > y] + \mathbb{P}[X > x] \mathbb{P}[Y > y] \quad - \text{PUOD} \\ &\geq \mathbb{P}[X \leq x] - (1 - \mathbb{P}[Y \leq y]) + (1 - \mathbb{P}[X \leq x])(1 - \mathbb{P}[Y \leq y]) \\ &\geq \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y] \quad - \text{PLOD} \end{aligned}$$

Cependant, cette équivalence n'est plus vraie pour $n > 2$, en effet soit \mathbf{X} prenant les valeurs $(1, 1, 1), (1, 0, 0), (0, 1, 0)$ et $(0, 0, 1)$ avec la probabilité $\frac{1}{4}$, alors \mathbf{X} est PUOD mais pas PLOD, on note que $\mathbb{P}[\mathbf{X} \leq \mathbf{0}] = 0$ et $\mathbb{P}[X_1 \leq 0] \mathbb{P}[X_2 \leq 0] \mathbb{P}[X_3 \leq 0] = \frac{1}{8}$.

En terme de lois de répartition, les variables seront PLOD si :

$$\begin{aligned} H(x_1, \dots, x_n) &\geq \prod_{i=1}^n F_i(x_i) \\ C(F_1(x_1), \dots, F_n(x_n)) &\geq \prod_{i=1}^n F_i(x_i) \end{aligned}$$

ceci est équivalent à, par un changement de variable

$$C(u_1, \dots, u_n) \geq u_1 \cdots u_n$$

ou encore, en terme d'ordre sur les copules

$$C \geq \Pi^n$$

Pour résumer et aborder d'autres notions de dépendances, citons :

Proposition 2 Soit H une loi jointe et C la copule associée alors :

- C est PUOD (PLOD) si et seulement si H est PUOD (PLOD) ;
- C est associée si et seulement si H est associée ;
- C est PDM si et seulement si H est PDM.

6.4.2 Mesures d'association multivariées

Soit \mathbf{U} un vecteur (U_1, \dots, U_n) à marges uniformes. Pour obtenir une mesure d'association multivariée, on procède comme ci-dessus en partant d'une mesure locale de dépendance. Par la proposition 2 ci-dessus, cette mesure n'est fonction que de la loi de \mathbf{U} . On définit donc de façon équivalente

$$p \triangleq \int \cdots \int_{[0;1]^n} \mathbb{P}[\mathbf{U} > \mathbf{u}] - \prod_{i=1}^n \mathbb{P}[U_i > u_i] \, du_1 \cdots du_n$$

Cette quantité est l'équivalent du tau pour deux variables. On peut écrire

$$p = \int \cdots \int_{[0;1]^n} \mathbb{P} [\mathbf{U} > \mathbf{u}] \, d u_1 \cdots d u_n = \mathbb{P} [\mathbf{U} > \mathbf{V}]$$

où \mathbf{V} est un vecteur aléatoire à n composantes uniformes et indépendantes. Or

$$\begin{aligned} \mathbb{P} [\mathbf{V} < \mathbf{U}] &= \int_{I^n} \mathbb{P} [\mathbf{V} < \mathbf{u}] \, d C(\mathbf{u}) \\ &= \int_{I^n} u_1 \cdots u_n \, d C(\mathbf{u}) \text{ car } V_i \text{ i.i.d. } V_i \rightsquigarrow \mathcal{U}_{[0;1]} \end{aligned}$$

puisque $\int_{I^n} \prod_{i=1}^n (1 - u_i) \, d u_1 \cdots d u_n = \left(\frac{1}{2}\right)^n$, on aboutit à l'expression suivante :

$$p = \int_{I^n} u_1 \cdots u_n \, d C(\mathbf{u}) - \left(\frac{1}{2}\right)^n \quad (\star)$$

On va normaliser cette quantité. Pour $C = \Pi^n$, on a $p = 0$. Pour $C = M$, on sait que :

$$\iint_{I^2} g(u, v) \, d M(u, v) = \int_0^1 g(u, u) \, d u$$

On montre ce résultat pour $n > 2$. Donc pour $C = M$, on a $p_n = \frac{1}{n+1} - \left(\frac{1}{2}\right)^n$. Du fait de l'ordre entre les copules ($C \leq M$), il vient que pour toute copule C , $p_C \leq \frac{1}{n+1} - \left(\frac{1}{2}\right)^n$. Afin d'avoir une mesure d'association, il faut normaliser \star et on aboutit à la définition suivante.

Définition 6 (Nelsen (1996)) Une mesure d'association multivariée ρ^+ dérivée de la dépendance PUOD est donnée par :

$$\rho^+ = \frac{2^n(n+1)}{2^n - n - 1} \int_{I^n} u_1 \cdots u_n \, d C(\mathbf{u}) - 1$$

On a regardé le cas multivarié ($n > 2$) mais pas le cas où l'on souhaite associer des vecteurs.

Ces notions ont été généralisées au cadre multidimensionnel par Chhetry et al. (1989). Il faut introduire une notion supplémentaire, celle de loi *conditionnellement croissante en suite* (*conditionally increasing in sequence (CIS)*).

Définition 7 (Conditionally increasing in sequence (CIS)) Soient n variables aléatoires unidimensionnelles X_1, \dots, X_n , on dit qu'elles sont *conditionnellement croissante en suite* (*conditionally increasing sequence (CIS)*) si :

$$X_i \uparrow_{st} (X_1, \dots, X_{i-1}), \quad i = 2, \dots, n$$

ou de manière équivalente, si $\mathbb{E} [\Phi(X_i) | X_1 = x_1, \dots, X_{i-1} = x_{i-1}]$ est croissante en (x_1, \dots, x_{i-1}) pour toute fonction croissante Φ .

Définition 8 (Setwise Positive Upper (Lower) Set Dependent (SPUSD - SPLSD))

On considère k vecteurs aléatoires $\mathbf{X}_1, \dots, \mathbf{X}_k$ de dimension m_1, \dots, m_k . On dit qu'ils sont Setwise Positively Upper Set Dependent (SPUSD) si :

$$\mathbb{P} \left[\bigcap_{i=1}^k \{\mathbf{X}_i \in U_i\} \right] \geq \prod_{i=1}^k \mathbb{P} [\mathbf{X}_i \in U_i]$$

Ils sont dits Setwise Positively Lower Set Dependent (SPLSD) si :

$$\mathbb{P} \left[\bigcap_{i=1}^k \{\mathbf{X}_i \in B_i\} \right] \geq \prod_{i=1}^k \mathbb{P} [\mathbf{X}_i \in B_i]$$

Théorème 5 Soient $\mathbf{U}_1, \dots, \mathbf{U}_k$ k vecteurs aléatoires de loi L et F_1, \dots, F_k k lois de répartition multivariées. Soit H la loi jointe des k vecteurs, alors si L est SPUSD (SPLSD) et chaque marge F_i est CIS alors H est SPUSD (SPLSD).

Toutes ces propriétés se démontrent par le théorème suivant.

Théorème 6 Soit $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k) = ((X_{11}, \dots, X_{1m_1}), \dots, (X_{k1}, \dots, X_{km_k}))$ k vecteurs de loi jointe continue, et $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ k vecteurs définis par :

$$\mathbf{Y} = ((g_{11}(X_{11}), \dots, g_{1m_1}(X_{1m_1})), \dots, (g_{k1}(X_{k1}), \dots, g_{km_k}(X_{km_k})))$$

où les fonctions g_{ij} sont strictement croissantes alors \mathbf{X} et \mathbf{Y} ont même fonction de linkage.

Ce résultat est encore vrai si on remplace $g_{ij}(X_{ij})$ par $g_{ij}(X_{i1}, \dots, X_{ij})$ où on a besoin uniquement que $g_{ij}(x_{i1}, \dots, x_{ij})$ est strictement croissante par rapport à la dernière variable.

6.5 La fonction de linkage

On s'intéresse au cas où X, Y sont des vecteurs aléatoires de lois F, G ; $F : \mathbb{R}^m \rightarrow [0; 1]$ et $G : \mathbb{R}^n \rightarrow [0; 1]$. On développe ainsi l'extension de ce qui précède au cas de marges multivariées de la fonction de copule. On désire écrire :

$$\forall (x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{R},$$

$$H(x_1, \dots, x_n, y_1, \dots, y_n) = C(F(x_1, \dots, x_n), G(y_1, \dots, y_n)),$$

où C vérifie les propriétés semblables à celles d'une copule. Or d'après Genest et al. (Genest et al. (1995)), une telle écriture n'est possible que pour des variables indépendantes dans le cas multidimensionnel. Si on impose aux marges d'avoir des densités alors on n'a pas cette implication. Ce résultat s'énonce ainsi :

Théorème 7 (Genest et al. (1995)) Soit $C : [0; 1]^2 \rightarrow [0; 1]$ une application, on suppose que l'équation :

$$H(x_1, \dots, x_m, y_1, \dots, y_n) = C(F(x_1, \dots, x_m), G(y_1, \dots, y_n))$$

définisse une loi à $m + n$ dimensions de marges,

$$\begin{aligned} H(x_1, \dots, x_m, \infty, \dots, \infty) &= F(x_1, \dots, x_m) \\ H(\infty, \dots, \infty, y_1, \dots, y_n) &= G(y_1, \dots, y_n) \end{aligned}$$

et cela pour tout entier m, n / $m + n \leq 3$ alors on a $C(u, v) = uv$ pour $u, v \in [0; 1]$.

Démonstration (Genest et al. (1995)). On suppose pour simplifier que $m = 2$ et $n = 1$; le résultat est indépendant des marges. On suppose que X_1, Y_1 sont des v.a. uniformes sur $[0; 1]$. Supposons que $X_2 = 1 - X_1$. Le triplet $((X_1, X_2), Y_1)$ a alors pour loi $C(\max(x_1 + x_2 - 1; 0), y)$, car :

$$\mathbb{P}[X_1 \leq a, X_2 \leq b] = \mathbb{P}[X_1 \leq a, 1 - b \leq X_1] = a - (1 - b)$$

Par la suite, si on choisit $(s, t) \in [0; 1]$ tel que $s + t \leq 1$ alors on a :

$$\mathbb{P}[s < X_1 \leq s + t, 1 - s < X_2 \leq 1, 0 < Y_1 \leq v] = C(s + t, v) - C(s, v) - C(t, v).$$

Puisque $X_2 = 1 - X_1$,

$$\mathbb{P}[s < X_1 \leq s + t, 1 - s < X_2 \leq 1, 0 < Y_1 \leq v] = \mathbb{P}[s < X_1 < s, 0 < Y_1 \leq v] = 0.$$

Ainsi la copule C vérifie l'équation,

$$C(s + t, v) = C(s, v) + C(t, v).$$

Donc C vérifie l'équation fonctionnelle de Cauchy et d'après le théorème 3, p.48 dans Aczél (1966), on obtient :

$$C(u, v) = u \times C(1, v) = u \times v.$$

■

Remarque 3 Lorsqu'on parle de copule multivariée, on considère des lois unidimensionnelles F_i , $1 \leq i \leq k$, et une loi jointe H et dans ce cadre on écrit :

$$H(x_1, \dots, x_k) = C(F_1(x_1), \dots, F_k(x_k))$$

Ceci est clairement distinct de l'écriture précédente. Cette approche est vue dans le cas de la régression multivariée dans Cuadras (1992), au sens où une variable endogène Y s'écrit comme fonction de plusieurs variables X_i .

Le théorème 7 montre que la notion usuelle de copules est inadaptée dans le cas de marges multivariées. D'autre part, cette approche est restrictive car l'écriture impose que si $x \neq x' \in \mathbb{R}^m$ tels que $F(x) = F(x')$ alors $H(x, y) = H(x', y)$ pour tout $y \in \mathbb{R}^n$. Il s'agit là, d'une justification de l'emploi des *fonctions de linkage*, puisqu'elle permet de décrire la dépendance entre des vecteurs aléatoires et non entre les variables aléatoires scalaires. La fonction de linkage doit être vue comme *une* extension multivariée de la notion de copule.

Pour présenter la notion de linkage, on présente le lemme suivant.

Lemme 6 Soit X_1, \dots, X_n n variables aléatoires de loi jointe F de marge F_i . On note $F_{i|1,2,\dots,i-1}$ la loi de $X_i|X_1, \dots, X_{i-1}$.

Si F est absolument continue (pour avoir la continuité des marges et lois conditionnelles) alors on peut définir les variables U comme suit

$$(U_1, \dots, U_n) \stackrel{\text{déf}}{=} \Psi_F(X_1, \dots, X_n).$$

Les variables aléatoires U_i sont de loi $\mathcal{U}_{[0;1]}$ et indépendantes entre elles, où la fonction Ψ_F est définie par :

$$\Psi_F(x_1, \dots, x_n) = (F_1(x_1), F_{2|1}(x_2|x_1), \dots, F_{n|1,\dots,n-1}(x_n|x_1, \dots, x_{n-1}))$$

Démonstration. Une première étape consiste à démontrer que pour toute variable aléatoire X de loi F , $F(X)$ suit une loi uniforme sur $[0; 1]$; on peut trouver une preuve très détaillée dans Rüschendorf et Valk (1993).

La seconde étape est de regarder la loi de U_2 conditionnée par U_1 . Tout d'abord par construction on a :

$$U_2 = F_{2|1}(X_2|F_1^{-1}(U_1)).$$

Donc on a

$$\begin{aligned} \mathbb{P}[U_2 \leq u_2|U_1 = u_1] &= \mathbb{P}[F_{2|1}(X_2|F_1^{-1}(U_1)) \leq u_2|U_1 = u_1] \\ &= \mathbb{P}[X_2 \leq F_{2|1}^{-1}(u_2)|U_1 = u_1] \\ &= F_{2|1}(F_{2|1}^{-1}(u_2)|F_1^{-1}(u_1)) \\ &= u_2. \end{aligned}$$

La variable aléatoire U_2 suit bien une loi uniforme et de plus U_2 et U_1 sont indépendants. On procède de même pour $i > 2$. ■

Pour l'instant on s'est placé dans le cadre initial de n variables aléatoires unidimensionnelles.

À présent, on considère $\mathbf{X}_1, \dots, \mathbf{X}_k$ k vecteurs aléatoires de loi jointe F . Chaque vecteur \mathbf{X}_i est de loi $F_i : \mathbb{R}^{m_i} \rightarrow [0; 1]$. On notera les « marges des marges » F_{ij} .

L'idée est de construire un vecteur \mathbf{U}_i , image de chaque vecteur \mathbf{X}_i par l'application Ψ_{F_i} . On est bien dans le cadre précédent ; à l'intérieur de chaque vecteur, on s'intéresse aux composantes unidimensionnelles. Une fois présentée cette idée, on peut définir la fonction de linkage.

Définition 9 (Linkage) Soit $\mathbf{X}_1, \dots, \mathbf{X}_k$ k vecteurs aléatoires de loi jointe F et avec les notations précédentes, alors la fonction de linkage L associée à F , ou de manière équivalente au k -uplet de vecteurs aléatoires, est la loi jointe du k -uplet $(\mathbf{U}_1, \dots, \mathbf{U}_k)$ défini par :

$$\mathbf{U}_i = \Psi_{F_i}(\mathbf{X}_i).$$

Remarque 4 *Il est à noter que la définition de la fonction de linkage de vecteurs aléatoires fait intervenir un ordre dans la construction. Il est évident que la fonction de linkage obtenue en modifiant la construction par une renumérotation de composantes de \mathbf{X} est différente de celle définie précédemment. Ce point ne sera pas abordé dans la suite, mais une construction faisant intervenir une permutation aléatoire peut résoudre cet obstacle.*

Pour illustrer cette idée, on va prendre un exemple simple.
Soit $\mathbf{X} = \begin{pmatrix} W_1 & Z_1 \\ W_2 & Z_2 \end{pmatrix}$ où la loi de X est :

$$X \rightsquigarrow \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_W & \rho & \rho \\ \rho_W & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho_Z \\ \rho & \rho & \rho_Z & 1 \end{pmatrix} \right)$$

alors la fonction de linkage est la loi jointe du vecteur

$$\begin{pmatrix} \Phi(W_1) \\ \Phi\left(\frac{W_2 - \rho_W W_1}{\sqrt{1 - \rho_W^2}}\right) \\ \Phi(Z_1) \\ \Phi\left(\frac{Z_2 - \rho_Z Z_1}{\sqrt{1 - \rho_Z^2}}\right) \end{pmatrix}$$

On voit que les transformations sont indépendantes de ρ mais il n'en est pas de même de la fonction de linkage L .

6.5.1 \star - Lien entre la copule et la fonction de linkage dans le cadre bivarié

On se place pour cela dans le cas à marges unidimensionnelles avec $X = (X_1, X_2)$, $X_i \in \mathbb{R}$ de loi F_i . La loi jointe est notée F .

Par définition, L est la loi jointe de $(U_1, U_2) = (\Psi_{F_1}(X_1), \Psi_{F_2}(X_2))$,

$$\begin{aligned} L(u_1, u_2) &= \mathbb{P} [F_1(X_1) \leq u_1, F_2(X_2) \leq u_2] \\ &= F (F_1^{-1}(u_1), F_2^{-1}(u_2)). \end{aligned}$$

Ici l'inversion est parfaitement définie.

D'autre part, on a $F(x_1, x_2) = C (F_1(x_1), F_2(x_2))$. Ainsi on a par inversion :

$$C(u_1, u_2) = F (F_1^{-1}(u_1), F_2^{-1}(u_2))$$

d'où :

$$\forall (u_1, u_2) \in [0; 1], C \equiv L.$$

6.5.2 * - Lien entre la fonction de linkage et la loi jointe multi-dimensionnelle

Soit $\mathbb{Z} = (\mathbf{X}, \mathbf{Y})$ de loi jointe H avec $\mathbf{X} \in \mathbb{R}^m$ de loi F et $\mathbf{Y} \in \mathbb{R}^n$ de loi G .
On a par construction, en notant L la fonction de linkage associée à H ,

$$\begin{aligned} L_{\mathbf{U}, \mathbf{V}}(\mathbf{u}, \mathbf{v}) &= L_{\Psi_F(\mathbf{X}), \Psi_G(\mathbf{Y})}(\mathbf{u}, \mathbf{v}) \\ &= \mathbb{P} [\Psi_F(\mathbf{X}) \leq \mathbf{u}, \Psi_G(\mathbf{Y}) \leq \mathbf{v}] \end{aligned}$$

ici \mathbf{u}, \mathbf{v} sont des éléments de \mathbb{R}^m et \mathbb{R}^n respectivement ; on considère qu'il s'agit d'*orthant* (extension des pavés bidimensionnels en dimension supérieure) pour donner un sens à « $\Psi_F(\mathbf{X}) \leq \mathbf{u}$ » .

Étant sur les orthants, on peut ramener les « inégalités globales » à des inégalités composantes par composantes en écrivant :

$$\begin{aligned} L_{\mathbf{U}, \mathbf{V}}(\mathbf{u}, \mathbf{v}) &= \mathbb{P} [F_1(X_1) \leq u_1, F_{2|1}(X_2|X_1) \leq u_2, \dots, G_1(Y_1) \leq v_1, \dots, \dots] \\ &= \mathbb{P} [X_1 \leq F_1^{-1}(u_1), \dots, Y_1 \leq G_1^{-1}(v_1), \dots] \\ &\stackrel{\text{déf}}{=} H \left(F_1^{-1}(u_1), F_{2|1}^{-1}(u_2|F_1^{-1}(u_1)), \dots, G_1^{-1}(v_1), \dots, \dots \right). \end{aligned}$$

Ainsi on a une relation entre L et H

$$L_{\mathbf{U}, \mathbf{V}}(\mathbf{u}, \mathbf{v}) = H \left(F_1^{-1}(u_1), F_{2|1}^{-1}(u_2|F_1^{-1}(u_1)), \dots, G_1^{-1}(v_1), \dots, \dots \right).$$

On voit que pour simuler la loi H , il est nécessaire d'avoir les lois marginales et celle conditionnelles et L .

Remarque 5 *Pour être précis, on doit écrire :*

$$\begin{aligned} L_{\Psi_F(\mathbf{X}), \Psi_G(\mathbf{Y})}(\mathbf{u}, \mathbf{v}) &= \mathbb{P} [F_1(X_1) \leq u_1, F_{2|1}(X_2|X_1) \leq u_2, \dots, G_1(Y_1) \leq v_1, \dots, \dots] \\ &= \mathbb{P} [\hat{X}_1 \leq F_1^{-1}(u_1), \dots, \hat{Y}_1 \leq G_1^{-1}(v_1), \dots] \\ &\stackrel{\text{déf}}{=} H \left(F_1^{-1}(u_1), F_{2|1}^{-1}(u_2|F_1^{-1}(u_1)), \dots, G_1^{-1}(v_1), \dots, \dots \right) \end{aligned}$$

car on a comme résultat d'après un lemme,

$$\left(\hat{X}_1, \hat{X}_2, \dots, \hat{Y}_1, \dots \right) \stackrel{\mathcal{L}}{=} (X_1, X_2, \dots, Y_1, \dots),$$

où $\left(\hat{X}_1, \hat{X}_2, \dots, \hat{Y}_1, \dots \right) \stackrel{\Delta}{=} (\Psi_{F_1}^*(U_1), \Psi_{F_2}^*(U_2), \dots)$ et Ψ_F^* représente l'« inverse » de Ψ_F :

$$\Psi_F^*(x_1, \dots, x_n) = \left(F_1^{-1}(x_1), F_{2|1}^{-1}(x_2|F_1^{-1}(x_1)), \dots \right).$$

Cette écriture est intéressante car on peut se ramener à une sorte de copule.
En effet, si on se place dans un cas simple $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ et $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, et $(\mathbf{X}, \mathbf{Y}) \rightsquigarrow H$ alors si

on évalue la loi sur l'orthant $A \times B$ avec $A =]-\infty, a_1] \times]-\infty, a_2]$ et $B =]-\infty, b_1] \times]-\infty, b_2]$, on a :

$$\begin{aligned} H(A, B) &= \mathbb{P} [\mathbf{X} \in A, \mathbf{Y} \in B] \\ &= \mathbb{P} [X_1 \leq a_1, X_2 \leq a_2, Y_1 \leq b_1, Y_2 \leq b_2]. \end{aligned}$$

On voudrait écrire alors qu'il existe une fonction f telle que

$$H(A, B) = f(F_1(a_1), F_2(a_2), G_1(b_1), G_2(b_2))$$

ici la fonction f est une application de $[0; 1]^4 \rightarrow [0; 1]$, et non pas de $[0; 1]^2 \rightarrow [0; 1]$ comme pour les copules. En fait, la fonction f existe toujours, il s'agit d'une *copule multivariée* ou *n-copule*.

Définition 10 Une *n-sous-copule* C'_n est une fonction telle que :

- $\text{Dom} C'_n = S_1 \times \cdots \times S_n$ où $\forall i, S_i \subset [0; 1]$ et $\{0; 1\} \in S_i$;
- C'_n est grounded et *n-croissante*, i.e.

$$V_{C'_n}(B) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} \cdots \Delta_{a_1}^{b_1} C'_n(t) \leq 0$$

- C'_n a des marges unidimensionnelles $C'_{n,k}$, $k = 1, \dots, n$ telles que $C_{n,k}(u) = u$.

Définition 11 Une *n-copule* est une *n-sous-copule* dont le domaine est $[0; 1]^n$.

Le théorème suivant est l'extension du théorème (2).

Théorème 8 Soit H une loi jointe *n-dimensionnelle* de marges F_i . Il existe une *n-copule* telle que :

$$\forall \mathbf{x} \in \bar{\mathbb{R}}^n / H(\mathbf{x}) = H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

De même que pour le cas bivarié, il existe des bornes pour les *n-copules*.

Proposition 3 Toute *n-copule* C_n est telle que :

$$W^n(\mathbf{u}) \leq C_n(\mathbf{u}) \leq M^n(\mathbf{u})$$

avec :

$$\begin{aligned} W^n(\mathbf{u}) &= \max\left(\sum_{i=1}^n u_i - n + 1; 0\right) \\ M^n(\mathbf{u}) &= \min(u_1, \dots, u_n) \end{aligned}$$

On doit faire une remarque différenciant les cas univarié et multivarié. En effet, on peut montrer que M^n est une *n-copule* mais qu'il n'en est pas de même pour W^n . Pour cela, on montre que la propriété de *n-croissance* est infirmée pour $n > 2$. On considère l'orthant défini par $[\frac{1}{2}; \mathbf{1}]$ où $\mathbf{1}$ désigne le vecteur $(1, \dots, 1)$, alors on peut montrer que :

$$V_{W^n}\left(\left[\frac{1}{2}; \mathbf{1}\right]\right) = 1 - \frac{n}{2}$$

ainsi pour $n > 2$, $V_{W^n}(B) < 0$.

6.5.3 ★ - Lien entre la loi jointe et la n -copule - Lien entre la n -copule et la fonction de linkage

Soit $I = [0; 1]$, on a les deux formules suivantes :

$$L_{\mathbf{U}, \mathbf{V}}(\mathbf{u}, \mathbf{v}) = H \left(F_1^{-1}(u_1), F_{2|1}^{-1}(u_2|F_1^{-1}(u_1)), \dots, G_1^{-1}(v_1), \dots, \dots \right)$$

et

$$C_n(u_1, \dots, u_n, v_1, \dots, v_n) = H \left(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, G_1^{-1}(v_1), G_2^{-1}(v_2), \dots \right)$$

d'où

$$L_{\mathbf{U}, \mathbf{V}}(\mathbf{u}, \mathbf{v}) = C_n(u_1, F_2(F_{2|1}^{-1}(u_2|F_1^{-1}(u_1))), \dots, v_1, G_2(G_{2|1}^{-1}(v_2|G_1^{-1}(v_1))), \dots)$$

On remarque que L est une application de $I^2 \times I^2$ dans $[0; 1]$ alors que C_n est une application de I^4 dans I . De plus, pour reconstituer la loi jointe, il est nécessaire d'avoir soit le uplet $\{C_n; F_1, \dots, F_n; G_1, \dots, G_n\}$, soit le uplet $\{L; F_1; F_{i|1, \dots, i-1}; G_1; G_{j|1, \dots, j-1}\}$.

On peut regarder l'ensemble $\mathcal{H}(f, g)$ défini par :

$$\mathcal{H}(f, g) = \left\{ H / F_2 \circ F_{2|1}^{-1}(u_2|F_1^{-1}(u_1)) = f(u_1, u_2), \dots, \right. \\ \left. \text{et } G_2 \circ G_{2|1}^{-1}(b_2|G_1^{-1}(b_1)) = g(b_1, b_2), \dots \right\} \quad (6.2)$$

où les points de suspension désignent les égalités entre les différentes composantes de C_n et H . Par construction toutes les lois jointes H de $\mathcal{H}(f, g)$ auront même copules et donc même fonction de linkage.

Un théorème établit cette implication.

Théorème 9 Soient $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$, $\mathbf{X}_i \in \mathbb{R}^{m_i}$ et $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_k)$, $\mathbf{Y}_i \in \mathbb{R}^{m_i}$, si \mathbf{X} et \mathbf{Y} ont même fonction de copule alors ils ont même fonction de linkage.

Démonstration. Voir Shaked et Shanthikumar (1999). ■

La réciproque n'est pas vraie. Il faut pour cela considérer W, Z deux variables aléatoires unidimensionnelles indépendantes. Par exemple, soit

$$\mathbf{X} = \left(\left(\begin{array}{c} Z \\ Z + W \end{array} \right), W \right) \text{ et } \mathbf{Y} = \left(\left(\begin{array}{c} Z \\ 2Z + W \end{array} \right), W \right)$$

alors on montre que \mathbf{X} et \mathbf{Y} ont même fonction de linkage et n'ont pas même copule.

Le résultat pour la fonction de linkage est obtenu en montrant que si \mathbf{X} a pour fonction de linkage L alors tout vecteur de la forme $\left(\left(\begin{array}{c} Z \\ g(Z, W) \end{array} \right), W \right)$ a pour fonction de linkage L si g est une fonction croissante en W (pas nécessairement en Z). On comprend bien ici que la fonction de linkage ne concerne que la dépendance inter-vecteur et non intra-vecteur comme la copule.

6.6 ★ - Une nouvelle copule C^*

On considère dans cette section une nouvelle copule. Elle servira à construire des familles de linkages comme nous le verrons dans la section suivante (paragraphe 6.7).

Toutes les preuves sont données dans la partie 6.8.

Soit \mathcal{P} l'ensemble des lois de répartition sur $I^4 = [0; 1]^4$, on considère l'application :

$$\Phi : \begin{array}{l} \mathcal{P} \rightarrow \mathcal{P} \\ F \mapsto C_F \end{array} \text{ où } C_F \text{ est la copule de la loi } F$$

Proposition 4 *Si $P \in \mathcal{P}$ est à marges uniformes alors $\Phi(P) = P$.*

Dans le même esprit que la proposition 4, on va démontrer une proposition relative à la fonction de linkage. Avant, on définit l'application Ψ par :

$$\Psi : \begin{array}{l} \mathcal{P} \rightarrow \mathcal{P} \\ F \mapsto L_F \end{array} \text{ où } L_F \text{ est la fonction de linkage de la loi } F$$

Proposition 5 *Soit $P \in \mathcal{P}$ telle que $P_2 = P_{2|1}$ et $P_4 = P_{4|3}$, alors $\Psi(P) = \Phi(P)$. Si on fait l'hypothèse supplémentaire de marges uniformes, alors $\Psi(P) = P$.*

Proposition 6 *La fonction de linkage est identique à la n -copule si et seulement si on a indépendance des composantes.*

Proposition 7 *Toute fonction de linkage est une n -copule.*

Ce résultat est en accord avec la proposition 6. En effet si on considère un couple de vecteurs ayant pour loi une fonction de linkage alors la proposition 4 conclut à l'égalité entre L et la copule du couple.

Lemme 7 *Soit C une n -copule alors toute application définie par :*

$$(u_1, \dots, u_n) \mapsto C(u_1, \dots, u_{i-1}, 1, \dots, 1, u_{j+1}, \dots, u_n).$$

Alors C est une k -copule ($k \leq n$).

ainsi on peut regarder $C(u, v, 1, 1)$ comme une copule et en particulier le théorème suivant.

Théorème 10 (Théorème 2.2.7 Nelsen (1998))

Soit C une copule. Pour tout $v \in I$, la dérivée $\frac{\partial C}{\partial u}(u, v)$ existe pour presque tout u et :

$$0 \leq \frac{\partial C}{\partial u}(u, v) \leq 1.$$

(On a un résultat similaire pour $\frac{\partial C}{\partial v}(u, v)$). De plus, les fonctions $u \mapsto \frac{\partial C}{\partial v}(u, v)$ et $v \mapsto \frac{\partial C}{\partial u}(u, v)$ sont définies et non décroissantes sur $[0; 1]$.

(On précise pour presque tout u car les bornes de Fréchet M et W , par exemple, n'admettent pas de dérivée.)

On rappelle que la loi marginale de \mathbf{X} est $C_X : (x, y) \mapsto C(x, y, 1, 1)$ et que la loi de X_2 conditionnellement à $X_1 = x_1$ est donnée par :

$$\mathbb{P} [X_2 \leq x_2 | X_1 = x_1] = \lim_{\Delta x_1 \downarrow 0} \frac{C_X(x_1 + \Delta x_1, x_2) - C_X(x_1, x_2)}{\Delta x_1} = \frac{\partial C_X(x_1, x_2)}{\partial x_1}$$

Lemme 8 Soit (X, Y) de loi F . On note C le copule associée. Alors :

$$\frac{\partial C}{\partial u}(u, F_Y(y))|_{u=F_X(x)} = F_{Y|X=x}(y)$$

Corollaire 1 Avec les notations précédentes, on a

$$\left[\frac{\partial C}{\partial u} \right]^{-1} (u, F_{Y|X=x}(y))|_{u=F_X(x)} = F_Y(y)$$

On peut généraliser ce résultat aux n -copules comme précédemment.

Soit un couple de vecteurs \mathbf{X}, \mathbf{Y} où $\mathbf{X} = (X_1, X_2)'$ et $\mathbf{Y} = (Y_1, Y_2)'$ de loi jointe C . Si les marges de (X, Y) sont uniformes alors on définit

$$\mathbb{P} [X_1 \leq x_1, X_2 \leq x_2, Y_1 \leq y_1, Y_2 \leq y_2] \stackrel{\text{déf}}{=} C(x_1, x_2, y_1, y_2)$$

Définition 12 (Copule étoilée) Soit C une copule, on définit sa version étoilée C^* par :

$$\forall u_1, u_2, v_1, v_2 \in [0; 1]^4, \\ C^*(u_1, u_2, v_1, v_2) = C \left(u_1, \left[\frac{\partial C}{\partial u_1} \right]^{-1} (u_1, u_2, 1, 1), v_1, \left[\frac{\partial C}{\partial v_1} \right]^{-1} (1, 1, v_1, v_2) \right). \quad (6.3)$$

D'après le théorème 10, cette application est parfaitement définie.

Ici, $\left[\frac{\partial C}{\partial u_1} \right]^{-1} (u_1, u_2, 1, 1)$ désigne l'inverse de la fonction $u_2 \mapsto \frac{\partial C}{\partial u_1}(u_1, u_2, 1, 1)$.

Proposition 8 L'application C^* est une n -copule.

Proposition 9 L'application C^* vérifie :

$$\forall (u_1, u_2) \in [0; 1]^2, C^*(u_1, u_2, 1, 1) = u_1 u_2 \\ \forall (v_1, v_2) \in [0; 1]^2, C^*(1, 1, v_1, v_2) = v_1 v_2$$

Proposition 10 L'application C^* est une fonction de linkage.

Théorème 11 L'application C^* et la fonction de linkage sont identiques.

On déduit ce théorème de la stabilité par fonction strictement croissante de la fonction de linkage.

6.7 ★ - Famille de fonctions de linkage

Nous exposons ici une famille de fonctions de linkage construite à partir de la copule C^* .

Exemple 4 *On regarde la famille FGM des n -copules*

$$\begin{aligned} C_\theta(\mathbf{u}) = & u_1 u_2 u_3 u_4 [1 + \{\theta_{12}(1 - u_1)(1 - u_2) + \theta_{13}(1 - u_1)(1 - u_3) \\ & + \theta_{14}(1 - u_1)(1 - u_4) + \theta_{23}(1 - u_2)(1 - u_3) + \theta_{24}(1 - u_2)(1 - u_4) \\ & + \theta_{34}(1 - u_3)(1 - u_4)\} + \{\theta_{123}(1 - u_1)(1 - u_2)(1 - u_3) \\ & + \theta_{124}(1 - u_1)(1 - u_2)(1 - u_4) + \theta_{134}(1 - u_1)(1 - u_3)(1 - u_4) \\ & + \theta_{234}(1 - u_2)(1 - u_3)(1 - u_4)\} \\ & + \theta_{1234}(1 - u_1)(1 - u_2)(1 - u_3)(1 - u_4)]. \end{aligned}$$

On a :

$$\begin{aligned} C_\theta^{12}(u_1, u_2) &= u_1 u_2 [1 + \theta_{12}(1 - u_1)(1 - u_2)] \\ C_\theta^{34}(u_3, u_4) &= u_3 u_4 [1 + \theta_{34}(1 - u_3)(1 - u_4)] \\ \frac{\partial C_\theta^{12}}{\partial u_1}(u_1, u_2) &= c_{u_1}(u_2) \\ c_{u_1}^{-1}(u_2) &= \frac{-1 - a + 2au_1 + \varepsilon \sqrt{\Delta_{12,a}}}{2(2au_1 - a)} & \varepsilon = \pm 1 \\ \Delta_{12,a} &= (1 + a - 2au_1)^2 + 4(2au_1 - a)u_2. \\ c_{v_1}^{-1}(v_2) &= \frac{-1 - e + 2ev_1 + \varepsilon \sqrt{\Delta_{34,e}}}{2(2ev_1 - e)} & \varepsilon = \pm 1 \end{aligned}$$

or on a une fonction positive uniquement pour $\varepsilon = +1$, on aboutit à la famille paramétrique

$$C^*(u_1, u_2, v_1, v_2) = C_\theta(u_1, c_{u_1}^{-1}(u_2), v_1, c_{v_1}^{-1}(v_2)).$$

6.8 Démonstrations

Démonstration de la proposition 1. Calcul direct.

$$\begin{aligned} \tau_{X,Y} &= \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \\ &= 2(\mathbb{P}[X_1 > X_2, Y_1 > Y_2] + \mathbb{P}[X_1 < X_2, Y_1 < Y_2]) - 1 \\ &= 4\mathbb{P}[X_1 > X_2, Y_1 > Y_2] - 1 \end{aligned}$$

et

$$\begin{aligned} \mathbb{P}[X_1 > X_2, Y_1 > Y_2] &= \iint_{\mathbb{R}^2} \mathbb{P}[X_2 < x, Y_2 < y] \, dC(F(x), G(y)) \\ &= \iint_{\mathbb{R}^2} C(F(x), G(y)) \, dC(F(x), G(y)) \\ &= \iint_{[0;1]^2} C(u, v) \, dC(u, v). \end{aligned}$$

■

Démonstration du lemme 5. Soit $X \perp\!\!\!\perp Y$. Alors on a pour tout $(x, y) \in \mathbb{R}^2$, $H(x, y) = F(x)G(y)$. Puisque les marges sont continues, la copule est unique et est déterminée par $C(u, v) = H(F^{-1}(u), G^{-1}(v)) = uv$.

Réciproquement, on suppose que $C = \Pi$. D'après la représentation due à *Sklar*, on a $H(x, y) = C(F(x), G(y)) = F(x)G(y)$, ce qui signifie que X et Y sont indépendants. ■

Démonstration du théorème 5. On a l'équivalence entre la propriété *Setwise Positive Upper Set Dependent (SPUSD)* et le fait que pour toute fonction positive croissante g_i on a :

$$\mathbb{E} \left[\prod_{i=1}^k g_i(\mathbf{X}_i) \right] \geq \prod_{i=1}^k \mathbb{E} [g_i(\mathbf{X}_i)].$$

Or par définition, on peut écrire que :

$$\mathbb{E} \left[\prod_{i=1}^k g_i(\mathbf{X}_i) \right] = \mathbb{E} \left[\prod_{i=1}^k g_i(\Psi_{F_i}^*(\mathbf{U}_i)) \right].$$

Par hypothèse, L est SPUSD donc on a

$$\begin{aligned} \mathbb{E} \left[\prod_{i=1}^k g_i(\mathbf{X}_i) \right] &\geq \prod_{i=1}^k \mathbb{E} [g_i(\Psi_{F_i}^*(\mathbf{U}_i))] \\ &\geq \prod_{i=1}^k \mathbb{E} [g_i(\mathbf{X}_i)]. \end{aligned}$$

Il reste à démontrer que la fonction $g_i \circ \Psi_{F_i}^*$ est croissante. On va prouver qu'il s'agit de la composition de fonctions croissantes.

Lemme 9 Soient X_1, \dots, X_n des v.a. unidimensionnelles de loi jointe H . Si H est *Conditionally Increasing in Sequence* alors $(u_1, \dots, u_n) \mapsto \Psi_H^*(u_1, \dots, u_n)$ est croissante en $(u_1, \dots, u_n) \in [0; 1]^n$.

■

Démonstration du théorème 6. voir Li et al. (1996). ■

Démonstration de la proposition 4. Par définition, on a :

$$\Phi(P)(x_1, x_2, x_3, x_4) = P(P_1(x_1), P_2(x_2), P_3(x_3), P_4(x_4)), \forall (x_1, x_2, x_3, x_4)$$

où P_i désigne la i^{e} marge de la loi P . Par hypothèse, P est à marges uniformes, donc pour tout i , $P_i \equiv Id$, ce qui clôt la démonstration. ■

Démonstration de la proposition 5. Soit (\mathbf{X}, \mathbf{Y}) de loi P , la fonction de linkage $\Psi(P)$ est la loi du couple de vecteurs défini par :

$$\left(\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \right) = \left(\begin{pmatrix} P_1(X_1) \\ P_{2|1}(X_2) \end{pmatrix}, \begin{pmatrix} P_3(Y_1) \\ P_{4|3}(Y_2) \end{pmatrix} \right).$$

L'hypothèse d'égalité entre les lois conditionnelles et les lois marginales impliquent que pour tout $(u_1, u_2, v_1, v_2) \in I^4$:

$$\begin{aligned} \Psi(P)(u_1, u_2, v_1, v_2) &= \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, V_1 \leq v_1, V_2 \leq v_2] \\ &= \mathbb{P}[P_1(X_1) \leq u_1, P_2(X_2) \leq u_2, P_3(Y_1) \leq v_1, P_4(Y_2) \leq v_2] \\ &\stackrel{\text{déf}}{=} \Phi(P)(u_1, u_2, v_1, v_2). \end{aligned}$$

Le résultat sur $\Psi(P) = P$ s'obtient avec l'adjonction de la proposition (4). ■

Démonstration de la proposition 6.

Pour simplifier, prenons $n = 4$:

$$L(F_1(X_1), F_{2|1}(X_2), G_1(Y_1), G_{2|1}(Y_2)) = C(F_1(X_1), F_2(X_2), G_1(Y_1), G_2(Y_2))$$

si on fait l'hypothèse que $L \equiv C$, alors on doit avoir :

$$\begin{cases} F_2(X_2) = F_{2|1}(X_2) \\ G_2(Y_2) = G_{2|1}(Y_2) \end{cases} \iff X_1 \perp\!\!\!\perp X_2 \text{ et } Y_1 \perp\!\!\!\perp Y_2$$

■

Démonstration de la proposition 7.

Soit L une fonction de linkage. Par définition, c'est d'une loi de probabilité sur I^n . Le théorème de *Sklar* affirme qu'il existe une n -copule C telle que :

$$L(x_1, \dots, x_n) = C(L_1(x_1), \dots, L_n(x_n)), \forall (x_1, \dots, x_n) \in I.$$

Dans cette formule, L_i désigne la i^{e} marge de L . Par construction, toutes les marges de L sont des lois uniformes sur I . Par conséquent $L_i \equiv Id$; la proposition est démontrée.

On peut aussi appliquer la proposition suivante pour démontrer cette proposition.

Une troisième méthode pourrait s'appuyer sur la caractérisation d'une n -copule. ■

Démonstration du théorème 10.

L'existence des dérivées partielles est la conséquence de la monotonie des fonctions $u \mapsto$

$C(u, a)$ et $v \mapsto C(a, v)$ (qui sont les sections *horizontales* et *verticales* de la copule C).
 Les inégalités sont issues de la relation :

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

Enfin, la copule étant une fonction 2-croissante, on sait que $u \mapsto C(u, v_2) - C(u, v_1)$ est non décroissante si $v_1 \leq v_2$. Par conséquent, sa dérivée $\frac{\partial[C(u, v_2) - C(u, v_1)]}{\partial u}$ est non négative. Donc l'application $u \mapsto \frac{\partial C(u, v)}{\partial u}$ est non décroissante. ■

Démonstration du lemme 8.

On a :

$$\begin{aligned} F_{Y|X=x}(y) &= \mathbb{P}[Y \leq y | X = x] \\ &= \mathbb{P}[F_Y(Y) \leq F_Y(y) | F_X(X) = F_X(x)] \\ &= \mathbb{P}[V \leq F_Y(y) | U = F_X(x)] \text{ en posant } U = F_X(X), V = F_Y(Y). \end{aligned}$$

Or pour tout $(u, v) \in I^2$, si $(U, V) \rightsquigarrow C$

$$\begin{aligned} \mathbb{P}[V \leq v | U = u] &= \lim_{\varepsilon \downarrow 0} \mathbb{P}[V \leq v | u \leq U \leq u + \varepsilon] \\ &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[u \leq U \leq u + \varepsilon, V \leq v]}{\mathbb{P}[u \leq U \leq u + \varepsilon]} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[U \leq u + \varepsilon, V \leq v] - \mathbb{P}[U \leq u, V \leq v]}{\varepsilon} \\ &= \lim_{\varepsilon \downarrow 0} \frac{C(u + \varepsilon, v) - C(u, v)}{\varepsilon} \\ &= \frac{\partial C}{\partial u}(u, v). \end{aligned}$$

on a donc

$$F_{Y|X=x}(y) = \frac{\partial C}{\partial u}(u, F_Y(y))|_{u=F_X(x)}.$$

On montre de même, pour une n -copule :

$$F_{X_i|X_j=x_j}(x_i) = \frac{\partial C}{\partial u_j}(1, \dots, 1, F_{X_i}(x_i), 1, \dots, 1, u_j, 1, \dots, 1)|_{u_j=F_{X_j}(x_j)}$$

■

Démonstration de la proposition 8.

Il faut montrer les assertions suivantes :

- C^* est une application de I^n dans I ;
 - pour tout $\mathbf{u} \in I^n$ ayant au moins une composante nulle alors $C^*(\mathbf{u}) = 0$;
 - pour tout $\mathbf{u} \in I^n / \mathbf{u} = (1, \dots, 1, u_k, 1, \dots, 1)$ alors $C^*(\mathbf{u}) = u_k$;
 - $\forall (\mathbf{a}, \mathbf{b}) \in I^n, \mathbf{a} \leq \mathbf{b}, V_{C^*}([\mathbf{a}, \mathbf{b}]) \geq 0$.
-

En fait, le couple (U^*, V^*) admet une loi de probabilité à marges uniformes. Par conséquent il s'agit d'une 2-copule. ■

Démonstration de la proposition 9.

On prend les notations classiques en notant $C = H$ puis $C(u, v, 1, 1) = F(u, v)$ et $C(1, 1, u, v) = G(u, v)$ et $\frac{\partial C}{\partial u}(u, F_2(y), 1, 1)|_{u=F_1(x)} = F_{2|1}(y)$. Les résultats s'appuient sur le fait que C est une copule. ■

Démonstration de la proposition 10.

La proposition précédente et les conditions aux marges permettent de conclure. ■

Démonstration du théorème 11.

Il y a unicité de la construction de linkage par :

$$L(u_1, u_2, v_1, v_2) = H \left(F_1^{-1}(u_1), F_{2|1}^{-1}(u_2), G_1^{-1}(v_1), G_{2|1}^{-1}(v_2) \right).$$

On peut aussi le prouver à l'aide du corollaire 1 en écrivant :

$$\begin{aligned} L(x_1, x_2, y_1, y_2) &= H \left(F_{X_1}^{-1}(x_1), F_{X_2|X_1=x_1}^{-1}(x_2), G_{Y_1}^{-1}(y_1), G_{Y_2|Y_1=y_1}^{-1}(y_2) \right) \\ &= C \left(F_{X_1} F_{X_1}^{-1}(x_1), F_{X_2} F_{X_2|X_1=x_1}^{-1}(x_2), \right. \\ &\quad \left. G_{Y_1} G_{Y_1}^{-1}(y_1), G_{Y_2} G_{Y_2|Y_1=y_1}^{-1}(y_2) \right) \\ &= C^* \left(x_1, \frac{\partial C}{\partial x_1}(F_{X_1} F_{X_1}^{-1}(x_1), F_{X_2} F_{X_2|X_1=x_1}^{-1}(x_2), 1, 1), \right. \\ &\quad \left. y_1, \frac{\partial C}{\partial y_1}(1, 1, G_{Y_1} G_{Y_1}^{-1}(y_1), G_{Y_2} G_{Y_2|Y_1=y_1}^{-1}(y_2)) \right). \end{aligned}$$

On va montrer que :

$$\frac{\partial C}{\partial u_1}(u_1, F_{X_2} F_{X_2|X_1=x_1}^{-1}(x_2), 1, 1)|_{u_1=F_{X_1} F_{X_1}^{-1}(x_1)} = x_2.$$

Pour cela, on s'appuie sur le lemme 8 qui permet d'écrire,

$$\frac{\partial C}{\partial x_1}(F_{X_1} F_{X_1}^{-1}(x_1), F_{X_2} F_{X_2|X_1=x_1}^{-1}(x_2), 1, 1) = F_{X_2|X_1=F_{X_1}^{-1}(x_1)} \left(F_{X_2|X_1=x_1}^{-1}(x_2) \right) = x_2$$

On peut démontrer ce théorème plus directement, en s'appuyant sur le théorème 6. Dans notre cas, les fonctions f_i sont les fonctions inverses des lois conditionnelles du vecteur \mathbf{X} (de même pour les fonctions g pour le vecteur \mathbf{Y}); donc elles satisfont les hypothèses du théorème.

On peut simplement utiliser un théorème moins fort (car on utilise uniquement des fonctions à une variable, théorème 3.4, Li et al. (1996)) pour déflater des marges et ramener le problème sur des vecteurs à composantes uniformes. Ainsi on obtient la n -copule sous-jacente, ensuite le calcul de la fonction de linkage mène à la copule étoilée.

On peut directement utiliser ce théorème pour obtenir comme loi de $(\mathbf{X}', \mathbf{Y}')$ la copule étoilée et par la proposition 6. ■

Chapitre 7

Lien entre la transformation du linkage et l'information mutuelle

Ce chapitre expose un résultat novateur liant la notion d'information mutuelle à celle de linkage. Avec ce résultat, l'information mutuelle peut être vue comme un indicateur de dépendance entre vecteurs aléatoires, sans tenir compte d'une quelconque dépendance au niveau interne de chaque vecteur.

7.1 Lien entre l'Information mutuelle et la Copule

On considère deux variables aléatoires X, Y unidimensionnelles réelles. H est leur loi jointe, admettant pour marges F, G . On sait d'après Sklar (1959) que l'on peut toujours définir une copule C , fonction de $[0; 1]^2 \rightarrow [0; 1]$, associée à H telle que :

$$\forall (x, y) \in \text{Dom } H, \quad H(x, y) = C(F(x), G(y)).$$

Pour simplifier les notations, on suppose que $\text{Dom } H = \mathbb{R}^2$.

On considère l'information mutuelle entre la loi jointe H et la loi jointe $F \otimes G$ où $dH = h \, d\lambda^2$ et $dF = f \, d\lambda$, $dG = g \, d\lambda$:

$$I(H, F \otimes G) = \int_{\mathbb{R} \times \mathbb{R}} \log \left[\frac{h(x, y)}{f(x)g(y)} \right] h(x, y) \, dx \, dy.$$

On va montrer que dans le cas de deux variables aléatoires unidimensionnelles, l'information mutuelle est une fonction uniquement de la copule et qu'il s'agit de l'entropie de celle-ci $\mathcal{H}(C)$,

$$\mathcal{H}(C) = \int_{\text{Dom } C} \log(c) c \, d\lambda \quad \text{avec} \quad dC \triangleq c \, d\lambda,$$

λ est une mesure sur $\text{Dom } C$.

Lemme 10 *Soient deux variables aléatoires réelles X, Y de loi jointe H admettant les marges F et G de densités f, g par rapport à la mesure de Lebesgue sur \mathbb{R} . Si on note C la copule associée au couple (X, Y) , i.e.*

$$\forall (x, y) \in \mathbb{R}^2, \quad H(x, y) = C(F(x), G(y)).$$

Si on note (U, V) un couple de variables uniformes ayant pour loi jointe C alors on a :

$$I(H, F \otimes G) = \int_{[0;1] \times [0;1]} \log \left[\frac{\partial^2 C(u, v)}{\partial u \partial v} \right] \frac{\partial^2 C(u, v)}{\partial u \partial v} \, d u \, d v.$$

Ce résultat peut être généralisé au cas multivarié. Il faut d'abord définir l'extension multivariée de l'information mutuelle.

On note H la loi jointe de n variables aléatoires unidimensionnelles ayant respectivement pour loi F_i avec $1 \leq i \leq n$. On note l'information mutuelle qui nous intéresse ici :

$$I(H, \otimes_{i=1}^n F_i) = \int_{\mathbb{R} \times \dots \times \mathbb{R}} \log \left[\frac{h(x_1, \dots, x_n)}{\prod_{i=1}^n f_i(x_i)} \right] h(x_1, \dots, x_n) \, d x_1 \dots d x_n.$$

Dans ce cadre, on considère la copule multivariée C^* associée à H , voir Sklar (1959), Nelsen (1998) pour l'existence de C^* . On peut écrire :

$$H(x_1, \dots, x_n) = C^*(F_1(x_1), \dots, F_n(x_n)).$$

On obtient alors le lemme suivant.

Lemme 11 Soient n variables aléatoires réelles (X_1, \dots, X_n) de loi jointe H admettant les marges F_i avec $1 \leq i \leq n$ ayant des densités par rapport à la mesure de Lebesgue sur \mathbb{R} . Si on note C^* la copule multivariée associée à (X_1, \dots, X_n) , i.e.

$$\forall (x, y) \in \mathbb{R}^2, H(x_1, \dots, x_n) = C^*(F_1(x), \dots, F_n(x_n))$$

et si on note (U_1, \dots, U_n) le n -uplet de variables uniformes ayant pour loi jointe C^* alors on a :

$$I(H, \otimes_{i=1}^n F_i) = \int_{[0;1] \times \dots \times [0;1]} \log [c^*(x_1, \dots, x_n)] c^*(x_1, \dots, x_n) \, d u_1 \dots d u_n$$

où

$$c^*(x_1, \dots, x_n) \triangleq \frac{d C^*}{d \lambda^n} = \frac{\partial^n C^*(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}.$$

On note que ces résultats montrent que l'information mutuelle entre la loi jointe et le produit tensoriel des marges correspond à l'entropie de la copule sous-jacente.

7.2 Information mutuelle dans le cas multidimensionnel

On s'intéresse ici à l'information mutuelle entre deux vecteurs aléatoires \mathbf{X}, \mathbf{Y} avec $\mathbf{X} = (X_1, \dots, X_d)$ et $\mathbf{Y} = (Y_1, \dots, Y_l)$. La loi jointe de (\mathbf{X}, \mathbf{Y}) est H , le vecteur \mathbf{X} a pour loi F , le vecteur \mathbf{Y} a pour loi G et chaque composante X_i admet la loi F_i et Y_j la loi G_j

pour $1 \leq i \leq d$ et $1 \leq j \leq l$.

On considère l'information mutuelle dans le cadre multidimensionnel définie par :

$$I(H, F \otimes G) = \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{h(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})g(\mathbf{y})} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^d, \mathbf{y} \in \mathbb{R}^l$$

Avec cette définition, on a le théorème suivant.

Théorème 12

L'information mutuelle $I(H, F \otimes G)$ avec les notations précédentes se décompose ainsi :

$$I(H, F \otimes G) = \mathcal{H}(C_{d+l}) - \{\mathcal{H}(C_X) + \mathcal{H}(C_Y)\}$$

où $\mathcal{H}(C_{d+l})$ est l'entropie de la copule multivariée reliant H à ses marges F_i, G_j , la quantité $\mathcal{H}(C_X)$ est l'entropie de la copule reliant F à ses marges F_i et $\mathcal{H}(C_Y)$ est l'entropie de la copule reliant G à ses marges G_j .

Si on note :

$$c_{d+l}(u_1, \dots, v_l) \triangleq \frac{d C^{d+l}}{d \lambda^d \otimes d \lambda^l} = \frac{\partial^{d+l} C_{d+l}(u_1, \dots, v_l)}{\partial u_1 \cdots \partial v_l}$$

on obtient :

$$I(H, \otimes_{i,j} F_i G_j) = \int_{[0;1]^{d+l}} \log [c_{d+l}(u_1, \dots, v_l)] c_{d+l}(u_1, \dots, v_l) \, d u_1 \cdots d v_l$$

sous une autre forme on a :

$$I(H, \otimes_{i,j} F_i G_j) = \mathcal{H}(C_{d+l}).$$

7.3 Fonction de linkage

On vient d'obtenir une formule de décomposition de l'information mutuelle entre la loi jointe et les lois marginales d'un couple de vecteurs aléatoires. On peut voir cette formule comme une décomposition de l'information totale $\mathcal{H}(C_{d+l})$ entre une information intra-vectorielle représentée par les termes $\mathcal{H}(C_X) + \mathcal{H}(C_Y)$ et un terme inter-vectoriel $I(H, F \otimes G)$. Il faut démontrer cette dernière assertion. Pour cela l'idée est de considérer un changement de variables intra-vectoriel et de montrer que le terme $I(H, F \otimes G)$ est conservé.

Lemme 12 *L'information mutuelle est invariante par transformation bijective interne au vecteur, i.e. si on considère deux vecteurs aléatoires \mathbf{X}, \mathbf{Y} de loi jointe H , de marge F, G et leurs images \mathbf{X}', \mathbf{Y}' par des transformations bijectives ϕ, ψ alors :*

$$I(H, F \otimes G) = I(H', F' \otimes G').$$

où H' est la loi jointe de $(\mathbf{X}', \mathbf{Y}')$, le vecteur \mathbf{X}' admettant la loi F et le vecteur \mathbf{Y}' la loi G .

Une fois cette assertion vérifiée, la dernière phase va consister à faire le lien avec la fonction de linkage introduite par Li et al. (1996).

Pour introduire cette notion, on considère n variables aléatoires unidimensionnelles (X_i) de loi jointe F avec $X_i \rightsquigarrow F_i$ pour $1 \leq i \leq n$. On va construire n variables aléatoires indépendantes et uniformes sur $[0; 1]$. Pour cela, on examine la transformation Ψ_F définie par :

$$\Psi_F : \mathbb{R}^n \rightarrow [0; 1]^n \\ (x_1, \dots, x_n) \mapsto (u_1, \dots, u_n) = (F_1(x_1), F_{2|1}(x_2), \dots, F_{n|1\dots n-1}(x_n))$$

Lemme 13 Soient les variables $(U_i)_{1 \leq i \leq n}$ définies par :

$$(U_1, \dots, U_n) = \Psi_F(X_1, \dots, X_n).$$

Alors les variables U_i sont uniformes sur $[0; 1]$ et indépendantes.

Lemme 14 Si on suppose que les lois F et G de chaque vecteur admettent des densités strictement positives alors la transformation du linkage Ψ_F est biunivoque.

En résumant, on obtient le théorème suivant.

Théorème 13

Soient deux vecteurs aléatoires (\mathbf{X}, \mathbf{Y}) de loi jointe H . On note sa fonction de linkage L_H et sa copule C^* , et C_X et C_Y les copules multidimensionnelles associées au vecteur \mathbf{X} et \mathbf{Y} respectivement. Si on définit l'information mutuelle $I(H, F \otimes G)$ alors on a le résultat suivant :

$$I(H, F \otimes G) = \mathcal{H}(L_H).$$

D'autre part, on a la décomposition suivante :

$$I(H, F \otimes G) = \begin{cases} \mathcal{H}(L_H) \\ \mathcal{H}(C_{d+l}) + [\mathcal{H}(C_X) + \mathcal{H}(C_Y)] \end{cases}.$$

On peut généraliser ce résultat en considérant p vecteurs aléatoires (\mathbf{X}_i) de loi F_i , le vecteur aléatoire \mathbf{X}_i est à valeurs dans \mathbb{R}^{d_i} , leur loi jointe est H . On note la fonction de linkage associée L_H , la loi du p -uplet $(\mathbf{U}_1, \dots, \mathbf{U}_p)$ défini dans Sklar (1959). On montre qu'on a de même :

$$I(H, \otimes_{i=1}^p F_i) = \begin{cases} \mathcal{H}(C_{d_1+\dots+d_p}) + \sum_{i=1}^p \mathcal{H}(C_{X_i}) \\ \mathcal{H}(L_H) \end{cases}.$$

Remarque 6 Un contre-exemple apparent est le cas suivant.

Soient deux variables aléatoires X, Y de loi H ayant pour marges F et G . On suppose que l'information mutuelle $I(H, F \otimes G)$ est non nulle. Si on effectue le changement de variables $U = F_X(x), V = F_{Y|X=x}(y)$ alors on obtient deux variables U, V indépendantes. Si on suppose ce changement de variables bijectif, il y a conservation de l'information mutuelle, ce qui est absurde, puisque l'information mutuelle de (U, V) est nulle.

Il s'agit d'un contre-exemple en apparence pour le théorème énoncé, car ici le changement n'est pas interne au vecteur (ici réduit à la dimension 1). Par conséquent, la transformation étant inter-vectorielle et non intra-vectorielle, le théorème 13 ne s'applique pas.

7.4 Remarques

1. Les résultats obtenus en multidimensionnel sont cohérents avec ceux en unidimensionnel, car comme montré précédemment, la fonction de linkage coïncide avec la copule usuelle.
2. Un cas particulier illustre l'absence de réciproque à l'implication :

$$L_H = L_{\tilde{H}} \Rightarrow \mathcal{H}(L_H) = \mathcal{H}(L_{\tilde{H}}), \text{ i.e. } \exists L_H \neq L_{\tilde{H}} / I(H, F \otimes G) = I(\tilde{H}, \tilde{F} \otimes \tilde{G}).$$

Il suffit de prendre :

$$(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{pmatrix} \xrightarrow{(\phi, \psi)} (\mathbf{X}', \mathbf{Y}') = \begin{pmatrix} X_2 & Y_1 \\ X_1 & Y_2 \end{pmatrix}$$

Li et al. (1996) ont montré que les fonctions de linkage associées aux couples (\mathbf{X}, \mathbf{Y}) et $(\mathbf{X}', \mathbf{Y}')$ diffèrent. Par contre un calcul simple montre qu'on a égalité des informations mutuelles, confirmant l'indépendance de I par rapport aux transformations internes aux vecteurs, alors qu'il n'en est pas de même de la fonction de linkage. Le point sous-jacent ici est la dépendance de la fonction de linkage par rapport à l'ordre des variables X_i et donc à l'ordre du conditionnement.

7.5 Preuves

Preuve du lemme 10

La première étape consiste en un changement de variable qui va nous ramener à la copule sous-jacente.

Dans l'expression de l'information, on pose $u = F(x)$ et $v = G(y)$, le jacobien est :

$$\det J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial F^{-1}(u)}{\partial u} & 0 \\ 0 & \frac{\partial G^{-1}(v)}{\partial v} \end{vmatrix} = \frac{1}{f[F^{-1}(u)]} \times \frac{1}{g[G^{-1}(v)]}.$$

Une fois la substitution effectuée, on obtient :

$$I(H, F \otimes G) = \int_{[0,1]^2} \log \left[\frac{h(F^{-1}(u), G^{-1}(v))}{f(F^{-1}(u))g(G^{-1}(v))} \right] h(F^{-1}(u), G^{-1}(v)) |\det J| \, du \, dv.$$

La seconde étape consiste à établir les relations entre les densités de (X, Y) et de (U, V) , où $(U, V) \rightsquigarrow C$. Par définition, on a :

$$H(x, y) = C(F(x), G(y)),$$

d'où

$$\begin{aligned} \frac{\partial H}{\partial x}(x, y) &= \frac{\partial C(u, G(y))}{\partial u} \Big|_{u=F(x)} \cdot \frac{\partial F(x)}{\partial x} \\ &= \frac{\partial C(u, G(y))}{\partial u} \Big|_{u=F(x)} \cdot f(x). \end{aligned}$$

L'ordre de dérivation n'importe pas car H admet des dérivées partielles secondes continues sur \mathbb{R} , donc H satisfait les hypothèses du théorème de *Schwarz*. Pour obtenir la dérivée, nous dérivons ensuite par rapport à y ,

$$\begin{aligned} \frac{\partial^2 H(x, y)}{\partial x \partial y} &= \frac{\partial}{\partial y} \left\{ \frac{\partial C(u, G(y))}{\partial u} \Big|_{u=F(x)} f(x) \right\} \\ &= \frac{\partial^2 C(u, v)}{\partial u \partial v} \Big|_{u=F(x), v=G(y)} f(x) \times \frac{\partial G(y)}{\partial y}, \end{aligned}$$

d'où

$$h(x, y) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \Big|_{u=F(x), v=G(y)} f(x)g(y).$$

On utilise ces relations dans l'expression de l'information mutuelle pour trouver l'expression suivante où la valeur absolue du jacobien a été ôtée car ce dernier est positif,

$$\begin{aligned} I(H, F \otimes G) &= \int_{[0;1]^2} \log \left[\frac{h(F^{-1}(u), G^{-1}(v))}{f(F^{-1}(u))g(G^{-1}(v))} \right] h(F^{-1}(u), G^{-1}(v)) \det J \, d u \, d v \\ &= \int_{[0;1] \times [0;1]} \log \left[\frac{\frac{\partial^2 C(F(F^{-1}(u)), G(G^{-1}(v)))}{\partial u \partial v} f(F^{-1}(u))g(G^{-1}(v))}{f(F^{-1}(u))g(G^{-1}(v))} \right] \\ &\quad \times \frac{\partial^2 C(F(F^{-1}(u)), G(G^{-1}(v)))}{\partial u \partial v} f(F^{-1}(u))g(G^{-1}(v)) \det J \, d u \, d v \\ &= \int_{[0;1] \times [0;1]} \log \left[\frac{\partial^2 C(u, v)}{\partial u \partial v} \right] \frac{\partial^2 C(u, v)}{\partial u \partial v} \, d u \, d v \end{aligned}$$

Preuve du lemme 11

La preuve est identique au cas de deux variables.

Preuve du théorème 12

On va procéder en deux étapes. La première est une simple manipulation algébrique de l'expression et la seconde s'appuie sur le paragraphe précédent.

Manipulation algébrique

On va introduire les quantités $\prod_{i=1}^d f_i(x_i)$ et $\prod_{j=1}^l g_j(y_j)$,

$$\begin{aligned}
 I(H, F \otimes G) &= \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{h(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})g(\mathbf{y})} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \\
 &= \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{h(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})g(\mathbf{y}) \frac{\prod_{i=1}^d f_i(x_i) \prod_{j=1}^l g_j(y_j)}{\prod_{i=1}^d f_i(x_i) \prod_{j=1}^l g_j(y_j)}} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \\
 &= \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{h(\mathbf{x}, \mathbf{y})}{\prod_{i=1}^d f_i(x_i) \prod_{j=1}^l g_j(y_j)} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \\
 &\quad - \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{f(\mathbf{x})}{\prod_{i=1}^d f_i(x_i)} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \right. \\
 &\quad \left. + \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{g(\mathbf{y})}{\prod_{j=1}^l g_j(y_j)} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \right\}.
 \end{aligned}$$

Pour les deux dernières intégrales, on peut écrire :

$$\int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{f(\mathbf{x})}{\prod_{i=1}^d f_i(x_i)} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} = \int_{\mathbb{R}^d} \underbrace{\int_{\mathbb{R}^l} h(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}}_{f(\mathbf{x})} \log \left[\frac{f(\mathbf{x})}{\prod_{i=1}^d f_i(x_i)} \right] \, d\mathbf{x}.$$

Ainsi :

$$\begin{aligned}
 \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{f(\mathbf{x})}{\prod_{i=1}^d f_i(x_i)} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} &= \int_{\mathbb{R}^d} \log \left[\frac{f(\mathbf{x})}{\prod_{i=1}^d f_i(x_i)} \right] f(\mathbf{x}) \, d\mathbf{x} \\
 \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{g(\mathbf{y})}{\prod_{j=1}^l g_j(y_j)} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} &= \int_{\mathbb{R}^l} \log \left[\frac{g(\mathbf{y})}{\prod_{j=1}^l g_j(y_j)} \right] g(\mathbf{y}) \, d\mathbf{y}.
 \end{aligned}$$

Introduction des copules

Sous cette forme on peut introduire les copules multidimensionnelles associées au d -uplet (X_1, \dots, X_d) notée C_X et la copule C_Y associée au l -uplet (Y_1, \dots, Y_l) .

La seconde étape s'appuie sur le résultat trouvé dans la partie précédente faisant intervenir la copule multidimensionnelle que l'on note ici C_{d+l} .

Preuve du lemme 12

On considère deux vecteurs aléatoires \mathbf{X}, \mathbf{Y} de loi jointe H , de marge F, G ; on considère deux vecteurs \mathbf{X}', \mathbf{Y}' obtenu par les changements de variables ϕ, ψ de la manière suivante :

$$\begin{pmatrix} X_1 & Y_1 \\ \vdots & \vdots \\ X_d & Y_d \end{pmatrix} = \begin{pmatrix} \phi_1(X_1, \dots, X_d) & \psi_1(Y_1, \dots, Y_d) \\ \vdots & \vdots \\ \phi_d(X_1, \dots, X_d) & \psi_d(Y_1, \dots, Y_d) \end{pmatrix} = \begin{pmatrix} X'_1 & Y'_1 \\ \vdots & \vdots \\ X'_d & Y'_d \end{pmatrix}$$

on prend le cas ici, où $d = l$.

On va montrer que $I(H, F \otimes G) = I(H', F' \otimes G')$. Ensuite on montrera que les transformations menant au linkage satisfont les hypothèse sur ϕ et ψ et qu'ainsi $I(H, F \otimes G) = \mathcal{H}(L_H)$.

On part de l'expression :

$$I(H, F \otimes G) = \int_{\mathbb{R}^d \times \mathbb{R}^l} \log \left[\frac{h(\mathbf{x}, \mathbf{y})}{f(\mathbf{x})g(\mathbf{y})} \right] h(\mathbf{x}, \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y}.$$

On effectue le changement de variable $\mathbf{x}' = \phi(\mathbf{x})$ défini par $x'_i = \phi_i(x_1, x_2)$ (de même pour \mathbf{y}' en remplaçant ϕ par ψ).

Sous l'hypothèse d'un changement biunivoque, on peut écrire :

$$\mathbf{x} = \phi^{-1}(\mathbf{x}') \text{ et } \mathbf{y} = \psi^{-1}(\mathbf{y}')$$

le jacobien d'une telle transformation s'écrit :

$$\det J = \begin{vmatrix} \dots & \vdots & \dots \\ \dots & \frac{\partial x_i}{\partial x'_i} & \dots \\ \dots & \dots & \dots \\ \dots & \frac{\partial y_j}{\partial y'_j} & \dots \\ \dots & \vdots & \dots \end{vmatrix} = \begin{vmatrix} \dots & \dots & \dots \\ \dots & \frac{\partial x_i}{\partial x'_i} & \dots \\ \dots & \dots & \dots \end{vmatrix} \times \begin{vmatrix} \dots & \dots & \dots \\ \dots & \frac{\partial y_j}{\partial y'_j} & \dots \\ \dots & \dots & \dots \end{vmatrix} = \det J_X \cdot \det J_Y.$$

On obtient une factorisation du jacobien par les déterminants des transformations car on ne modifie que la structure interne des vecteurs. On peut substituer ces expressions dans celle de l'information mutuelle,

$$I(H, F \otimes G) = \int_{\phi(\mathbb{R}^d) \times \Psi(\mathbb{R}^d)} \log \left[\frac{h(\phi^{-1}(\mathbf{x}'), \psi^{-1}(\mathbf{y}'))}{f(\phi^{-1}(\mathbf{x}'))g(\psi^{-1}(\mathbf{y}'))} \right] \cdot h(\phi^{-1}(\mathbf{x}'), \psi^{-1}(\mathbf{y}')) |\det J_X \det J_Y| \, d\mathbf{x}' \, d\mathbf{y}' \quad (7.1)$$

La seconde étape consiste à identifier les densités des transformations notées $\tilde{h}, \tilde{f}, \tilde{g}$. On a les relations suivantes :

$$\begin{aligned} \tilde{f}(\mathbf{x}') &= f(\phi^{-1}(\mathbf{x})) |\det J_X| \\ \tilde{g}(\mathbf{y}') &= g(\psi^{-1}(\mathbf{y})) |\det J_Y| \\ \tilde{h}(\mathbf{x}', \mathbf{y}') &= h(\phi^{-1}(\mathbf{x}), \psi^{-1}(\mathbf{y})) |\det J_X \det J_Y| \end{aligned}$$

on voit qu'alors l'expression précédente de l'information mutuelle devient :

$$\begin{aligned} I(H, F \otimes G) &= \int_{\phi(\mathbb{R}^d) \times \Psi(\mathbb{R}^d)} \log \left[\frac{\tilde{h}(\mathbf{x}', \mathbf{y}')}{\tilde{f}(\mathbf{x}') \tilde{g}(\mathbf{y}')} \right] \tilde{h}(\mathbf{x}', \mathbf{y}') \, d\mathbf{x}' \, d\mathbf{y}' \\ &= I(H', F' \otimes G'). \end{aligned}$$

Preuve du lemme 13

On considère pour simplifier le cas où $n = 2$ et (X_1, X_2) admettent une densité jointe f . On va montrer que les variables définies par

$$(U_1, U_2) = (F_1(X_1), F_{2|X_1}(X_2))$$

sont des variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[0; 1]$. Tout d'abord, il faut montrer que de telles variables sont correctement définies par l'expression précédente, et pour cela définir la loi conditionnelle. On définit $F_{2|X_1=x_1}(x_2)$ par :

$$\forall(x_1, x_2), F_{2|X_1=x_1}(x_2) = \lim_{\varepsilon \rightarrow 0} F_{2|X_1 \in \Delta_{x_1}^\varepsilon}(x_2) \text{ où } \Delta_{x_1}^\varepsilon = [x_1 - \varepsilon; x_1 + \varepsilon]$$

or en s'appuyant sur la formule des probabilités conditionnelles, on peut écrire :

$$\begin{aligned} F_{2|X_1 \in \Delta_{x_1}^\varepsilon}(x_2) &= \frac{\mathbb{P}[X_1 \in \Delta_{x_1}^\varepsilon, X_2 \leq x_2]}{\mathbb{P}[X_1 \in \Delta_{x_1}^\varepsilon]} \\ &= \frac{\frac{1}{2\varepsilon} \mathbb{P}[X_1 \in \Delta_{x_1}^\varepsilon, X_2 \leq x_2]}{\frac{1}{2\varepsilon} \mathbb{P}[X_1 \in \Delta_{x_1}^\varepsilon]}. \end{aligned}$$

Si $x \rightarrow \mathbb{P}[X_1 \in \Delta_x^\varepsilon]$ est continue en x_1 alors :

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \mathbb{P}[X_1 \in \Delta_{x_1}^\varepsilon] = f_{X_1}(x_1)$$

et si la fonction $t \rightarrow \int_{-\infty}^{x_2} f(t, y) \, dy$ est continue en x_1 alors

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_{x_1 - \varepsilon}^{x_1 + \varepsilon} \int_{-\infty}^{x_2} f(x, t) \, dt \, dx = \int_{-\infty}^{x_2} f(x_1, t) \, dt.$$

Supposons que $f_{X_1}(x_1) > 0$ et définissons $\mathbb{P}[X_2 \leq x_2 | X_1 = x_1]$ comme la limite quand ε tend vers 0 de $\mathbb{P}[X_2 \leq x_2 | X_1 \in \Delta_{x_1}^\varepsilon]$, on obtient finalement :

$$\begin{aligned} F_{2|X_1=x_1}(x_2) &= \frac{1}{f_{X_1}(x_1)} \int_{-\infty}^{x_2} f(x_1, t) \, dt \\ &= \int_{-\infty}^{x_2} f_{X_2|X_1=x_1}(x_1, t) \, dt \end{aligned}$$

La fonction $f_{X_2|X_1=x_1} : t \rightarrow \frac{f(x_1, t)}{f_{X_1}(x_1)}$ est bien une densité de probabilité puisqu'elle est positive et vérifie :

$$\int_{-\infty}^{+\infty} f_{X_2|X_1=x_1}(x_1, t) \, dt = \frac{1}{f_{X_1}(x_1)} \int_{-\infty}^{+\infty} f(x_1, t) \, dt = \frac{1}{f_{X_1}(x_1)} \times f_{X_1}(x_1) = 1.$$

Ensuite, le second point consiste à montrer que U_1 et U_2 sont indépendantes et de loi uniforme (Rosenblatt (1952)). L'outil pour obtenir ce résultat est le changement de variable dans l'intégrale de *Stieljes*. On s'appuie sur Widder (1947).

Théorème 14 (Widder (1947) - Th. 11a) *Si f est continue, μ est à variation bornée et $x \rightarrow \beta(x)$ continue strictement croissante sur $[a; b]$ alors*

$$\int_a^b f(x) \, d\mu(x) = \int_{\beta(a)}^{\beta(b)} f(\beta(x)) \, d\mu(\beta(x)).$$

Démonstration. On s'appuie sur la définition de l'intégrale et sur l'existence de la limite. En effet, l'intégrale de *Stieljes* est par définition :

$$\int_a^b f(x) d\mu(x) = \lim_{\delta \rightarrow 0} \sum_{i=0}^{n-1} f(\xi_i) [\mu(x_{i+1}) - \mu(x_i)] \text{ avec } x_i \leq \xi_i \leq x_{i+1}$$

où $a = x_0 < x_1 < \dots < x_n = b$ est une subdivision. Si la limite existe et est indépendante de la subdivision et du choix des points ξ_i alors la limite est l'intégrale de *Stieljes*.

Le premier point est l'existence de la limite dans notre cas, ceci est réglé par le théorème suivant :

Théorème 15 (Widder (1947) - Th. 4a) *Si f est continue et μ est à variation bornée sur $[a, b]$ alors l'intégrale de *Stieljes* de f par rapport à μ sur $[a, b]$ existe.*

On va effectuer le changement de variable sur la somme finie puis prendre la limite. On a par définition :

$$\int_a^b f(x) \, d\mu(x) = \lim_{\delta \rightarrow 0} \sum_{i=0}^{n-1} f(x_i) [\mu(x_{i+1}) - \mu(x_i)]$$

on a

$$\sum_{i=0}^{n-1} f(x_i) [\mu(x_{i+1}) - \mu(x_i)] = \sum_{i=0}^{n-1} f(\beta(t_i)) [\mu(\beta(t_{i+1})) - \mu(\beta(t_i))] \text{ où } x_i = \beta(t_i)$$

et par définition

$$\lim_{\delta \rightarrow 0} \sum_{i=0}^{n-1} f(\beta(t_i)) [\mu(\beta(t_{i+1})) - \mu(\beta(t_i))] = \int_{\beta(a)}^{\beta(b)} f(\beta(t)) \, d\mu(\beta(t)).$$

On prouve ainsi l'égalité des deux intégrales et donc le résultat. ■

Théorème 16 (Widder (1947) - Th. 11b) *Si f est continue, μ est continue strictement croissante et si η est l'inverse μ alors :*

$$\int_a^b f(x) \, d\mu(x) = \int_{\mu(a)}^{\mu(b)} f(\eta(x)) \, dx.$$

Ces théorèmes s'appliquent ici car nous avons pour μ des fonctions de répartition absolument continue F_{X_1} et $F_{X_2|X_1}$ donc il s'agit d'une fonction à variation bornée (Billingsley (1995)).

$$\begin{aligned}
\mathbb{P}[U_1 \leq u_1, U_2 \leq u_2] &= \mathbb{P}[F_1(X_1) \leq u_1, F_{2|X_1}(X_2) \leq u_2] \\
&= \int_{t_1 / F_1(t_1) \leq u_1} \int_{t_2 / F_{2|X_1=t_1}(t_2) \leq u_2} d F_{2|X_1=t_1}(t_2) d F_1(t_1) \\
&= \int_0^{u_1} \int_0^{u_2} d t_2 d t_1 \\
&= u_1 \cdot u_2
\end{aligned}$$

Preuve du lemme 14

Il faut montrer que sous des hypothèses acceptables, les fonctions de la forme $u_i = \phi_i(x_1, \dots, x_{i-1}, x)$ sont bijectives. Pour la fonction de linkage, les fonctions u_i sont les transformations marginales (pour $i = 1$) et conditionnelles (pour $i > 1$) appliquées sur les variables initiales X_i .

Une condition suffisante est de supposer que la densité de chaque vecteur est strictement positive. En effet, sous cette hypothèse, on a :

$$f_{X_i|X_1, \dots, X_{i-1}} = \frac{f_{X_1, \dots, X_i}}{f_{X_1, \dots, X_{i-1}}} = \frac{\int \cdots \int_{d-i} f(x_1, \dots, x_d)}{\int \cdots \int_{d-i+1} f(x_1, \dots, x_n)} > 0 \text{ par hypothèse}$$

donc en intégrant, la loi marginale est strictement croissante et continue donc inversible.

Preuve du théorème 13

Il suffit de s'appuyer sur les différents lemmes démontrés. En effet, de par le lemme (12), on a l'égalité :

$$I(H, F \otimes G) = I(H', F' \otimes G')$$

où H' est la loi jointe de (\mathbf{U}, \mathbf{V}) . Le vecteur \mathbf{U} admet la loi F' et le vecteur \mathbf{V} la loi G . Or dans le cas du linkage, les vecteurs \mathbf{U} et \mathbf{V} ont des composantes uniformes et indépendantes. Par conséquent $F \equiv \mathcal{U}_{[0;1]}^{\otimes d}$ et $G \equiv \mathcal{U}_{[0;1]}^{\otimes l}$ et par définition $H' \equiv L_H$. L'égalité devient donc :

$$\begin{aligned}
I(H, F \otimes G) &= I(L_H, \mathcal{U}_{[0;1]}^{\otimes d} \otimes \mathcal{U}_{[0;1]}^{\otimes l}) \\
&= \int \cdots \int \log \left[\frac{l_H(\mathbf{u}, \mathbf{v})}{1 \times \cdots \times 1} \right] l_H(\mathbf{u}, \mathbf{v}) d \mathbf{u} d \mathbf{v} \\
&= \mathcal{H}(L_H)
\end{aligned}$$

1007. LIEN ENTRE LA TRANSFORMATION DU LINKAGE ET L'INFORMATION MUTUELLE

Chapter 8

Bivariate Cox models and dependence analysis

This paper introduces a new class of Cox models for bivariate data. We discuss the role of copulae and the effect of covariates. It is shown that dependence indicators can easily be regressed on the covariates when the bivariate hazard baseline is positively quadrant dependent. The role of extreme-value copulae is enlightened in this respect. As an example we regress the Spearman's rho on covariate inside this model.

8.1 Some model of dependence for duration data

Let (X, Y) be non-negative random variables defined on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), P)$.

Suppose each individual in a homogeneous population is subject to two failure times X and Y which are both observed. Assume also that X and Y are both absolutely continuous (a.c.) non negative random variables (r.v's) with joint cdf (cumulative distribution function) $H_{X,Y}(x, y) \equiv \mathbb{P}[X \leq x, Y \leq y]$ and pdf (probability density function) $h_{X,Y}(x, y)$. The corresponding sdf (survival distribution function) is $\bar{H}_{X,Y}(x, y) \equiv \mathbb{P}[X > x, Y > y]$, and the margins of $H_{X,Y}$ will be denoted by H_X and H_Y , from which $\bar{H}_{X,Y}(x, y) = 1 - H_X(x) - H_Y(y) + H_{X,Y}(x, y)$. The hazard rate of X , also known as the instantaneous or age specific failure rate is

$$\lambda_X(x) \equiv \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[x < X \leq x + \Delta | X > x]}{\Delta}$$

which may conveniently be written as

$$\lambda_X(x) = -\frac{d \log \bar{H}_X(x)}{dx}. \quad (8.1)$$

8.2 Cox models

8.2.1 Univariate Cox models

A Cox model is a well-recognised statistical technique for exploring the relationship between the survival of a patient and several explanatory variables. Let Z_{ij} be the j th

covariate of the i th person, where $i = 1, \dots, n$ and $j = 1, \dots, p$. It is natural to think of the set of covariates as forming an $n \times p$ matrix, and we use Z_i to denote the covariate vector for subject i , that is, the i th row of the matrix. We can consider time-varying or time-fixed covariates. We note X the duration studying. A very important aspect is the study of the impacts of exogeneous variables Z . The most famous models to handle this is the Cox model. It writes :

The Cox model specifies the hazard for individual i as

$$\lambda_{i|\mathbf{Z}=\mathbf{z}_i}(x) = \lambda^0(x)\Phi(\beta, \mathbf{z}_i)$$

where λ^0 is an unspecified nonnegative function of time called the *baseline hazard* and β is a $p \times 1$ column vector of coefficients. The main idea is that the baseline failure time is altered by values of \mathbf{Z} . In the term *baseline failure time*, we mean a average failure time in medical surveys, but we also mean a base failure time, or a base failure rate λ^0 , from which the observations differ from because of values of covariates.

Example 1 In many cases, we take $\Phi(z) = \exp(\alpha z)$ and $z \in \mathbb{R}^1$. So, the sign of α indicate if the failure time is shortened ($\alpha < 0$), lengthened ($\alpha > 0$) or identical ($\alpha = 0$).

Because the hazard ratio for two subjects with fixed covariate vectors \mathbf{Z}_i and \mathbf{Z}_j ,

$$\frac{\lambda_{i|x}}{\lambda_{j|x}} = \frac{\lambda^0(x)\Phi(\beta, \mathbf{Z}_i)}{\lambda^0(x)\Phi(\beta, \mathbf{Z}_j)} = \frac{\Phi(\beta, \mathbf{Z}_i)}{\Phi(\beta, \mathbf{Z}_j)},$$

is constant over time, the model is also know as the *proportional hazards* model.

We conclude this short introduction by the problem of estimation of β . Cox (1972) introduced the notion of *partial likelihood* function. To be precise, we must distinguish the case of untied and tied response X_i .

In the case of untied failure time and uncensored data, the partial likelihood has the form

$$L(\beta; (X_1, \dots, X_n), \mathbf{Z}_1, \dots, \mathbf{Z}_n) = \prod_{i=1}^n \frac{\Phi(\beta; \mathbf{Z}_i)}{\sum_{j \in R(X_i)} \Phi(\beta; \mathbf{Z}_j)}$$

where $R(X_i)$ be the risk set at time X_i , $R(X_i) = \{j : X_j \geq X_i\}$.

In the case of tied failure time data, two adjusted partial likelihood have been proposed. Let $X_{(1)} < X_{(2)} < \dots < X_{(k)}$ be the k ordered event times. Let the set \mathcal{D}_i consist of the d_i individuals who failed at the time X_i . Denote $\mathbf{S}_i = \sum_{l \in \mathcal{D}_i} \mathbf{Z}_l$.

Breslow (1974) suggests a partial likelihood of

$$L_B(\beta; (X_1, \dots, X_n), \mathbf{Z}_1, \dots, \mathbf{Z}_n) = \prod_{i=1}^k \frac{\Phi(\beta; \mathbf{S}_i)}{\sum_{j \in R(X_i)} \Phi(\beta; \mathbf{Z}_j)}$$

Efron (1977) proposed an alternative partial likelihood of

$$L_E(\beta; (X_1, \dots, X_n), \mathbf{Z}_1, \dots, \mathbf{Z}_n) = \prod_{i=1}^k \frac{\Phi(\beta; \mathbf{S}_i)}{\prod_{j=1}^{d_i} \left[\sum_{l \in R(X_i)} \Phi(\beta; \mathbf{Z}_l) - \frac{j-1}{d_i} \sum_{l \in \mathcal{D}_i} \Phi(\beta; \mathbf{Z}_l) \right]}$$

Note that when the number of ties is small, Breslow's and Efron's likelihoods are quite close. The Breslow likelihood is usually preferred to Efron's since is is quite straight forward to compute. Of course, if no ties occur at the event times, both likelihood functions reduce to the regular Cox's partial likelihood.

8.2.2 Bivariate Cox models

Now, we observed the couple of failure time (X, Y) .

The term of *multivariate survival data* covers the field where independence between survival times cannot be assumed. Two main types of data structures exist,

- parallel data;
- longitudinal data.

For the first type, the dependence appears between simultaneous measures, or in the second one the dependence is temporal one. Often, the second type is rather considered stochastic processes data, or *multiple time data*.

We will study only the first type, *parallel data*, (Hougaard (2000) p.9).

Regression models aim at modeling dependence upon explanatory variables. In the proportional hazard model the cause specific hazard functions satisfy

$$\begin{cases} \lambda_{X_{ij}}(x) = \lambda_{X_{ij}}^0(x)\Phi(\mathbf{z}_i) \\ \lambda_{Y_{ij}}(y) = \lambda_{Y_{ij}}^0(y)\Psi(\mathbf{z}_i), \end{cases} \quad (8.2)$$

where the data is (x_{ij}, y_{ij}, z_i) , $j = 1, \dots, n_i$, $i = 1, \dots, n$, and for each i the observed covariate \mathbf{z}_i belongs to \mathbb{R}^d . Often the positive functions $\Phi(z)$ and $\Psi(z)$ are assumed to be parametric functions, with the standard $\Phi(z) = \exp(\alpha'z)$ and $\Psi(z) = \exp(\beta'z)$, $\alpha, \beta \in \mathbb{R}^d$.

Investigators are rarely interested only in marginal behaviors as described in (8.2). This model ignores the dependence between X and Y and can only be of poor interest for practical applications.

With the extension to bivariate cases, we must introduce a *conditional* hazard rate. Let

$$\lambda_{Y|X=x}(y) \equiv \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[y < Y \leq y + \Delta | X = x, Y > y]}{\Delta} \quad (8.3)$$

and

$$\lambda_{Y|X>x}(y) \equiv \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}[y < Y \leq y + \Delta | X > x, Y > y]}{\Delta}. \quad (8.4)$$

Modeling the multivariate dependency through (8.3) and (8.4) leads to different types of models, namely

$$\begin{cases} \lambda_{X_i}(x) = \lambda_{X_i}^0(x)\Phi(z_i) \\ \lambda_{Y_i|X_i=x}(y) = \lambda_{Y_i|X_i=x}^0(y)\Psi(z_i) \end{cases} \quad (8.5)$$

or

$$\begin{cases} \lambda_{X_i}(x) = \lambda_{X_i}^0(x)\Phi(z_i) \\ \lambda_{Y_i|X_i>x}(y) = \lambda_{Y_i|X_i>x}^0(y)\Psi(z_i). \end{cases} \quad (8.6)$$

variables.

We are interested mainly in positive dependency between the two variables. Both models characterize the property that the failure of one component puts an extra load on the other component, as for example in studies involving a two organs system; for example, time to failure for several similar human organs, like time to blindness of the

right and left eye. The same situation arises for kidneys, although it is more difficult to follow the individual status for each kidney. Such data are not common in medicine, but similar data are particularly relevant in industrial applications.

Direct approaches based on regression type models cannot deserve our purpose. Indeed, consider for example a model defined through

$$\begin{cases} X = f(z, U) \\ Y = g(z, V) \end{cases}$$

with $f(z, \cdot)$ and $g(z, \cdot)$ strictly increasing for all z . then following Nelsen (1998) Th. 2.4.3., (X, Y) has same copula as (U, V) for all z , which motivates this paper.

With these new models, a important question appears: Are such models mathematically valid ? That is, Is there a cdf H^z for the r.v. (X, Y) and a baseline cdf F^0 which are compatible with them ? Ignoring technicalities, this is indeed always the case for Model (8.5). Model (8.6) is valid under certain restriction on the baseline hazard, as will be seen later. Model (8.5) has been widely studied, e.g. DeMasi (1994). Indeed the usual paradigm in regression analysis is conditioning upon the observed value of the variable X . Note however that in the present setting the covariate is in fact z . The main difficulty with the first model is that marginally it does not satisfy (8.2), the Cox paradigm, which trivially holds true for the model (8.6) with the setting $x = 0$.

Before to study this question, we look into the Cox paradigm for the first model.

8.2.3 Model based on $\lambda_{Y|X=x}$

The model (8.5) (see DeMasi (1994)) has been widely studied. There exists several reasons for this model. First, it always exists a distribution function H^z . Moreover, it is relevant to compute the likelihood, see DeMasi (1994) for details.

But, two drawbacks can be done.

1. Indeed the usual paradigm in regression analysis is conditioning upon the observed value of the variable X_{ij} . Note however that in the present setting the covariate is in fact z .

2. The main difficulty with the first model is that marginally it does not satisfy (8.2), the Cox paradigm,

$$\begin{aligned} \lambda_Y^z(y) &= \lim_{\Delta} \frac{1}{\Delta} \mathbb{P}_{H_Y^z} [y < Y \leq y + \Delta | Y > y] \\ &= \lim_{\Delta} \frac{1}{\Delta} \int \mathbb{P}_{H_{X,Y}^z} [y < Y \leq y + \Delta | Y > y, X = x] d H_X^z(x) \\ &= \lim_{\Delta} \frac{1}{\Delta} \int \lambda_{Y|X=x}^z(y) d H_X^z(x) \\ &\stackrel{8.5}{=} \lim_{\Delta} \frac{1}{\Delta} \int \lambda_{Y|X=x}^0(y) \Psi(z) d H_X^z(x) \\ &= \Psi(z) \lim_{\Delta} \frac{1}{\Delta} \int \lambda_{Y|X=x}^0(y) d H_X^z(x) \end{aligned}$$

This paradigm trivially holds true for the model (8.6) with the setting $x = 0$,

$$\lambda_{Y_i|X_i>x} \mathbf{z}_i(y) = \lambda_{Y_i|X_i>x}^0(y) \Psi(\mathbf{z}_i),$$

we obtain $\lambda_{Y_i} \mathbf{z}_i(y) = \lambda_{Y_i}^0(y) \Psi(\mathbf{z}_i)$, so this models satisfies the Cox paradigm for Y .

8.2.4 Statistical estimation

For statistical estimation both models have respective interest. The likelihood function in the first one can easily be written, and hence the parameters can be estimated by partial likelihood maximization. The second model allows a straightforward estimation of the parameters, which are defined independently upon the dependence structure induced by the model, setting $x = 0$ in the second equation, and applying standard univariate estimation for the parameters of the functions Φ and Ψ . Clearly model (8.6) leads to an easy description of the sdf of the r.v. (X, Y) , while (8.5) is more adequate to handle the properties of the pdf. This difference motivates our interest in model (8.6). Since the sdf can easily be written as a function of the baseline and the covariate, it represents a natural model for the regression of basic dependence indices on the covariate.

Let us now sketch the main results which we present in connection with model (8.6):

1. The class of totally positive distributions of order two (TP2) is stable under the model, and the model is properly defined when the hazard baseline F^0 is TP2. This class appears quite naturally as the one under which the model is properly defined, and it is appropriate for the description of positive dependence between its margins.
2. Since the TP2 property of a multivariate cdf refers only to its copula, model (8.6) describes the changes of the baseline copula induced by the covariate. Also this implies that the model is valid independently upon the marginal distributions. Only the structure of dependence is involved in the domain of validity of the model.
3. A large subclass of copulae satisfies the TP2 property, namely the *extreme value copulae*, which are parametrized through a class of *dependence functions* which are easy to estimate.
4. Any TP2 df is positive quadrant dependent (PDQ). We can show that to some extent, any PQD copula can be well approximated by an extreme value copula, which legitimates the restriction of the model to such class, which is also stable under the model. Inside this class the covariate z acts in an explicit and easy way.
5. All dependence parameters and indicators are captured through the copula. For example the celebrated Spearman's ρ , or the Kendall's τ , can be written explicitly as integral functional of the copula. Inside the class of extreme value copulae, their expression is even simpler. Thus, we can model the way these parameters depend on the covariate.

8.3 Some useful facts in bivariate dependence

Throughout this section, F generically denotes a cdf on \mathbb{R}^2 , and F_X, F_Y denotes its margins. We describe the toolbox which is needed in order to study the bivariate model (8.6). For simplicity, we consider only distributions which are absolutely continuous.

8.3.1 Max-id distribution - Min-id distribution - Positive Quadrant Dependence (PQD)

Definition 1 A bivariate cdf F is minimum infinitely divisible (min-id), if for all $t > 0$, \bar{F}^t is a sdf. Analogously, F is maximum infinitely divisible (max-id) if for all $t > 0$, F^t is a cdf.

Note that all distributions on the real line are both min-id and max-id. This is not true on the square — the mixed second derivative of F^t may fail to be positive.

Definition 2 (Totally positive of order 2 (TP2)) A non-negative function F is totally positive of order 2 (TP2) if for all $x_1 < x_2, y_1 < y_2$ $F(x_1, y_1)F(x_2, y_2) \geq F(x_1, y_2)F(x_2, y_1)$. If F is a differentiable function, F is TP2 iff :

$$\frac{\partial F}{\partial x}(x, y) \times \frac{\partial F}{\partial y}(x, y) \leq \frac{\partial^2 F}{\partial x \partial y}(x, y) \times F(x, y).$$

Proof. The proof of the second point could be found in Resnick (1987), p.254, Proposition 5.3. ■

Definition 3 (Positive Quadrant Dependent) Let (X, Y) r.v. with joint cdf $F, (X, Y)$ or F is positive quadrant dependent (PQD) if :

$$\mathbb{P}_{F_{X,Y}} [X > x, Y > y] \geq \mathbb{P}_{F_X} [X > x] \mathbb{P}_{F_Y} [Y > y] \quad \text{for all } (x, y) \in \mathbb{R}^2$$

or in other terms

$$F_{X,Y}(x, y) \geq F_X(x)F_Y(y).$$

Lemma 1 If F is TP2 then F is Positive Quadrant Dependent (PQD), i.e. for all (x, y) , $F(x, y) \geq F_X(x)F_Y(y)$.

Proof. Theorem 2.3 p.26 Joe (1997). ■

The following theorem holds.

Theorem 1 Let F be a bivariate cdf, then :

(i) F is max-id iff F is TP2;

(ii) F is min-id iff \bar{F} is TP2.

Proof. voir Joe (1997), Th 2.6 p.30 . ■

Copulas - Empirical Copulas

The study of copulas and their applications in statistics is a rather modern phenomenon. Copula was first employed by Sklar (1959), but they appear implicitly in earlier work on dependence by many authors. We give only two famous one. *Hoeffding* used them to study nonparametric measures of association such as Spearman's rho in his papers (1940, 1941). Deheuvels (1978, 1979, 1981b,a) used empirical dependence functions (*i.e.* empirical copulas) to estimate the population copula and to construct various nonparametric tests of independence.

Let (X, Y) be two rv's with a joint c.d.f F and marginal distributions F_X and F_Y .

Definition 4 (Copula) *A two-dimensional copula is a function C from $I^2 = [0; 1]^2$ to I with the following properties :*

- for every $u, v \in I$,

$$C(u, 1) = u \quad \text{and} \quad C(1, v) = v.$$

- C is grounded *i.e.*

$$C(u, 0) = C(0, v) = 0$$

- C is a 2-increasing function, *i.e.* for every $u_1, u_2, v_1, v_2 \in I$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Theorem 2 (Sklar (1959)) *Let F be a joint distribution function with margins F_X and F_Y , then there exists a copula C such that for all $x, y \in \mathbb{R}$,*

$$F(x, y) = C(F_X(x), F_Y(y))$$

If F_X and F_Y are continuous, then C is unique; otherwise C is uniquely determined on $\text{Im } F_X \times \text{Im } F_Y$. Conversely, if C is a copula and F_X and F_Y are distribution functions, then the function F defined above is a joint distribution function with margins F_X and F_Y .

Proof.Nelsen (1998). ■

Their dependence properties are summarized through the copula function C , defined on $I \equiv [0, 1] \times [0, 1]$ by:

$$C(u, v) = F(F_X^{\leftarrow}(u), F_Y^{\leftarrow}(v)).$$

where F_X^{\leftarrow} is the inverse of F_X defined by

$$F_X^{\leftarrow}(u) = \inf \{x : F_X(x) \leq u\}$$

Example 2 (Examples of copulas) *We give here some examples of copulas,*

- if X and Y are independent then $C(u, v) = \Pi(u, v) = uv$;
- copula of positive comonotony, *i.e.* it exists an increasing function f such that $Y = f(X)$, then $C(u, v) = M(u, v) = \min(u, v)$;

- *copule of negative comonotony*, i.e. it exists a decreasing function f such that $Y = f(X)$, then $C(u, v) = W(u, v) = \max(u + v - 1, 0)$.

Theorem 3 (Fréchet-Hoeffding bounds) *Let C be a copula, then for every $(u, v) \in \text{Dom} C$,*

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v). \quad (8.7)$$

Proof.Nelsen (1998). ■

We now introduce the notion of empirical copula.

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ a n -sample of $[X, Y]$ with joint distribution F , F_X , F_Y are margins. We define empirical joint distribution F_n by

$$F_n(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x, Y_i \leq y\}}$$

Empirical margins is also defined by

$$F_{X,n}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \quad F_{Y,n}(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq y\}}$$

By using the definition of the generalized inverse, we consider $F_{X,n}^{\leftarrow}$ (and $F_{Y,n}^{\leftarrow}$). So now we can define empirical copula C_n by

$$C_n(u, v) = F_n(F_{X,n}^{\leftarrow}(u), F_{Y,n}^{\leftarrow}(v)).$$

Since $F_{X,n}$ is right continuous, we have the equivalence

$$x \leq F_{X,n}^{\leftarrow}(u) \text{ iff } F_{X,n}(x) \leq u$$

We rewrite the empirical copula

$$\begin{aligned} C_n(u, v) &= F_n(F_{X,n}^{\leftarrow}(u), F_{Y,n}^{\leftarrow}(v)) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq F_{X,n}^{\leftarrow}(u); Y_i \leq F_{Y,n}^{\leftarrow}(v)\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq F_{X,n}^{\leftarrow}(u); Y_i \leq F_{Y,n}^{\leftarrow}(v)\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{F_{X,n}(X_i) \leq u; F_{Y,n}(Y_i) \leq v\}} \end{aligned}$$

We consider the *rank* of X_i among X_1, \dots, X_n , denoted r_{X_i} ; then $r_{X_i} = nF_{X,n}(X_i)$. So, the empirical copula could be written as the following form that we will use

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{r_{X_i} \leq nu; r_{Y_i} \leq nv\}}.$$

8.3.2 Survival Copulas

In many applications, the random variables of interest represent the lifetimes of individuals or objects in some population.

In the following, let (U, V) be the bivariate r.v. $(F_X(X), F_Y(Y))$. Clearly, $C(u, v) = \mathbb{P}[U \leq u, V \leq v]$. We define the survival copula of (X, Y) , noted \hat{C} .

Definition 5 (Survival copula) *Let (X, Y) be r.v. with joint distribution H and copula C , then the survival copula \hat{C} , is defined for every $u, v \in I$:*

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

Care should be taken not to confuse the survival copula $\hat{C}(u, v)$ with the joint survival function \bar{C} for these two uniform $(0, 1)$ -r.v. whose joint distributions is C . Note that $\bar{C}(u, v) = \mathbb{P}[U > u, V > v]$, so we have the following relation :

$$\bar{C}(u, v) = \hat{C}(1 - u, 1 - v).$$

As seen above, the function \hat{C} have nice properties.

Lemma 2 *Let (X, Y) be two r.v. with joint distribution H , margins are H_X, H_Y and copula is C . We have :*

- \hat{C} is a copula;
- \hat{C} is the copula associated to the joint survival distribution \bar{H} , i.e. for every $x, y \in \mathbb{R}^+$,

$$\bar{H}(x, y) = \hat{C}(\bar{H}_X(x), \bar{H}_Y(y)).$$

As for copula, we can define empirical survival copula, defined by

$$\hat{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{n-r_{X_i} \leq nu; n-r_{Y_i} \leq nv\}}.$$

8.3.3 Copulas - Max-infinitely divisible Distribution

The following simple lemma express the intuitive fact that TP2, max-id, and min-id are properties of the copula, and not of the margins, and therefore reflect properties of the dependence between X and Y .

Lemma 3 *Let H be a bivariate cdf,*

- (i) *H is max-id, iff C is max-id;*
- (ii) *H is min-id, iff \hat{C} is max-id.*

Proof. Suppose F is min-id. We calculate the expressions in the Definition 2 as functions of the copula. We have, denoting \bar{F}_X and \bar{F}_Y the margins of the sdf \bar{F}

$$\begin{aligned}\frac{\partial}{\partial x} \bar{H}(x, y) &= \frac{\partial}{\partial u} \hat{C}(u, \bar{H}_Y(y)) \Big|_{u=\bar{H}_X(x)} \frac{\partial}{\partial x} \bar{H}_X(x) \\ \frac{\partial}{\partial y} \bar{H}(x, y) &= \frac{\partial}{\partial v} \hat{C}(\bar{H}_X(x), v) \Big|_{v=\bar{H}_Y(y)} \frac{\partial}{\partial y} \bar{H}_Y(y) \\ \frac{\partial^2}{\partial xy} \bar{H}(x, y) &= \frac{\partial^2}{\partial u \partial v} \hat{C}(u, v) \Big|_{u=\bar{H}_X(x), v=\bar{H}_Y(y)} \frac{\partial}{\partial x} \bar{H}_X(x) \frac{\partial}{\partial y} \bar{H}_Y(y).\end{aligned}$$

Since $\frac{\partial}{\partial x} \bar{H}_X(x) \frac{\partial}{\partial y} \bar{H}_Y(y) \geq 0$, a change of variables yields:

$$\frac{\partial}{\partial u} \hat{C}(u, v) \frac{\partial}{\partial v} \hat{C}(u, v) \leq \frac{\partial^2}{\partial u \partial v} \hat{C}(u, v) \hat{C}(u, v)$$

we conclude that \hat{C} is TP2 by Theorem 1. ■

Remark 1 *If F has a density which is TP2 (implying that F is also TP2) then its copula inherits of the same property as checked by direct calculation.*

8.3.4 Extreme-value Copulas

Introduction to extreme-value theory

We give here a short introduction to extreme-value theory in order to introduce the max-stable laws and the extreme-value copulas. For a complete introduction to this theory, see Galambos (1987); Resnick (1987).

Let (X, Y) be iid random variables with distribution H , we consider $(X_i, Y_i)_{1 \leq i \leq n}$ a sample of size n . Let $M_n^X = \max_{1 \leq i \leq n} X_i$ and $M_n^Y = \max_{1 \leq i \leq n} Y_i$, be the componentwise maxima.

Let

$$H_n(z_1, z_2) = \mathbb{P} [M_n^X < z_1, M_n^Y < z_2].$$

Multivariate extreme value (MEV) distributions come from limits (in law) of $\left(\frac{M_n^X - a_n^X}{b_n^X}, \frac{M_n^Y - a_n^Y}{b_n^Y} \right)$ where $(a_n^X, b_n^X), (a_n^Y, b_n^Y)$ are some quantities dependent upon distribution of (X, Y) . It means

$$H_n(a_n^X + b_n^X X, a_n^Y + b_n^Y Y) \rightarrow H(x). \quad (8.8)$$

First if a limiting distribution exists, then each univariate margin must be in the generalized extreme value (GEV) distribution family $(H^*(z, \gamma))$.

Let H_X^* and H_Y^* be the limit margins, and suppose H^* the limiting distribution, then it can be written in the form

$$\begin{aligned}H^*(z_1, z_2) &= \lim_n H^n(a_n^X + b_n^X z_1, a_n^Y + b_n^Y z_2) \\ &= \lim_n \mathbb{P} [M_n^X \leq a_n^X + b_n^X z_1, M_n^Y \leq a_n^Y + b_n^Y z_2] \\ &= C^*(H_X^*(z_1, \gamma_1), H_Y^*(z_2, \gamma_2)).\end{aligned}$$

where C^* is a copula, more precisely a MEV copula.

By taking a sample of size kn , we give necessary conditions for MEV copulas, Galambos (1987), p. 293 Theorem 5.2.1.

Theorem 4 *If H is such that, with some sequences, $\mathbf{a}_n, \mathbf{b}_n$, (8.8) holds then the copula C^* of the limit distribution H^* satisfies*

$$C^*(u, v)^t = C^*(u^t, v^t),$$

for all $t > 0$ and for every $u, v \in I$.

Max-stable Distribution

Definition 6 (Max-stable distribution) *A multivariate distribution H is max-stable if for each $t > 0$,*

$$H^t(x_1, \dots, x_p) = H(a_{1t} + b_{1t}x_1, \dots, a_{pt} + b_{pt}x_p)$$

for some vectors $\mathbf{a}_t, \mathbf{b}_t$.

By this definition, for every $t > 0$, H^t is a distribution function and hence every max-stable distribution is max-id.

Proposition 11 *The class of multivariate extreme value distributions is precisely the class of max-stable distribution functions with nondegenerate marginals.*

Proof. see Resnick (1987), Proposition 5.9 p.264. ■

The next result gives sufficient and necessary conditions, see Galambos (1987) Theorem 5.2.4 p.294.

Theorem 5 *Let C be an extreme-value copula on I^p , then :*

$$C(u_1, \dots, u_p)^t = C(u_1^t, \dots, u_p^t),$$

for all $t > 0$ and for every $u_i \in I, i = 1, \dots, p$.

Proof. Theorem 5.2.4 is based on Theorem 5.2.1 where the index is integer, but following algebras in Joe (1997), p.173, we obtain the result for k any positive real. ■

Of course, C is not a extreme value distribution, because its marginals is univariate and hence they are not MEV distributions.

Extreme-values copulas (EVC)

Among various classes of copulas, the family of *extreme value* copulas (EVC) is of peculiar interest.

The next lemma gives the relation between the extreme-value copula and max-id notion seen in (8.3.1), p.106.

Lemma 4 *All extreme-value copula are max-id distributions.*

(Joe (1997), paragraphe 6.2.1).

An important result of EVC is *the Pickands' representation*; see Pickands (1981). It appears that EVC's can be described through a function of only one variable; as such its complexity is weaker compared to the whole class of copulas. Another class enjoying this property is the *Archimedean* class of copulas. We will not consider Archimedean copulas here any more.

The following fact holds. See Joe (1997), Theorems 6.3 and 6.4.

Theorem 6 *A copula C is an EVC if and only if*

$$C(u, v) \equiv \exp \left[(\log uv) A \left(\frac{\log v}{\log uv} \right) \right] \quad (8.9)$$

where the function A is convex on $[0, 1]$, $A(0) = A(1) = 1$, and $1 \geq A(t) \geq \max(t, 1 - t)$.

The function A is usually referred to as the *dependence function*, or as the *Pickands function* of the copula C . Similarly to Lemma 3 we have:

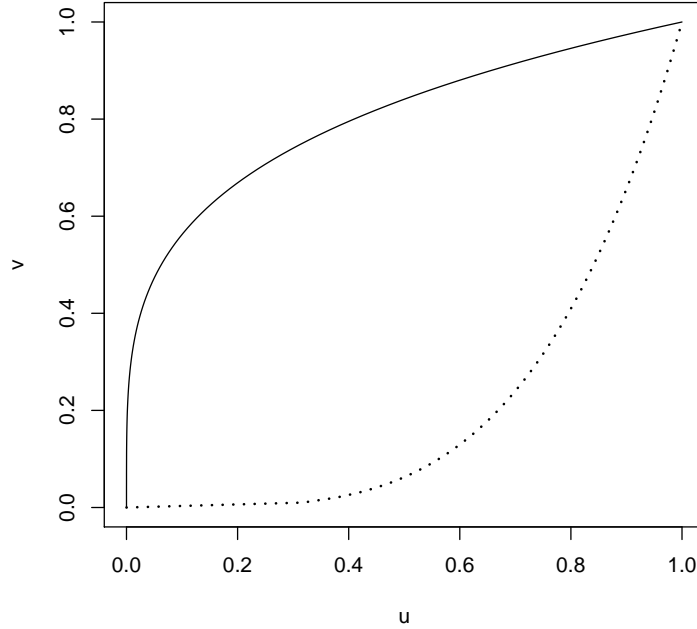
Lemma 5 *If F is a cdf with EVC copula and Pickands' function A , so is F^t for any $t > 0$.*

Proof. Let $F_{Xt} = F_X^t$ and $F_{Yt} = F_Y^t$ be the marginal cdf's of F^t . The copula of F^t is

$$\begin{aligned} C_t(u, v) &= F^t(F_{Xt}^{\leftarrow}(u), F_{Yt}^{\leftarrow}(v)) \\ &= F^t(F_X^{\leftarrow}(u^{1/t}), F_Y^{\leftarrow}(v^{1/t})) \\ &= C^t(u^{1/t}, v^{1/t}) \\ &= C(u, v). \end{aligned}$$

■

To conclude, by lemma 2, since the survival copula is a copula hence a distribution, we could talk about extreme-value copula when we talk about survival copula.

Figure 8.1: Parametrization de I^2 .

8.4 Approximation of a PQD copula by Extreme-Value Copula

In this paragraph, we show some results about approximation of a positive quadrant copula by an extreme-value copula.

First, let us parameterize the points in I in an adequate way. For any s in $[0, 1]$ the arc $\{(v^{1/s-1}, v) : v \in [0, 1]\}$ is denoted by \mathfrak{s} .

The choice of this parametrization is motivated by the following observation. This arc is the set of points (u, v) on which any dependence function A is constant and equals $A(s)$.

$$A\left(\frac{\log v}{\log uv}\right) = A(s), \quad \text{pour } (u, v) \in \mathfrak{s}.$$

The family of the arcs \mathfrak{s} covers I .

If C is an EVC, by algebras, we can write,

$$A(s) = \frac{s}{\log v} \log C(v^{1/s-1}, v).$$

So, for any copula C define on \mathfrak{s} the function

$$\mathbb{A}(s, v) \equiv \frac{s}{\log v} \log C(v^{1/s-1}, v). \quad (8.10)$$

The function \mathbb{A} mimics the dependence function A . When C is an EVC, then $\mathbb{A}(s, v) \equiv A(s)$ as seen by direct substitution. Using \mathbb{A} the copula C , analogously to (8.9), write

$$C(u, v) = \exp \left[\log(uv) \mathbb{A} \left(\frac{\log v}{\log uv}, v \right) \right].$$

In order to construct a dependence function close to \mathbb{A} , independent to v , we consider average of the values of \mathbb{A} on \mathfrak{s} through $\bar{\mathbb{A}}(s) \equiv \int_0^1 \mathbb{A}(s, v) dv$. Obviously other choices are possible which do not change in any way the results in the Proposition hereafter.

Define now

$$\tilde{C}(v^{1/s-1}, v) = \exp \left[\frac{\log v}{s} \bar{\mathbb{A}}(s) \right] \quad (8.11)$$

which by a change of variables is also $\tilde{C}(u, v) = \exp \left[\log(uv) \bar{\mathbb{A}} \left(\frac{\log v}{\log uv} \right) \right]$. We assume

(H1) C is PQD

(H2) the function $x \rightarrow x \frac{\partial C(x, y)}{\partial x}$ is non increasing for all y in $[0, 1]$.

For completeness define $\mathbb{A}(0, v) = 1$, $\mathbb{A}(s, 1) = 1$ and $\mathbb{A}(1, v) = 1$.

Proposition 12 *Under (H1) and (H2) the copula \tilde{C} is an EVC, and for all s in $[0, 1]$,*

$$\sup_{(u, v) \in \mathfrak{s}} \left| C(u, v) - \tilde{C}(u, v) \right| \leq \frac{1}{2} \text{osc}_{\mathfrak{s}} \mathbb{A}$$

where $\text{osc}_{\mathfrak{s}} F$ denotes the oscillation of the function \mathbb{A} on the arc \mathfrak{s} .

Proof. We will first prove the following results pertaining to the function $\mathbb{A}(s, v)$ defined above. For all v in $]0, 1[$,

1. $\mathbb{A}(s, v) \geq \max(s, 1 - s)$ by the upper Frechet bound.
2. $\lim_{s \rightarrow 0} \mathbb{A}(s, v) = \lim_{s \rightarrow 1} \mathbb{A}(s, v) = 1$ by the above two statements.
3. $\mathbb{A}(s, v) \leq 1$, since C is PQD.
4. The mapping $s \rightarrow \mathbb{A}(s, v)$ is convex: The second derivative of \mathbb{A} with respect to s is

$$\begin{aligned} \frac{\partial^2}{\partial s^2} \mathbb{A}(s, v) &= \frac{v^{\frac{1}{s}-1} \log v}{s^3 \left[C \left(v^{\frac{1}{s}-1}, v \right) \right]^2} \left\{ \frac{\partial}{\partial s} C \left(v^{\frac{1}{s}-1}, v \right) C \left(v^{\frac{1}{s}-1}, v \right) \right. \\ &\quad \left. - v^{\frac{1}{s}-1} \left(\frac{\partial}{\partial s} C \left(v^{\frac{1}{s}-1}, v \right) \right)^2 + v^{\frac{1}{s}-1} \frac{\partial^2}{\partial s^2} C \left(v^{\frac{1}{s}-1}, v \right) C \left(v^{\frac{1}{s}-1}, v \right) \right\}. \end{aligned}$$

For detailed calculus, see the end of this chapter.

Denote $C_X(u, v) = \frac{\partial}{\partial u} C(u, v)$ and $C_{XX}(u, v) = \frac{\partial^2}{\partial u^2} C(u, v)$. For any $(x, y) \in [0; 1]^2$,

set (s, v) such that $v = x$ and $y = v^{1/s-1}$. Convexity of the mapping $s \mapsto \mathbb{A}(s, v)$ for all $v \in]0; 1[$ implies

$$\frac{C_X(x, y)}{C(x, y)} - x \frac{C_X(x, y)^2}{C(x, y)^2} + x \frac{C_{XX}(x, y)}{C(x, y)} \leq 0.$$

The left hand side in the above bracket is the derivative with respect to the variable x of:

$$x \frac{C_X(x, y)}{C(x, y)} = \frac{\mathbb{P}[X \leq x] \mathbb{P}[Y \leq y | X = x]}{\mathbb{P}[Y \leq y, X \leq x]} = \frac{\mathbb{P}[Y \leq y | X = x]}{\mathbb{P}[Y \leq y | X \leq x]}.$$

We conclude applying (H2) .

By all this we can extend \mathbb{A} on I . For any value of v in $]0, 1[$ the resulting extension $s \rightarrow \mathbb{A}(s, v)$ has all the properties of a dependence function. Since all those properties are preserved under convex combination, as the one which defines the function $\bar{\mathbb{A}}$ it follows by the Pickands' characterization of EVC's that \tilde{C} is indeed itself an EVC.

For the upper bound on $\alpha \equiv \sup_{(u,v) \in \mathfrak{s}} |C(u, v) - \tilde{C}(u, v)|$, we use (8.10) and (8.11) together with the inequality $|e^{-a} - e^{-b}| \leq \max(e^{-a}, e^{-b}) |b - a|$ to write

$$\alpha \leq \sup_{s,v} \left| \exp \left(\frac{\log v}{s} \min(\mathbb{A}(v^{1/s-1}, v), \bar{\mathbb{A}}(s)) \right) (\mathbb{A}(v^{1/s-1}, v) - \bar{\mathbb{A}}(s)) \right|.$$

Since $\inf_{v,s} \mathbb{A}(v^{1/s-1}, v) \leq \inf_s \bar{\mathbb{A}}(s)$, with

$$\begin{aligned} \inf_{s,v} \mathbb{A}(v^{1/s-1}, v) &= \inf_{s,v} \frac{s \log(C(v^{1/s-1}, v))}{\log v} \\ &\geq \inf_{s,v} \frac{s}{\log v} \log(\min(v^{1/s-1}, v)) \\ &\geq \inf_{s,v} s \log \max_s(\exp(1/s - 1), \exp(1)) \\ &\geq \max_s(1 - s, s) = 1/2. \end{aligned}$$

■

The upper bound in Proposition 12 is indeed 0 when C is an EVC. The graphs show that the bound can be large with respect to the true distance between the copulas as measured on the arcs.

We represent on these graphs, Fig. 8.2, the error between a PQD copula and its approximation (dotted line) and the upper bound of the former proposition (black line). This abscissa is s , which parameterizes the arcs in I , as describes above.

The first example is

$$1. C(u, v) = \exp \left\{ 1 - [(1 - \log u)^\theta + (1 - \log v)^\theta - 1]^{1/\theta} \right\} \text{ with } \theta = 2;$$

$$2. C(u, v) = \left\{ 1 + [(1/u - 1)^\theta + (1/v - 1)^\theta]^{1/\theta} \right\}^{-1} \text{ with } \theta = 6.$$

The errors are about 0.05 for the first one and 0.15 for the second one.

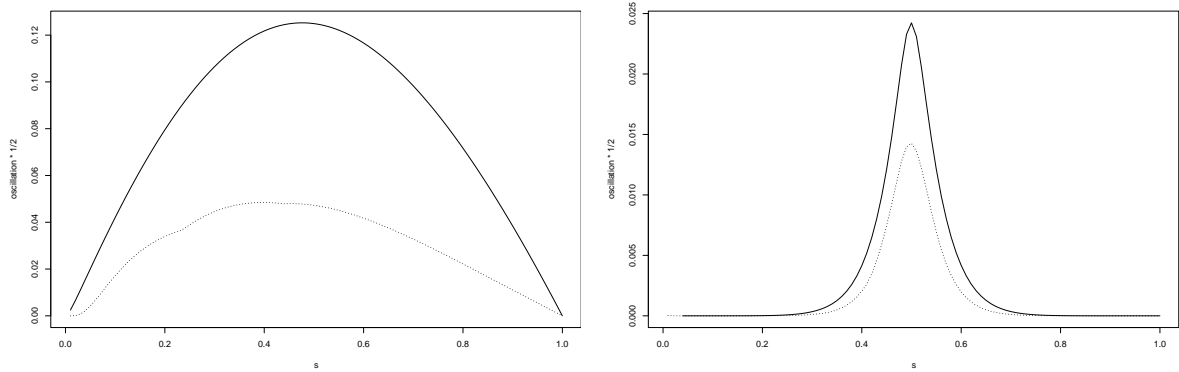


Figure 8.2: Oscillation

Remark 2 We obtain above the following condition $x \mapsto x \frac{C_X(x,y)}{C(x,y)}$ is non increasing for all y on $[0; 1]$. We can write this quantity

$$\begin{aligned} x \frac{C_X(x,y)}{C(x,y)} &= \frac{x \mathbb{P}[Y \leq y | X = x]}{\mathbb{P}[X \leq x, Y \leq y]} \\ &= \frac{\mathbb{P}[X \leq x] \mathbb{P}[Y \leq y | X = x]}{\mathbb{P}[X \leq x, Y \leq y]} \\ &= \frac{\mathbb{P}[Y \leq y | X = x]}{\mathbb{P}[Y \leq y | X \leq x]} \end{aligned}$$

We consider the stochastically increasing order.

Y is stochastically increasing in X or the conditional distribution $F_{2|1}$ is stochastically increasing if

$$\mathbb{P}[Y > y | X = x] = 1 - F_{2|1}(y|x) \uparrow x \text{ for all } y.$$

So from above algebras, a sufficient condition is

- (i) $x \mapsto \mathbb{P}[Y \leq y | X = x]$ is non increasing;
- (ii) $x \mapsto \mathbb{P}[Y \leq y | X \leq x]$ is non decreasing.

The condition (i) is equivalent to SI order and the condition (ii) is denoted Left-tail decreasing (LTD) From Joe (1997), Th 2.3(a), (SI implies LTD), so a sufficient condition is Y is SI in X .

The uniform convergence of \bar{A}_n is proved in the first section of the next chapter.

8.5 Coherence

We return to the main problem of the new model.

Not all baseline survival d.f.'s $F_{X,Y}^0$ define a coherent model, so that λ_X and $\lambda_{Y|X>x}$ are the marginal and conditional specific cause hazards for some bivariate sdf \bar{H}^z under a given covariate z . We conclude from the first equation of (8.6) that $\bar{H}_X^z(x) = \bar{F}_X^{0\Phi(z)}(x)$. By the second equation, plugging $x = 0$, we get $\bar{H}_Y^z(y) = \bar{F}_Y^{0\Psi(z)}(y)$. The model is coherent if $\bar{F}_X^{0\Phi(z)}(x)\bar{F}_{Y|X>x}^{0\Psi(z)}(y)$ defines a sdf.

Notice that

$$\begin{aligned}\bar{H}_{X,Y}^z(x,y) &= \left(\bar{F}_X^{0\Psi(z)}(x)\bar{F}_{Y|X>x}^{0\Psi(z)}(y)\right)\bar{F}_X^{0\Phi(z)-\Psi(z)}(x) \\ &= \bar{F}_{X,Y}^{0\Psi(z)}(x,y)\bar{F}_X^{0\Phi(z)-\Psi(z)}(x)\end{aligned}\tag{8.12}$$

which is indeed a sdf when $\Phi(z) \geq \Psi(z)$ and $\bar{F}_{X,Y}^{0\Psi(z)}$ is a sdf.

Note however that $F_{X,Y}^0(x,y)^t$ is indeed a s.d.f. for all $t \geq 1$. Hence when $\Psi(z) \geq 1$, (8.6) is defined when $\Phi(z) \geq \Psi(z)$ without further restrictions pertaining to $\bar{F}_{X,Y}^0$.

Remark 3 When $\Phi(z) \geq \Psi(z) \geq 1$ then model (8.6) is properly defined whenever $\bar{F}_{X,Y}^0$ is PQD or not.

In the sequel, we only consider cases when $\bar{F}_{X,Y}^0$ is TP2, which allows to consider any positive values for $\Phi(z)$ and $\Psi(z)$. Note however that $F_{X,Y}^t(x,y)$ is indeed a s.d.f. for all $t \geq 1$. Hence when $\Psi(z) \geq 1$, (8.6) is defined when $\Phi(z) \geq \Psi(z)$ without further restrictions.

As said, not all bivariate survival d.f.'s $\bar{F}_{X,Y}^0$ are such that for all positive t , $\bar{F}_{X,Y}^0{}^t$ is a sdf. Min-infinite divisibility of the baseline hazard seems to be a natural assumption here. Assume therefore that

$$F_{X,Y}^0 \text{ is min-infinitely divisible} \tag{H}$$

Then by (8.12), $\bar{H}_{X,Y}^z$ is a sdf as soon as $\Phi(z) \geq \Psi(z) \geq 0$.

Let us consider the case when $0 \leq \Phi(z) \leq \Psi(z)$. Analogously with (8.6) the model may then be written

$$\begin{cases} \lambda_Y(y) = \lambda_Y^0(y)\Psi(z) \\ \lambda_{X|Y>y}(x) = \lambda_{X|Y>y}^0(x)\Phi(z) \end{cases}\tag{8.13}$$

permuting the role of X and Y . In a similar way to the above we have that

$$\bar{H}_{X,Y}^z(x,y) = \bar{F}_{X,Y}^{0\Phi(z)}(x,y)\bar{F}_Y^{0\Psi(z)-\Phi(z)}(y).$$

is a proper sdf. To summarize the above arguments we state:

Let the model be defined by (8.6) if $\Phi(z) \geq \Psi(z)$ and by (8.13) if $\Phi(z) \leq \Psi(z)$. Call (M) the model defined in this way. When (H) holds then for all z , $\bar{H}_{X,Y}^z$ is a sdf and

$$\begin{aligned}\bar{H}_{X,Y}^z(x,y) &= \mathbf{1}\{\Phi(z) \geq \Psi(z)\}\bar{F}_{X,Y}^{0\Psi(z)}(x,y)\bar{F}_X^{0\Phi(z)-\Psi(z)}(x) \\ &\quad + \mathbf{1}\{\Phi(z) \leq \Psi(z)\}\bar{F}_{X,Y}^{0\Phi(z)}(x,y)\bar{F}_Y^{0\Psi(z)-\Phi(z)}(y).\end{aligned}\tag{8.14}$$

Min-infinite divisibility of the baseline will also make any $H_{X,Y}^z$ min-infinitely divisible, showing that this class is *stable* under (M). Indeed for any positive t

$$\begin{aligned} \bar{H}_{X,Y}^z{}^t(x, y) &= \mathbf{1}\{\Phi(z) \geq \Psi(z)\} \bar{F}_{X,Y}^0{}^{t\Psi(z)}(x, y) \bar{F}_X^0{}^{t(\Phi(z)-\Psi(z))}(x) \\ &\quad + \mathbf{1}\{\Phi(z) \leq \Psi(z)\} \bar{F}_{X,Y}^0{}^{t\Phi(z)}(x, y) \bar{F}_Y^0{}^{t(\Psi(z)-\Phi(z))}(y), \end{aligned}$$

which still is a sdf.

More precisely,

Lemma 6 *Let H and F satisfying (M). We have the following equivalences*

- *if F is d.f. and max.i.d. then H is max.i.d.;*
- *if H is d.f. and max.i.d. then F is max.i.d. .*

Proof. If F is d.f. and max.i.d then H is max.i.d. directly. Now, we consider H max.i.d. and F satisfying (M). So we have by algebras

$$\begin{aligned} \bar{F}_{X,Y}^0(x, y) &= H_{X,Y}^z(x, y) \frac{1}{\bar{\Psi}(z)} \bar{F}_X^0 \frac{\Psi(z)-\Phi(z)}{\bar{\Psi}(z)} \\ &= H_{X,Y}^z(x, y) \frac{1}{\bar{\Psi}(z)} \bar{H}_X^0 \frac{\Psi(z)-\Phi(z)}{\bar{\Psi}(z)\Phi(z)} \\ &= H_{Y|X>x}^z(y) \frac{1}{\bar{\Psi}(z)} \bar{H}_X^0 \frac{1}{\Phi(z)} \end{aligned}$$

and we conclude. ■

The interest of such a lemma is to build a test $H0: F$ is max.i.d. vs $H1: F$ is not max.i.d.. Indeed, in practice, we have only sample of distribution H , so with this lemma, we can deduce the property on F directly from test based on sample of distribution H .

By Lemma 3, min-infinite divisibility is not a property of the cdf but of its copula. Formula (8.14) can be written for copulae through

$$\begin{aligned} \hat{C}_{H^z}(u, v) &= \mathbf{1}\{\Phi(z) \geq \Psi(z)\} u^{\frac{\Phi(z)-\Psi(z)}{\Phi(z)}} \hat{C}_{F^0} \left(u^{\frac{1}{\Phi(z)}}, v^{\frac{1}{\Psi(z)}} \right)^{\Psi(z)} \\ &\quad + \mathbf{1}\{\Phi(z) \leq \Psi(z)\} u^{\frac{\Psi(z)-\Phi(z)}{\Psi(z)}} \hat{C}_{F^0} \left(u^{\frac{1}{\Psi(z)}}, v^{\frac{1}{\Phi(z)}} \right)^{\Phi(z)}. \end{aligned} \quad (8.15)$$

Although (H) is in full accordance with univariate Cox models it is only a sufficient condition for coherence. The following example illustrates this fact.

Example 3 *Let $\hat{C}_{F_{XY}^0}(u, v) = uv \exp\{-\theta \log u \log v\}$, with $\theta \in [0; 1]$; this is the Gumbel-Barnett family which may not always be TP2, since not PQD for some values of the parameter. Using (8.15) and assuming $\Phi(z) \geq \Psi(z)$ we obtain*

$$\hat{C}_{H_{XY}^z}(u, v) = uv \exp \left\{ -\frac{\theta}{\Phi(z)} \log u \log v \right\}$$

which still is a Gumbel-Barnett copula when $\theta/\Phi(z)$ belongs to $[0; 1]$, see Nelsen (1998) p. 94.

This example shows that (H) is indeed the only acceptable condition for coherence. Otherwise the baseline hazard F_{XY}^0 defines a coherent model only for specific values of the covariate. This motivates our interest in good classes of min-infinitely divisible copulae which we intend to regress on the covariate z . Among all subclasses of TP2 copulae, the EVC's enjoy nice properties as seen in Proposition 12 above, since they can nicely approximate PQD copulae and are parametrized through a smooth function of only one variable. When F_{XY}^0 has an EVC $\hat{C}_{F_{XY}^0}$ with Pickands' function A then, denoting $\hat{C}_{H_{XY}^z}$ the survival copula of H_{XY}^z and using (8.15), we have

$$\hat{C}_{H_{XY}^z}(u, v) = \exp \left[\log(uv) B^z \left(\frac{\log v}{\log uv} \right) \right]$$

with

$$B^z(s) = \frac{\Phi(z) - \Psi(z)}{\Phi(z)}(1-s) + \left[\frac{\Psi(z)}{\Phi(z)}(1-s) + s \right] A \left(\frac{1}{\frac{\Psi(z)}{\Phi(z)} \frac{1}{s} - \left(\frac{\Psi(z)}{\Phi(z)} - 1 \right)} \right)$$

when $\{\Phi(z) \geq \Psi(z)\}$, and a similar expression under $\{\Phi(z) \leq \Psi(z)\}$. To summarize, after some algebra, noting $K(z) = \min \left(\frac{\Phi(z)}{\Psi(z)}, \frac{\Psi(z)}{\Phi(z)} \right)$, we have

$$B^z(s) = [1 - K(z)](1-s) + [K(z)(1-s) + s] A \left(\frac{s}{\frac{K(z)}{1-s} + s} \right). \quad (8.16)$$

B^z is a convex function defined on $[0, 1]$ which satisfies $B^z(0) = B^z(1) = 1$ and $\max(s, 1-s) \leq B^z(s) \leq 1$ because $A(t) \leq 1$ and $K(z) \leq 1$.

Lemma 7 *The function $s \mapsto B^z(s)$ defined above satisfies :*

- $B^z(s) \leq 1$ for all $s \in [0, 1]$ and for all $z \in \mathbb{R}$;
- $B^z(s) \geq \max(s, 1-s)$ for all $s \in [0, 1]$ and for all $z \in \mathbb{R}$.

Proof. The proof of the first assertion is direct since $A(t) \leq 1$. From (8.16) and the fact that A is a dependence function, we write :

$$\begin{aligned} B^z(s) &\geq [1 - K(z)](1-s) \\ &\quad + [K(z)(1-s) + s] \max_s \left\{ \frac{s}{K(z)(1-s) + s}; \frac{K(z)(1-s)}{K(z)(1-s) + s} \right\} \\ B^z(s) &\geq [1 - K(z)](1-s) \\ &\quad + \max_s \{s; K(z)(1-s)\} \end{aligned}$$

We must prove that $B^z(s) \geq s$ for all z and $B^z(s) \geq (1-s)$ for all z . Since by definition, $\max_s \{s; K(z)(1-s)\} \geq s$ and $0 \leq s \leq 1$, $K(z) \leq 1$, then

$$B^z(s) - s = [1 - K(z)](1-s) + \max_s \{s; K(z)(1-s)\} - s \geq 0$$

We form $B^z(s) - (1 - s)$ and consider two cases. First, if $s \geq K(z)(1 - s)$ for all z , then $B^z(s) \geq [1 - K(z)](1 - s) + s - (1 - s) = -K(z)(1 - s) + s \geq 0$ by hypothesis. If $s \leq K(z)(1 - s)$ then we obtain $B^z(s) - (1 - s) \geq 0$ directly. So we conclude the result. ■

This basic result shows that $\hat{C}_{H_{XY}^z}$ is an EVC with Pickands' function B^z . We have proved that the class of EVC's is *stable* under (M).

Notice that we do not require that F_{XY}^0 has a bivariate extreme value distribution, since only its copula should be an EVC. As such F_{XY}^0 can have arbitrary margins, and still fits in (M).

From (8.16), we deduce a transitive expression that links $B^{z'}$ and B^z for any z and z' . It holds

$$B^{z'}(t) = \left[1 - \tilde{K}(z, z')\right] (1 - t) + \left[\tilde{K}(z, z')(1 - t) + t\right] B^z \left(\frac{1}{\frac{\tilde{K}(z, z')}{t} - [\tilde{K}(z, z') - 1]} \right), \quad (8.17)$$

where

$$\tilde{K}(z, z') = K(z') \cdot K(z)^{-1} \quad (8.18)$$

independently of the Pickands' function of the baseline. This formula can be seen as a kind of expression of the proportional hazard property, which links two hazard rates independently on the baseline.

When the covariate acts equally on X and Y , i.e. $\Phi(z) = \Psi(z)$ for all z , then $B^z(s) = A(s)$ for all values of s . Thus, the copula of H_{XY}^z equals that of the baseline hazard F_{XY}^0 — the dependency structure of X and Y should not be altered through (M). Only the marginal distributions of X and Y in that case reflect the role of the covariate.

Figure 8.3 illustrates the transition formula (8.16). The baseline copula is the Gumbel-Hougaard extreme-value copula with $\theta = 3$. We use $\Phi(z) = \exp(\alpha_1 z + \alpha_0)$, with $(\alpha_0, \alpha_1) = (1, 2)$, and $\Psi(z) = \exp(\beta_1 z + \beta_0)$, with $(\beta_0, \beta_1) = (0.5, 1.5)$.

The dependence function A is the line, the dotted one represents the dependence function B^z for $z = 0.2$, and the dashed one is the dependence function for $z = 2$.

One consequence of the model is that the covariate has an impact on the marginal distribution, but also on the copula. If $\Phi(z) = \Psi(z)$, the copula does not change. However, if $\Phi(z) > 1 = \Psi(z)$, then a change in the distribution of Z , which made X stochastically smaller, is not followed by a corresponding change in the distribution of Y ; thus it should follow by a change in their dependency, however this is measured.

8.6 Covariates and estimation

A proper statistical study of the model (M) must include covariates. Let z be a variable which can get a finite number of values. Suppose we have n_i independent observations $(X_j^{z_i}, Y_j^{z_i})_{j=1}^{n_i}$ from of the bivariate r.v. (X^{z_i}, Y^{z_i}) with unknown cdf $H_{XY}^{z_i}$.

We suggest the following estimation procedure. First estimate consistently the function Φ using the sample of the X^z values, and Ψ using the samples of the Y^z 's. This provides consistent estimates of $\Phi(z)$ and $\Psi(z)$ for any z . The standard maximum Cox

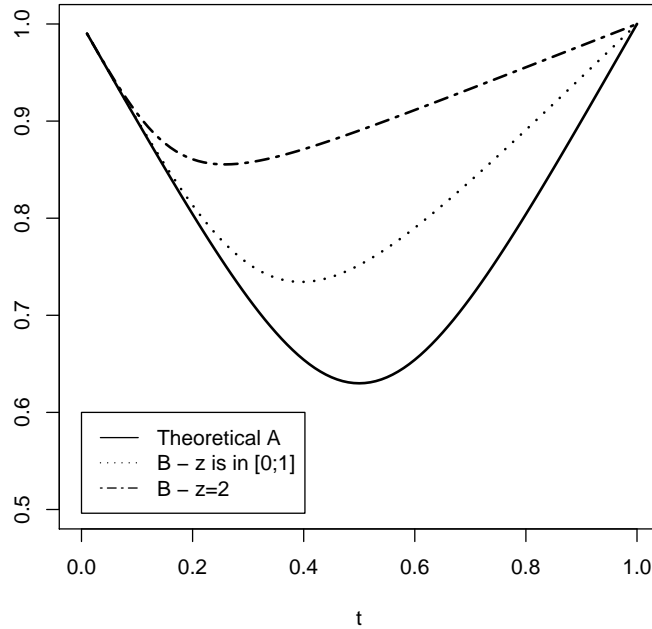


Figure 8.3: Illustration of formula (8.16)

partial likelihood estimator can be used at this stage, if the Cox model is assumed to hold. Define next $\tilde{K}(z, z')$ as in (8.18) where $\Phi(z)$ and $\Psi(z)$ are substituted by their estimators. The asymptotic normality of the parameters is trivial. Their joint distribution can be easily assessed using an asymptotic linearization of the estimators.

For each z_i estimate consistently the Pickands' function B^{z_i} of \hat{C}_{XY}^z . Standard estimates found in the literature suppose that the marginal distributions of (X^{z_i}, Y^{z_i}) are known, which we do not assume to hold. We therefore proceed as follows. First we estimate \hat{C}_{XY}^z through the non parametric estimate

$$\widehat{C}_n^{z_i}(u, v) \equiv \frac{1}{n} \sum_{j=1}^{n_i} 1 \{n - \text{rank}(X_j^{z_i}) \leq un; n - \text{rank}(Y_j^{z_i}) \leq vn\}$$

which is known to be strongly consistent and asymptotic normal; see Van Der Vaart and Wellner (1996).

We then use the same trick as in Section 2 which was designed in order to provide an EVC approximation of a PQD copula. For s in $[0, 1]$ define a function \mathbb{A}_n on the arc \mathfrak{s} by

$$\mathbb{A}_n(v^{\frac{1}{s}-1}, v) \equiv \frac{s}{\log v} \log \widehat{C}_n^{z_i}(v^{\frac{1}{s}-1}, v).$$

Average \mathbb{A}_n on \mathfrak{s} and get

$$B_n^{z_i}(s) \equiv \bar{\mathbb{A}}_n(s) \equiv \min \left\{ 1, \max \left(\frac{1}{k_{n_i}} \sum_{l=1}^{k_{n_i}} \mathbb{A}_n(v_l^{\frac{1}{s}-1}, v_l); s; 1 - s \right) \right\} \quad (8.19)$$

for $0 < v_1 < \dots < v_{k_{n_i}} < 1$ a uniform grid on $[0, 1]$. The reason to introduce the extra max operation is to get a properly defined estimate on the segment $[0, 1]$. It is easily checked that $B_n^{z_i}(0) = B_n^{z_i}(1) = 1$ and by construction $B_n^{z_i}(s) \geq \max(s, 1 - s)$. When k_{n_i} tends to infinity together with n_i and when \hat{C}_{XY}^z is an EVC continuity arguments show that $B_n^{z_i}(s)$ point-wisely converges a.s. to $B^{z_i}(s)$. Asymptotic properties of those estimators are postponed to a further work.

We illustrate 8.19 in Fig. 8.4. The dataset consists in $n = 200$ i.i.d. points from the Gumbel-Hougaard extreme-value copula $\hat{C}(u, v) = \exp \left\{ - \left[(-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} \right\}$ with $\theta = 3$. The estimate B_n^z (referred to as "empirical estimator") is compared to the classical Pickands' estimate of the dependence function.

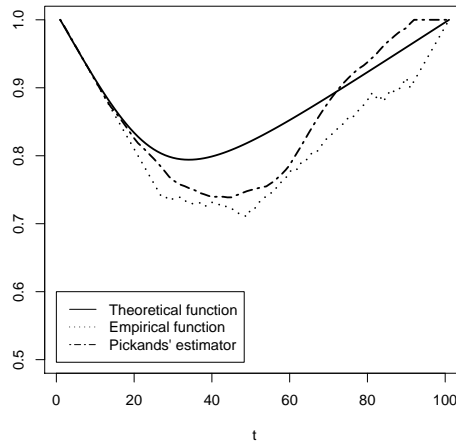


Figure 8.4: Two estimators of B^z

Let z be some value of the covariate for which no data are observed. Use (8.17) for each z_i so that

$$B^z(s) = \left[1 - \tilde{K}(z_i, z) \right] (1 - s) + \left[\tilde{K}(z_i, z)(1 - s) + s \right] B^{z_i} \left(\frac{1}{\frac{\tilde{K}(z_i, z')}{s} - \left[\tilde{K}(z_i, z) - 1 \right]} \right)$$

holds for any i .

Define the forecast of $B^z(s)$ through the interpolation

$$B_n^z(s) = \sum_{i=1}^k \frac{W(|z - z_i|)}{\sum_{j=1}^k W(|z - z_j|)} \left\{ \left[1 - \tilde{K}(z_i, z) \right] (1 - s) + \left[\tilde{K}(z_i, z)(1 - s) + s \right] B_n^{z_i} \left(\frac{1}{\frac{\tilde{K}(z_i, z')}{s} - \left[\tilde{K}(z_i, z) - 1 \right]} \right) \right\} \quad (8.20)$$

where $W(|z - z_i|)$ is positive and tends to 0 as the distance between z and z_i increases. This is done in order to give more weight to the B^{z_i} functions for z_i close to z .

We consider a data set consisting of 10 values of the covariate z . The baseline hazard has a Gumbel-Hougaard survival EVC. For each one we simulated 200 couples (X_i, Y_i) in the following way: $\Phi(z) = \exp(\alpha_1 z + \alpha_0)$, $\Psi(z) = \exp(\beta_1 z + \beta_0)$, $\alpha_1 = 3$, $\alpha_0 = 2$, $\beta_1 = 2$, $\beta_0 = 1$, and $\theta = 3$. We used formula (8.16) to define the copulae for each z_i , $1 \leq i \leq 10$. The values of the z_i 's range between 0 and 1. The prediction is set for $z = 1.2$.

We used formula (8.20) with $W(x) = \exp(-2x)$. In Figure 8.5, the dashed line is the forecasted B function when the B^{z_i} 's are estimated through Pickands' method. The dotted line is similar, but the $B_n^{z_i}$'s are the estimates defined by (8.19). The solid line is the theoretical B^z as deduced from the model.

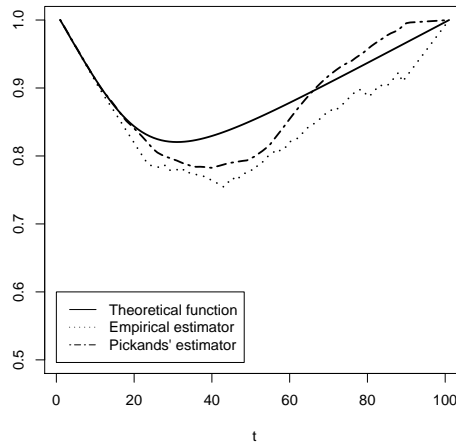


Figure 8.5: Forecasted dependence functions

8.7 Application - Regression of Spearman's rho on covariates

We illustrate the above procedure regressing the Spearman correlation coefficient on covariate, taking F_{XY}^0 the Gumbel-Hougaard copula, defined by

$$F_{XY}^0(x, y) \equiv \exp \left\{ - \left[(-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} \right\}$$

with $\theta \geq 1$. This copula is an EVC; see Nelsen (1998) p. 94.

First, we recall some properties of Spearman's rho in the context of extreme-value distribution. See e.g. Hürlimann (2003) for details. As usual, let C be a copula, and \hat{C} its survival copula. We have

$$\rho_S = 12 \iint C(u, v) dudv - 3 = 12 \int_0^1 \frac{dt}{(1 + A(t))^2} - 3,$$

and simple calculation leads to $\rho_S = 12 \iint \hat{C}(u, v) du dv - 3$.

We compute ρ_S^z for several values of z through its empirical counterpart

$$\rho_S^z = 12 \int \frac{dt}{[1 + B_n^z(t)]^2} - 3 \quad (8.21)$$

where B_n^z is given in (8.20).

Ten values of the covariate z have been considered. The baseline hazard has a Gumbel-Hougaard survival EVC. For each value we simulate 500 couples (X_i, Y_i) in the following way: $\Phi(z) = \exp(\alpha_1 z + \alpha_0)$, $\Psi(z) = \exp(\beta_1 z + \beta_0)$ $\alpha_1 = 3, \alpha_0 = 2, \beta_1 = 2, \beta_0 = 1, \theta = 3$. We use equation (8.16) to define the copulae for each z_i , $1 \leq i \leq 10$. The simulation step uses the above algorithm, and $W(x) = \exp(-2x)$. The computation is illustrated in Figure 8.6. The black line is ρ_S^z . The dashed is the forecast (for z 's in the range of the z_i 's). The dots are the empirical values on the samples for the z_i 's.

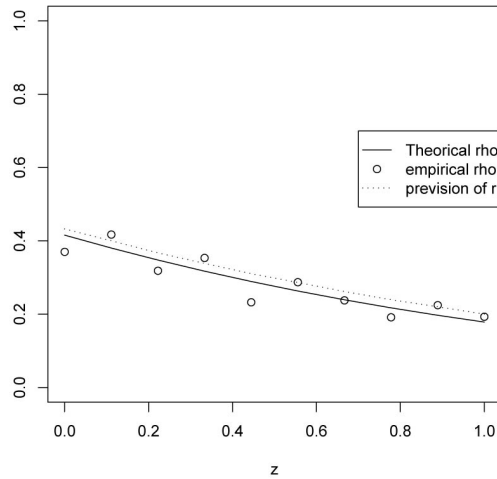


Figure 8.6: Forecasted Spearman's rho

8.8 A case study

The model has been used in order to evaluate the tolerance of car drivers with respect to speed regulation.

In contrast with the ACC (Adaptive Cruise Control), the ISA (Intelligent Speed Adaptation) systems, in the french version (LAVIA), automatically limits the vehicle speed to the local prescribed speed. This system is based on GPS and on two in-vehicles sensors (an odometer to measure distances and a gyrometer to measure direction), so that the position of the vehicle is know quite accurately at each moment, see Figure 8.7.

At any time the driver may use the "kick-down", pushing the accelerator at full stroke, which temporarily disables the LAVIA. As soon as the vehicle speed returns below the speed limit, the LAVIA is enabled again.

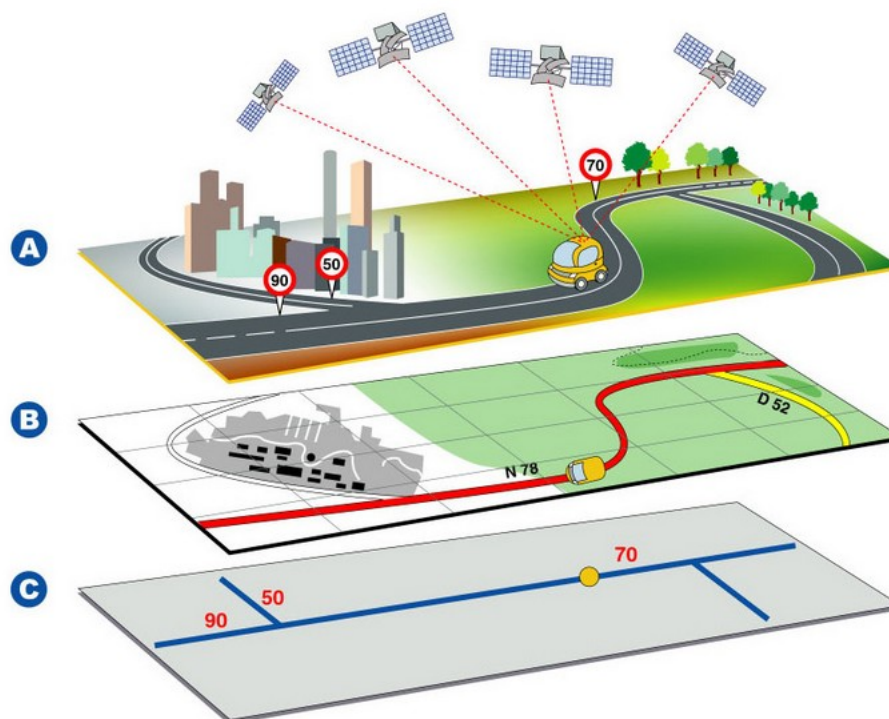


Figure 8.7: The LAVIA system - ©LAVIA project

All these technologies ask the problem of acceptance from the driver's viewpoint. We consider the acceptance or no-acceptance (kick-down) of the LAVIA according to the legal speed limit.

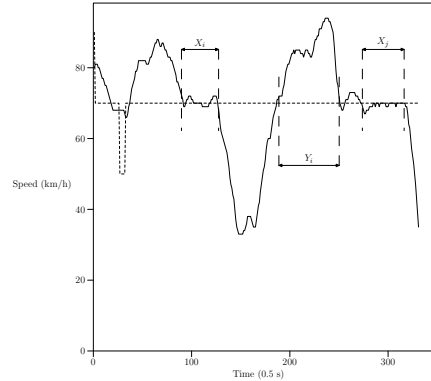
On a single drive, let X_i denote the i th interval of kilometers where the speed of the vehicle is restricted to the legal speed by the LAVIA and Y_i be the i th distance of unrestricted limit over the legal speed; hence Y_i measures some interval of kilometers for which the driver has made the kickdown.

We consider distance and not time, because in order to compare the results for different speed limit, only distance is comparable.

Set $X = \sum_i X_i$ and $Y = \sum_i Y_i$, the total restricted (unrestricted) amount of distance with (resp. without) the LAVIA system really working. X measures the total amount of distance when the driver is constrained by the LAVIA while Y indicates the distance when he exceeds the legal speed voluntarily. The notion of acceptance is captured in the data set (X, Y) on various drives, and the covariate is the average prescribed speed on the drive.

In practice, we have considered 128 drives in a urban zone close to Paris (France). The values of the covariate Z are 37 km/h , 45 km/h , 48 km/h and 50 km/h and for each value of the covariate we have considered 32 drives.

The Cox model marginal hypothesis cannot be rejected for X or for Y , according to the test by Andersen et al. (1993) implemented in R software (see Table 8.1). The null hypothesis H_0 : the hazard rates are proportional, can't be rejected.

Figure 8.8: Data X, Y

	ρ	χ^2	p
Z	-0.119	2.01	0.157
Z	-0.151	3.11	0.0778

Table 8.1: Test of proportional hazard rates

We consider here the Cox model with $\Phi(z) = \exp(\alpha_0 + \alpha_1 z)$ and $\Psi(z) = \exp(\beta_0 + \beta_1 z)$.

So, we calculate with our model the evolution of Spearman's rho between the two distances in function of the mean of prescribed speed of one way.

Figure (8.8) presents both the empirical estimates of the Spearman's rank correlation coefficient for each value of the forecasted speed, as well as forecasted by model (The bias is due to some informatical problems).

We see that on real data, our method performs well. The *evolution* of the Spearman's rho between the two distances in function of exogeneous variable Z is in accordance with plausible interpretation.

8.9 Detailed proofs

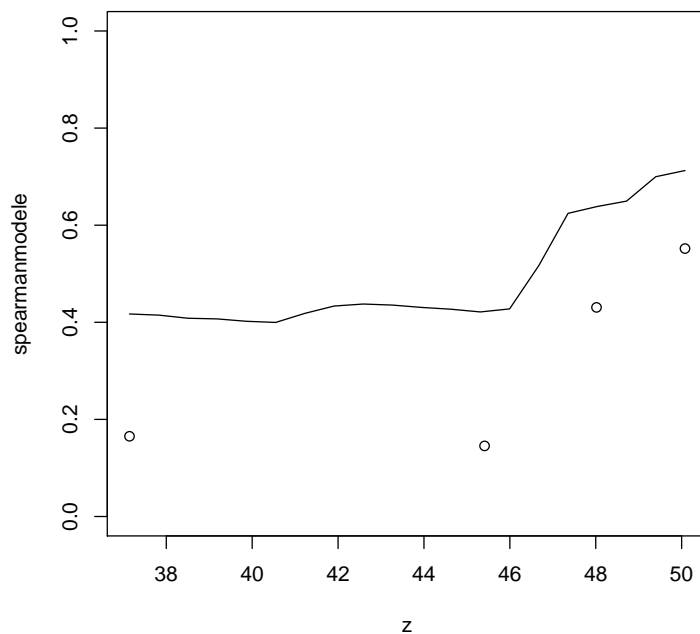
8.9.1 Approximation of a PQD copula by Extreme-Value Copula

Let

$$\mathbb{A}(s, v) = \frac{s}{\log v} \log C(v^{1/s-1}, v).$$

The first derivative is, with the notation $C_X(u, v) = \frac{\partial}{\partial u} C(u, v)$

$$\frac{\partial}{\partial s} \mathbb{A}(s, v) = \frac{1}{\log v} \log C(v^{1/s-1}, v) - \frac{v^{1/s-1} C_X(v^{1/s-1}, v)}{s C(v^{1/s-1}, v)}$$



Hence we obtain :

$$\frac{\partial}{\partial s} [\log C(v^{1/s-1}, v)] = -\frac{v^{1/s-1} \log v C_X(v^{1/s-1}, v)}{s^2 C(v^{1/s-1}, v)}.$$

For the second derivative, the calculus is

$$\begin{aligned} \frac{\partial^2}{\partial s^2} \mathbb{A}(s, v) &= \frac{1}{\log v} \frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} v^{1/s-1} \left(-\frac{\log v}{s^2} \right) \\ &\quad - \frac{1}{s} \frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} v^{1/s-1} \left(-\frac{\log v}{s^2} \right) \\ &\quad - v^{1/s-1} \frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} \left(-\frac{1}{s^2} \right) - \frac{v^{1/s-1}}{s} \frac{\partial}{\partial s} \left(\frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} \right) \end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial s} \left(\frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} \right) &= \frac{C_{XX}(v^{1/s-1}, v) v^{1/s-1} \left(-\frac{\log v}{s^2}\right) C(v^{1/s-1}, v)}{C(v^{1/s-1}, v)^2} \\
&\quad - \frac{C_X(v^{1/s-1}, v) \frac{\partial}{\partial s} C(v^{1/s-1}, v)}{C(v^{1/s-1}, v)^2} \\
&= \frac{C_{XX}(v^{1/s-1}, v) v^{1/s-1} \left(-\frac{\log v}{s^2}\right)}{C(v^{1/s-1}, v)} \\
&\quad - \frac{C_X(v^{1/s-1}, v)^2 v^{1/s-1} \left(-\frac{\log v}{s^2}\right)}{C(v^{1/s-1}, v)^2}
\end{aligned}$$

hence

$$\begin{aligned}
\frac{\partial^2}{\partial s^2} \mathbb{A}(s, v) &= \frac{1}{\log v} \frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} v^{1/s-1} \left(-\frac{\log v}{s^2}\right) \\
&\quad + \frac{v^{1/s-1} \log v C_X(v^{1/s-1}, v)}{s^3 C(v^{1/s-1}, v)} \\
&\quad + \frac{v^{1/s-1} \log v C_X(v^{1/s-1}, v)}{s^2 C(v^{1/s-1}, v)} \\
&\quad - \frac{v^{1/s-1} \log v}{s} \left[\frac{C_{XX}(v^{1/s-1}, v) v^{1/s-1} \left(-\frac{\log v}{s^2}\right)}{C(v^{1/s-1}, v)} \right. \\
&\quad \left. - \frac{C_X(v^{1/s-1}, v)^2 v^{1/s-1} \left(-\frac{\log v}{s^2}\right)}{C(v^{1/s-1}, v)^2} \right].
\end{aligned}$$

So we obtain the result

$$\begin{aligned}
\frac{\partial^2}{\partial s^2} \mathbb{A}(s, v) &= \frac{v^{1/s-1} \log v}{s^3} \left[\frac{C_X(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} - \frac{C_{XX}(v^{1/s-1}, v) v^{1/s-1} \left(-\frac{\log v}{s^2}\right)}{C(v^{1/s-1}, v)} \right. \\
&\quad \left. + \frac{C_X(v^{1/s-1}, v)^2 v^{1/s-1} \left(-\frac{\log v}{s^2}\right)}{C(v^{1/s-1}, v)^2} \right]
\end{aligned}$$

8.9.2 Study of the estimator

After some notations, we prove the convergence of the prior estimator $A_n(t)$ to the theoretical Pickands' function in the case of extreme-value copula. We conclude with a result on the approximation of the Spearman's rho.

Notations

Let C be an extreme-value copula. Our estimator of A is defined up to some fixed specified range on $[0; 1]$. Indeed, boundary effects impose to define it on $[\delta_n; 1 - \delta_n]$, for $(\delta_n)_n$

satisfying :

$$\left\{ \begin{array}{l} \delta_n \xrightarrow[n \rightarrow +\infty]{} 0 \\ \delta_n^{-2\theta-1} \left[\log \left(\frac{n}{\log \log n} \right) \right]^{-1} \xrightarrow[n \rightarrow +\infty]{} 0, \text{ for some } \theta > 0 \end{array} \right. \quad (\star)$$

This means that $(\delta_n)_n$ should converge to 0 slowly enough. For such δ_n , set $I_{\delta_n} = [\delta_n; 1 - \delta_n]^2$, a sequence of subdomains in I with (δ_n) as in (\star) , I_{δ_n} contains enough observations to make C_n close to C in I_{δ_n} as will appear in the proofs of our results.

For $0 < s < 1$, we recall that \mathfrak{s} denotes the arc $\mathfrak{s}_s := \{(u, v) \in [0; 1]^2; u = v^{1/s-1}\}$. As we can see in fig. 8.9.2, we consider only the arcs \mathfrak{s}_s such that s is in $\mathcal{S}_{\delta_n} = \{s / \mathfrak{s}_s \cap I_{\delta_n} \neq \emptyset\}$. As shown the schema below, we can express this restriction \mathcal{S}_{δ_n} as a interval $[\underline{s}_{\delta_n}; \bar{s}_{\delta_n}]$. For the dashed arcs, we obtain the condition $\delta_n = (1 - \delta_n)^{1/\underline{s}_{\delta_n}-1}$ so :

$$\underline{s}_{\delta_n} = \frac{\log(1 - \delta_n)}{\log[\delta_n(1 - \delta_n)]} \xrightarrow[n \rightarrow +\infty]{} 0$$

The dotted arcs illustrate the condition $(1 - \delta_n) = \delta_n^{\frac{1}{\bar{s}_{\delta_n}}-1}$ so :

$$\bar{s}_{\delta_n} = \frac{\log \delta_n}{\log[\delta_n(1 - \delta_n)]} \xrightarrow[n \rightarrow +\infty]{} 1.$$

So, we will consider only $s \in [\underline{s}_{\delta_n}; \bar{s}_{\delta_n}]$ (when δ_n goes to 0, \underline{s}_{δ_n} goes to 0 and \bar{s}_{δ_n} goes to 1).

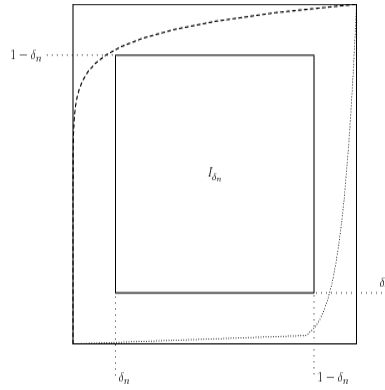


Figure 8.9: Parametrisation of $[0; 1]^2$

With C an extreme-value copula, we define the following quantities for $t \in [0; 1]$:

$$A(t) = \int_0^1 \frac{t}{\log v} \log C(v^{1/t-1}, v) \, dv$$

$$A_n(t) = \int_{\delta_n}^{1-\delta_n} \frac{\underline{s}_{\delta_n} + t(\bar{s}_{\delta_n} - \underline{s}_{\delta_n})}{\log v} \log C_n \left(v^{\left[\frac{1}{\underline{s}_{\delta_n} + t(\bar{s}_{\delta_n} - \underline{s}_{\delta_n})} \right]^{-1}}, v \right) \, dv$$

The aim of this section is to show the uniform convergence of A_n to A .

Uniform convergence of $A_n(t)$ to $A(t)$

In this section, we prove the uniform convergence of $A_n(t)$ to $A(t)$ on $[0; 1]$ for all $(\delta_n)_n$ satisfying (\star) ,

Proposition 13 *For any sequence $(\delta_n)_n$ satisfying (\star) , we have :*

$$\lim_{n \rightarrow +\infty} \sup_{t \in [0; 1]} |A_n(t) - A(t)| = 0$$

Let

$$\begin{aligned} \overline{A(t)} &= \int_0^1 \frac{\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})}{\log v} \log C \left(v^{\frac{1}{\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})}}^{-1}, v \right) dv \\ \overline{\overline{A(t)}} &= \int_{\delta_n}^{1-\delta_n} \frac{\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})}{\log v} \log C \left(v^{\frac{1}{\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})}}^{-1}, v \right) dv = \overline{A(t)}(1 - 2\delta_n) \end{aligned}$$

since C is an extreme-value copula, the integrand does not depend on v . Moreover with the same reason, $\overline{A(t)} = A(\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n}))$, hence

$$\begin{aligned} \sup_{t \in [0; 1]} |A_n(t) - A(t)| &\leq \sup_t |A_n(t) - \overline{\overline{A(t)}}| + \sup_t |\overline{\overline{A(t)}} - \overline{A(t)}| + \sup_t |\overline{A(t)} - A(t)| \\ &= T_1 + T_2 + T_3 \end{aligned}$$

We show that $T_2 = \sup_t |\overline{\overline{A(t)}} - \overline{A(t)}|$ converges to 0; indeed $\overline{\overline{A(t)}} - \overline{A(t)} = 2\delta_n A(\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})) \leq 2\delta_n \sup_t A(t)$. From (\star) and $A(t) \leq 1$, this upper bound tends to 0 as n goes to $+\infty$.

The term T_3 could be written as $\sup_t |A(\underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})) - A(t)|$. As \underline{s}_{δ_n} converges to 0 and \overline{s}_{δ_n} converges to 1, and since A is continuous on $[0; 1]$, this quantity converges to 0.

We now show that T_1 converges to 0. First to simplify notations, we put $s = \underline{s}_{\delta_n} + t(\overline{s}_{\delta_n} - \underline{s}_{\delta_n})$. We get, since $\sup_t s \leq 1$,

$$\sup_t |A_n(t) - \overline{\overline{A(t)}}| \leq \sup_{s \in [\underline{s}_{\delta_n}; \overline{s}_{\delta_n}]} \int_{\delta_n}^{1-\delta_n} \frac{s}{|\log v|} |\log C_n(v^{1/s-1}, v) - \log C(v^{1/s-1}, v)| dv.$$

To bound this quantity, we use the inequality $|\log(1+x)| \leq 3/2|x|$ for $-1/2 \leq x \leq 1/2$ which can be used for n large enough and yields

$$\begin{aligned} \sup_{s \in [\underline{s}_{\delta_n}; \overline{s}_{\delta_n}]} \int_{\delta_n}^{1-\delta_n} \frac{s}{|\log v|} |\log C_n(v^{1/s-1}, v) - \log C(v^{1/s-1}, v)| dv &\leq \\ \sup_{s \in [\underline{s}_{\delta_n}; \overline{s}_{\delta_n}]} \sup_{v \in I_{\delta_n}} \frac{3}{2} \left| \frac{C_n(v^{1/s-1}, v) - C(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} \right| \int_{\delta_n}^{1-\delta_n} \frac{dv}{|\log v|}. & \end{aligned}$$

Since $-\log v \leq (1-v)/v$ for positive v , we get :

$$\int_{\delta_n}^{1-\delta_n} \frac{dv}{|\log v|} \leq -\log \delta_n$$

As $I_{\delta_n} \subset I = [0; 1]^2$, we write :

$$\sup_{s \in [\underline{s}_{\delta_n}, \bar{s}_{\delta_n}], v \in I_{\delta_n}} \left| \frac{C_n(v^{1/s-1}, v) - C(v^{1/s-1}, v)}{C(v^{1/s-1}, v)} \right| \leq \sup_{I^2} |C_n(v^{1/s-1}, v) - C(v^{1/s-1}, v)| \sup_{s \in [\underline{s}_{\delta_n}, \bar{s}_{\delta_n}], v \in I_{\delta_n}} \left| \frac{1}{C(v^{1/s-1}, v)} \right|.$$

One may check (refer to Theorem 3.1 in Deheuvels (1979)), that there exists a constant κ (depending upon C only) such that, with probability 1,

$$\limsup_{n \rightarrow +\infty} \sqrt{\frac{n}{\log \log n}} \sup_{0 \leq u, v \leq 1} |C_n(u, v) - C(u, v)| = \kappa < \infty.$$

So, we get :

$$\sup_{(u, v) \in I^2} |C_n(u, v) - C(u, v)| = \mathcal{O} \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ a.s..}$$

We now get an upper bound for the supremum $\sup_{s \in [\underline{s}_{\delta_n}, \bar{s}_{\delta_n}], v \in I_{\delta_n}} \left| \frac{1}{C(v^{1/s-1}, v)} \right|$. Since C is an extreme-value copula, it holds from Pickands' representation that

$$\sup_{s \in [\underline{s}_{\delta_n}, \bar{s}_{\delta_n}], v \in I_{\delta_n}} \frac{1}{C(v^{1/s-1}, v)} \leq \sup_{s \in [\underline{s}_{\delta_n}, \bar{s}_{\delta_n}]} \delta_n^{-A(s)/s} \leq \delta_n^{-1/\underline{s}_{\delta_n}}$$

which entails

$$\sup_s |A_n(s) - \overline{A(s)}| \leq (-\log \delta_n) \delta_n^{-1/\underline{s}_{\delta_n}} \mathcal{O} \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ a.s..}$$

As $(-\log \delta_n) = o((\delta_n)^{-1/\underline{s}_{\delta_n}})$ and from (\star) , the terms $(-\log \delta_n) \delta_n^{-1/\underline{s}_{\delta_n}} \mathcal{O} \left(\left(\frac{\log \log n}{n} \right) \right)$ converge to 0.

This prove the uniform convergence of $A_n(t)$ to $A(t)$ for $t \in [0; 1]$ and $(\delta_n)_n$ satisfying (\star) .

Chapter 9

Estimation of the dependence function of a bivariate extreme-value copula in case of unknown margins

We consider in this chapter a new estimator of the dependence function of a bivariate extreme-value copula. Following Pickands (1981); Tiago de Oliveira (1997); Hall and Tajvidi (2000); Capéraà et al. (1997); Abdous and Ghoudi (2005) proposed various estimators under the hypothesis of known margins. Our estimator does not depend of the margins, following an idea given in the end of Capéraà et al. (1997). In the second section, we prove its uniform convergence and study a functional version of this convergence. To conclude, we provide various results of simulations in order to prove the performance of our estimator.

9.1 Notations

Let (X, Y) be two absolutely continuous random variables (r.v.'s) with joint cumulative distribution function (c.d.f.) $H_{X,Y}(x, y) = \mathbb{P}[X \leq x, Y \leq y]$ and marginals distributions F, G ($F(x) = H(x, +\infty)$ and $G(y) = H(+\infty, y)$).

Sklar (1959) introduced the notion of copula of a bivariate distribution H , which is defined on $I^2 = [0; 1]^2$ through

$$C(u, v) := H(F^{-1}(u), G^{-1}(v)). \quad (9.1)$$

Reciprocally, given a c.d.f. C on I^2 and two c.d.f.'s F and G on \mathbb{R} (a.c or not), Sklar's theorem asserts that any c.d.f. H on \mathbb{R}^2 can be defined through (9.1) with $F^{-1}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}$ (resp. G^{-1}), the so-called generalized inverse.

Among all copulae on I^2 , the subclass of extreme-value copulae (e.v.c.) deserves attention. Indeed, they capture the dependence structure of all bivariate c.d.f.'s which occur as the limiting types of point maxima of the form $(\max X_i, \max Y_i)$, where (X_i, Y_i) is a n -independent sample of (X, Y) .

In the following, we consider (X_i, Y_i) a n -independent sample with bivariate extreme-value distribution.

The celebrated Pickands' representation (Pickands (1981)) theorem asserts that whenever C is an e.v.c., there exists a function A defined on $[0; 1]$ such that

$$C(u, v) = \exp \left\{ \log(uv) A \left(\frac{\log(v)}{\log(uv)} \right) \right\}, \quad (9.2)$$

for $(u, v) \in \text{int } I^2$, with $A(0) = A(1) = 1$, A convex and $\max(s, 1-s) \leq A(s) \leq 1$ for all s in $[0; 1]^2$.

We remark that for all $s \in]0, 1]$, it holds

$$\begin{aligned} A(s) &= \frac{s}{\log v} \log C(v^{1/s-1}, v) \\ &= -\log C(e^{s-1}, e^{-s}) \end{aligned}$$

where we have used $C^t(u, v) = C(u^t, v^t)$ for positive t ($t = -\frac{s}{\log v}$), a property which is equivalent to extreme-value copula. This expression is the starting point for the definition of an estimate of the function A . We extend the definition of A to $s = 0$ by continuity, setting $A(0) = 1$.

Many estimators of the Pickands' function A do exist; see Abdous and Ghoudi (2005) and Capéraà et al. (1997) for details. They all assume that the underlying bivariate sample has extreme value univariate marginals; a remark in Capéraà et al. (1997) quotes that this hypothesis may be relaxed somehow. This is of some relevant interest; indeed extreme value copulae do not only matter for bivariate extreme behavior, as may be expected from the naming. In Broniatowski et al. (2007), it has been proved that e.v.c. appear naturally as a convenient approximation class for Positive Quadrant dependency (PQD) and may play a role for the extension of the well-known Cox model for bivariate models. In this respect, the above hypothesis on the marginals of the observed sample is irrelevant and indeed should be relaxed. The aim of the present paper is to introduce a new estimate for the Pickands' function without this assumption. This new estimate is simple, and enjoys good limit properties. This note is organized as follows. Section 2 presents the estimate and states its limit properties. In section 3, a number of simulation results aim at the comparison with classical estimates, both when the marginals are e.v. or not.

9.2 The form of the estimate and its limit properties

Define the plug-in version of A through

$$A_n(s) = -\log C_n(e^{s-1}, e^{-s}), \quad (9.3)$$

for $0 \leq s \leq 1$, where C_n is the empirical counterpart of C , defined on an i.i.d. sample $(X_i, Y_i)_{i=1}^n$, with common e.v.c. C , with Pickands' function A . The empirical copulae C_n is defined through

$$C_n(u, v) = H_n(F_n^{-1}(u), G_n^{-1}(v)).$$

We start with some definitions.

Let C be a copula, we define W_C a bivariate Wiener process satisfying

$$\begin{aligned}\mathbb{E}[W_C(u, v)] &= 0 \\ \mathbb{E}[W_C(u, v)W_C(u', v')] &= C(u \wedge u', v \wedge v').\end{aligned}$$

for all (u, v, u', v') in I .

Let further B_C be the C -Brownian bridge defined by (Fermanian et al. (2004)) :

$$B_C(u, v) = W_C(u, v) - C(u, v)W_C(1, 1)$$

which is centered and satisfies

$$\mathbb{E}[B_C(u, v)B_C(u', v')] = C(u \wedge u', v \wedge v') - C(u, v)C(u', v').$$

We define a tied-down Brownian bridge G_C (Fermanian et al. (2004); Deheuvels et al. (2006)) by

$$G_C(u, v) = B_C(u, v) - \frac{\partial}{\partial x}C(u, v)B_C(u, 1) - \frac{\partial}{\partial y}C(u, v)B_C(1, v)$$

and

$$\begin{aligned}\mathbb{E}[G_C(u, v)G_C(u', v')] &= C(u \wedge u', v \wedge v') - C(u, v)C(u', v') \\ &- \frac{\partial}{\partial u}C(u', v')[C(u \wedge u', v) - C(u, v)u'] - \frac{\partial}{\partial u}C(u, v)[C(u \wedge u', v') - C(u', v')u] \\ &- \frac{\partial}{\partial v}C(u', v')[C(u, v \wedge v') - C(u, v)v'] - \frac{\partial}{\partial v}C(u, v)[C(u', v \wedge v') - C(u', v')v] \\ &+ \frac{\partial}{\partial u}C(u, v)\frac{\partial}{\partial u}C(u', v')[u \wedge u' - uu'] + \frac{\partial}{\partial v}C(u, v)\frac{\partial}{\partial v}C(u', v')[v \wedge v' - vv'] \\ &+ \frac{\partial}{\partial u}C(u, v)\frac{\partial}{\partial v}C(u', v')[C(u, v') - uv'] + \frac{\partial}{\partial v}C(u, v)\frac{\partial}{\partial u}C(u', v')[C(u', v) - u'v].\end{aligned}$$

Finally, for all $s \in [0; 1]$, set

$$D(s) = \frac{G_C(e^{s-1}, e^{-s})}{C(e^{s-1}, e^{-s})}. \quad (9.4)$$

We state that A_n defined by (9.3) is a uniformly consistent estimate of A .

Proposition 14 *It holds*

$$\lim_{n \rightarrow +\infty} \sup_{s \in [0; 1]} |A_n(s) - A(s)| = 0. \quad a.s.$$

Proof. We rely on Deheuvels (1979), Th 3.1:

There exists some κ (depending upon C only) such that, with probability 1,

$$\limsup_{n \rightarrow +\infty} \sqrt{\frac{n}{\log \log n}} \sup_{0 \leq u, v \leq 1} |C_n(u, v) - C(u, v)| = \kappa < \infty.$$

Write

$$\sup_s |A_n(s) - A(s)| \leq \sup_{s \in [0,1]} \left| \log C_n(e^{s-1}, e^{-s}) - \log C(e^{s-1}, e^{-s}) \right|.$$

Use the inequality $|\log(1+x)| \leq |3x/2|$ for $-1/2 \leq x \leq 1/2$ (Abramowitz and Stegun (1965)), since, for n large,

$$\sup_{0 \leq u, v \leq 1} |C_n(u, v) - C(u, v)| \rightarrow 0 \quad \text{a.s.}$$

Also note that $C(e^{s-1}, e^{-s}) = \exp(-A(s)) \geq e^{-1}$ since C is positive quadrant dependent, *i.e.* $C(u, v) \geq uv$. Therefore

$$\sup_s |A_n(s) - A(s)| \leq \sup_{s \in [0,1]} 3/2 \left| \frac{C_n(e^{s-1}, e^{-s}) - C(e^{s-1}, e^{-s})}{C(e^{s-1}, e^{-s})} \right|.$$

This entails

$$\sup_{s \in [0,1]} \left| \frac{C_n(e^{s-1}, e^{-s}) - C(e^{s-1}, e^{-s})}{C(e^{s-1}, e^{-s})} \right| \leq \sup_{I^2} |C_n(u, v) - C(u, v)| \sup_{s \in [0,1]} \frac{1}{C(e^{s-1}, e^{-s})}.$$

Since $C(e^{s-1}, e^{-s}) = \exp\{-A(s)\}$ and $A(s) \leq 1$, we get

$$\sup_{s \in [0,1]} \left| \frac{C_n(e^{s-1}, e^{-s}) - C(e^{s-1}, e^{-s})}{C(e^{s-1}, e^{-s})} \right| \leq e \sup_{I^2} |C_n(u, v) - C(u, v)|,$$

which concludes the proof. ■

9.3 Asymptotic Law of the process $t \rightarrow \sqrt{n} [A_n(s) - A(s)]$

Theorem 7 *We have*

$$(s \rightarrow \sqrt{n} [A_n(s) - A(s)]) \rightsquigarrow (s \rightarrow D_C(s)), \quad n \rightarrow +\infty$$

where D is a Gaussian process defined by :

$$D(s) = \frac{G_C(e^{s-1}, e^{-s})}{C(e^{s-1}, e^{-s})},$$

and G_C is the Gaussian process defined above and \rightsquigarrow designates the weak convergence on $\ell^\infty([0; 1]^2)$.

$s \rightarrow D(s)$ is a Gaussian process with mean 0 and covariance given by :

$$\begin{aligned} \text{Cov}(D(s_1), D(s_2)) &= \frac{1}{C(e^{s_1-1}, e^{-s_1})C(e^{s_2-1}, e^{-s_2})} \times \\ &\left\{ C(e^{s_1-1} \wedge e^{s_2-1}, e^{-s_1} \wedge e^{-s_2}) - C(e^{s_1-1}, e^{-s_1})C(e^{s_2-1}, e^{-s_2}) \right. \\ &- \frac{\partial}{\partial u} C(e^{s_2-1}, e^{-s_2}) [C(e^{s_1-1} \wedge e^{s_2-1}, e^{-s_1}) - C(e^{s_1-1}, e^{-s_1})e^{s_2-1}] \\ &- \frac{\partial}{\partial v} C(e^{s_2-1}, e^{-s_2}) [C(e^{s_1-1}, e^{-s_1} \wedge e^{-s_2}) - C(e^{s_1-1}, e^{-s_1})e^{-s_2}] \\ &- \frac{\partial}{\partial u} C(e^{s_1-1}, e^{-s_1}) [C(e^{s_1-1} \wedge e^{s_2-1}, e^{-s_2}) - C(e^{s_2-1}, e^{-s_2})e^{s_1-1}] \\ &+ \frac{\partial}{\partial u} C(e^{s_1-1}, e^{-s_1}) \frac{\partial}{\partial u} C(e^{s_2-1}, e^{-s_2}) [e^{s_1-1} \wedge e^{s_2-1} - e^{s_1-1}e^{s_2-1}] \\ &+ \frac{\partial}{\partial u} C(e^{s_1-1}, e^{-s_1}) \frac{\partial}{\partial v} C(e^{s_2-1}, e^{-s_2}) [C(e^{s_1-1}, e^{-s_2}) - e^{s_1-1}e^{-s_2}] \\ &- \frac{\partial}{\partial v} C(e^{s_1-1}, e^{-s_1}) [C(e^{s_2-1}, e^{-s_1} \wedge e^{-s_2}) - C(e^{s_2-1}, e^{-s_2})e^{-s_1}] \\ &+ \frac{\partial}{\partial v} C(e^{s_1-1}, e^{-s_1}) \frac{\partial}{\partial u} C(e^{s_2-1}, e^{-s_2}) [C(e^{s_2-1}, e^{-s_1}) - e^{-s_1}e^{s_2-1}] \\ &\left. + \frac{\partial}{\partial v} C(e^{s_1-1}, e^{-s_1}) \frac{\partial}{\partial v} C(e^{s_2-1}, e^{-s_2}) [e^{-s_1} \wedge e^{-s_2} - e^{-s_1}e^{-s_2}] \right\} \end{aligned}$$

where $\frac{\partial}{\partial u} C(e^{s-1}, e^{-s})$ means $\frac{\partial}{\partial u} C(u, v)|_{u=e^{s-1}, v=e^{-s}}$.

Theorem 7 is a plain consequence of the δ -method applied to the following convergence result, due to Fermanian et al. (2004), which we recall hereafter for completeness.

Theorem 8 *Let (X, Y) two r.v. with joint distribution H . Suppose that H has continuous marginal distribution functions and that the copula has continuous partial derivatives. Then the empirical copula process $\{\sqrt{n}[C_n(u, v) - C(u, v)], 0 \leq u, v, \leq 1\}$ converges weakly to the Gaussian process $\{G_C(u, v), 0 \leq u, v \leq 1\}$ in $\ell^\infty([0; 1]^2)$.*

Since C is an extreme-value copulae, C is positive-quadrant dependent, which yields $C(e^{t-1}, e^{-t}) \geq e^{-1} > 0$. Slutsky's theorem, with the function $\Phi : x \mapsto \log x$, which is differentiable at $C(e^{t-1}, e^{-t})$ for all $t \in [0; 1]$, then yields

$$\sqrt{n} [\log C_n(e^{t-1}, e^{-t}) - \log C(e^{t-1}, e^{-t})] \xrightarrow{\mathcal{L}} \frac{G_C(e^{t-1}, e^{-t})}{C(e^{t-1}, e^{-t})},$$

for all $t \in [0; 1]$. Since the limit process is Gaussian, it is also a sub-gaussian process for the standard deviation semi-metric $d(s, t) = \sigma(X_s - X_t)$ (Van Der Vaart and Wellner (1996), p.101). So, we have the inequality $\mathbb{P}[|X_s - X_t| > x] \leq 2e^{-1/2x^2/d^2(s,t)}$ for $x > 0$, which concludes on the tightness of the process $X_s = \sqrt{n}[A_n(s) - A(s)]$. This ends the proof of the law of the process $t \rightarrow \sqrt{n}[A_n(s) - A(s)]$.

9.4 Comparison with other estimators

We consider three estimates of A ,

- Pickands

$$A_P(s) = n \left[\sum_{i=1}^n \min \left(\frac{X_i}{s}, \frac{Y_i}{1-s} \right) \right]^{-1}.$$

- Capéraà-Fougères-Genest

$$A_{CFG}(s) = \exp \left\{ p(s) \int_0^s \frac{H_n(z) - z}{z(1-z)} dz - [1 - p(s)] \int_s^1 \frac{H_n(z) - z}{z(1-z)} dz \right\}$$

where H_n is the empirical distribution function of $Z = \frac{\log U}{\log UV}$ with uniform transforms $U = F(X)$, $V = G(Y)$ (in the case of known margins) and $\bar{X} = n^{-1} \sum_i X_i$, $\bar{Y} = n^{-1} \sum_i Y_i$.

- our estimate

$$A_{BD}(s) = -\log C_n(e^{s-1}, e^{-s}).$$

None of them satisfies the inequality $\max(s, 1-s) \leq A_*(s) \leq 1$. In practice, the software package *evd* written by *Alex Stephenson* based on *Coles (2001)*, provides estimates of the form $\tilde{A}_*(s) = \min(1, \max(s, 1-s, A_*(s)))$. We adopt the same modification for A_{BD} . Simulations hereafter are based on the Gumbel-Hougaard (also called Gumbel's copula) copula

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

with $0 < \alpha < 1$.

We consider a sample $(X_i, Y_i)_{i=1}^n$ with common Gumbel copulae and non extreme-value marginals, namely $F(x) = x^2$ on $[0; 1]$ and $F_Y = F_X$, with $n = 500$ (Gumbel margins) and $n = 1000$ for non extreme-value margins. The result is in Figure 9.1 for Gumbel marginals and in Figure 9.2 for non-extreme marginals. On both figures, a confidence interval (gray solid line) and the lower bound ($t \rightarrow \min(t, 1-t)$) are drawn according to Theorem 7.

Next we simulate N samples $(X_i, Y_i)_{i=1}^n$ of various sizes n .

$$e_*(n) = \frac{1}{N} \sum_{l=1}^N \sup_t |A(t) - A_*^n(t)|$$

We estimate the uniform error committed when estimating A . We have considered the case when the margins obey an extreme value (Gumbel) distributions.

Note that the confidence bound shown in Figure 9.1 does not fit with the true one in neighborhoods of 0 and 1, due to the substitution of A_{BD} by \tilde{A}_{BD} .

We note that in the case of extreme-value margins, the uniform error is slightly larger than its competitors ones, since no information pertaining to margins is used. At the contrary, in the case of non extreme-value margins, our estimate performs well for large

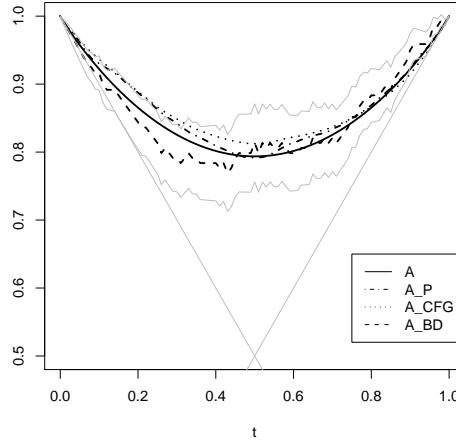


Figure 9.1: Gumbel marginals - Gumbel-Hougaard copula ($\theta = 1.5$)

T_j	20	50	100	500	1000
$A_{Pickands}$	0.1046568	0.06418146	0.05083636	0.02141466	0.01548567
A_{Hall}	0.1046544	0.064186612	0.05083797	0.02141506	0.01548622
A_{CFG}	0.08800316	0.05500164	0.03951114	0.01673942	0.0126947
A_{BD}	0.1799266	0.1184944	0.0861421	0.03719185	0.02534665

Table 9.1: $\sup |\hat{A}_*(s) - A(s)|$ - case of extreme-value margins

sample size; the competitors also behave satisfactorily; this may suggest that their consistency can be proved without any requirements on the margins. However their uniform errors decay very slowly (even, for Pickands' estimate, does not decay) with respect to the sample size.

To conclude, we show a practical comparison of the limit variance of the Capéraà-Fougères-Genest estimator A_{CFG} and the variance of the limiting process in Theorem 7. From Capéraà et al. (1997), the limit variance is

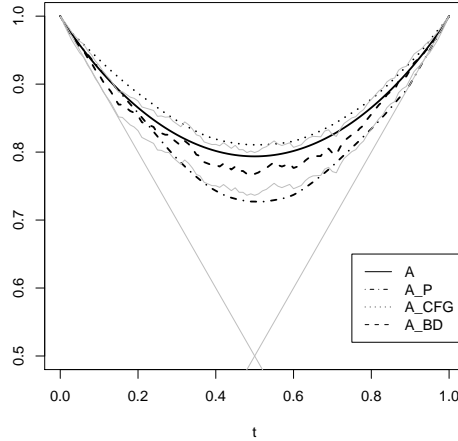
$$\Gamma(t) = p(t)^2 \Gamma_0(t, t) + [1 - p(t)]^2 \Gamma_1(t, t) - 2p(t) [1 - p(t)] V(t) \tag{9.5}$$

with

$$\Gamma_0(t, t) = \int_0^t \int_0^t \frac{H(u \wedge v) - H(u)H(v)}{uv(1-u)(1-v)} \, du \, dv \tag{9.6}$$

$$\Gamma_1(t, t) = \int_t^1 \int_t^1 \frac{H(u \wedge v) - H(u)H(v)}{uv(1-u)(1-v)} \, du \, dv \tag{9.7}$$

$$V(t) = - \int_0^t \int_t^1 \frac{H(u) [1 - H(v)]}{uv(1-u)(1-v)} \, du \, dv \tag{9.8}$$

Figure 9.2: Non-extreme value margins - Gumbel-Hougaard copula ($\theta = 1.5$)

T_j	20	50	100	500	1000
$A_{Pickands}$	0.1196512	0.08234023	0.06676055	0.04849252	0.05648585
A_{Hall}	0.1059388	0.07351504	0.005461706	0.03566127	0.03333635
A_{CFG}	0.1327330	0.07713294	0.05553333	0.03991047	0.03990171
A_{BD}	0.1735227	0.1178186	0.0875918	0.0359993	0.02596318

Table 9.2: $\sup \left| \hat{A}_*(s) - A(s) \right|$ - case of non-extreme value margins

where H is the distribution function of $Z = \frac{\log(U)}{\log(UV)}$ and $(U_i, V_i) = (F(X_i), G(Y_i))$ with F, G the known margins of X, Y .

In figures 9.3 and 9.4, the solid line is the limit variance of our estimator, the gray dotted line is the limit variance for the weight function $p(t) = 1 - t$ and the black dotted line is the limit variance for the optimal weight function $p(t) = \frac{\Gamma_1(t,t) - V(t)}{\Gamma_0(t,t) + \Gamma_1(t,t) - 2V(t)}$, Capéraà et al. (1997).

The limit variance of our estimator is less than the limit variance of CFG estimator, even in the case of the optimal choice for the weight function. The limit variance of our estimator is four times less than the limit variance of CFG estimator is the particular point $t = 0.5$, where the variability is maximal.

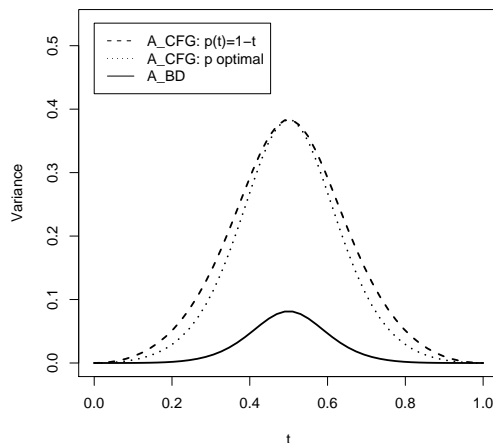
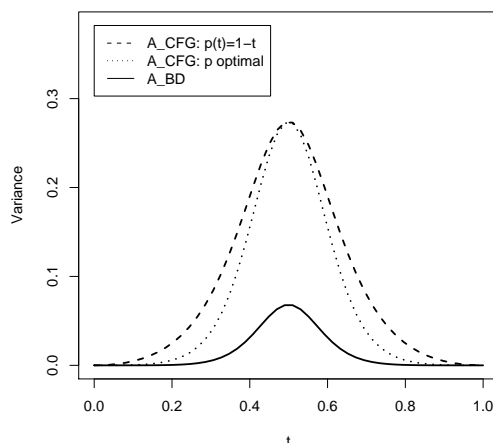
Figure 9.3: Gumbel-Hougaard copula with $\theta = 4$ 

Figure 9.4: Galambos copula with $\theta = 4 - (A(t) = 1 - [t^{-\theta} + (1-t)^{-\theta}]^{-1/\theta})$ $C(u, v) = uv \exp \left\{ [(-\log u)^{(-\theta)} + (-\log v)^{(-\theta)}]^{(-1/\theta)} \right\}$

9.5 Discussion

Definition 9.3 uses a small amount of the information provided by the sample since the value of A at point s is related only to the empirical copula at $(u = e^{s-1}, v = e^{-s})$. Since A is constant on the arc $\mathfrak{s} = \{(u, v) \in [0, 1]^2, u = v^{1/s-1}\}$, a competitor of our estimator (see (9.3)) can be defined through

$$\tilde{A}_n(s) = -\log \iint_{(u,v) \in \mathfrak{s}} C_n(u, v) \, du \, dv. \quad (9.9)$$

It can be shown quite easily that $\lim_{n \rightarrow +\infty} \sup_s |\tilde{A}_n(s) - A(s)| = 0$ a.s., a statement in the same vein as Proposition 14. The limit distribution of a normalized version of the process $\tilde{A}_n(s)$ seems quite complex. Further numerical simulations of $\tilde{A}_n(s)$ do not prove that its performance is substantially higher than that of $A_n(s)$. Indeed, the variability of $\tilde{A}_n(s)$ exceeds that of $A_n(s)$, since the averaging in (9.9) yields erratic behaviors of \tilde{A}_n due to small values of both u and v on the arc \mathfrak{s} .

Another amelioration of our estimator can be to use a smoothing version of empirical copula, in order to avoid the irregularity of A_{BD} .

Graphiques

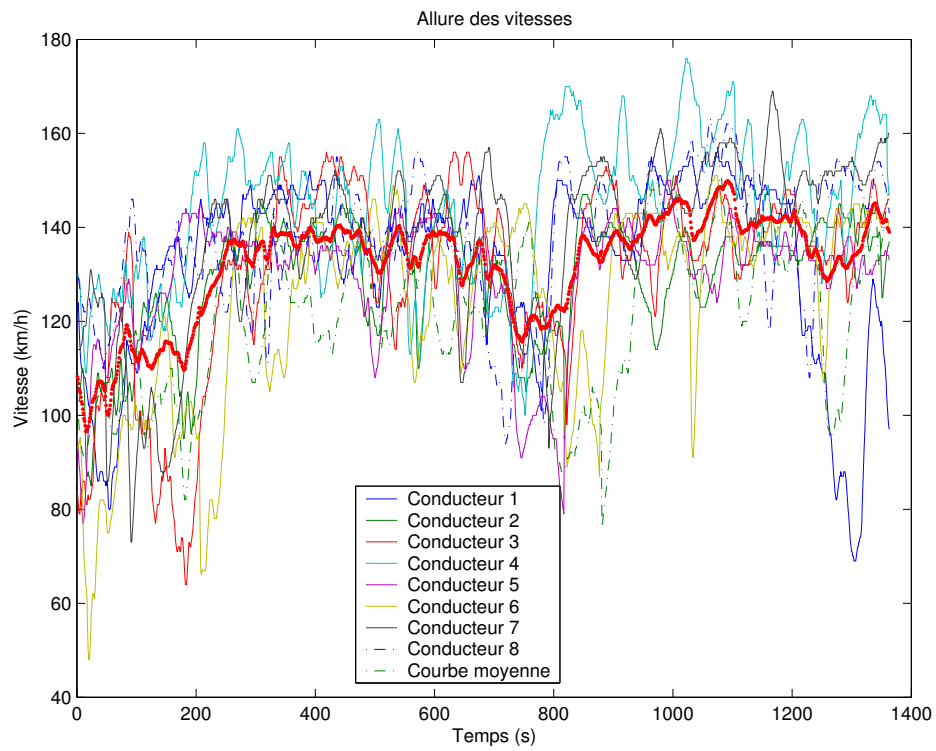


Figure 5: Allure des vitesses pratiquées par les conducteurs

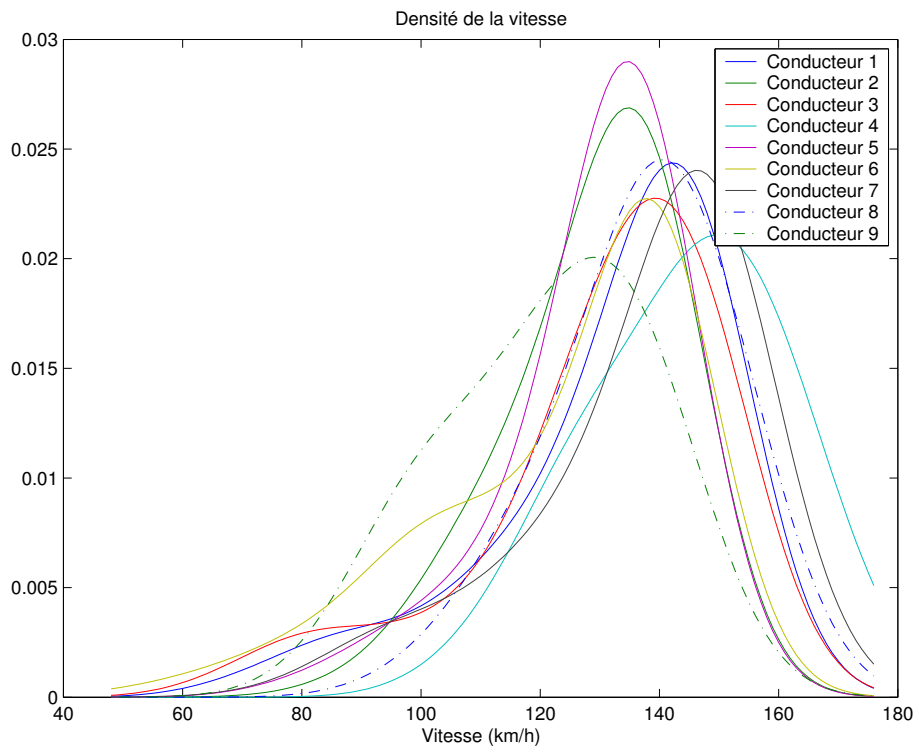


Figure 6: Densité des vitesses de chaque conducteurs

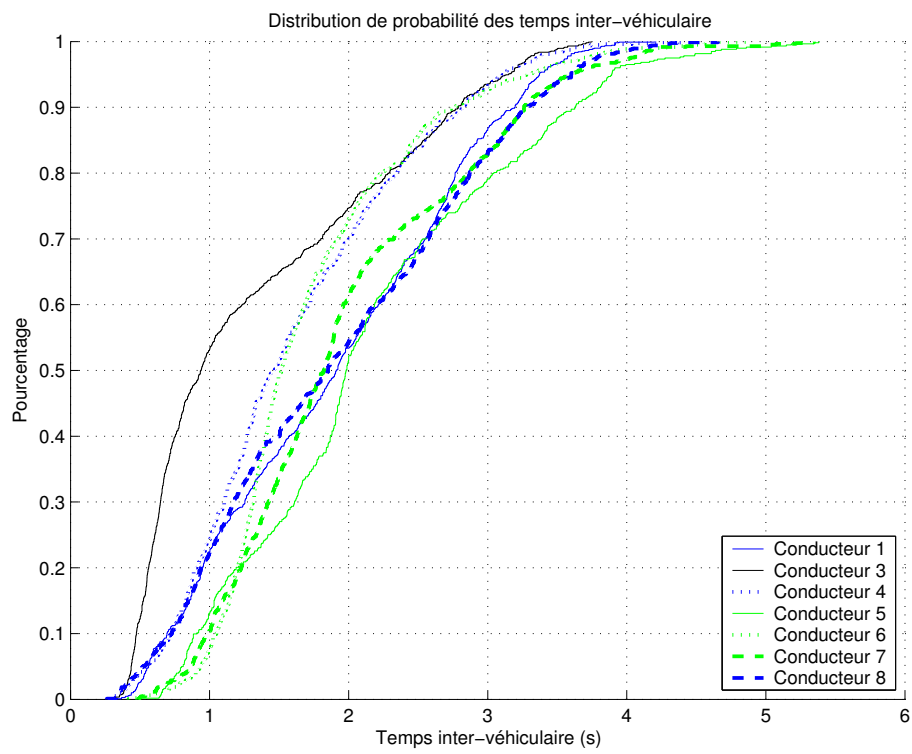


Figure 7: Distribution des temps intervéhiculaires (s)

Input-Output Hidden Markov Model

from M. Broniatowski, A. Depire and Y. Ritov.

.1 Main Results

Consider an unobserved states X_1, X_2, \dots , observations Y_1, Y_2, \dots and observed control (or context) variables C_1, C_2, \dots . To simplify notations, let $X_{i:j} = (X_i, X_{i+1}, \dots, X_j)$, with similar notation for all other processes.

We assume :

A1 Given the context, the state process $X_{1:\infty}$ is Markovian :

$$\mathbb{P}[X_{t+1} = x | X_{1:t}, C_t] = T_\theta(X_t, x; C_t), \quad \theta \in \Theta.$$

A2 The series Y_1, Y_2, \dots given X_1, X_2, \dots is of *i.i.d.* random variables. The conditional distribution of $Y_{1:n}$ has a density $\prod_{i=1}^n g_\theta(y_i | X_i, C_i)$ with respect to some product measure. Let $G_\theta(\cdot | \cdot, \cdot)$ be the probability measure which correspond to $g_\theta(\cdot | \cdot, \cdot)$.

A3 The series $C_{1:\infty}$ causally depends on the Y s :

$$\mathcal{L}(C_t | C_{1:t-1}, Y_{1:n}, X_{1:n}) = \mathcal{L}(C_t | C_{1:t-1}, Y_{1:t}), \quad t \leq n$$

As a result of Assumption A3, we can consider inference conditioned on context, such that the context is considered fixed and non-stochastic. We also assume the following structural regularity conditions :

A4 The state X_t and the control C_t , $t = 1, 2, \dots$, can get a finite number of values, $X_t \in \{\xi_1, \dots, \xi_K\}$ and $C_t \in \{\zeta_1, \dots, \zeta_M\}$.

A5 Θ is an open set of an euclidean space. If for any $\theta, \theta' \in \Theta$, $T_\theta(\cdot, \cdot; \cdot) = T_{\theta'}(\cdot, \cdot; \cdot)$ and $g_\theta(\cdot | \cdot) = g_{\theta'}(\cdot | \cdot)$ then $\theta = \theta'$.

A6 $\omega = \min_{\xi_1, \xi_2, \zeta, \theta} T_\theta(\xi_1, \xi_2; \zeta) > 0$.

Note that we don't specify or parametrize the distribution of the context variables. As a result, we consider only the marginal (or partial) likelihood of the observed Y s, defined recursively by :

$$\begin{aligned} \log L_\theta(Y_1, \dots, Y_n | C_{1:n}) &= \log p_\theta(Y_1, \dots, Y_n | C_{1:n}) \\ &= \sum_{t=1}^n \log p_\theta(Y_t | Y_{1:t-1}, C_{1:n}) \end{aligned}$$

with

$$\begin{aligned}
p_\theta(Y_t|Y_{1:t-1}, C_{1:n}) &= \lim_{\Delta} \frac{1}{\Delta} \mathbb{P}_\theta \left[Y_t \leq \tilde{Y}_t \leq Y_t + \Delta | Y_{1:t-1}, C_{1:n} \right], \tilde{Y}_t \text{ iid copies of } Y_t \\
&= \sum_{k=1}^K \lim_{\Delta} \frac{1}{\Delta} \mathbb{P}_\theta \left[Y_t \leq \tilde{Y}_t \leq Y_t + \Delta, X_t = \xi_k | Y_{1:t-1}, C_{1:n} \right] \text{ (finite summation)} \\
&= \sum_{k=1}^K \lim_{\Delta} \frac{1}{\Delta} \mathbb{P}_\theta \left[Y_t \leq \tilde{Y}_t \leq Y_t + \Delta | X_t = \xi_k, Y_{1:t-1}, C_{1:n} \right] \mathbb{P}_\theta [X_t = \xi_k | Y_{1:t-1}, C_{1:n}] \\
&= \sum_{k=1}^K g_\theta(Y_t; \xi_k, c_t) \mathbb{P}_\theta [X_t = \xi_k | Y_{1:t-1}, C_{1:n}]
\end{aligned} \tag{10}$$

With markovien hypothesis, we obtain :

$$L_\theta(Y_t|Y_{1:t-1}, C_{1:n}) = \sum_{k=1}^K g_\theta(Y_t; \xi_k, c_t) \mathbb{P}_\theta [X_t = \xi_k | Y_{1:t-1}, C_{1:n}].$$

Let $\mathfrak{B}_t(k) = \mathbb{P}_\theta [X_t = \xi_k | Y_{1:t-1}, C_{1:n}]$, it is defined recursively :

$$\mathfrak{B}_t(\xi_k) = \frac{\sum_{l=1}^K \mathfrak{B}_{t-1}(\xi_l) T(\xi_l, \xi_k, c_t) g_\theta(Y_{t-1}, \xi_l, c_{t-1})}{\sum_{k'=1}^K \sum_{l=1}^K \mathfrak{B}_{t-1}(\xi_l) g_\theta(Y_{t-1}, \xi_l, c_{t-1}) T(\xi_l, \xi_{k'}, c_t)} \tag{11}$$

This relation is obtained as following

$$\begin{aligned}
\mathbb{P}_\theta [X_t = \xi_k | Y_{1:t-1}, C_{1:n}] &= \sum_{l=1}^K \mathbb{P}_\theta [X_t = \xi_k, X_{t-1} = \xi_l | Y_{1:t-1}, C_{1:n}] \\
&= \sum_{l=1}^K \frac{\mathbb{P}_\theta [X_t = \xi_k, X_{t-1} = \xi_l, Y_{1:t-1}, C_{1:n}]}{\mathbb{P}_\theta [Y_{1:t-1}, C_{1:n}]} \\
&= \sum_{l=1}^K \mathbb{P}_\theta [X_t = \xi_k | X_{t-1} = \xi_l, Y_{1:t-1}, C_{1:n}] \frac{\mathbb{P}_\theta [X_{t-1} = \xi_l, Y_{1:t-1}, C_{1:n}]}{\mathbb{P}_\theta [Y_{1:t-1}, C_{1:n}]} \\
&= \sum_{l=1}^K T_\theta(\xi_l, \xi_k; c_t) \frac{\mathbb{P}_\theta [X_{t-1} = \xi_l, Y_{1:t-1}, C_{1:n}]}{\mathbb{P}_\theta [Y_{1:t-1}, C_{1:n}]} \\
&= \sum_{l=1}^K T_\theta(\xi_l, \xi_k; c_t) \frac{\mathbb{P}_\theta [Y_{t-1} | X_{t-1} = \xi_l, Y_{1:t-2}, C_{1:n}] \mathbb{P}_\theta [Y_{1:t-2}, X_{t-1} = \xi_l, C_{1:n}]}{\mathbb{P}_\theta [Y_{1:t-1}, C_{1:n}]} \\
&= \sum_{l=1}^K T_\theta(\xi_l, \xi_k; c_t) g_\theta(Y_{t-1}, \xi_l, c_{t-1}) \mathbb{P}_\theta [X_{t-1} = \xi_l | Y_{1:t-2}, C_{1:n}] \frac{\mathbb{P}_\theta [Y_{1:t-2}, C_{1:n}]}{\mathbb{P}_\theta [Y_{1:t-1}, C_{1:n}]} \\
&= \sum_{l=1}^K T_\theta(\xi_l, \xi_k; c_t) g_\theta(Y_{t-1}, \xi_l, c_{t-1}) \mathfrak{B}_{t-1}(l) \frac{1}{\mathbb{P}_\theta [Y_{t-1} | Y_{1:t-2}, C_{1:n}]}
\end{aligned}$$

The last factor does not depend on l , so

$$\begin{aligned} \mathbb{P}_\theta [Y_{t-1}|Y_{1:t-2}, C_{1:n}] &= \sum_{q=1}^K \mathbb{P}_\theta [Y_{t-1}, X_{t-1} = \xi_q | Y_{1:t-2}, C_{1:n}] \\ &= \sum_{q=1}^K g_\theta(Y_{t-1}, \xi_q, c_{t-1}) \mathbb{P}_\theta [X_{t-1} = \xi_q | Y_{1:t-2}, C_{1:n}] \\ &= \sum_{q=1}^K \mathfrak{B}_{t-1}(\xi_q) g_\theta(Y_{t-1}, \xi_q, c_{t-1}) \end{aligned}$$

To conclude, we calculate

$$\begin{aligned} \mathbb{P}_\theta [Y_{t-1}|Y_{1:t-2}, C_{1:T}] &= \sum_{q=1}^K \mathfrak{B}_{t-1}(\xi_q) g_\theta(Y_{t-1}, \xi_q, c_{t-1}) \\ &= \sum_{k'=1}^K \sum_{l=1}^K \mathfrak{B}_{t-1}(\xi_l) g_\theta(Y_{t-1}, \xi_l, c_{t-1}) T(\xi_l, \xi_{k'}, c_t) \end{aligned}$$

In essence, this is the likelihood of the Y s conditional on the observed contexts. Note that $\mathfrak{B}_t(\xi_k, \theta)$ depends explicitly on the context $C_{1:t}$.

Then the log-likelihood is given by :

$$\begin{aligned} \ell(\theta, Y_{1:n}|C_{1:n}) &= \sum_{t=1}^n \log L_t(\theta, Y_t|Y_{1:t-1}, C_{1:n}) \\ &= \sum_{t=1}^n \ell(\theta, Y_t|C_{1:n}, Y_{1:t-1}) \end{aligned} \tag{12}$$

where, with some abuse of notation, $\ell(\theta, W)$ is the log-likelihood function for θ after observing the random variable W .

We want to prove in this text that the log-likelihood equation has quadratic approximation in the neighborhood of θ_0 ; with curvature given by the observed information for θ . As a result, we will obtain that the likelihood equation has an asymptotical normality and efficient root in the neighborhood of the truth. Therefore, if the MLE (Maximum Likelihood Estimator) is \sqrt{n} -consistent, it is efficient. The existence of a \sqrt{n} -consistent estimator is guaranteed if $\min_k \sum_{t=1}^n \mathbb{1}_{\{C_t = \zeta_k\}} \xrightarrow{P} \infty$, and the model is identified by the distribution of any fixed length chain (as implied by Assumptions A5 and A6), we can always consider the sum of the log-likelihoods of the sub-chains with a fixed length k .

Our proof depends heavily on the ideas and methods in Bickel et al. (1998) and Bickel et al. (2002). The seminal work of any analysis of HMMs is Baum and Petrie (1966). This was restricted to finite distributions. Bickel and Ritov (1996) proved that under some conditions, the HMM model with finite state space and general observation space satisfy the LAN condition, and hence MLE-like estimators are efficient. Bickel et al. (1998) extend the result considerably and proved that in fact the MLE is efficient. Finally,

Bickel et al. (2002) gave a general bounds on the derivatives of the log-likelihood function of such models.

The generalization of the bounds given in Bickel et al. (2002) would be immediate and simple. Extending the results of Bickel and Ritov (1996) and Bickel et al. (1998) should be done with more care since the presentation of these results was heavily depended on the stationnarity (or at least asymptotic stationarity) of the hidden Markov chain. A property our model lacks without additional assumptions on the stochastic behavior of the context variables.

We need the following differentiability conditions. Here \mathbb{E}_0 is the expectation under the truth.

A7 There is a neighborhood $V \subset \Theta$ of the true

$$\mathbb{E}_{\theta_0} \left[\sup_{\theta \in \Theta} \max_x \left| \frac{\partial^k}{\partial \theta^k} \log g_\theta(Y_t|x) \right|^3 \right] < A$$

for all $k \leq 3$. We further assume as usual that derivative and expectations commute.

We need another assumptions that will ensure that all combinations of the arguments of T will occur frequently. Otherwise, there would be no hope to estimate θ within the \sqrt{n} -rate.

A8 For any $j \in \{1, \dots, M\}$, $\liminf_n n^{-1} \sum_{t=1}^n \mathbf{1}_{\{C_t=\zeta_j\}} > 0$ with probability 1. Moreover, for any $\varepsilon > 0$, there is $\eta > 0$ such that:

$$S_\eta = \left\{ y : \inf_c \frac{\min_x g_\theta(y|x, c)}{\max_x g_\theta(y|x, c)} > \eta \right\}$$

we have $\lim_{\eta \rightarrow 0} \min_{x,c} G_\theta(S_\eta|x, c) = 1$.

This condition is satisfied if Y is x plus normal noise. More generally, if $\max_{x,c} g_\theta(y|x, c) \min_x g_\theta(y|x, c)$ is bounded on compacts. Without this condition, either the C_t may visit ζ_1 to infrequently so that there would not enough information to estimate $T(\cdot, \cdot; \zeta_1)$ at the \sqrt{n} -rate, or if the second part of the assumption is not satisfied, we may have that ξ_1 is visited frequently enough, but only when the *a priori* probability that $X_t = \xi_1$ is null, so that $T(\xi_1, \cdot; \zeta_1)$ can be estimated.

The conclusion we draw from this assumption is the following.

Lemma 8 *If Assumptions A6 and A8 are satisfied, then for every $i = 1, \dots, K$ and $j = 1, \dots, M$,*

$$\liminf n^{-1} \sum_{t=1}^n \mathbf{1}_{\{X_t=\xi_i, C_t=\zeta_j\}} > 0$$

with probability 1.

Proof. All the equalities in this proof hold only with probability 1. Let $\varepsilon = \liminf_n n^{-1} \sum_{t=1}^n \mathbb{1}_{\{C_t = \zeta_j\}}$, and let τ_1, τ_2, \dots , be the visits of $\{C_t\}$ at ζ_j . We prove the lemma by contradiction. Suppose the lemma is not true and $\liminf n^{-1} \sum_{m=1}^n \mathbb{1}_{\{X_{\tau_m} = \xi_i\}} = 0$. Note that for every η , if $Y_t \in S_\eta$, then $\mathfrak{B}_\theta^{1:t}(\xi_i) > n\omega$, consequently :

$$\begin{aligned} 0 &= \liminf_m \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{\{X_{\tau_k} = \xi_i\}} \\ &= \liminf_m \frac{1}{m} \sum_{k=1}^m \mathfrak{B}_\theta^{1:\tau_k}(\xi_i) \\ &\geq n\omega \liminf_m \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{\{Y_{\tau_k} \in S_\eta\}} \end{aligned}$$

hence

$$\begin{aligned} 1 &= \liminf_m \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{\{Y_{\tau_k} \notin S_\eta\}} \\ &\leq \liminf_m \frac{\tau_m}{m} \frac{1}{\tau_m} \sum_{t=1}^{\tau_m} \mathbb{1}_{\{Y_t \notin S_\eta\}} \\ &\leq \liminf_m \frac{\tau_m}{m} \left(1 - \min_x G_\theta(S_\eta|x)\right) \\ &\leq \frac{1}{\varepsilon} \left(1 - \min_x G_\theta(S_\eta^c(x))\right) \end{aligned}$$

However, by Assumptions A8, the RHS converges to 0 as $\eta \rightarrow 0$ and we obtained a contradiction. The lemma follows. ■

In the sequel we will use A_1, A_2, A_3 to denote constants that may vary from equation to equation. To simplify notation and to avoid considering directed derivatives, Hessians, etc, we will assume that $\Theta \subset \mathbb{R}$. Let $\dot{\ell}$ and $\ddot{\ell}$ denote the first and second derivatives of ℓ with respect to θ .

The following is an almost immediate consequence of our conditions.

Theorem 9 *Under the above assumptions, and for every sequence $c_{1:n}$ of the context variables, there is a constant $A < \infty$ which depends only on the bounds in Assumption A7 such that*

$$\mathbb{E}_{\theta_0} \left[\sup_{\theta \in \Theta} \max_{k \leq 3} \left| \frac{\partial^k}{\partial \theta^k} \ell(\theta, Y_{1:n}, c_{1:n}) \right| \right] \leq nA$$

where the expectation is taken over the distribution of $Y_{1:n}$ given the context $c_{1:n}$.

The proof follows essentially from Theorems 2.1 and 2.2 of Bickel et al. (2002), since the fact that the chain $X_{1:n}$ is not stationary given the context has no impact on mixing bounds used in the proof of these theorems, and the bounds are only related to the conditional distribution of $X_{1:n}$ given the observable variable. This chain is exponential ergodic with a uniform mixing in view of Assumptions A6.

We conclude from Theorem 9, that

$$\ell(\theta, Y_{1:n}|c_{1:n}) - \ell(\theta_0, Y_{1:n}|c_{1:n}) = (\theta - \theta_0)\dot{\ell}(\theta_0, Y_{1:n}|c_{1:n}) + \frac{1}{2}(\theta - \theta_0)^2\ddot{\ell}(\theta_0, Y_{1:n}|c_{1:n}) + \nu_\theta. \quad (13)$$

where $\sup_{\theta \in V} |n^{-1}(\theta - \theta_0)^{-3}\nu_\theta| = \mathcal{O}_P(1)$. Equation 13, at least heuristically, leads to the validity of the quadratic expansion, at least in a $n^{-1/2}$ neighborhood of θ_0 .

The following proposition would be helpful.

Proposition 15 *Under the above conditions :*

1.

$$\mathbb{E}_{\theta_0} \left[\sup_{\theta \in \Theta} \max_{k \leq 3} \left| \frac{\partial^k}{\partial \theta^k} \log \mathfrak{B}_\theta^{1:t}(x) \right| \right] \leq A_1$$

2. For some $\rho < 1$ and all $0 \leq k \leq 3$, $1 \leq s \leq t \leq n$, contexts $c_{1:n}$ and x ,

$$\left| \frac{\partial^k}{\partial \theta^k} [\mathbb{P}[X_t = \xi | Y_{s:t}, c_{s:t}, X_{s+1} = \xi_1] - \mathbb{P}[X_t = \xi | Y_{1:t}, c_{1:t}]] \right| < A_2 \rho^{t-s}.$$

The proof, again, follows Bickel et al. (2002) Theorem 2.2. The second part of the Proposition (15) is essentially restatement of Bickel et al. (2002) Theorem 2.2 (ii), while the first part follows Theorem 2.2 (i) with the identification of Y_1 with X_1 and shifting of the time scale.

To simplify notation, we write $\mathbb{P}_\theta[X_t = \xi | Y_{s:t}, C_{s:t}]$ for $\mathbb{P}[X_t = \xi | Y_{s:t}, C_{s:t}, X_{s+1} = \xi_1]$, and similarly for other cases of conditioning on short past.

The summands on the expansion of $\ell(\theta, Y_{1:n}|c_{1:n})$ as on the RHS of (12) depend on their past only through $\mathfrak{B}_\theta^t(\cdot) = \mathbb{P}_\theta[X_t = \cdot | Y_{1:t-1}; c_{1:t}]$, see (11). The consequence of Proposition (15) is, therefore, that the covariances between the summands of both $\dot{\ell}(\theta_0, Y_{1:n}|c_{1:n})$ and $\ddot{\ell}(\theta_0, Y_{1:n}|c_{1:n})$ decreases exponentially with the time distance and therefore the variances of this terms is of order n , as needed for the quadratic approximation in (13) to be valid in an $n^{-1/2}$ neighborhood of the truth.

Further more, the fact that $\dot{\ell}$ has mean 0 and variance with is equal to negative of the mean of $\ddot{\ell}$ follows directly from the commutability part of Assumption A7. So the LAMN (locally asymptotically mixed normal) property would follow if we show that CLT applies to $n^{-1/2}\dot{\ell}(\theta, Y_{1:n}|c_{1:n})$, and LLN to $n^{-1}\ddot{\ell}(\theta, Y_{1:n}|c_{1:n})$.

Let $\mathfrak{B}_\theta^{s:t}(\cdot) = \mathbb{P}_\theta[X_t = \cdot | Y_{s:t}; C_{s:t}, X_{s+1} = \xi_1]$, an immediate consequence of Proposition 15 is the following :

Lemma 9 *Let $\tau = \tau_n$ be such that $\tau_n/n \rightarrow 0$ and $\tau_n/\log n \rightarrow +\infty$. Then*

$$\begin{aligned} \sum_{t=1}^n \dot{\ell}(\theta, Y_t | Y_{1:t-1}, C_{1:t}) &= \sum_{t=1}^n \dot{\ell}(\theta, Y_t | Y_{s_t:t-1}, C_{s_t:t}) + \mathcal{O}_p(n\rho^{\tau_n}) \\ \sum_{t=1}^n \ddot{\ell}(\theta, Y_t | Y_{1:t-1}, C_{1:t}) &= \sum_{t=1}^n \ddot{\ell}(\theta, Y_t | Y_{s_t:t-1}, C_{s_t:t}) + \mathcal{O}_p(n\rho^{\tau_n}) \end{aligned}$$

where $s_t = \max\{1, t - \tau\}$.

Proof. Write

$$\begin{aligned} \dot{\ell}(\theta, Y_t|Y_{1:t-1}, C_{1:t}) - \dot{\ell}(\theta, Y_t|Y_{s_t:t-1}, C_{s_t:t}) &= \frac{\sum_x \mathfrak{B}_\theta^{1:t}(x) g_\theta(Y_t|x) + \sum_x \mathfrak{B}_\theta^{1:t}(x) \dot{g}_\theta(Y_t|x)}{\sum_x \mathfrak{B}_\theta^{1:t}(x) g_\theta(Y_t|x)} \\ &\quad - \frac{\sum_x \mathfrak{B}_\theta^{s_t:t}(x) g_\theta(Y_t|x) + \sum_x \mathfrak{B}_\theta^{s_t:t}(x) \dot{g}_\theta(Y_t|x)}{\sum_x \mathfrak{B}_\theta^{s_t:t}(x) g_\theta(Y_t|x)} \end{aligned}$$

this difference can be expanded to six terms as usual. These terms can be bounded by applying Proposition 15 and Assumptions A6 and A7. We give now only two examples of these terms. Note that $\min_x \mathfrak{B}_\theta^{s_t:t}(x) \geq \omega = \min T(\cdot, \cdot)$. Hence

$$\begin{aligned} \left| \frac{\sum_x \mathfrak{B}_\theta^{1:t}(x) g_\theta(Y_t|x) \sum_x (\mathfrak{B}_\theta^{1:t}(x) - \mathfrak{B}_\theta^{s_t:t}(x)) g_\theta(Y_t|x)}{\sum_x \mathfrak{B}_\theta^{1:t}(x) g_\theta(Y_t|x) \sum_x \mathfrak{B}_\theta^{s_t:t}(x) g_\theta(Y_t|x)} \right| &\leq \max \frac{\mathfrak{B}_\theta^{1:t}(x)}{\mathfrak{B}_\theta^{s_t:t}(x)} \max \frac{\mathfrak{B}_\theta^{1:t}(x) - \mathfrak{B}_\theta^{s_t:t}(x)}{\mathfrak{B}_\theta^{1:t}(x)} \\ &\leq A_1 A_2 \omega^{-1} \rho^\tau \end{aligned}$$

and

$$\begin{aligned} \left| \frac{\sum_x [\mathfrak{B}_\theta^{1:t}(x) - \mathfrak{B}_\theta^{s_t:t}(x)] \dot{g}_\theta(Y_t|x)}{\sum_x \mathfrak{B}_\theta^{1:t}(x) g_\theta(Y_t|x)} \right| &\leq \max \frac{\mathfrak{B}_\theta^{1:t}(x) - \mathfrak{B}_\theta^{s_t:t}(x)}{\mathfrak{B}_\theta^{1:t}(x)} \max_x \frac{\dot{g}_\theta(Y_t|x)}{g_\theta(Y_t|x)} \\ &\leq A_2 \omega^{-1} \rho^\tau \max_x \frac{\dot{g}_\theta(Y_t|x)}{g_\theta(Y_t|x)}, \end{aligned}$$

where A_1 and A_2 are the same constants as in Proposition 15. We can bound, therefore

$$\left| \dot{\ell}(\theta, Y_t|Y_{1:t-1}, C_{1:t}) - \dot{\ell}(\theta, Y_t|Y_{s_t:t-1}, C_{s_t:t}) \right| < A \rho^\tau \frac{\dot{g}_\theta(Y_t|x)}{g_\theta(Y_t|x)}.$$

Consequently,

$$\begin{aligned} \left| \sum_{t=1}^n \dot{\ell}(\theta, Y_t|Y_{1:t-1}, C_{1:t}) - \sum_{t=1}^n \dot{\ell}(\theta, Y_t|Y_{s_t:t-1}, C_{s_t:t}) \right| &\leq A \rho^\tau \sum_{t=1}^n \frac{\dot{g}_\theta(Y_t|x)}{g_\theta(Y_t|x)} \\ &= \mathcal{O}_p(n \rho^{\tau n}) \end{aligned}$$

By LLN for random functions of a Markovian state, Saulis and Statulevičius (1991), and Assumption A7. The proof for the approximatio of $\ddot{\ell}$ is similar. ■

Lemma 10 *Let*

$$I_t(\theta, Y_{1:t}, C_{1:t-1}) = \sum_{s=1}^t \mathbb{E} \left[\dot{\ell}^2(\theta, Y_s|Y_{1:s-1}, C_{1:s-1}) | Y_{1:s-1}, C_{1:s-1} \right],$$

then

$$I_t^{-1/2}(\theta, Y_{1:t}, C_{1:t-1}) \dot{\ell}(\theta, Y_{1:t}|C_{1:t}) \xrightarrow{D} N(0, 1)$$

and

$$n^{-1} \ddot{\ell}(\theta, Y_{1:n}|C_{1:n}) + n^{-1} I_t(\theta, Y_{1:t}, C_{1:t-1}) \xrightarrow{p} 0.$$

Proof. The process $\sum_{t=1}^n \dot{\ell}(\theta, Y_t | Y_{1:t-1}, C_{1:t})$, $n = 1, 2, \dots$, is a martingale. Moreover its quadratic variation process is $\langle \dot{\ell}(\theta, Y_{1:t}, C_{1:t}) \rangle = I_t(\theta, Y_{1:t}, C_{1:t-1})$.

However, a hiding mechanism always reduces the information, and therefore,

$$\Delta I_t(\theta, Y_{1:t}, C_{1:t-1}) \leq \mathbb{E} \left[\dot{\ell}^2(\theta, Y_t, X_t | X_{t-1}, C_{t-1}) | X_{t-1}, C_{t-1} \right]$$

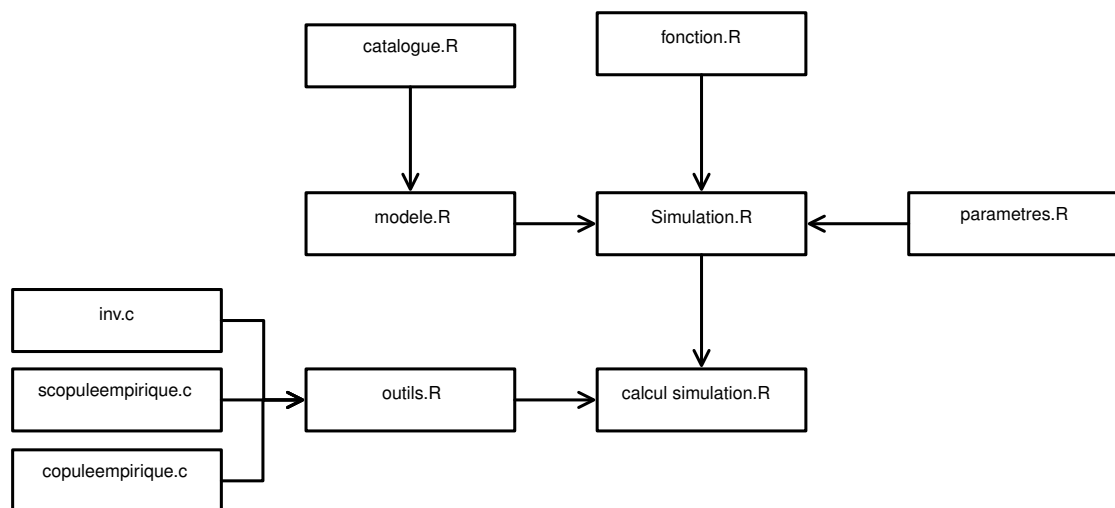
where for any sequence a_1, a_2, \dots , $\Delta a_t = a_t - a_{t-1}$. Hence, we can apply to deduce the first claim of the lemma.

For the second part note that by Assumption A7, $\mathbb{E} \ddot{\ell}(\theta, Y_t | Y_{1:t-1}, C_t) = -\mathbb{E} \left[\dot{\ell}^2(\theta, Y_t | Y_{1:t-1}, C_t) \right]$.

■

Package R - Modèles de Cox

.2 Programmes principaux



.2.1 Simulation d'un échantillon suivant le modèle de Cox – Simulation.R

```

1 # simulation.R
3 # modele de Cox conditionnel
# lambda_X(t|z)=lambda_0(t)*Phi(z)
5 # lambda_Y(t|x,z)=lambda_0(t)*Psi(z)
7 # La simulation a t teste via cox.ph du package survival
# l'estimation des paramtres est correcte par cox.ph
9
# simulation des variables X,Y pour une sequence Z
11 rm(list=ls())
13 profiling=FALSE
15
16 if(profiling) { Rprof("simulation.out") }
17 if(!profiling) { print("Pas de profiling du programme.") }
19 #####
# valeur des parametres
21 #####
23 source('parametres.R')
25 # loi de base, parametre, parametre de Phi, Psi, Copule et Derivee de C
print("Les marges sont fixes dans modele.R")
27 source('modele.R')
29 # Fonction Phi, Psi
print("Les fonctions Phi et Psi sont fixes dans fonction.R")
31 source('fonction.R')
33 #####
35 # Simulation des echantillons X,Y
37 #####
39 # generation des exogenes
nz<-4
41 Z<-seq(0,1,length=nz)
43 # La taille de sous-classes ne sont pas identiques
# pour une question de simplicite, on donne un vecteur de taille nz
45 # o la valeur dans la ime colonne donne l'index de d part et non pas la taille de la classe, que l'on obtient par dif.
# echantillon de la forme (X((ncl[l]+1):ncl[l+1]),Y((ncl[l]+1):ncl[l+1])) associ Z[l]
47 # avec le vecteur ncl [0 ...], il doit commencer par 0 et donc avoir nz+1 cases
# On note Ncltotal la taille totale (equivalent de n*nz dans le cas simple)
49 ncl<-rep(0,nz+1)
51
cas_simple<-1
53 if (cas_simple==1)
{
55     print("Taille identique dans chaque classe.")
ncl<-20
57     ncl<-seq(0,n*nz,by=n)
rm(n)
59
tcl<-rep(0,nz)
61     for(i in 1:length(tcl))
tcl[i]<-ncl[i+1]-ncl[i]
63 } else
{
65     # cas de repartition non identique
# a renseigner la main
67     # ON RENTRE l'EFFECTIF DE CHAQUE CLASSE
# LE NOMBRE DE CLASSE DOIT ETRE EN ACCORD AVEC LA TAILLE DE Z
69     print("Taille non identique dans chaque classe")
.)
71     tcl<-c(10,20,9)
if(length(tcl)!=nz) stop("Erreur de specification des effectifs de chaque classe. Incohérent avec le nombre de")
73
ncl<-rep(0,length(tcl)+1)
75     for(i in 1:length(tcl))
ncl[i+1]<-ncl[i]+tcl[i]
77 }
79 # dtermination du vecteur tcl qui donne la taille de chaque classe
# toutes les valeurs sont identiques dans le cas d'une rpartition identique
81
83 if(cas_simple==1) {
# cas classique o les classes sont de taille identique
85     Ncltotal<-100*nz
} else
87 {

```

```

89      # avec les notations, la taille totale est entr dans la derniere
      # colonne de ncl
      Ncltotal<-ncl[length(ncl)]
91 }
93 X<-rep(0,Ncltotal)
94 Y<-rep(0,Ncltotal)
95
96 source('calcul_simulation.R')
97
98 #####
99 # RESULTATS
100 #####
101
102 if (profiling) {
103     Rprof(NULL)
104     print("Statistique de occupation")
105     print(summaryRprof('simulation.out'))
106 }
107
108 print("Generation d'une sequence Z de taille")
109 print(nz)
110 print("Generation d'un echantillon (X,Y) de taille")
111 if (cas_simple==1)
112 {
113     # cas de repartition identique
114     # on prend par exemple le dernier
115     print(tcl[length(tcl)])
116 } else
117 {
118     # cas de repartition non identique
119     print("Nombre de classes"); print(nz)
120     print("Effectif de chaque classe"); print(tcl)
121 }
122
123 #print(" Verification:")
124 #print(" plage de X"); print(c(min(X),max(X)))
125 #print(" plage de Y"); print(c(min(Y),max(Y)))
126 save(X,Y,Z,ncl, file="simulation.dat")
127 print("Donnees sauvegardees dans simulation.dat")
128 print("nom de variables: X, Y et Z")
129 print(" Param tres")
130 print(" alpha0"); print(alpha0);
131 print(" alpha1"); print(alpha1);
132 print(" beta0"); print(beta0);
133 print(" beta1"); print(beta1);
134
135 #####
136 # GRAPHIQUE
137 #####
138
139 # Trac pour un Z fix
140
141 # Choix du Z
142 l<-ceiling(runif(1,min=1,max=length(Z)))
143 print("indice de Z (de l' echantillon) et valeur de Z"); print(c(l,Z[l]))
144
145 z0<-Z[l]
146 XX<-X[(ncl[l]+1):(ncl[l+1])]
147 YY<-Y[(ncl[l]+1):(ncl[l+1])]
148
149 x11()
150 print("Resultats graphiques")
151 par(mfrow=c(2,2))
152 hist(XX)
153 box(which="figure")
154 hist(YY)
155 box(which="figure")
156 plot(XX,YY)
157 title("Nuage de points (X,Y) d'un Z")
158 box(which="figure")
159 zz<-seq(min(Z),max(Z),by=0.01)
160 trace<-Psi(zz,beta0,beta1)/Phi(zz,alpha0,alpha1)
161 plot(zz,trace,type='l',xlab="z",ylab="Psi/Phi")
162 title('Psi/Phi - ratio inferieur u1')
163 box(which="figure")
164
165 # liste_ev contient les etiquettes des copules de v.e
166 # cette liste est renseigne dans Catalogue.R
167 if (sum(type==liste_ev))
168 {
169     print("Trac de la fonction de dependance theorique")
170     x11()
171     t<-seq(0,1,length=100)
172     Atrace<-rep(0,length(t))
173     for (i in 1:length(t)) Atrace[i]<-ACatalogue(t[i],theta,type)
174     plot(Atrace,type='l')
175     title('Fonction de dependance de base Copule v.e.')
176 }
177
178 # x11()
179 # par(mfcol=c(2,1))

```

```

# Phi<-Phi(zz, alpha0, alpha1)
181 # plot(zz, Phi, type='l', xlab="z", ylab="Phi")
# title('Fonction Phi')
183 # box(which="figure")
# Psi<-Psi(zz, beta0, beta1)
185 # plot(zz, Psi, type='l', xlab="z", ylab="Psi")
# title('Fonction Psi')
187 # box(which="figure")

189 # COURBE DE SURVIE THEORIQUE ET EMPIRIQUE
# ABSENCE DE CENSURE
191
193 x11()
par(mfcol=c(2,1))
plot(HXbar)
195 xx<-seq(0,max(XX)+1,length=100)
hatS<-function(XX,xx) sum(XX>xx)/length(XX)
197 hatSXv<-rep(0,length(xx))
for(i in 1:length(xx))
199   hatSXv[i]<-hatS(XX,xx[i])
points(xx,hatSXv,type='l',lty=3,pch=-1)
201 title("Courbe de survie de X - Loi HX")
box(which="figure")
203 plot(HYbar)
yy<-seq(0,max(YY)+1,length=100)
205 hatSYv<-rep(0,length(yy))
for(i in 1:length(yy))
207   hatSYv[i]<-hatS(YY,yy[i])
points(yy,hatSYv,type='l',lty=3,pch=-1)
209 title("Courbe de survie de Y - Loi HY")
box(which="figure")

```

Fichier de paramètres – parametres.R

```

#####
2 # valeur des paramètres
#####
4
# loi de base
6 lambda<-1.5
eta<-1.5
8
# parametre pour les fonctions Phi et Psi
10
# voir la definition de Psi en cas de changement de alpha
12 # evidemment il faut des COEFFICIENTS DISTINCTS ENTRE ALPHA ET BETA
# SINON AUCUNE MODIFICATION
14 # dans le cas de fonctions identiques (voir fonction.R), il FAUT que
# alpha_1 > beta_1
16 alpha0<-1
alpha1<-3
18 beta0<-1
beta1<-2
20
# selon la famille la plage de theta varie
22 theta<-1.5
type<-"GH"
24 print("Type de copule")
print(type)

```

Fichier de spécification du modèle – modele.R

```

1 # modele.R
3 # Pour les definitions des copules
source('catalogue.R')
5
#####
7 #####
# PARTIE MODIFIABLE
#####
9 #####
11 #####
13 # Definition des fonctions pour X
#####
15 # Definition des fonctions pour X - LOI DE BASE
17
FXbar<-function(x)
19 {
21   # utile uniquement a titre illustratif dans simulation, courbe de survie
FXbar<-exp(-(x/lambda)^eta)
23 }
FXbarInv<-function(u)
25 {
27   # Forme de la premiere marge de la loi de base
FXbarInv<-lambda*(-log(u))^(1/eta)

```

```

}
29 #####
31 # Definition des fonctions pour Y
32 #####
33 # Definition des fonctions pour Y - LOI DE BASE
35 FYbar<-function(y)
37 {
39     # utile uniquement a titre illustratif dans simulation, courbe de survie
40     FYbar<-exp(-(y/lambda)^eta)
41 }
42 FYbarInv<-function(v)
43 {
44     # Forme de la seconde marge de la loi de base
45     FYbarInv<-lambda*(-log(v))^(1/eta)
46 }
47 #####
49 # COPULE DE SURVIE
50 #####
51 SCopule<-function(u,v)
52 {
53     # Copule de survie
54     # ne pas etre definie par u+v-1+C(1-u,1-v) car il faut que SCopuel soit TP2 ou ve
55     # et il n'y a pas de resultats du genre, si Copule est TP2 alors SCopule est TP2.
56     # Il s'agit de deux concepts distincts
57     SCopule<-CopuleCatalogue(u,v,theta,type)
58 }
59
60 DSCopule<-function(u,v)
61 {
62     # derive par rapport u de la fonction SCopule
63     DSCopule<-DCopuleCatalogue(u,v,theta,type)
64 }
65
66 #####
67 # PARTIE NON MODIFIABLE
68 #####
69
70 # Definition des fonctions pour X - LOI DU MODELE
71
72 HXbar<-function(x)
73 {
74     # courbe de survie pour le Z choisi au-dessus
75     # NON MODIFIABLE
76     # intervient uniquement graphiquement dans simulation.R
77     HXbar<-FXbar(x)^Phi(z0,alpha0,alpha1)
78 }
79
80 HXbarInv<-function(u,z)
81 {
82     # NON MODIFIABLE
83     HXbarInv<-FXbarInv(u^(1/Phi(z,alpha0,alpha1)))
84 }
85
86 HXInv<-function(u,z)
87 {
88     # NON MODIFIABLE
89     HXInv<-HXbarInv(1-u,z)
90 }
91
92 # Definition des fonctions pour Y - LOI DU MODELE
93
94 HYbar<-function(y)
95 {
96     # courbe de survie pour le Z choisi au-dessus
97     # NON MODIFIABLE
98     # intervient uniquement graphiquement dans simulation.R
99     HYbar<-FYbar(y)^Psi(z0,beta0,beta1)
100 }
101
102 HYbarInv<-function(v,z)
103 {
104     # NON MODIFIABLE
105     HYbarInv<-FYbarInv(v^(1/Psi(z,beta0,beta1)))
106 }
107
108 HYInv<-function(v,z)
109 {
110     # NON MODIFIABLE
111     HYInv<-HYbarInv(1-v,z)
112 }
113

```

Fichier des modèles disponibles – catalogue.R

```
1 # Catalogue de copules
```

```

# avec eventuellement leur fonction de dependance associe
3 # et leur derive pour la simulation
# Liste de copules de base
5
CopuleCatalogue<-function(u,v,theta,type)
7 {
9     # Famille de Gumbel-Hougaard pour Nelsen
10    # Famille de Gumbel (B6) pour Joe
11    # Valeurs-extremes
12    # Theta >= 1
13    if(type == "GH")
14    {
15        if (theta < 1) stop("Erreur de valeur de theta pour GH.")
16        CopuleCatalogue<-exp(-((-log(u))^theta+(-log(v))^theta)^(1/theta))
17    }
18    # Famille de Gumbel ou Logistique
19    # VALEURS EXTREMES
20    # IDENTIQUE A LA FAMILLE precedente, hormis le changement de theta en 1/theta
21    # 0<=alpha<=1
22    # A(s)=(s^(1/alpha)+(1-s)^(1/alpha))^alpha
23    else
24    if(type == "GL0")
25    {
26        if((theta < 0)|| (theta > 1)) stop("Erreur de valeur pour la copule de Gumbel Logistique.")
27        CopuleCatalogue<-exp( ( (log(u))^(1/theta)+(log(v))^(1/theta) )^theta )
28    }
29
30    # Copule de Galambos
31    # copule B7 (Joe)
32    # VALEURS EXTREMES
33    # Theta > 0
34    else
35    if(type == "GL")
36    {
37        if ((theta < 0)) stop("Erreur de valeur de theta pour GL, theta > 0.")
38        CopuleCatalogue<-(u*v)*exp((-log(u))^(-theta)+(-log(v))^(-theta))^(1/theta)
39    }
40
41    # Copule de Ali-Mikhail-Haq
42    # Theta dans [-1;1]
43    # PQD si theta > 0 donc dans le modele de Cox CN theta > 0 (car TP2 => PQD)
44    # par le programme TP2.R, cette copule est TP2 si theta > 0
45    # voir Nelsen p. 81
46    else
47    if(type == "AMH")
48    {
49        if((theta >= 0)&&(theta <= 1)) print("Famille TP2.")
50        if ((theta < 0)|| (theta > 1)) stop("Erreur de valeur de theta pour AMH.")
51        CopuleCatalogue<-(u*v)/(1-theta*(1-u)*(1-v))
52    }
53
54    # Famille de Gumbel
55    # Theta dans [0;1]
56    # voir Nelsen p.29
57    # ATTENTION FAMILLE NON PQD, donc NON TP2, sauf si theta < 0
58    #else
59    if(type == "G")
60    {
61        if ((theta < 0)|| (theta > 1)) stop("Erreur de valeur de theta pour G.")
62        CopuleCatalogue<-u*v-1+(1-u)*(1-v)*exp(-theta*log(1-u)*log(1-v))
63    }
64
65    # Famille de Gumbel-Barnett (survie de Gumble)
66    # Theta dans [0;1]
67    # voir Nelsen p.94
68    # ATTENTION FAMILLE NON PQD, donc NON TP2, sauf si theta < 0
69    #else
70    if(type == "GB")
71    {
72        if ((theta < 0)|| (theta > 1)) stop("Erreur de valeur de theta pour GB.")
73        CopuleCatalogue<-u*v*exp(-theta*log(u)*log(v))
74    }
75
76    else
77    if(type == "C13")
78    {
79        # Nelsen p.94
80        # 4.2.13
81        # theta > 0
82        # PQD et TP2 si theta >= 1
83        if ((theta < 0)) stop("Erreur de valeur de theta pour C13, theta > 0.")
84        CopuleCatalogue<-exp(1-((1-log(u))^theta+(1-log(v))^theta-1)^(1/theta))
85    }
86
87    else
88    if(type == "C6")
89    {
90        # Nelsen p.94
91        # 4.2.6
92        # theta > 1
93        if ((theta < 1)) stop("Erreur de valeur de theta pour C6, theta > 0.")
94        CopuleCatalogue<-1-((1-u)^theta+(1-v)^theta-(1-u)^theta*(1-v)^theta)^(1/theta)
95    }

```

```

95     else
96         if(type == "C12")
97             {
98                 # Nelsen p.94
99                 # 4.2.12
100                # theta > 1
101                if ((theta < 1)) stop("Erreur de valeur de theta pour C13, theta > 0.")
102                CopuleCatalogue <- 1 / (1 + ((1/u - 1)^theta + (1/v - 1)^theta)^(1/theta))
103            }
104     else
105         if(type == "Clayton")
106             {
107                 # Famille de Clayton
108                 # Nelsen p.94
109                 # 4.2.1
110                 # theta >= -1 et theta != 0
111                 # PQD et TP2 si theta > 0
112                 if (theta < -1) stop("Erreur de valeur de theta pour C1, theta > -1.")
113                 CopuleCatalogue <- max(u^(-theta) + v^(-theta) - 1, 0)^(1/theta)
114             }
115     else
116         {
117             print("Type de copule non defini, choix:")
118             print("GH: Gumbel-Hougaard (Nelsen) - Gumbel (B6) (Joe)")
119             print("GL0: Gumbel Logistique")
120             print("GL: Galambos (B7)")
121             print("AMH: Ali-Mikhail-Haq")
122             print("C16, C12, C13")
123             print("G: Gumbel - ni TP2 ni PQD")
124             stop("Refaire votre choix")
125         }
126
127 DCopuleCatalogue <- function(u, v, theta, type)
128 {
129     # Derivee par rapport u de la copule
130     # Utile pour la simulation dans le cas non v.e.
131     # sinon on utilise la fonction A et Aprime
132
133     # Famille de Gumbel-Hougaard
134     # on donne l'expression ici si on utilise cette famille pour le modele de Cox
135     # Theta >= 1
136     if (type == "GH") DCopuleCatalogue <- CopuleCatalogue(u, v, theta, type = "GH") * ((-log(u))^(theta - 1) / u * (-log(u))^(theta - 1) / v)
137     else
138         if (type == "GL") DCopuleCatalogue <- 1 / u * CopuleCatalogue(u, v, theta, type = "GL") * (1 - ((-log(u))^(theta - 1) + (-log(v))^(theta - 1)))
139     else
140         if (type == "GL0") DCopuleCatalogue <- CopuleCatalogue(u, v, theta, type = "GL0") * theta * ((log(u))^(1/theta) + (log(v))^(1/theta))
141     else
142         if (type == "AMH") DCopuleCatalogue <- (v - theta * v + theta * v^2) / (1 - theta * (1 - u) * (1 - v))^2
143     else
144         if (type == "C13") DCopuleCatalogue <- CopuleCatalogue(u, v, theta, type = "C13") * (-1/theta * ((1 - log(u))^(theta - 1) + (1 - log(v))^(theta - 1)))
145
146     # Famille de Gumbel
147     # Theta dans [0; 1]
148     # else
149     #     if (type == "G") DCopuleCatalogue <- 1 - (1 - v) * exp(-theta * log(1 - u) * log(1 - v)) * (1 - theta * log(1 - v))
150
151     else
152         {
153             print("Type de copule non defini, choix:")
154             print("GH: Gumbel-Hougaard (Nelsen) - Gumbel (B6) (Joe)")
155             print("GL0: Gumbel Logistique")
156             print("GL: Galambos (B7)")
157             print("AMH: Ali-Mikhail-Haq")
158             print("G: Gumbel - ni TP2 ni PQD")
159             stop("Refaire votre choix")
160         }
161 }
162
163 liste_ev <- c("GH", "GL0", "GL")
164
165 ACatalogue <- function(t, theta, type)
166 {
167     # Definition des fonctions de dependance pour les copules de v.e.
168
169     # Famille de Gumbel-Hougaard
170     # Theta >= 1
171     if (type == "GH") ACatalogue <- t * ((1/t - 1)^theta + 1)^(1/theta)
172     else
173         # Famille de Gumbel-Logistique
174         # 0 <= Theta <= 1
175         if (type == "GL0") ACatalogue <- (t^(1/theta) + (1 - t)^(1/theta))^theta
176     else
177         # Famille de Galambos
178         # Theta > 0
179         # ancienne valeur ERRONEE if (type == "GL") ACatalogue <- ((-t)^(-theta) + (-1 + t)^(-theta))^(1/theta)
180         if (type == "GL") ACatalogue <- 1 - (t^(-theta) + (1 - t)^(-theta))^(1/theta)
181     else
182         {
183             print("Type de copule non defini, choix:")
184             print("GH: Gumbel-Hougaard (Nelsen) - Gumbel (B6) (Joe)")
185         }
186 }

```

```

187     print("GL0: Gumbel Logistique")
188     print("GL: Galambos (B7)")
189     stop("Refaire votre choix")
190   }
191 }
AprimeCatalogue<-function(t, theta, type)
193 {
194   # Derivee par rapport t de la fonction A
195
196   # Famille de Gumbel-Hougaard
197   # Theta >= 1
198   if (type=="GH") AprimeCatalogue<-((1/t)^(theta+1))^(1/theta-1)*((1/t)^(theta+1)-1/t*(1/t)^(theta-1))
199   else
200     # Famille de Gumbel-Logistique
201     # 0<=Theta<= 1
202     if (type=="GL0") AprimeCatalogue<-(t^(1/theta)+(1-t)^(1/theta))^(theta-1)*(t^(1/theta)-1)^(1/theta-1)
203   else
204     # Famille de Galambos
205     # Theta<= 1
206     #ancienne valeur ERRONEE if (type=="GL") AprimeCatalogue<-((-t)^(-theta)+(-1+t)^(-theta))^(1/theta-1)*(-t)^(-theta-1)
207     if (type=="GL") AprimeCatalogue<-(t^(-theta)+(1-t)^(-theta))^(1/theta-1)*((1-t)^(-theta)-1)^(1/theta-1)
208   else
209     {
210       print("Type de copule non defini, choix:")
211       print("GH: Gumbel-Hougaard (Nelsen) - Gumbel (B6) (Joe)")
212       print("GL0: Gumbel Logistique")
213       print("GL: Galambos (B7)")
214       stop("Refaire votre choix")
215     }
216 }

```

Fichier de spécification des marges – fonction.R

```

# fonction.R
2
#####
4 #####
# Definition des fonctions
6 #####
8 #####
# Phi>Psi
10
Phi<-function(z, a0, a1)
12 {
13   Phi<-exp(a1*z+a0)
14   #Phi<-a1*z+a0
15 }
Psi<-function(z, b0, b1)
18 {
19   Psi<-exp(b1*z+b0)
20   #Psi<-b1*z+b0
21 }
K<-function(Phi, Psi, z, z0, a0, a1, b0, b1)
24 {
25   # valeur de K de z0 au point z
26   K<-(Psi(z, b0, b1)/Phi(z, a0, a1))*(Phi(z0, a0, a1)/Psi(z0, b0, b1))
27 }

```

Routine de calcul – calculsimulation.R

```

# calcul_simulation.R
2
source('outils.R')
4
#####
6 #####
# Simulation des echantillons X,Y
8 #####
10 #####
# on genere un couple (U,V) dont la copule de survie est donnee
12
for(l in 1:nz)
14 {
15   #####
16   # Generation des variables X
17   #####
18
19   # DEFINITION de n dans le cas de rpartition non identique, afin de rutiliser
20   # les formules issues du cas de rpartition identique
21   n<-tcl[1]
22
23   U<-runif(n)
24   XX<-rep(0, n)

```



```

26   for(i in 1:n)
27     {
28       XX[i]<-HXInv(1-U[i],Z[1])
29     }
30   X[(ncl[1]+1):(ncl[1+1])<-XX
31
32   #####
33   # Generation des variables Y
34   #####
35
36   T<-runif(n)
37   V<-rep(0,n)
38
39   # Constitution de l'echantillon UV, de copule C-H(z)
40
41   # on veut obtenir (U,V) de copule C
42   # pour cela, on va obtenir (1-U,1-V) de copule \hat{C}
43   # donc on genere (U,T) uniforme
44   # X est la transformation inverse de U
45   # on obtient (1-U,1-V) de copule \hat{C} par l'algorithme suivant
46
47   for(i in 1:n)
48     {
49       V[i]<-InversionZ(U[i],Z[1],T[i])
50     }
51
52   YY<-rep(0,n)
53   for(i in 1:n)
54     {
55       YY[i]<-HYInv(1-V[i],Z[1])
56     }
57   Y[(ncl[1]+1):(ncl[1+1])<-YY
58 }

```

Routines de calcul – outils.R

```

1 dyn.load("../SRC/copuleempirique.so")
2 dyn.load("../SRC/scopuleempirique.so")
3 dyn.load("../SRC/inv.so")
4
5 # Copule empirique
6
7 Copuleempirique<-function(X,Y,u,v)
8 {
9   .C("copuleempirique",
10     as.double(X),
11     as.double(Y),
12     as.integer(length(X)),
13     as.double(u),
14     as.double(v),
15     as.double(sort(X)),
16     as.double(sort(Y)),
17     A=as.double(1)
18   )$A
19 }
20
21 Copuleempiriqueetoile<-function(X,Y,u,v)
22 {
23   .C("copuleempiriqueetoile",
24     as.double(rank(X)/length(X)),
25     as.double(rank(Y)/length(Y)),
26     as.integer(length(X)),
27     as.double(u),
28     as.double(v),
29     A=as.double(1)
30   )$A
31 }
32
33 SCopuleempirique<-function(X,Y,u,v)
34 {
35   .C("scopuleempirique",
36     as.double(rank(X)/length(X)),
37     as.double(rank(Y)/length(Y)),
38     as.integer(length(X)),
39     as.double(u),
40     as.double(v),
41     A=as.double(1)
42   )$A
43 }
44
45 #####
46 # Definition des fonctions utiles
47 #####
48
49 cuv<-function(u,v)
50 {
51   facteur<-A(log(v)/log(u*v))-log(v)/log(u*v)*Aprime(log(v)/log(u*v))
52   resultat<-Copule(u,v)*1/u*facteur
53   # car la derivee est bornee par 1
54   if(resultat=="NaN") cuv<-1
55 }

```

```

55     else cuv<-resultat
56   }
57   cuvz<-function(u,v,z)
58 {
59     # formule obtenue par derivation de la relation entre  $\hat{C}_H^z$  et  $\hat{C}_F$ 
60     coef<-(Phi(z,alpha0,alpha1)-Psi(z,beta0,beta1))/Phi(z,alpha0,alpha1)
61     termel<-coef*u^(coef-1)*SCopule(u^(1/Phi(z,alpha0,alpha1)),v^(1/Psi(z,beta0,beta1)))^Psi(z,beta0,beta1)
62     facteur<-DSCopule(u^(1/Phi(z,alpha0,alpha1)),v^(1/Psi(z,beta0,beta1)))*(1/Phi(z,alpha0,alpha1))*u^(1/Phi(z,alpha0,alpha1))
63     terme2<-u^coef*Psi(z,beta0,beta1)*SCopule(u^(1/Phi(z,alpha0,alpha1)),v^(1/Psi(z,beta0,beta1)))^(Psi(z,beta0,beta1))
64     cuv<-termel+terme2
65   }
66 }
67
68 #####
69 # Inversion
70 #####
71
72 Inversion1<-function(u,t)
73 {
74     InvC<-function(f,guess=c(1e-6,1-1e-6),tol=1e-5)
75     {
76         f.check<-function(x) {
77             x<-f(x)
78         }
79         .Call("Finv",body(f.check),as.double(guess),as.double(tol),new.env())
80     }
81 }
82 # guess c(0,1) - fonction non decroissante
83 G<-function(v) cuv(u,v)-t
84
85 Inversion1<-InvC(G)
86 }
87
88 Inversion2<-function(u,t)
89 {
90     InvC<-function(f,guess=c(1-1e-6,1e-6),tol=1e-5)
91     {
92         f.check<-function(x) {
93             x<-f(x)
94         }
95         .Call("Finv",body(f.check),as.double(guess),as.double(tol),new.env())
96     }
97 }
98
99 # modification de guess c(1,0) - car la fonction est decroissante (je crois)
100 G<-function(v) Copule(u,v)/u-t
101
102 Inversion2<-InvC(G)
103 }
104
105 # Phase de test
106 # InversionR<-function(u,t)
107 # {
108 #     pas<-0.001
109 #     v<-1-pas
110 #     a<-cuv(u,v)
111 #     if(a=="NaN") stop("Erreur")
112 #     while((cuv(u,v)>t)&&(v>0))
113 #     {
114 #         v<-v-pas
115 #     }
116 #     InversionR<-v
117 # }
118
119 InversionZ<-function(u,z,t)
120 {
121     InvC<-function(f,guess=c(1e-6,1-1e-6),tol=1e-5)
122     {
123         f.check<-function(x) {
124             x<-f(x)
125         }
126         .Call("Finv",body(f.check),as.double(guess),as.double(tol),new.env())
127     }
128 }
129
130 G<-function(v) cuvz(u,v,z)-t
131
132 InversionZ<-InvC(G)
133 }
134
135 #####
136 # algorithme Efron
137 #####
138
139 efron<-function(X,Z,ncl)
140 {
141     # Format de X: vecteur de taille n*nz
142     # Format de Z: vecteur de taille nz
143
144     # Test pour eviter les redondances
145     DV<-rep(0,length(X))
146     dd<-rep(0,length(X))

```

```

147 XTT<-rep(0,length(X))
    ZTT<-rep(0,length(X))
149 # gonflement de Z pour avoir les dimensions de X
    # par recopie
151 nz<-length(Z)

153 # creation d'un vecteur ZZ qui est une recopie de Z au format de X
    for(l in 1:nz) ZTT[(ncl[l]+1):(ncl[l+1])]<-Z[l]
155
156 ZZ<-matrix(0,length(X),length(X))
157 k<-0

159 for(i in 1:length(X))
    {
161     emplacement<-rep(X[i],length(X))==X
        dd[i]<-sum(emplacement)
163     if(dd[i]==1)
        {
165         k<-k+1
            XTT[k]=X[i]
167         ZZ[k,1]=ZTT[i]
            DV[i]=1
169         }
        else
171         {
            if(DV[i]==0)
173             {
                # Motif premiere fois rencontr
                # mais avec egalites
175                 k<-k+1
                    XTT[k]=X[i]
177                 DV<-DV+emplacement
                    # repere les indices d'egalite
179                 j=1
                    trouve<-0
                    while(trouve<dd[i])
183                     {
                        if(emplacement[j]==TRUE)
185                         {
                            trouve<-trouve+1
                            ZZ[k,trouve]=ZTT[j]
187                         }
                            j<-j+1
189                     }
                    }
191         }
    }
193 }

195 XType<-rep(0,k)
    for(i in 1:k)
197     XType[i]<-XTT[i]

199 dX<-rep(0,k)
    for(i in 1:k)
201     dX[i]<-sum(ZZ[i,]!=0)

203 maxNbZ<-max(dX)
    ZType<-matrix(0,k,maxNbZ)
205 for(i in 1:k)
        for(j in 1:maxNbZ)
207             ZType[i,1:maxNbZ]<-ZZ[i,1:maxNbZ]

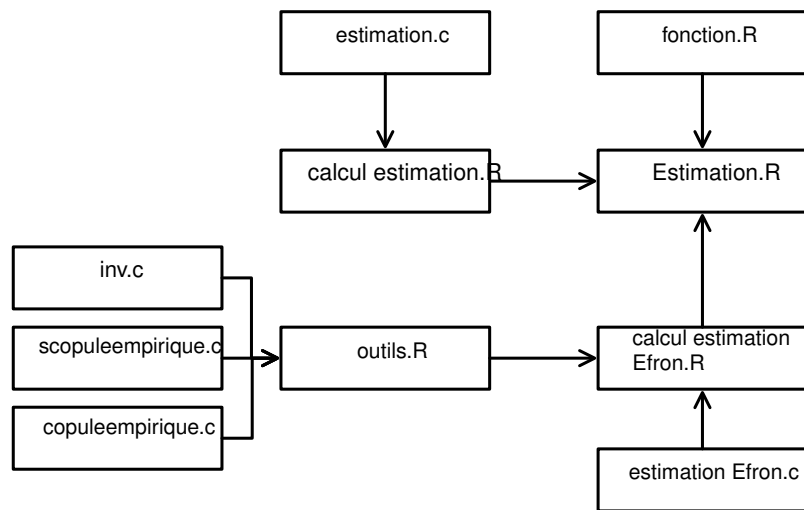
209 # reorganisation sous forme de vecteur de Z
    # a cause de l'impossibilite de passer une matrice la fonction C
211 # d'ou reorganisation sous la forme d'un vecteur

213 Zbis<-rep(0,length(X))
    pos<-1
215 for(i in 1:k)
        for(j in 1:dX[i])
217             {
                Zbis[pos]<-ZType[i,j]
219                 pos<-pos+1
            }

221 rm(ZZ,dd,trouve,emplacement,DV,XTT,pos,Zbis,maxNbZ,ZTT)
223 # reste XType ZType dX k

225 sortie<-list()
    sortie[[1]]<-XType
227 sortie[[2]]<-ZType
    sortie[[3]]<-dX
229 sortie[[4]]<-k
    sortie
231 }

```



.2.2 Estimation des coefficients marginaux du modèle de Cox – Estimation.R

```

1 # Estimation.R
3 # modele de Cox conditionnel
# lambda_X(t|z)=lambda_0(t)*Phi(z)
5 # lambda_Y(t|x,z)=lambda_0(t)*Psi(z)
7 # L'estimation a t valid par l'emploi de cox.ph du package Survival
# les estimations coincident SAUF EN CAS D'EGALITE DES X (ou Y)
9
# estimation en s'appuyant sur l'echantillon X,Y pour une sequence Z
11 # des quantites alpha1, alpha0, beta1, beta0
13 rm(list=ls())
15 load("simulation.dat")
17 # dans le cas rel, il faut avoir dj dtermin ncl
# le format est X, Y, Z avec la donne ncl
19
if(!(exists("ncl")))
21 {
# ncl n'existe pas
23 stop("NCL n'existe pas.")
# creation du ncl
25 # a priori ici on est dans le cas de donnees reelles
# il faut que le format des donnees soit: X,Y,Z de taille identique
27 # et non pas X,Y de taille N1 et Z de taille N2 plus petit comme dans
# le cas d'une simulation
29 }
31 source('fonction.R')
33 profiling=FALSE
if (profiling) { Rprof("estimation.out") }
35 if (!(profiling)) { print("Pas de profiling du programme.") }
37 # valeurs initiales pour calculer la vraisemblance partielle
39 # REMARQUE DANS LE CAS OU Phi et Psi sont des exponentielles de la forme
# Phi(z)=exp(alpha0+z*alpha1), alpha0 n'est pas estimable par la vraisemblance partielle
41 # un calcul montre que le terme exp(alpha0) se met en facteur en haut et en bas et donc
# se simplifie, rendant impossible la factorisation
43 # d'ou la possibilite de les mett 0
alpha0initiale<-1
45 alpha1initiale<-3
beta0initiale<-1
47 beta1initiale<-2
49 # bornes
minalpha0<-alpha0initiale-1
51 maxalpha0<-alpha0initiale+1
minalpha1<-alpha1initiale-1
53 maxalpha1<-alpha1initiale+1
55 minbeta0<-beta0initiale-1
maxbeta0<-beta0initiale+1
57 minbeta1<-beta1initiale-1
maxbeta1<-beta1initiale+1
59
# precision
61 nba0<-50
nba1<-50
63
nbb0<-50
65 nbb1<-50
67 print("Precision:")
print("alpha0:")
69 print((maxalpha0-minalpha0)/nba0)
print("alpha1:")
71 print((maxalpha1-minalpha1)/nba1)
print("beta0:")
73 print((maxbeta0-minbeta0)/nbb0)
print("beta1:")
75 print((maxbeta1-minbeta1)/nbb1)
77 # Affichage
print("Plage de calcul")
79 print("Plage pour alpha0")
print(c(minalpha0, maxalpha0))
81 print("Plage pour alpha1")
print(c(minalpha1, maxalpha1))
83 print("Plage pour beta0")
print(c(minbeta0, maxbeta0))
85 print("Plage pour beta1")
print(c(minbeta1, maxbeta1))
87

```

```

# Determination du programme appeler en fonction
89 # du cas d'egalite ou pas

91 rangX<-rank(X)
# on utilise la moyenne car il faut tenir compte de l'ensemble du
93 # tableau rang
if(mean(rangX==floor(rangX))==1)
95 {
# Pas d'egalite donc methode classique Cox
97 print("Methode_Cox.")
source('calcul_estimation.R')
99 } else
{
# Egalite donc Efron
101 print("Methode_Efron.")
103 source('calcul_estimation_Efron.R')
}

105 if (profiling) {
107 Rprof(NULL)
print("Statistique_d'occupation")
109 print(summaryRprof('estimation.out'))
}

111 #####
113 print("Estimations_obtenues")
print("alpha0")
115 print(hatalpha0)
print("alpha1")
117 print(hatalpha1)
print("beta0")
119 print(hatbeta0)
print("beta1")
121 print(hatbeta1)

123 graphique<-FALSE
if(graphique)
125 {
# TRACE
127 x11()
par(mfrow=c(1,2))
129 persp(PLX,theta=45,nticks=5,zlim=range(PLX,finite=TRUE),ticktype="detailed",col="lightblue",xlab="alpha_0",ylab="alpha_1",
title('Surface_pour_alpha0,alpha1'))
131 box(which="figure")
contour(PLX)
133 box(which="figure")

135 #x11()
par(mfrow=c(1,2))
137 persp(PLY,theta=45,nticks=5,zlim=range(PLY,finite=TRUE),ticktype="detailed",col="lightblue",xlab="beta_0",ylab="beta_1",
title('Surface_pour_beta0,beta1'))
139 box(which="figure")
contour(PLY)
141 box(which="figure")

143 x11()
par(mfrow=c(1,2))
145 image(seq(minalpha0,maxalpha0,by=pasa0),seq(minalpha1,maxalpha1,by=pasa1),PLX,col=terrain.colors(50),xlab="alpha_0",ylab="alpha_1",
title('Surface_pour_alpha0,alpha1'))
147 box(which="figure")
image(seq(minbeta0,maxbeta0,by=pasb0),seq(minbeta1,maxbeta1,by=pasb1),PLY,col=terrain.colors(50),xlab="beta_0",ylab="beta_1",
149 title('Surface_pour_beta0,beta1'))
box(which="figure")
151 }

153 # SAUVEGARDE
save(hatalpha0,hatalpha1,hatbeta0,hatbeta1,file="estimation.dat")
155 print("Donnees_sauvegardees_dans_estimation.dat")
print("nom_de_variables:_hatalpha0,_hatalpha1,_hatbeta0,_hatbeta1")

```

Routine de calcul – calcul estimation.R

```

# calcul_estimation-C.R
2
# modele de Cox conditionnel
4 # lambda_X(t|z)=lambda_0(t)*Phi(z)
# lambda_Y(t|x,z)=lambda_0(t)*Psi(z)
6 #
# estimation en s'appuyant sur l'echantillon X,Y pour une sequence Z
8 # des quantites alpha1, alpha0, beta1, beta0

10 dyn.load("../SRC/estimation.so")

12 # A OPTIMISER DANS LE PRINCIPE ORDONNER
Vraisemblancepartielle<-function(X,Z,ncl,G,c0,c1)
14 {
nz<-length(Z)
16
ZS<-array(0,dim=c(length(X),1))
18 ZZ<-array(0,dim=c(length(X),1))
rang<-sort.list(X)

```

```

20
21     # creation d'un vecteur ZZ qui est une recopie de Z au format de X
22     for(1 in 1:nz) ZZ[(ncl[1]+1):(ncl[1+1])]<-Z[1]
23     # on ordonne les Z avec le classement des X
24     ZS<-ZZ[rang]
25
26     # Appel de la routine en C
27     VPC<-function(f,ZSZ,tailleZS=length(ZS))
28     {
29         f.check<-function(x) {
30             x<-f(x)
31         }
32         .Call("VPC",body(f.check),as.double(ZSZ),as.integer(tailleZS),new.env())
33     }
34 GG<-function(z) G(z,c0,c1)
35
36 Vraisemblancepartielle<-VPC(GG,ZS)
37 }
38
39 # pas
40 pasa0<-(maxalpha0-minalpha0)/nba0
41 pasa1<-(maxalpha1-minalpha1)/nba1
42
43 pasb0<-(maxbeta0-minbeta0)/nbb0
44 pasb1<-(maxbeta1-minbeta1)/nbb1
45
46 # Tableau possedant 2 indices
47 PLX<-array(0,dim=c((nba0+1),(nba1+1)))
48 PLY<-array(0,dim=c((nbb0+1),(nbb1+1)))
49 for(ia0 in 1:(nba0+1))
50     for(ia1 in 1:(nba1+1))
51     {
52         PLX[ia0,ia1]<-Vraisemblancepartielle(X,Z,ncl,Phi,minalpha0+(ia0-1)*pasa0,minalpha1+(ia1-1)*pasa1)
53     }
54 for(ib0 in 1:(nbb0+1))
55     for(ib1 in 1:(nbb1+1))
56     {
57         PLY[ib0,ib1]<-Vraisemblancepartielle(Y,Z,ncl,Psi,minbeta0+(ib0-1)*pasb0,minbeta1+(ib1-1)*pasb1)
58     }
59
60 #####
61
62 indicemax<-function(T)
63 {
64     # renvoie la ligne et la colonne
65     # donnant le maximum du tableau T
66     l1<-which.max(T)
67     if ((l1 %% dim(T)[1]) == 0)
68     {
69         l1<-dim(T)[1]
70     } else
71     {
72         l1<-(l1 %% (dim(T)[1]))
73     }
74
75     cc<-((which.max(T)-l1) %% dim(T)[1]) + 1)
76     coefficient<-c(l1,cc)
77 }
78
79
80 indiceX<-indicemax(PLX)
81
82 hatalpha0<-minalpha0+(indiceX[1]-1)*pasa0
83 hatalpha1<-minalpha1+(indiceX[2]-1)*pasa1
84
85 indiceY<-indicemax(PLY)
86
87 hatbeta0<-minbeta0+(indiceY[1]-1)*pasb0
88 hatbeta1<-minbeta1+(indiceY[2]-1)*pasb1
89
90 dyn.unload("../SRC/estimation.so")

```

Routine de calcul – calcul estimation Efron.R

```

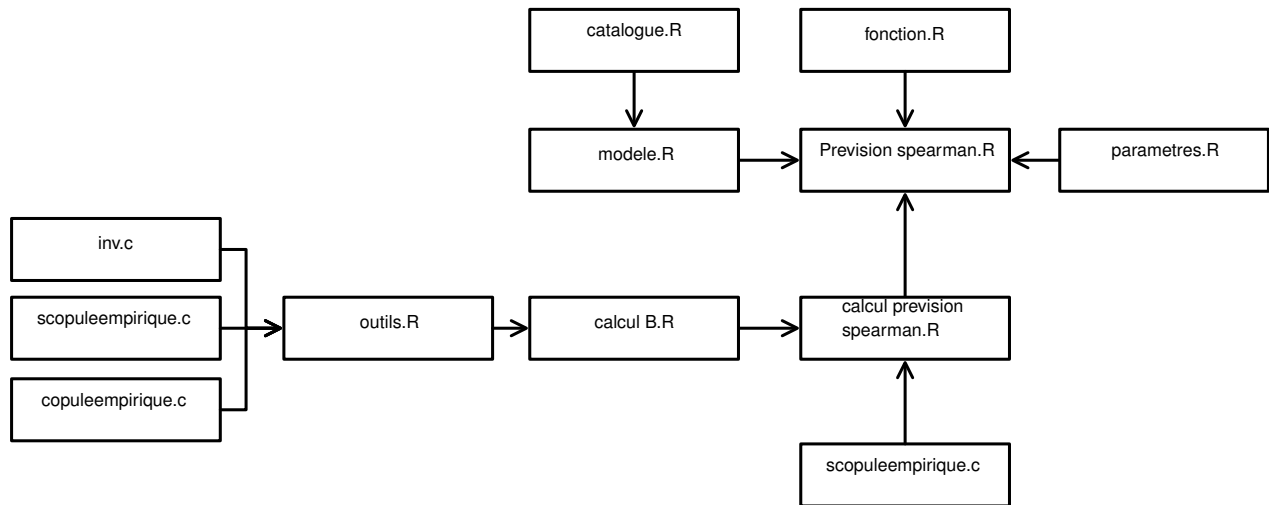
dyn.load("../SRC/estimation_Efron.so")
2
3 source('outils.R')
4
5 resultat<-efron(X,Z,ncl)
6 XType<-resultat[[1]]
7 ZType<-resultat[[2]]
8 dX<-resultat[[3]]
9 kx<-resultat[[4]]
10
11 resultat<-efron(Y,Z,ncl)
12 YType<-resultat[[1]]
13 ZYType<-resultat[[2]]
14 dY<-resultat[[3]]
15 ky<-resultat[[4]]
16
17 VEfron<-function(XType,ZType,dXType,k,G,c0,c1)

```

```

18 {
19     # XType est la version reduite de X initial
20     # reduite dans le sens, sans recopie
21     #XS<-sort(XType)
22     rang<-sort.list(XType)
23     dXS<-dXType[rang]
24     # on ordonne les Z avec le classement des X
25     # on ordonne par blocs
26     ZS<-matrix(0,k,max(dXType))
27     ZS<-ZType[rang,]
28
29     # S est utile uniquement si Phi et Psi sont exponentielles
30     S<-rep(0,length(XType))
31     for(i in 1:length(XType))
32     # S[i]<-sum(ZS[i,])
33
34     VPCEfron<-function(f,ZT,dXT,tailleS)
35     {
36         f.check<-function(x) {
37             x<-f(x)
38         }
39         .Call("VPCEfron",body(f.check),as.matrix(ZT),as.integer(dXT),as.integer(tailleS),new.env())
40     }
41
42     GK<-function(z) G(z,c0,c1)
43     VPEfron<-VPCEfron(GG,ZS,dXS,length(XType))
44 }
45
46 # pas
47 pasa0<-(maxalpha0-minalpha0)/nba0
48 pasa1<-(maxalpha1-minalpha1)/nba1
49
50 pasb0<-(maxbeta0-minbeta0)/nbb0
51 pasb1<-(maxbeta1-minbeta1)/nbb1
52
53 # Tableau possedant 2 indices
54 PLX<-array(0,dim=c((nba0+1),(nba1+1)))
55 PLY<-array(0,dim=c((nbb0+1),(nbb1+1)))
56 for(ia0 in 1:(nba0+1))
57     for(ia1 in 1:(nba1+1))
58     {
59         #PLX[ia0,ia1]<-Vraisemblancepartielle(X,Z,Phi,minalpha0+(ia0-1)*pasa0,minalpha1+(ia1-1)*pasa1)
60         PLX[ia0,ia1]<-VPEfron(XType,ZXType,dX,kx,Phi,minalpha0+(ia0-1)*pasa0,minalpha1+(ia1-1)*pasa1)
61     }
62
63 for(ib0 in 1:(nbb0+1))
64     for(ib1 in 1:(nbb1+1))
65     {
66         PLY[ib0,ib1]<-VPEfron(YType,ZYType,dY,ky,Psi,minbeta0+(ib0-1)*pasb0,minbeta1+(ib1-1)*pasb1)
67     }
68
69 #####
70 indicemax<-function(T)
71 {
72     # renvoie la ligne et la colonne
73     # donnant le maximum du tableau T
74     ll<-which.max(T)
75     if ((ll%%dim(T)[1])==0)
76     {
77         ll<-dim(T)[1]
78     } else
79     {
80         ll<-(ll%%(dim(T)[1]))
81     }
82
83     cc<-(((which.max(T)-ll)%/dim(T)[1])+1)
84     coefficient<-c(ll,cc)
85 }
86
87
88 indiceX<-indicemax(PLX)
89
90 hatalpha0<-minalpha0+(indiceX[1]-1)*pasa0
91 hatalpha1<-minalpha1+(indiceX[2]-1)*pasa1
92
93 indiceY<-indicemax(PLY)
94
95 hatbeta0<-minbeta0+(indiceY[1]-1)*pasb0
96 hatbeta1<-minbeta1+(indiceY[2]-1)*pasb1
97
98 dyn.unload("../SRC/estimation-Efron.so")

```

.2.3 Prevision du rho spearman – Prevision spearman.R

```

# modele de Cox conditionnel
2 # lambda_X(t|z)=lambda_0(t)*Phi(z)
# lambda_Y(t|x,z)=lambda_0(t)*Psi(z)
4 #
# estimation de B(z,t)
6 # en s'appuyant sur l'expression entre deux fonctions B induite
# par le modele
8 # On fait intervenir les ratios issu du modele sous la notation K

10 rm(list=ls())
  print("Prevision du Rho de Spearman")
12
  profiling=FALSE
14 if (profiling) { Rprof("prevision.out") }
  if (!(profiling)) { print("Pas de profiling du programme.") }
16
# les donnees contiennent:
18 # echantillon (X,Y) et Z dans simulation.R
# les estimations des coefficients issues de estimation.R
20
  load("simulation.dat")
22   load("estimation.dat")

24 cas_reel<-0
  if (cas_reel==0)
26 {
    source('parametres.R')
28   print("Cas de donnees simules, on considere")
    print("le cadre des valeurs extremes")
30   print("la famille")
    print(type)
32 }else
  print("Cas de donnees reelles.")
34
# CHARGEMENT DES MODULES
36 # le module modele est utile pour le trac de la courbe thorique
  source('modele.R')
38 source('fonction.R')

40 print("Estimations utilisees")
  print("alpha0 alpha1")
42 print(c(hatalpha0, hatalpha1))
  print("beta0 beta1")
44 print(c(hatbeta0, hatbeta1))

46 print("MODIFICATION DES VALEURS DES PARAMETRES ALPHA et BETA - PHASE DE TEST")
  hatalpha0<-alpha0
48 hatalpha1<-alpha1
  hatbeta0<-beta0
50 hatbeta1<-beta1
  print("alpha0 alpha1")
52 print(c(hatalpha0, hatalpha1))
  print("beta0 beta1")
54 print(c(hatbeta0, hatbeta1))
56 #####

```

```

# TRACE du rho de Pearson en fonction de Z
58 #####

60 # On va tracer trois courbes
# la premiere est la courbe theorique
62 # la seconde est le nuage de points issu, pour z fix, de l'estimation par R du tau sur (X,Y)
# la troisieme est issu du modele, i.e. le tau issu de hatB
64
nb<-5
66 print("Taille du sous-echantillonnage sur Z")
print(nb)
68
z<-seq(min(Z),max(Z),length=nb)
70 # print("TRAFFIC pour avoir les memes z que l'empirique")
# z<-Z
72 # nb<-length(z)

74 source('calcul_previson_spearman.R')

76 #####
# RESULTATS
78 #####

80 print("Resultats graphiques")
#x11()
82 if(cas_reel==0)
{
84     postscript("previson_spearman_psfrag.ps")
# si simulation-estimation-previson
86     plot(z,spearmanth,ylim=c(0,1),type='l',col='black',ylab="")
points(Z,spearmanempirique,type='p',col='black')
88     points(z,spearmanmodele,type='b',lty=3,pch=-1,col='black')
legend(0,0.3,c("Theoretical_rho_label_longlonglong","empirical_rho_label_longlonglong","previson_of_rho_label_longlonglong"),bty="n")
90     #title("Spearman's rho")
dev.off()
92     save(spearmanth,spearmanempirique,spearmanmodele,file="spearman.dat")
print("Erreur entre le coefficient spearman theorique et celui fourni par le modele")
94     erreur<-max(abs(spearmanth-spearmanmodele))
print(erreur)

96     if(length(spearmanth)==length(spearmanempirique))
98     {
100         print("Erreur entre le coefficient spearman theorique et celui empirique")
erreur<-max(abs(spearmanth-spearmanempirique))
102         print(erreur)
} else
104 {
# si cas reel: estimation-previson
106     # x11()
pdf("previson_spearman_Cas-reel.pdf")
108     plot(z,spearmanmodele,ylim=c(0,1),type='l',col='black')
points(Z,spearmanempirique,type='p',col='black')
110     # title("Spearman's rho")
dev.off()
112 }

114 if (profiling) {
Rprof(NULL)
116     print("Statistiques d'occupation")
print(summaryRprof('previson.out'))
118 }

```

Routine de calcul – calcul prevision spearman.R

```

# calcul_previson_spearman.R
2
# modele de Cox conditionnel
4 # lambda_X(t|z)=lambda_0(t)*Phi(z)
# lambda_Y(t|x,z)=lambda_0(t)*Psi(z)
6 #
# estimation de B(z,t)
8 # en s'appuyant sur l'expression entre deux fonctions B induite
# par le modle
10 # On fait intervenir les ratios issu du modele sous la notation K
12 dyn.load("../SRC/scopuleempirique.so")
14 source('calcul_B.R')

16 #####

18 # On va tracer trois courbes
# la premiere est la courbe theorique
20 # la seconde est le nuage de points issu, pour z fix, de l'estimation par R du tau sur (X,Y)
# la troisieme est issu du modele, i.e. le tau issu de hatB
22
#####
24
# if(cas_reel==0){
26 #     # pour eviter du calcul inutile

```

```

#      print("cadre des copules de valeurs extremes pour la theorique.")
28 #      spearmanth<-rep(0,nb)
#      for(i in 1:nb)
30 #      {
#          B<-function(t,z)
32 #          {
#              K<-Psi(z,hatbeta0,hatbeta1)/Phi(z,hatalpha0,hatalpha1)
34 #              terme1<-(1-K)*(1-t)
#              facteur1<-K*(1-t)+t
36 #              arg<-K/t-(K-1)
#              facteur2<-AClassique(1/arg,choix="survie")
38 #              terme2<-facteur1*facteur2
#              B<-terme1+terme2
40 #          }
#      F<-function(t)
42 #      {
#          F<-1/(1+Bz(t,z[i]))^2
44 #      }
#      J1<-integrate(F,lower=0,upper=1,subdivision=1000)
46 #
#      J2<-0
48 #      tt<-seq(0,1,length=1000)
#      for(k in 1:length(tt))
50 #          J2<-J2+F(tt[k])
#      J2<-J2/(length(tt)-1)
52 #      print("Rho th| integrale (integrate) - Rho th| integrale(calcul main)")
#      print(c(J1$value,12*J1$value-3,J2,12*J2-3))
54 #      spearmanth[i]<-12*J1$value-3
#      }
56 # }

#####
nz<-length(Z)
60
spearmanempirique<-rep(0,nz)
62
for(i in 1:nz)
64     spearmanempirique[i]<-cor(X[(ncl[i]+1):(ncl[i+1])],Y[(ncl[i]+1):(ncl[i+1])],use="complete.obs",method="spearman")
# print("Spearman empirique - valeur du rho")
66 # print(spearmanempirique)

#####
# spearmanmodele<-rep(0,nb)
70 # print("Spearman modele - valeur de l'integrale")
# for(i in 1:nb)
72 # {
#     F<-function(t)
74 #     {
#         F<-1/(1+hatBZHE(t,z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1))^2
76 #     }
#
#     # pour des raison inconnues, l'integrale directe via R ne donne pas
78 #     # de bons resultats
#     #J1<-integrate(F,lower=0.0001,upper=0.9999,subdivisions=10000)
80 #
#     K<-0
#     ecart<-1e-4
82 #     pas<-1e-3
#     ti<-seq(ecart,1-ecart,by=pas)
84 #     for(k in 1:length(tt))
#         K<-K+F(ti[k])
86 #     K<-K*pas
#
#     #print("Rho mod| integrale (integrate) - Rho mod| integrale(calcul main)")
90 #     #print(c(J1$value,12*J1$value-3,J,12*J-3))
#     print("Rho mod| integrale(calcul main)")
92 #     print(c(J,12*J-3))
#     spearmanmodele[i]<-12*J-3
94 # }

# NOUVELLE ECRITURE
98 spearmanmodele<-rep(0,nb)
# if(cas_reel==0)
100 {
#     print("cadre des copules de valeurs extremes pour la theorique.")
102 #     spearmanth<-rep(0,nb)
# }
104 # print("Spearman modele - valeur de l'integrale")
# for(i in 1:nb)
106 # {
#     print("Boucle sur Z"); print(i)
108 #     # TEST pour savoir si Phi(z)>=Psi(z)
#     if(Phi(z[i],hatalpha0,hatalpha1)<Psi(z[i],hatbeta0,hatbeta1))
110 #     {
#         stop("Erreur sur l'ordre de Phi et Psi.")
112 #         # inversion de X et Y selon le modele de l'article
#
114 #         XX<-X
#         YY<-Y
116 #     }
# }
118 # DEFINITION DES FONCTIONS UTILES POUR LE CAS THEORIQUE

```

```

120   if (cas_reel==0)
121   {
122       Bz<-function(t,z)
123       {
124           K<-Psi(z,hatbeta0,hatbeta1)/Phi(z,hatalpha0,hatalpha1)
125           terme1<-(1-K)*(1-t)
126           if (terme1 < 0) stop("Erreur sur le sens de Phi et Psi.")
127           facteur1<-K*(1-t)+t
128           arg<-K/t-(K-1)
129           facteur2<-AClassique(1/arg,choix="survie")
130           terme2<-facteur1*facteur2
131           Bz<-terme1+terme2
132       }
133       Fth<-function(t)
134       {
135           Fth<-1/(1+Bz(t,z[i]))^2
136       }
137
138   # DEFINITION POUR LE MODELE, DEFINI TOUT LE TEMPS
139   Fmod<-function(t)
140   {
141       Fmod<-1/(1+hatBZHE(t,z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1))^2
142   }
143   Fmod1<-function(t)
144   {
145       Fmod1<-1/(1+hatAsimplifieZHE(t,z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1))^2
146   }
147   # Fmod2<-function(t)
148   # {
149   #     Fmod2<-1/(1+hatAnZHE(t,z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1))^2
150   # }
151
152   FmodPickands<-function(t)
153   {
154       FmodPickands<-1/(1+hatBZHEPickands(t,z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1))^2
155   }
156   Jmod<-0
157   Jmod1<-0
158   # Jmod2<-0
159   JmodPickands<-0
160   ecart<-1e-4
161   pas<-1e-2
162   tt<-seq(ecart,1-ecart,by=pas)
163   if (cas_reel==0) Jth2<-0
164
165   norme_sup<-0
166   norme_sup_inv<-0
167   norme_sup1<-0
168   norme_sup_inv1<-0
169   # norme_sup2<-0
170   # norme_sup_inv2<-0
171   norme_sup_Pickands<-0
172   norme_sup_inv_Pickands<-0
173   for (k in 1:length(tt))
174   {
175       # COMPARAISON DE L'ESTIMATION DE LA FONCTION DE DEPENDANCE
176       # POUR CHAQUE Z
177       # ON DETERMINE LA NORME SUP ENTRE Bz et hatBZHE
178       if (cas_reel==0)
179       {
180           norme_sup<-max(norme_sup,abs(Bz(tt[k],z[i])-hatBZHE(tt[k],z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1)))
181       }
182       norme_sup_inv<-max(norme_sup_inv,abs(Fth(tt[k])-Fmod(tt[k])))
183       norme_sup1<-max(norme_sup1,abs(Bz(tt[k],z[i])-hatAsimplifieZHE(tt[k],z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1)))
184       norme_sup_inv1<-max(norme_sup_inv1,abs(Fth(tt[k])-Fmod1(tt[k])))
185       # norme_sup2<-max(norme_sup2,abs(Bz(tt[k],z[i])-hatAnZHE(tt[k],z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1)))
186       # norme_sup_inv2<-max(norme_sup_inv2,abs(Fth(tt[k])-Fmod2(tt[k])))
187       norme_sup_Pickands<-max(norme_sup_Pickands,abs(Bz(tt[k],z[i])-hatBZHEPickands(tt[k],z[i],Z,X,Y,ncl,Phi,Psi,hatalpha0,hatalpha1,hatbeta0,hatbeta1)))
188       norme_sup_inv_Pickands<-max(norme_sup_inv_Pickands,abs(Fth(tt[k])-FmodPickands(tt[k])))
189
190       Jth2<-Jth2+Fth(tt[k])
191   }
192   # CALCUL DE L'INTEGRALE
193   Jmod<-Jmod+Fmod(tt[k])
194   Jmod1<-Jmod1+Fmod1(tt[k])
195   # Jmod2<-Jmod2+Fmod2(tt[k])
196   JmodPickands<-JmodPickands+FmodPickands(tt[k])
197
198   print("Boucle sur (en%) et boucle sur (en%)"); print(c(i,100*i/nb,k,100*k/length(tt)))
199   affichage<-"NON"
200   if ((cas_reel==0)&&(affichage=="OUI"))
201   {
202       print("Norme sup entre fonction"); print(norme_sup);
203       print("Norme sup entre fonction Pickands"); print(norme_sup_Pickands);
204   }

```

```

212 #           print("Norme sup entre fonction (simplifie)"); print(norme_sup1);
           print("Norme sup entre fonction (complique)"); print(norme_sup2);
214 # print("Norme sup entre inverse des carr "); print(norme_sup_inv);
           # print("Norme sup entre inverse des carr Pickands "); print(norme_sup_inv_Pickands);
216 # print("Norme sup entre inverse des carr (simplifie)"); print(norme_sup_inv1);
           # print("Norme sup entre inverse des carr (complique)"); print(norme_sup_inv2);
218     }
220     }
222     {
224       Jth2<-Jth2*pas
       JmodPickands<-JmodPickands*pas
226     }
228     Jmod<-Jmod*pas
           Jmod1<-Jmod1*pas
           # Jmod2<-Jmod2*pas
230     # pour des raison inconnues, l'integrale directe via R ne donne pas
           # de bons resultats
232     # J1<-integrate(Fmod, lower=0.0001, upper=0.9999, subdivisions=10000)
           # print("Rho mod|integrale (integrate) - Rho mod|integrale (calcul main)")
234     # print(c(J1$value, 12*J1$value-3, J, 12*J-3))
236     if (cas_reel==0)
238     {
           # Jth1<-integrate(Fth, lower=0, upper=1, subdivision=1000)
           # print("Rho th|integrale (integrate) - Rho th|integrale (calcul main)")
240     # print(c(Jth1$value, 12*Jth1$value-3, Jth2, 12*Jth2-3))
           # spearmanth[i]<-12*Jth1$value-3
242     #
244 # PAR LE CALCUL DE SOMME d'OU Jth2 et pas Jth1
           spearmanth[i]<-12*Jth2-3
246     }
           # print("Ecart entre integrale - Pickands - simplifie - complique"); print(c(abs(Jmod-Jth2), abs(Jth2-JmodPickands), abs(J
248 # print("Ecart entre integrale - Pickands - simplifie "); print(c(abs(Jmod-Jth2), abs(Jth2-JmodPickands), abs(Jmod1-Jth2)));
           # print("Ecart entre rho - Pickands - simplifie "); print(c(abs(12*Jmod-12*Jth2), abs(12*Jth2-12*JmodPickands), abs(12*Jmod1-
250 # print("Rho mod|integrale (calcul main)")
           # print(c(Jmod, 12*Jmod-3))
           spearmanmodele[i]<-12*Jmod1-3
252     }
254 }
dyn.unload("../SRC/scopuleempirique.so")

```

Routine de calcul d'estimation de la fonction de dépendance non paramétrique – calcul B.R

```

1 # calcul_B.R
3 source('outils.R')
5 #####
7 #####
9 # Definition de la fonction A pour les copules de ve
11 AClassique<-function(t, choix)
12 {
13     # copule de survie de valeur extreme
           # donc on peut prendre n'importe quelle valeur pour v
15     # ATTENTION, ne pas ecrire log(C(exp(1-t), exp(t)))
           # car la propriete C^t(u,v)=C(u^t, v^t) n'est valid que pour t>0
17     if(choix=="directe") AClassique <- -log(Copule(exp(t-1), exp(-t)))
19     if(choix=="survie") AClassique <- -log(SCopule(exp(t-1), exp(-t)))
21 }
23
25 Aempirique<-function(t, U, V, v)
26 {
27     Aempirique <- t*log(v)*log(Copuleempirique(X, Y, v^(1/t-1), v))
29 }
31 #####
           # DIFFERENTS ESTIMATEURS DE A (ou B)
           # formule de transition adapte
33 #####
35 poids<-function(z0, zz, Z)
36 {
37     Fct<-function(x) exp(-2*x)
39     nz<-length(Z)
           somme<-0
41     for(i in 1:nz)

```

```

43     {
44         somme<-somme+Fct( abs(z0-Z[i]))
45     }
46     poids<-Fct( abs(z0-zz))/somme
47 }
48 # VERSION INITIALE
49 hatBZE<-function(s,X,Y,choix)
50 {
51     # estimation de la fonction de dependance B de z (z dans l'echantillon)
52     # Z n'apparait pas ici, en fait Z est fix donc il s'agit de
53     # l'echantillon (X,Y) associe un Z
54     # pour s'affranchir de la dependance par rapport v
55     # on integre sur v
56
57     v<-seq(0.01,0.99,length=100)
58     terme<-rep(0,length(v))
59     for(i in 1:length(v))
60     {
61         if(choix=="directe") a<-Copuleempirique(X,Y,v[i]^(1/s-1),v[i])
62         if(choix=="survie") a<-SCopuleempirique(X,Y,v[i]^(1/s-1),v[i])
63         terme[i]<-terme[i]+ifelse(a>0,log(a)/log(v[i]),0)
64     }
65     # le choix de median peut s'avrer meilleur
66     termeM<-mean(terme)
67     # l'integrale donne
68     #termeM<-terme/(length(v)-1)
69
70     # modification pour obtenir une fonction non nulle pour des s petits, a cause de la copule empirique
71     hatBZE<-min(1,max(s,1-s,s*termeM))
72 }
73
74 hatBZHE<-function(t,z,Z,X,Y,ncl,Phi,Psi,a0,a1,b0,b1)
75 {
76     # valeur de B de z (valeur non incluse necessairement dans les valeurs de Z)
77     # au point t
78     # on integre sur tous les Z de l'echantillon
79     nz<-length(Z)
80
81     hatB1<-0
82     for(i in 1:nz)
83     {
84         Kv<-K(Phi,Psi,z,Z[i],a0,a1,b0,b1)
85         terme1<-(1-Kv)*(1-t)
86         facteur1<-Kv*(1-t)+t
87         arg<-Kv/t-(Kv-1)
88         #facteur2<-Bempirique(1/arg,X[((i-1)*n+1):(i*n)],Y[((i-1)*n+1):(i*n)],type="survie")
89         facteur2<-hatBZE(1/arg,X[(ncl[i]+1):(ncl[i+1])],Y[(ncl[i]+1):(ncl[i+1])],choix="survie")
90         terme2<-facteur1*facteur2
91         hatB1<-hatB1+poids(z,Z[i],Z)*(terme1+terme2)
92     }
93     hatBZHE<-hatB1
94 }
95
96 #####
97 # Version simplifiee
98 #####
99
100 Ansimplifie<-function(t,X,Y,choix="directe")
101 {
102     if(choix=="directe") a<-Copuleempirique(X,Y,exp(t-1),exp(-t))
103     if(choix=="survie") a<-SCopuleempirique(X,Y,exp(t-1),exp(-t))
104     Ansimplifie<-min(1,max(t,1-t,-log(a)))
105 }
106
107 AnsimplifieV2<-function(t,X,Y,choix="directe")
108 {
109     if(choix=="directe")
110     {
111         integrale<-0
112         pas<-seq(exp(-t),1,length=100)
113         for(i in 1:length(pas))
114             integrale<-integrale+Copuleempirique(X,Y,pas[i]**(1/t-1),pas[i])
115         a<-integrale/(length(pas)-1)
116     }
117     if(choix=="survie")
118     {
119         integrale<-0
120         pas<-seq(exp(-t),1,length=100)
121         for(i in 1:length(pas))
122             integrale<-integrale+SCopuleempirique(X,Y,pas[i]**(1/t-1),pas[i])
123         a<-integrale/(length(pas)-1)
124     }
125     AnsimplifieV2<-min(1,max(t,1-t,-log(a)))
126 }
127
128 AnsimplifieV3<-function(t,X,Y,choix="directe")
129 {
130     integrale<-0
131     pas<-seq(0,1,length=100)
132     for(i in 1:length(pas))

```

```

135     {
136         if(choix=="directe")
137         {
138             argument<-Copuleempirique(X,Y,pas[i]**(1/t-1),pas[i])
139             integrale<-integrale+log(ifelse(argument>0,argument,0))
140         }
141         if(choix=="survie")
142         {
143             argument<-SCopuleempirique(X,Y,pas[i]**(1/t-1),pas[i])
144             integrale<-integrale+log(ifelse(argument>0,argument,0))
145         }
146     }
147 }
148
149 hatAsimplifieZHE<-function(t,z,Z,X,Y,ncl,Phi,Psi,a0,a1,b0,b1)
150 {
151     # valeur de B de z (valeur non incluse necessairement dans les valeurs de Z)
152     # au point t
153     # on integre sur tous les Z de l'echantillon
154     nz<-length(Z)
155
156     hatB1<-0
157     for(i in 1:nz)
158     {
159         Kv<-K(Phi,Psi,z,Z[i],a0,a1,b0,b1)
160         terme1<-(1-Kv)*(1-t)
161         facteur1<-Kv*(1-t)+t
162         arg<-Kv/t-(Kv-1)
163         #facteur2<-Bempirique(1/arg,X[((i-1)*n+1):(i*n)],Y[((i-1)*n+1):(i*n)],type="survie")
164         facteur2<-Ansimplifie(1/arg,X[(ncl[i]+1):(ncl[i+1])],Y[(ncl[i]+1):(ncl[i+1])],choix="survie")
165         terme2<-facteur1*facteur2
166         hatB1<-hatB1+poids(z,Z[i],Z)*(terme1+terme2)
167     }
168     hatAsimplifieZHE<-hatB1
169 }
170
171 #####
172 # Version avec delta_n
173 #####
174
175 deltan<-function(n) deltan<-1/sqrt(log(n))
176
177 minsxn<-function(n) minsxn<-log(1-deltan(n))/log(deltan(n)*(1-deltan(n)))
178
179 maxsxn<-function(n) maxsxn<-log(deltan(n))/log(deltan(n)*(1-deltan(n)))
180
181 An<-function(t,X,Y,choix="directe")
182 {
183     pas<-1e-2
184     somme<-0
185
186     nech<-length(X)
187
188     # la grille evolue avec deltan
189     if(deltan(nech)>(1-deltan(nech))) stop("La taille de l'echantillon est trop faible.")
190     else v<-seq(deltan(nech),(1-deltan(nech)),by=pas)
191
192     for(i in 1:length(v))
193     {
194         T1<-minsn(nech)+t*(maxsxn(nech)-minsn(nech))
195         if(choix=="directe") a<-Copuleempirique(X,Y,v[i]^(1/T1-1),v[i])
196         if(choix=="survie") a<-SCopuleempirique(X,Y,v[i]^(1/T1-1),v[i])
197         T2<-ifelse(a>0,log(a),0)
198         somme<-somme+T1/log(v[i])*T2
199     }
200     # modification possible avec l'incorporation de min(1,max(s,1-s,...))
201     An<-min(1,max(t,1-t,somme*pas))
202 }
203
204 # Drift du processus sqrt(n)[An(t)-A(t)]
205 drift<-function(t,n,G)
206 {
207     # G est de la forme
208     # G<-ACatalogue()
209     T1<-minsn(n)+t*(maxsxn(n)-minsn(n))
210     drift<-sqrt(n)*(G(T1)*(1-2*deltan(n))-G(t))
211 }
212
213 hatAnZHE<-function(t,z,Z,X,Y,ncl,Phi,Psi,a0,a1,b0,b1)
214 {
215     # valeur de B de z (valeur non incluse necessairement dans les valeurs de Z)
216     # au point t
217     # on integre sur tous les Z de l'echantillon
218     nz<-length(Z)
219
220     hatB1<-0
221     for(i in 1:nz)
222     {
223         Kv<-K(Phi,Psi,z,Z[i],a0,a1,b0,b1)
224         terme1<-(1-Kv)*(1-t)
225         facteur1<-Kv*(1-t)+t

```

```

227     arg<-Kv/t-(Kv-1)
      #facteur2<-Bempirique(1/arg,X[((i-1)*n+1):(i*n)],Y[((i-1)*n+1):(i*n)],type="survie")
229     facteur2<-An(1/arg,X[(ncl[i]+1):(ncl[i+1])],Y[(ncl[i]+1):(ncl[i+1])],choix="survie")
      terme2<-facteur1*facteur2
231     hatB1<-hatB1+poids(z,Z[i],Z)*(termel+terme2)
      }
233   hatAnZHE<-hatB1

235 #####

237 # hatBZHE<-fonction(t,z,Z,X,Y,ncl,Phi,Psi,a0,a1,b0,b1)
# {
239 #   # valeur de B de z (valeur non incluse necessairement dans les valeurs de Z)
#   # au point t
241 #   # on integre sur tous les Z de l'echantillon
#   nz<-length(Z)
243 #
#   hatB1<-0
245 #   for(i in 1:nz)
#   {
247 #     Kv<-K(Phi,Psi,z,Z[i],a0,a1,b0,b1)
#     termel<-(1-Kv)*(1-t)
249 #     facteur1<-Kv*(1-t)+t
#     arg<-Kv/t-(Kv-1)
251 #     #facteur2<-Bempirique(1/arg,X[((i-1)*n+1):(i*n)],Y[((i-1)*n+1):(i*n)],type="survie")
#     facteur2<-hatBZE(1/arg,X[(ncl[i]+1):(ncl[i+1])],Y[(ncl[i]+1):(ncl[i+1])],choix="survie")
253 #     terme2<-facteur1*facteur2
#     hatB1<-hatB1+poids(z,Z[i],Z)*(termel+terme2)
255 #   }
#   hatBZHE<-hatB1
257 # }

259 # Pour la comparaison dans le programme borne
# et plus precisement dans calcul_prevision_spearman_IC.R
261
263 library(evd)
265 hatBZEPickands<-fonction(s,X,Y,choix)
# {
267 #   # ATTENTION voir le probleme de la survie
#   data<-matrix(0,length(X),2)
#   data[,1]<-X
269 #   data[,2]<-Y
#   # l'option convex=TRUE permet de prendre le convexe minorant et donc d'avoir
271 #   # un estimateur convexe
#   hatBZEPickands<-abvnonpar(s,data,method="pickands",plot=FALSE,convex=FALSE)
273 # }

275 hatBZHEPickands<-fonction(t,z,Z,X,Y,ncl,Phi,Psi,a0,a1,b0,b1)
# {
277 #   # valeur de B de z (valeur non incluse necessairement dans les valeurs de Z)
#   # au point t
279 #   # on integre sur tous les Z de l'echantillon
#   nz<-length(Z)
#   hatB1<-0
281 #   for(i in 1:nz)
#   {
283 #     Kv<-K(Phi,Psi,z,Z[i],a0,a1,b0,b1)
#     termel<-(1-Kv)*(1-t)
285 #     facteur1<-Kv*(1-t)+t
#     arg<-Kv/t-(Kv-1)
287 #     facteur2<-hatBZEPickands(1/arg,X[(ncl[i]+1):ncl[i+1]],Y[(ncl[i]+1):ncl[i+1]],choix="survie")
#     terme2<-facteur1*facteur2
289 #     hatB1<-hatB1+poids(z,Z[i],Z)*(termel+terme2)
291 #   }
#   hatBZHEPickands<-hatB1
293 # }

295 ACFGmoi<-fonction(tt,X,Y)
# {
297 #   # le parametre tolerance est introduit par
#   # resoudre les problemes de calculs de l'integrale
299 #   # dans A0n et A1n quand le denominateur s'annule

301   tolerance<-1e-5
#   finesse<-200
303   #print("reglages effectuer pour le calcul correct des transformations de marges.")
#   U<-exp(-1/X) # ces expressions sont issues du choix mar=c(1,1,1) dans la simulation vi le packahe evd
305   V<-exp(-1/Y)
#   Z<-log(U)/log(U*V)
307
#   Hn<-fonction(z)
309   {
#     n<-length(Z)
#     somme<-0
311     for(i in 1:n)
#       somme<-somme+(Z[i]<=z)
313     Hn<-somme/n
315   }
317   A0n<-fonction(tt)

```



```

319     {
320         integrale<-0
321         pas<-seq((0+tolerance), tt, length=finesse)
322         for(i in 1:length(pas))
323             integrale<-integrale+(Hn(pas[i])-pas[i])/(pas[i]*(1-pas[i]))
324         A0n<-exp(integrale)
325     }
327     A1n<-function(tt)
328     {
329         integrale<-0
330         pas<-seq(tt,(1-tolerance), length=finesse)
331         for(i in 1:length(pas))
332             integrale<-integrale+(Hn(pas[i])-pas[i])/(pas[i]*(1-pas[i]))
333         integrale<-integrale/(finesse-1)
334         A1n<-exp(-integrale)
335     }
337     p<-function(tt) 1-tt
339     ACFGmo<-exp(p(tt)*log(A0n(tt))+(1-p(tt))*log(A1n(tt)))
}

```

.3 Librairies C

.3.1 Copule empirique – copuleempirique.c

```

void copuleempiriqueetoile(double *U, double *V, int *taille, double *u, double *v, d
2 // Version de la pseudo copule empirique – correspond C_n* de Fermanian (2004)
    double s;
4     int i, j;
    int n=*taille;
6     double Uc=*u;
    double Vc=*v;
8
    s=0.0;
10    for(i=0; i<n; i++)
    {
12        s=s+(U[i]<=Uc)*(V[i]<=Vc);
    }
14    *valeur=s/(*taille);
}
16 // double Fn1(double *Xs, int *taille, double x) {
18 // // OK
    //     double s;
20 //     int i=0;
    //     int n=*taille;
22 //
    //     s=0.0;
24 //     while((Xs[i]<=x)&&(i<n))
    //     {
26 //         s=s+1;
    //         i=i+1;
28 //     }
    //     return s/(n*1.0);
30 // }

```

```

32 // pour verification dans un programme R
33 // void Fn(double *Xs,int *taille , double *x, double *valeur) {
34 //     *valeur=Fn1(Xs,taille ,( *x));
35 // }
36 //
37 // double FnInv1(double *Xs,int *taille , double u) {
38 // // OK
39 //     int i=( *taille)-1;
40 //
41 //     if(u<=1/(( *taille)*1.0)) return Xs[0];
42 //     else {
43 //         if(u>1) return 10;
44 //         else {
45 //             while(Fn1(Xs,taille ,Xs[i])>=u)
46 //                 i=i-1;
47 //             return Xs[i+1];
48 //         }
49 //     }
50 // }

51
52 double FnInv2(double *Xs,int *taille , double u) {
53 // OK
54 //     int i=( *taille);
55
56 //     if(u<=1/(( *taille)*1.0)) return Xs[0];
57 //     else {
58 //         if(u>1) return 10;
59 //         else {
60 //             while((i/(( *taille)*1.0)>=u))
61 //                 i=i-1;
62 //             return Xs[i];
63 //         }
64 //     }
65 // }
66
67 //verification de la fonction precedente
68 // void FnInv1P(double *Xs,int *taille , double *u,double *resultat){
69 //     *resultat=FnInv1(Xs,taille ,( *u));
70 // }

71
72 //verification de la fonction precedente
73 void FnInv2P(double *Xs,int *taille , double *u,double *resultat){
74     *resultat=FnInv2(Xs,taille ,( *u));
75 }
76
77 // void copuleempirique1(double *X,double *Y,int *taille ,double *x,double *y,dou
78 //     double s;

```

```

//      int i;
80 //      int n=*taille;
//      double xx,yy;
82 //
//      s=0.0;
84 //      xx=FnInv1(Xs,taille,(*x));
//      yy=FnInv1(Ys,taille,(*y));
86 //      for(i=0; i<n; i++)
//      {
88 //          s=s+(X[i]<=xx)*(Y[i]<=yy);
//      }
90 //      *valeur=s/((*taille)*1.0);
// }
92
void copuleempirique(double *X,double *Y,int *taille ,double *x,double *y,double
94     double s;
//      int i;
96     int n=*taille;
//      double xx,yy;
98
//      s=0.0;
100     xx=FnInv2(Xs,taille,(*x));
//      yy=FnInv2(Ys,taille,(*y));
102     for(i=0; i<n; i++)
//      {
104         s=s+(X[i]<=xx)*(Y[i]<=yy);
//      }
106     *valeur=s/((*taille)*1.0);
}

```

.3.2 Estimation des paramètres marginaux du modèle de Cox - implémentation de la méthode d'Efron au cas d'égalité

Estimation usuelle – estimation.c

```

1 #include <R.h>
// #include <Rdefines.h>
3
SEXP mkans(double x)
5 {
//     SEXP ans;
7     PROTECT(ans = allocVector(REALSXP,1));
//     REAL(ans)[0]=x;
9     UNPROTECT(1);
//     return ans;
11 }
13 double feval(double x, SEXP f, SEXP rho)

```

```

15     {
16         defineVar(install("x"), mkans(x), rho);
17         return(REAL(eval(f, rho))[0]);
18     }
19 SEXP VPC(SEXP f, SEXP ZR, SEXP nzR, SEXP rho)
20     {
21         int i, j;
22         double somme;
23         double produit;
24         int taille=INTEGER(nzR)[0];
25         double* Z=REAL(ZR);
26
27         produit=0.0;
28
29         /* optimisation du calcul */
30         /* on calcule la somme, denominateur, une fois
31          * puis on soustrait a chaque pas, le terme en trop
32          * */
33         somme=0.0;
34         for(i=0; i<taille; i++)
35             somme+=feval(Z[i], f, rho);
36         for(i=0; i<taille; i++)
37         {
38             if(i>0) somme=somme-feval(Z[i-1], f, rho);
39             produit+=log(feval(Z[i], f, rho))-log(somme);
40         }
41         return mkans(produit);
42     }

```

Estimation par la méthode d'Efron – estimation Efron.c

```

1 #include <R.h>
2 #include <Rdefines.h>
3 #include <Rinternals.h>
4
5 #define RMATRIX(m, i, j) (REAL(m)[ INTEGER(GET_DIM(m))[0]*(j)+(i) ])
6
7 SEXP mkans(double x)
8     {
9         SEXP ans;
10        PROTECT(ans = allocVector(REALSXP, 1));
11        REAL(ans)[0]=x;
12        UNPROTECT(1);
13        return ans;
14    }
15
16 double feval(double x, SEXP f, SEXP rho)
17     {

```

```

        }
67 // printf("somme Efron pour i=%d (DI=%d): %f\n", (i+1), DI[i], somme2)

69 denominateur=0.0;

71 precedent=0.0;
        if(i>0){
73         for(l=0;l<DI[i-1];l++)
            precedent=precedent+feval(RMATRIX(ZR,i-1,l),f,rho)
75         }
        // printf("pour i=%d - precedent: %f\n", (i+1), precedent);
77         somme1=somme1-precedent;
        // printf("d'ou la somme: %f\n", somme1);

79         for(j=0;j<DI[i];j++)
81         {
            denominateur+=log(somme1-((j*1.0+1)-1)/DI[i]*somme2);
83 // printf("calcul j:%d\n", (j+1));
            // printf("somme1 %f, somme2: %f\n", somme1, somme2);
85 // coef=((j*1.0+1)-1)/DI[i];
            // printf("j=%d, DI[%d]=%d, coefficient %f\n", (j+1), (i+1), DI[i], coef);
87 // printf("verif %f\n", somme1-((j*1.0+1)-1)/DI[i]*somme2);
            }
89         somme+=sommeLog-denominateur;
        // i=taille; printf("saut\n");
91     }
    // printf("somme finale: %f\n", somme);
93     return mkans(somme);
}

```

.3.3 Inversion d'un fonction – inv.c

```

#include <R.h>
2 #include <Rdefines.h>

4 SEXP mkans(double x)
{
6     SEXP ans;
    PROTECT(ans = allocVector(REALSXP,1));
8     REAL(ans)[0]=x;
    UNPROTECT(1);
10    return ans;
}

12 double feval(double x, SEXP f, SEXP rho)
14 {
    defineVar(install("x"), mkans(x), rho);
16    return(REAL(eval(f, rho))[0]);
}

```

```

}
18 SEXP Finv(SEXP f, SEXP guesses, SEXP stol, SEXP rho)
20 {
    double x0=REAL(guesses)[0], x1=REAL(guesses)[1], tol=REAL(stol)[0];
22     double f0, f1, fc, xc;

24     if(tol <=0.0) error("tolerance negative");
    f0=feval(x0, f, rho);
26     f1=feval(x1, f, rho);
    if(f0==0.0) return mkans(x0);
28     if(f1==0.0) return mkans(x1);
    if(f0*f1 >0.0) error("x[0] and x[1] have the same sign");
30     for(;;) {
        xc=0.5*(x0+x1);
32         if(fabs(x0-x1) < tol) return mkans(xc);
        fc=feval(xc, f, rho);
34         if(fc==0) return mkans(xc);
        if(f0*fc >0.0){
36             x0=xc; f0=fc;}
        else {
38             x1=xc; f1=fc;}
    }
40 }

```

.3.4 Copule de survie empirique – scopuleempirique.c

```

1 void scopuleempirique(double *U, double *V, int *taille, double *u, double *v, double
    double s;
3     int i, j;
    int n=*taille;
5     double Uc=*u;
    double Vc=*v;
7
    s=0.0;
9     for(i=0; i<n; i++)
    {
11         s=s+(1-U[i]<=Uc)*(1-V[i]<=Vc);
    }
13     *valeur=s/(*taille);
}

```

Package R - Estimation de la fonction de dépendance d'une copule bivariée de valeurs extrêmes

.4 Estimation de la fonction de dépendance dans le cadre non paramétrique – estimation A test.R

```

# VERSION POUR LES TESTS A GRANDE ECHELLE
2 library(evd)

4 # Modele des copules
  source('catalogue.R')
6 source('calcul_B.R')

8 # Replications possibles
  # pour tester sur plusieurs echantillons
10 print("Nombre de replications")
  print(taille_test)
12
14 print("Taille d'echantillon pour chaque replication")
  print(taille)

16 seuil_alpha<-0.05

18 # Declaration des variables
  ePickands<-rep(0,length(taille_test))
20 ePickandsmarge<-rep(0,length(taille_test))
  eHall<-rep(0,length(taille_test))
22 eCFG<-rep(0,length(taille_test))
  #eEmpirique<-rep(0,length(taille_test))
24 #eEmpiriquen<-rep(0,length(taille_test))
  eEmpiriquens<-rep(0,length(taille_test))
26 eEmpiriquensV2<-rep(0,length(taille_test))

28 for(test in 1:taille_test)
  {
30   print("Replication")
    print(test)
32
34   Copule<-function(u,v) CopuleCatalogue(u,v,theta,type)
    DUCopule<-function(u,v) DCopuleCatalogue(u,v,theta,type)
    DVCopule<-function(u,v) DCopuleCatalogue(v,u,theta,type) # CAR SYMETRIE
36   A<-function(t) ACatalogue(t,theta,type)
    Aprime<-function(t) AprimeCatalogue(t,theta,type)
38
39   # Definition des marges qui ne jouent en fait aucun role
40   # Elles peuvent jouer car les estimateurs par dfaut
41   # suppose que les marges sont de valeurs extremes
42   #F<-function(x) (0<=x<=1)*x**2 #exp(-(x^(-alpha)))
44
45   FInv<-function(u)
46   {
47     #FInv<-sqrt(u)
48
49     # Lois limites
50     # 1. Gumbel
51     # (domaine d'attraction: normale, lognormal, exponentielle)
52     FInv<-log(-log(u))
53
54     # 2. Frechet (gamma>0)
55     # (domaine d'attraction: student, Cauchy, Pareto)
56     #gamma<-10
    #if((u<0)||(u>1)){print(u); stop('erreur de plage pour u.')}

```

```

58     #FInv<-(-log(u))**(-1/gamma)
60     # 3. Weibull (gamma>0)
61     # (domaine d'attraction: loi beta, loi uniforme)
62     #gamma<-2
63     #if((u<0)|| (u>1)){print(u); stop('erreur de plage pour u.')}
64     #FInv<-(-log(u))**(1/gamma)
65 }
66 #G<-function(v) F(v)
67
68 GInv<-function(v)
69 {
70     GInv<-FInv(v)
71 }
72
73
74
75 # Cadre classique
76 Simulation<-function(taille)
77 {
78     choix<-1
79
80     # Sorties (X,Y)
81     if(choix==1)
82     {
83         print("Simulation_moi")
84         S<-matrix(0,taille,2)
85         U<-runif(taille)
86         T<-runif(taille)
87         V<-rep(0,taille)
88         VR<-rep(0,taille)
89         for(j in 1:taille)
90         {
91             V[j]<-Inversion1(U[j],T[j])
92         }
93
94         # Transformation de marges
95         S[,1]<-FInv(U)
96         S[,2]<-GInv(V)
97     }
98     else
99     {
100        # les marges ne sont pas uniformes
101        print("Simulation_via_la_librarie_evd")
102        S<-rbvevd(taille,dep=1/theta,model="log")
103    }
104    Simulation<-S
105 }
106
107
108 ApickandsR<-function(t,X,Y)
109 {
110     data<-matrix(0,length(X),2)
111     data[,1]<-X
112     data[,2]<-Y
113     # l'option convex=TRUE permet de prendre le convexe minorant et donc d'avoir
114     # un estimateur convexe
115     ApickandsR<-abvnonpar(t,data,method="pickands",plot=FALSE,convex=FALSE)
116 }
117
118 AHallR<-function(t,X,Y)
119 {
120     data<-matrix(0,length(X),2)
121     data[,1]<-X
122     data[,2]<-Y
123     AHallR<-abvnonpar(t,data,method="hall",plot=FALSE,convex=FALSE)
124 }
125
126 ATiagoR<-function(t,X,Y)
127 {
128     data<-matrix(0,length(X),2)
129     data[,1]<-X
130     data[,2]<-Y
131     AHallR<-abvnonpar(t,data,method="tdo",plot=FALSE,convex=FALSE)
132 }
133
134 ACFGR<-function(t,X,Y)
135 {
136     data<-matrix(0,length(X),2)
137     data[,1]<-X
138     data[,2]<-Y
139     ACFGR<-abvnonpar(t,data,method="cfg",plot=FALSE,convex=FALSE)
140 }
141
142 #####
143 # Comparaison des estimations
144 #####
145
146 continue<-TRUE
147 tentative<-1
148 while(continue)
149 {

```

```

150     ES<-Simulation(taille)
151     X<-ES[,1]
152     Y<-ES[,2]
153     if(max(Y)!="NaN")
154         continue<-FALSE
155     else
156         tentative<-tentative+1
157 }
158 print("Tentative (pour éviter une erreur lors de l'inversion des Y)")
159 print(tentative)
160
162 t<-seq(0,1,by=0.01)
163 vv<-seq(0.1,0.9,by=0.01)
164
165 ApickandsRV<-rep(0,length(t))
166 AHallRV<-rep(0,length(t))
167 ACFGRV<-rep(0,length(t))
168 #AempiriqueV<-rep(0,length(t))
169 # AempiriqueVn<-rep(0,length(t))
170 AempiriqueVns<-rep(0,length(t))
171 AempiriqueVnsV2<-rep(0,length(t))
172 Binf<-rep(0,length(t))
173 Bsup<-rep(0,length(t))
174
175 for(i in 1:length(t))
176 {
177     ApickandsRV[i]<-ApickandsR(t[i],X,Y)
178     AHallRV[i]<-AHallR(t[i],X,Y)
179     ACFGRV[i]<-ACFGR(t[i],X,Y)
180
181     #AempiriqueV[i]<-hatBZE(t[i],X,Y,choix="directe")
182     # AempiriqueVn[i]<-An(t[i],X,Y)
183     AempiriqueVns[i]<-Ansimplifie(t[i],X,Y)
184     AempiriqueVnsV2[i]<-AnsimplifieV2(t[i],X,Y)
185
186     # INTERVALLE DE CONFIANCE
187     u1<-exp(t[i]-1); v1<-exp(-t[i])
188     var1<-Copule(u1,v1)-Copule(u1,v1)**2
189     var2<-Copule(u1,v1)*(DUCopule(u1,v1)*(1-u1)+DVCopule(u1,v1)*(1-v1))
190     var3<-DUCopule(u1,v1)**2*u1*(1-u1)+DVCopule(u1,v1)**2*v1*(1-v1)
191     var4<-2*DUCopule(u1,v1)*DVCopule(u1,v1)*(Copule(u1,v1)-u1*v1)
192     variance<-((var1-2*var2+var3+var4)/(Copule(u1,v1)**2))
193     Binf[i]<-AempiriqueVns[i]-sqrt(variance)/sqrt(taille)*qnorm(1-seuil_alpha/2)
194     Bsup[i]<-AempiriqueVns[i]+sqrt(variance)/sqrt(taille)*qnorm(1-seuil_alpha/2)
195 }
196
197 erreurPickands<-0
198 erreurHall<-0
199 erreurCFG<-0
200 #erreurE<-0
201 # erreurEn<-0
202 erreurEns<-0
203 erreurEnsV2<-0
204
205 for(i in 1:length(t))
206 {
207     # Pickands
208     erreur<-abs(A(t[i])-ApickandsRV[i])
209     if(!is.nan(erreur)) erreurPickands<-max(erreurPickands,erreur)
210
211     # Hall
212     erreur<-abs(A(t[i])-AHallRV[i])
213     if(!is.nan(erreur)) erreurHall<-max(erreurHall,erreur)
214
215     # CFG
216     erreur<-abs(A(t[i])-ACFGRV[i])
217     if(!is.nan(erreur)) erreurCFG<-max(erreurCFG,erreur)
218
219     # Empirique
220     #erreur<-abs(A(t[i])-AempiriqueV[i])
221     #if(!is.nan(erreur)) erreurE<-max(erreurE,erreur)
222
223     # Empirique en n
224     erreur<-abs(A(t[i])-AempiriqueVn[i])
225     if(!is.nan(erreur)) erreurEn<-max(erreurEn,erreur)
226
227     # Empirique simplifiée
228     erreur<-abs(A(t[i])-AempiriqueVns[i])
229     if(!is.nan(erreur)) erreurEns<-max(erreurEns,erreur)
230
231     erreur<-abs(A(t[i])-AempiriqueVnsV2[i])
232     if(!is.nan(erreur)) erreurEnsV2<-max(erreurEnsV2,erreur)
233 }
234
235 ePickands[test]<-erreurPickands
236 eHall[test]<-erreurHall
237 eCFG[test]<-erreurCFG
238 # eEmpirique[test]<-erreurE
239 # eEmpiriquen[test]<-erreurEn
240 eEmpiriquens[test]<-erreurEns
241 eEmpiriquensV2[test]<-erreurEnsV2

```

```

242
244 # GRAPHISME
      {
246     print("Graphisme")
      ARV<-rep(0,length(t))
248     for(i in 1:length(t)) ARV[i]<-A(t[i])

250     #x11()
      #bitmap(nom)
252     pdf(nom)
      plot(t,ARV,type='l',col='black',ylim=c(0.5,1),ylab="",lwd=2,lty=1)

254     points(t,ApickandsRV,type='l',col='gray',lwd=2,lty=4)
      # points(t,AHallRV,type='l',col='green',pch=3,lty=3,lwd=2)
256     points(t,ACFGRV,type='l',col='gray',lwd=2,lty=3)
      # points(t,AempiriqueV,type='l',col='blue',lty=1,lwd=2)

260     # points(t,AempiriqueVn,type='p',col='blue',lty=1)
      points(t,AempiriqueVns,type='l',col='black',lwd=2,lty=2)
262     points(t,AempiriqueVnsV2,type='l',col='red',lwd=2,lty=2)
      # INTERVALLE DE CONFIANCE
264     points(t,Binf,type='l',lty=1,col='gray')
      points(t,Bsup,type='l',lty=1,col='gray')

266     points(t,t,type='l',lty=1,col='gray')
268     points(t,1-t,type='l',lty=1,col='gray')

270     # legend(0.77,0.75,c("Theoretical_label","label_Pickands_long","label_Hall_long","label_CFG_long","label_Empiri
      legend(0.75,0.65,c("A","A-P","A-CFG","A-BD","A-BD-V2"),col=c('black','gray','gray','black','red'),lty=c(1,4,3,2
272     dev.off()
      }
274 } # fin de la boucle sur replication

276 print("Theta_□(parametre_□copule)")
      print(theta)
278 #print("Alpha (parametre loi)")
      #print(alpha)

280
      print("Norme_□sup")
282 print("Pickands")
      print(mean(ePickands))
284 print("Hall")
      print(mean(eHall))
286 print("CFG")
      print(mean(eCFG))
288 #print("Empirique")
      #print(mean(eEmpirique))
290 # print("Empirique en n")
      # print(mean(eEmpiriquen))
292 print("Empirique_□simplifie")
      print(mean(eEmpiriquens))
294 print("Empirique_□simplifie_□V2")
      print(mean(eEmpiriquensV2))

296
298 dyn.unload("../SRC/inv.so")
      dyn.unload("../SRC/copuleempirique.so")
      dyn.unload("../SRC/scopuleempirique.so")
300 # FIN

```

Bibliographie

- Abdous, B. and Ghoudi, K. (2005). Non-parametric estimators of multivariate extreme dependence functions. *J. Nonparametr. Stat.*, 17(8) :915–935.
- Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*, volume 55. Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.
- Aczél, J. (1966). *Lectures on functional equations and their applications*. Mathematics in Science and Engineering, Vol. 19. Academic Press, New York. Translated by Scripta Technica, Inc. Supplemented by the author. Edited by Hansjorg Oser.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Aron, M., Biecheler, M.-B., and Peytavin, J.-F. (1999). Temps Intervéhiculaire sur auto-route interurbaine. *RTS*, 64(1) :NC.
- Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37 :1554–1563.
- Bengio, Y. and Frasconi, P. (1995). An input Output HMM Architecture. In Tesauro, G., Touretzky, D., and Leen, T., editors, *Advances in Neural Information Processing Systems*, pages 427–434. MIT Press, Cambridge.
- Bengio, Y. and Frasconi, P. (1996). Input/Output HMMs for Sequence Processing. *IEEE Transactions on Neural Networks*, 7(5) :1231–1249.
- Bengio, Y., Lauzon, V. P., and Ducharme, R. (2001). Experiments on the Application of IOHMMs to Model Financial Returns Series. *IEEE Transactions on Neural Networks*, 12(1) :113–123.
- Bickel, P. J. and Ritov, Y. (1996). Inference in hidden Markov models. I. Local asymptotic normality in the stationary case. *Bernoulli*, 2(3) :199–228.
- Bickel, P. J., Ritov, Y., and Rydén, T. (1998). Asymptotic Normality of the Maximum-Likelihood Estimator for General Hidden Markov Models. *The Annals of Statistics*, 26(4) :1614–1635.
- Bickel, P. J., Ritov, Y., and Rydén, T. (2002). Hidden Markov model likelihoods and their derivatives behave like i.i.d. ones : Details. *IHP*, 38 :825–846.
-

- Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons.
- Brackstone, M. and McDonald, M. (2003). Driver Behaviour and Traffic Modelling. Are we looking at the Right Issues? In IEEE, editor, *Proc. of the IEEE Intelligent Vehicles Symposium 2003*, Columbus, Ohio, U.S.A.
- Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30 :89–99.
- Broniatowski, M., Depire, A., and Ritov, Y. (2007). Bivariate cox models and dependence analysis. *Journal of Multivariate Analysis (submitted)*, 1 :21p. submitted.
- Bui, V. A. P. (1996). Modélisation par Chaînes de Markov cachées du comportement de conduite sur autoroute. Master's thesis, École Polytechnique.
- Capéraà, P., Fougères, A.-L., and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3) :567–577.
- Carsten, O. (2000). Driver Assistance Systems : How Can We Know They Are Safe? In *ORACLE Symposium*, London.
- Carsten, O. (2003). Human Aspects of Transport Telematics and Road Safety. In *Liikenneturvallisuuuden tulevaisuuden näkymiä*, Helsinki.
- Chhetry, D., Sampson, A. R., and Kimeldorf, G. (1989). Concepts of setwise dependence. *Probab. Eng. Inform. Sci.*, 3 :367–380.
- Churchill, G. A. and Lazareva, B. (1999). Bayesian Restoration of a Hidden Markov Chain with Applications to DNA Sequencing. *Journal of Computational Biology*, 6(2) :261–277.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.
- Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34 :187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.
- Cuadras, C. M. (1992). Probability Distributions with Given Multivariate Marginals and Given Dependence Structure. *Journal of Multivariate Analysis*, 42 :51–66.
- Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publication de l'Institut de Statistique de l'Université de Paris*, 23(3-4) :1–36.
- Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés - Un test non paramétrique d'indépendance. *Académie Royale de Belgique - Bulletin de la Classe des Sciences - série 5*, 65 :274–292.
-

-
- Deheuvels, P. (1981a). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.*, 11(1) :102–113.
- Deheuvels, P. (1981b). A Kolmogorov-Smirnov type test for independence and multivariate samples. *Rev. Roumaine Math. Pures Appl.*, 26(2) :213–226.
- Deheuvels, P., Peccati, G., and Yor, M. (2006). On quadratic functionals of the Brownian sheet and related processes. *Stochastic Process. Appl.*, 116(3) :493–538.
- DeMasi, R. (1994). *Proportional Hazards Models for Multivariate Failure Time Data with Generalized Competing Risks*. PhD thesis, University of North Carolina.
- Efron, B. (1977). The efficiency of Cox’s likelihood function for censored data. *J. Amer. Statist. Assoc.*, 72(359) :557–565.
- Endsley, M. (1996). Towards a theory of situation awareness in dynamic systems. *Human Factors*, NC :NC.
- Fancher, P. and Ervin, R. (1998). Adaptive Cruise Control Field Operational Test. Technical report, UMTRI.
- Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5) :847–860.
- Galambos, J. (1987). *The asymptotic theory of extreme order statistics*. Robert E. Krieger Publishing Co. Inc., Melbourne, FL, second edition.
- Genest, C., Molina, J. J. Q., and Lallena, J. A. R. (1995). De l’impossibilité de construire des lois à marges multidimensionnelles données à partir de copules. *Compte-Rendu de l’Académie des Sciences de Paris*, 1(320) :723–726.
- Greenberg, H. (1959). An analysis of traffic capacity. *Ops. Res.*, 7 :79–85.
- Greenshields, B. D. (1934). A study of traffic capacity. *Highway Res. Proc.*, 14 :448–474.
- Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(5) :835–844.
- Hoedemaeker, M. and Brookhuis, K. A. (1999). Driving with an adaptive cruise control (ACC). *Transportation Research Part F : Psychology and Behaviour*, 3 :95–106.
- Hougaard, P. (2000). *Analysis of multivariate survival data*. Statistics for Biology and Health. Springer-Verlag, New York.
- Hürlimann, W. (2003). Hutchinson-Lai’s conjecture for bivariate extreme value copulas. *Statist. Probab. Lett.*, 61(2) :191–198.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall.
- Knospe, W., Santen, L., Schadschneider, A., and Schreckenberg, M. (2002). A realistic two-lane traffic model for highway traffic.
-

- Kopf, M. (2002). Advanced Driver Assistance Systems Research within INVENT.
- Kuge, N., Yamamura, T., Shimoyama, O., and Liu, A. (2000). A driver behavior recognition method based on a driver model framework. *SAE Technical Papers*, 01-0349 :NC.
- Larraga, M. E. (2002). Cellular Automata for One-Lane Traffic Flow Modelling : Safety and Automation. Traffic Forum.
- Leroux, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.*, 40(1) :127–143.
- Li, H., Scarsini, M., and Shaked, M. (1996). Linkages : A tool for the Construction of Multivariate distributions with given nonoverlapping multivariate marginals. *Journal of Multivariate Analysis*, 56 :20–41.
- Monseur, M. and Malaterre, G. (1970). Prise de risque en automobile : étude expérimentale d'un parcours routier. *Cahier d'études ONSER*, 27 :NC.
- Nagel, K. and Schreckenberg, M. (1992). Cellular automaton models for freeway traffic. *J. Physique I*, 2 :2221–2229.
- Nelsen, R. B. (1996). *Distributions with Fixed Marginals and Related Topics*, volume 28, chapter Nonparametric measures of multivariate association, pages 223–232. IMS Lecture Notes.
- Nelsen, R. B. (1998). *An Introduction to Copulas*, volume 139 of *Lectures Notes in Statistics*. Springer-Verlag.
- Nilsson, L. (1995). Safety effects of adaptive cruise control in critical traffic situations. In *Proceedings of the Second World Congress on ITS, Yokohama, Japon*.
- OECD (1990). Behavioural adaptations to changes in the road transport system. Technical report, Road Transport Research Organisation for Economic Cooperation and Development.
- Orselli, J. and Chanaron, J.-J. (1999). L'Automatisation progressive de la conduite : de la boîte manuelle vers "l'autoroute automatisée" – état de la recherche en octobre 1999. Technical report, Predit.
- Pickands, III, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, volume 49, pages 859–878, 894–902. With a discussion.
- Prestl, W., Sauer, T., Steinle, J., and Tschernoster, O. (2000). The BMW active cruise control ACC. Technical report, SAE Technical Papers.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. In *IEEE*, volume 77, pages 257–286.
-

-
- Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York.
- Richardson, J. H., Ward, N. J., Fairclough, S. H., and Graham, R. (1996). PROMETHEUS/DRIVE AICC Safety Assessment : Basic Simulator. Technical report, DoT Vehicle Standards and Engineering Division Contract DPU 9/81/1 , HUSAT Research Institute, UK.
- Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23 :470–472.
- Rüschendorf, L. and Valk, V. D. (1993). On regression representations of stochastic processes. *Stochastic Processes and Applications*, 46 :183–198.
- Rydén, T. (1995). Estimating the order of hidden Markov models. *Statistics*, 26(4) :345–354.
- Saad, F. and Schnetzler, B. (1994). Un modèle conceptuel du comportement des conducteurs en conduite en file sur autoroute. Technical report, INRETS/DRAST.
- Saad, F. and Villame, T. (1996). Assessing New Driving Support Systems : Contributions of an Analysis of Drivers' Activity in Real Situations. In *Third Annual World Congress on Intelligent Transport Systems*.
- Saad, F. and Villame, T. (1999). Intégration d'un nouveau système d'assistance dans l'activité des conducteurs d'automobile. In *Sécurité et cognition*, volume 6, pages 105–114. Hermes.
- Saulis, L. and Statulevičius, V. A. (1991). *Limit theorems for large deviations*, volume 73 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht. Translated and revised from the 1989 Russian original.
- Shaked, M. and Shanthikumar, J. G. (1994). Multivariate Stochastic Orders. In *Stochastic Orders and their applications*, pages 113–151. Academic Press.
- Shaked, M. and Shanthikumar, J. G. (1999). Nouvelle version du chapitre 9 - Positive Dependence Orders. unpublished.
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231.
- Stephenson, A. (2003). Simulating multivariate extreme value distributions of logistic type. *Extremes*, 6(1) :49–59.
- Strat, Y. L. and Carrat, F. (1999). Monitoring Epidemiologic Surveillance Data using Hidden Markov Models. *Statistics in Medicine*, 18 :3463–3478.
- Tiago de Oliveira, J. (1997). *Statistical Analysis of Extremes*. Pendor.
-

- Vaart, A. W. V. D. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.
- West, R. J., Elander, J., and French, D. J. (1992). Decision making, personality and driving style as correlates of individual accident risk. Technical report, Report presented to Road Safety Division, Transport and Road Research Laboratory.
- Widder (1947). The Stieljes Integral. In *The Laplace Transform*. Princeton University.
- Ygnace, J.-L. and Benouar, H. (1998). Analyse prospective du concept d'autoroute automatique et les enjeux pour la sécurité, état de l'art aux USA et acceptabilité à la situation française. Technical report, INRETS/LESCOT.
- Zou, X. and Levinson, D. (2000). Simulation and Analysis of Mixed Adaptive Cruise Control Manual Traffic. Technical report, Department of Civil Engineering, University of Minnesota.
-