

# Copules et modèles avec variable de fragilité pour des données de survie multivariée

## Copulas and frailty models in multivariate survival data

Nesrine Idiou<sup>1</sup>, Fatah Benatia<sup>1</sup>, Mounir Mesbah<sup>2</sup>

<sup>1</sup>Laboratory LMA, Université de Biskra, Biskra, Algeria, nesrine.idiou@univ-biskra.dz, f.benatia@univ-biskra.dz,

<sup>2</sup>Laboratoire de Probabilités, Statistique et Modélisation (LPSM, UMR 8001), Université de Sorbonne, Paris 75005, France. mounir.mesbah@sorbonne-universite.fr

**RÉSUMÉ.** L'étude des copules et de leurs applications en statistique, en particulier en Biostatistique, est récente. Les copules jouent un rôle très important dans la modélisation de la structure de dépendance entre les lois marginales et la loi jointe d'un vecteur de variables aléatoires. Dans cet article, nous présentons une synthèse des travaux récents portant sur cette théorie et ses applications à l'analyse des données de survie multivariée. Enfin, à titre d'illustration, une application sur des données de survie bivariable issue de la littérature est présentée, et analysée par la procédure Proc Copula du logiciel SAS.

**ABSTRACT.** The study of copulas and their applications in statistics, and in particular in biostatistics, is recent. Copulas play a important role in the modeling of the dependence structure between the marginal distributions and the joint distribution of a vector random variable. In this article, we present a synthesis of recent works on this theory and its applications to the analysis of multivariate survival data. Finally, an application on bivariate survival data from the literature analyzed by the Proc Copula procedure of the SAS software, is given to illustrate a such approach.

**MOTS-CLÉS.** Copules, Analyse de survie, Modèle à fragilité, Dépendance, Copules Archimédiennes.

**KEYWORDS.** Copulas Survival Analysis, Frailty model, Dependence, Archimedian Copulas.

## Introduction

Récemment, une attention considérable a été accordée au problème de l'inférence sur les copules, le terme copule (copula) vient du mot latin "*copŭlae*", qui signifie liaison, lien ou union. Parmi les travaux statistiques les plus importants en théorie des copules, on peut citer ceux de (Hoeffding,1940,1941)[17] [18], qui a utilisé les copules pour étudier les mesures d'association nonparamétrique (rho de Spearman, voir section 2). Il a ainsi obtenu des inégalités optimales, fournissant des bornes supérieures et inférieures pour des cas particuliers de copules, citées dans le théorème dit Bornes de Fréchet-Hoeffding (1957,[12]). Les monographies de Deheuvels en (1979)[9], Cook et Johnson (1981)[7], Cherubini et al.(2004)[5], Nelsen(2006)[28], Joe(1997)[20] et Genest(1993)[14] résument dans une certaine mesure les résultats dans ce domaine.

En gros, une fonction copule est une fonction qui joint ou couple une fonction de distribution multivariée à ses fonctions de distributions marginales univariées, ce qui est traduit par Sklar(1959)[30] dans un théorème qui porte son nom. Il montre que, sous certaines conditions, il existe une fonction de copule unique  $C$  telle que :

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

Cette fonction joue un rôle très important dans la modélisation de la dépendance dans divers domaines de la statistique (finance, actuariat, et plus récemment biologie et santé, etc). Dans ce contexte et surtout dans le domaine de la santé pour l'étude des liens entre les dates d'occurrence d'une maladie, ses possibles dates de guérison, de rechute, et de décès. Plusieurs modèles de survie multivariée prenant en compte

la dépendance entre des variables aléatoires se basent sur la notion de copules, souvent sans y faire explicitement référence. La nature de ces problèmes en analyse de survie, amènent à construire des familles de fonctions de survie multivariée à partir des fonctions de survie marginales (univariées).

L'objectif de cet article est de présenter l'approche "copules" dans la modélisation de la survie multivariée. Cette approche apparaît implicitement dans Clayton (1978)[6] qui est l'un des premiers à proposer un modèle d'association bivariée pour l'analyse de survie, et ensuite dans Marshall and Olkin [1988][26].

Dans le contexte épidémiologique, les modèles avec fragilité sont des modèles faisant intervenir un paramètre individuel. Ces modèles épidémiologiques supposent, qu'un sujet peut être plus fragile qu'un autre et donc courir un risque plus grand de mort ou de la survenue d'un autre événement pathologique. Ces modèles avec fragilité introduits par Lancaster (1979)[22] utilisent un modèle de risque proportionnel. Vaupel et al (1979) [33] ont suggéré l'application du modèle gamma, allié à un autre modèle de fragilité. Les distributions gamma ont été utilisées en raison de leur attractivité calculatoire. Elles sont bien connues et de densités simples. L'application de ces modèles dans la survie est contemporaine (Clayton, 1978)[6]. L'idée principale est d'introduire une dépendance entre les d-temps de survie  $T_1, \dots, T_d$ , en utilisant une variable aléatoire non observée  $W$ .

L'article est organisé de la manière suivante. Dans la première section, nous présentons la définition des copules ainsi que leurs propriétés élémentaires. Dans la deuxième section nous présentons les mesures d'association entre deux variables aléatoires les plus utilisées ainsi que leur expression en terme de copules bivariées. Ensuite, dans la troisième section nous présentons le cas des familles de copules paramétriques, en détaillant les plus utilisées, ainsi que leurs propriétés fondamentales. La quatrième section est consacrée à une présentation succincte des fonctions de survie univariées et multivariées, et des copules qui leur sont spécifiques. Dans la cinquième section, nous considérons l'approche des modèles de survie multivariée avec variable de fragilité, et présentons les copules associées à ce type de modèles. Dans la sixième section, nous utilisons la procédure Proc Copula du logiciel SAS pour analyser des données réelles de durées récurrentes en hémodialyse.

## 1. Les copules

L'objectif essentiel de cette partie est de présenter la définition des copules ainsi que leurs propriétés élémentaires. Pour faciliter la tâche, nous nous limiterons ensuite à l'étude des copules bivariées.

Soit  $(X_1, \dots, X_d)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ , de fonction de répartition jointe  $F$  et de marginales  $F_i$  pour  $i = 1, \dots, d$ . Sklar en 1959[30] montre que, sous certaines conditions (entre autres,  $F$  continue), il existe une fonction unique notée  $C$  (pour Copule), telle que

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Cette représentation traduit les relations intrinsèques entre les composantes en réduisant/normalisant toutes les variables à des variables uniformes. Considérons un couple de variables aléatoires  $(X_1, X_2)$  de fonctions de distribution marginales  $F_1(x_1) = P(X_1 \leq x_1)$  et  $F_2(x_2) = P(X_2 \leq x_2)$  respectives et une distribution jointe  $F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$ .

Pour chaque couple  $(x_1, x_2)$  on peut associer trois nombres  $F_1(x_1)$ ,  $F_2(x_2)$  et  $F(x_1, x_2)$  appartenant à

l'intervalle  $[0, 1]$ . En d'autres termes pour chaque paire de nombres réels  $(x_1, x_2)$  correspond un point  $(F_1(x_1), F_2(x_2))$  dans le carré unité  $[0, 1] \times [0, 1]$  et à son tour ce point fait correspondre un nombre réel  $F(x_1, x_2)$  appartenant à l'intervalle  $[0, 1]$ . Nous allons alors montrer que cette correspondance qui attribue la valeur de la distribution à chaque couple de fonctions de distributions marginales est en effet une fonction (appelée copule).

Dans tout ce qui suit  $I$  désignera l'intervalle unité  $I = [0, 1]$ .

**DEFINITION 1**– Une copule bidimensionnelle ou 2-copule ou simplement copule bivariée  $C$  est une fonction de  $I^2 = [0, 1]^2$  dans  $I$  possédant les propriétés suivantes :

1. Pour tout  $(u, v) \in I^2$  on a :

$$C(u, 0) = C(0, v) = 0, \quad C(u, 1) = u \text{ et } C(1, v) = v. \quad [1]$$

2.  $C$  est 2-croissante c'est à dire :  $\forall (u_1, u_2, v_1, v_2) \in I^4$  tels que  $u_1 \leq u_2$ , et  $v_1 \leq v_2$ ,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0, \quad [2]$$

Ainsi, toute copule est une répartition sur  $I^2$  dont les répartitions marginales sont de loi uniforme sur  $I = [0, 1]$ .

### 1.1. Propriétés élémentaires des copules bivariées

Soit  $C_1$  et  $C_2$  deux copules.

#### — Ordre

**DEFINITION 2**– On dit que la copule  $C_1$  est plus petite que la copule  $C_2$ , (ou bien que  $C_2$  est plus grande que  $C_1$ ) et on note  $C_1 < C_2$  si on a :

$$C_1(u, v) \leq C_2(u, v) \text{ pour tout } (u, v) \in [0, 1]^2.$$

#### — Convexité et concavité

**DEFINITION 3**– Une copule est dite concave si on a :

$$C(\alpha a + (1 - \alpha)c, \alpha b + (1 - \alpha)d) \geq \alpha C(a, b) + (1 - \alpha)C(c, d),$$

pour tout  $\alpha, a, b, c, d$  dans  $[0, 1]$ . Elle est dite convexe si on a :

$$C(\alpha a + (1 - \alpha)c, \alpha b + (1 - \alpha)d) \leq \alpha C(a, b) + (1 - \alpha)C(c, d),$$

pour tout  $\alpha, a, b, c, d$  dans  $[0, 1]$ .

#### — Dérivées partielles

**PROPOSITION 1**– Les dérivées partielles d'une copule  $C(u, v)$  existent presque partout pour tout  $u, v \in [0, 1]$ , et vérifient :

$$0 \leq \partial_u C(u, v) \leq 1 \text{ et } 0 \leq \partial_v C(u, v) \leq 1.$$

## — Densité d'une copule

Soient  $X_1, X_2$  deux variables aléatoires continues, on note par  $f$  la fonction de densité jointe associée à leur distribution jointe  $F$ , et par  $f_1$  et  $f_2$  les fonctions de densités marginales de  $X_1, X_2$  respectivement associées à  $F_1$  et  $F_2$ , leurs fonctions de distribution marginales respectives.

PROPOSITION 2— La densité  $c(F_1(x_1), F_2(x_2))$  associée à la copule  $C(F_1(x_1), F_2(x_2))$  est définie par

$$c(F_1(x_1), F_2(x_2)) = \frac{\partial_{u,v}^2 C(F_1(x_1), F_2(x_2))}{F_1'(x_1)F_2'(x_2)} = \frac{f(x_1, x_2)}{f_1(x_1)f_2(x_2)}.$$

### 1.2. Le théorème de Sklar et ses applications

Le résultat fondamental de la théorie des copules est le théorème de Sklar ([30]). Il établit le lien entre la définition formelle ci-dessus des copules et les distributions des variables aléatoires, permettant ainsi l'application des copules dans la modélisation statistique.

THEOREM 1 (SKLAR, 1959)— Soit  $F$  une distribution bivariée dont les distributions marginales sont  $F_1$  et  $F_2$ , alors il existe une copule bivariée  $C$  telle que :

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \forall (x_1, x_2) \in \mathbb{R}^2. \quad [3]$$

1. Si les distributions marginales  $F_1$  et  $F_2$  sont continues, alors  $C$  est unique.
2. Réciproquement si  $C$  est une copule et  $F_1$  et  $F_2$  des distributions univariées, alors la fonction  $F$  définie par (3) est une distribution conjointe dont les marges sont  $F_1$  et  $F_2$ .

Pour la démonstration de ce théorème, voir Nelsen, [28]. Grâce au théorème de Sklar, on définit les copules à partir d'un couple de variables aléatoires de la façon suivante :

DEFINITION 4— Soient  $F_1, F_2$  les fonctions de répartition de deux variables aléatoires  $X_1$  et  $X_2$  respectivement, et  $C$  une copule bivariée définie sur  $[0, 1]^2$ , alors  $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$  est une fonction de répartition jointe de  $X_1, X_2$  dont les marginales sont  $F_1$  et  $F_2$ . La copule  $C$  est dite copule de  $X_1, X_2$  et sera notée  $C_{X_1, X_2}$ .

DEFINITION 5— Soit  $F$  une fonction de répartition univariée. Le quantile de  $F$  (l'inverse généralisée) est une fonction, notée  $F^{-1}$  de domaine de définition  $[0, 1]$  telle que :

$$F^{-1}(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\}$$

PROPOSITION 3— Soit  $X$  une variable aléatoire de répartition  $F$  :

1. Si  $U$  est une variable aléatoire uniforme sur  $[0, 1]$ , alors  $F^{-1}(U) \sim F$ .
2. Si  $F$  est continue, alors  $F(X) \sim U$ .

COROLLARY 1— Soit  $F$  une fonction de répartition bidimensionnelle de fonctions de répartition marginales  $F_1$  et  $F_2$ . Alors la copule  $C$  associée à  $F$ , en utilisant la proposition 3, est donnée par :

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)) \text{ pour tout } (u, v) \in [0, 1]^2.$$

### 1.3. Les bornes de Fréchet-Hoeffding

Toute copule  $C$  admet une borne inférieure et une borne supérieure déterminées par le théorème suivant :

**THEOREM 2 (FRECHET-HOEFFDING)**– Soit  $C$  une copule, alors :  $\forall u, v \in [0, 1]$ ,

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq M(u, v) = \min(u, v),$$

$W$  et  $M$  sont les bornes inférieure et supérieure de Fréchet-Hoeffding.

De l'utilisation du théorème de Sklar pour deux variables aléatoires  $X_1, X_2$  de fonction de répartition jointe  $F$  et de marginales  $F_1$  et  $F_2$  respectivement, et du théorème de Fréchet-Hoeffding, on a pour tout  $x_1$  et  $x_2 \in \overline{\mathbb{R}}$ ,

$$\max(F_1(x_1) + F_2(x_2) - 1, 0) \leq F(x_1, x_2) \leq \min(F_1(x_1), F_2(x_2)). \quad [4]$$

$M$  et  $W$  sont en fait des copules, car les bornes sus-citées sont elles-mêmes des fonctions de répartition jointes. Elles sont appelées bornes de Fréchet-Hoeffding pour une fonction de répartition jointe  $F$  et des marginales  $F_1$  et  $F_2$ .

### 1.4. Copules multivariées

Dans la section précédente on a présenté les copules de dimension 2. Dans cette section on va discuter le cas multivarié plus général, c'est à dire les copules de dimension  $d > 2$ . La généralisation des définitions et théorèmes se fait de la même manière que le cas bivarié.

**DEFINITION 6**– Une copule  $d$ -dimensionnelle est une fonction  $C$  de  $[0, 1]^d$  dans  $[0, 1]$  ayant les propriétés suivantes :

- Si  $u \in [0, 1]^d$ , a au moins une coordonnée  $u_i = 0, i = 1, \dots, d$ , alors  $C(u) = 0$ .
- Si les coordonnées sont toutes égales à 1 sauf la coordonnée  $u_i$ , alors  $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ .

#### **Théorème de Sklar multivarié**

Ce théorème a la même importance et le même rôle que dans le cas bivarié.

**THEOREM 3**– Soit  $F$  une fonction de répartition  $d$ -dimensionnelle de fonctions de répartition marginales  $F_1, \dots, F_d$ . Alors, il existe une  $d$ -copule  $C$  telle que pour tout  $x \in \mathbb{R}_d$ ,

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad [5]$$

1. Si les fonctions  $F_1, \dots, F_d$  sont continues, alors  $C$  est unique.
2. Inversement, si  $C$  est une  $d$ -copule et si  $F_1, \dots, F_d$  sont des fonctions de répartition univariées, alors la fonction  $F$  définie (5) est une fonction de répartition  $d$ -dimensionnelle de marginales  $F_1, \dots, F_d$ .

**COROLLARY 2**– Soit  $F$  une fonction de répartition  $d$ -dimensionnelle de fonctions de répartition marginales  $F_1, \dots, F_d$ . Alors, la copule  $C$  associée à  $F$  est donnée par :

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \text{ pour tout } u \in [0, 1]^d$$

### 1.4.1. Propriétés élémentaires

#### THEOREM 4– *Continuité uniforme*

Une copule  $C$  est uniformément continue sur son domaine. En particulier, pour tout  $u, v$  dans  $[0, 1]^d$ , nous avons :

$$|C(u) - C(v)| \leq \sum_{i=1}^n |v_i - u_i|.$$

#### THEOREM 5– *Dérivées partielles*

Soit  $C$  une copule. Les dérivées partielles de  $C$  existent presque partout (p.p.), pour tout  $i = 1, \dots, d$  et tout  $u \in [0, 1]^d$ , nous avons :

$$0 \leq \partial_{u_i} C(u) \leq 1, \quad u = (u_1, \dots, u_d).$$

De plus, les fonctions  $u \rightarrow \partial_{u_i} C(u), i = 1, \dots, d$  sont non décroissantes p.p.

#### THEOREM 6– *Densité d'une copule multivariée*

Si  $f$  et  $f_i$  sont les densités associées à  $F$  et  $F_i$  respectivement, alors la densité  $c$  associée à la copule  $C$  est définie par :

$$c(F_1(x_1), \dots, F_d(x_d)) = \frac{\partial_{u_1, \dots, u_d}^d C(F_1(x_1), \dots, F_d(x_d))}{F_1'(x_1) \dots F_d'(x_d)},$$

*c'est à dire :*

$$c(u_1, \dots, u_d) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{\prod_{i=1}^d f_i(F_i^{-1}(u_i))}.$$

Si la fonction de répartition multivariée  $F$  est absolument continue, grâce au théorème de Sklar, on peut exprimer sa fonction de densité  $f$  en fonction de  $c$  et des densités marginales  $f_1, \dots, f_d$  par :

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i). \quad [6]$$

## 2. Mesures de dépendance

La notion de dépendance a été largement abordée par de nombreux auteurs en statistique et en probabilité. L'étude de la dépendance entre deux ou plusieurs variables sert à évaluer la liaison qui peut exister entre ces variables. La manière traditionnelle d'évaluer la dépendance entre deux ou plusieurs variables est connue par l'utilisation du coefficient de corrélation linéaire. Cette mesure de dépendance est efficace lorsque la dimension est petite mais elle peut être trompeuse quand la dimension augmente. Trois mesures de dépendance sont généralement utilisées dans la littérature. Nous les présentons ci après.

## 2.1. Le coefficient de corrélation linéaire de Pearson

Introduit en 1896 par Pearson ([29]), le coefficient de corrélation linéaire est un indicateur facile à calculer. Il mesure la relation linéaire entre les valeurs de deux variables aléatoires réelles  $X$  et  $Y$  ayant chacune une variance finie. Il est noté généralement par  $\rho_P(X, Y)$  ou parfois par  $Cor_P(X, Y)$  et donné par :

$$\rho_P(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

où  $Cov(X, Y)$ , désigne la covariance entre les variables aléatoires  $X$  et  $Y$ ,  $Var(X)$  et  $Var(Y)$  désignent les variances respectives de  $X$  et  $Y$ .

- Si cet indicateur existe, il est toujours compris entre -1 et 1.
- Une valeur de -1 ou 1 indique une dépendance linéaire parfaite, négative ou positive entre  $X$  et  $Y$ .

La corrélation de Pearson peut être exprimée en fonction des copules comme l'indique la proposition suivante :

**PROPOSITION 4**— Si  $C$  désigne la copule des variables  $X$  et  $Y$ , la corrélation de Pearson  $\rho_P$  a pour expression :

$$\rho_P(X, Y) = \frac{1}{\sqrt{Var(X)Var(Y)}} \int_0^1 \int_0^1 (C(u, v) - uv) dF^{-1}(u) dG^{-1}(v),$$

où  $F$  et  $G$  sont les fonctions de répartition respectives des variables  $X$  et  $Y$ .

Si  $H$  est la distribution jointe de  $X$  et  $Y$ , nous aurons :

$$Cov(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (H(x, y) - F(x)G(y)) dx dy.$$

Il suffira d'appliquer le changement de variables,  $u = F(x)$  et  $v = G(y)$  pour démontrer la proposition. On remarque que la copule ne permet pas de déduire la valeur du coefficient de corrélation. A cet effet, les distributions marginales sont nécessaires.

Si  $(x_1, y_1), \dots, (x_n, y_n)$  constitue un ensemble d'observations des variables jointes  $X$  et  $Y$ . L'estimation de  $\rho_P(X, Y)$  est donnée par le coefficient de corrélation empirique :

$$r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \text{ avec } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ et } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

## 2.2. Mesures de concordance

Pour introduire le tau de Kendall et le rho de Spearman, on considère d'abord la notion de concordance présentée dans la définition suivante.

DEFINITION 7– a) *Réalisations concordantes* : deux réalisations  $(x, y)$  et  $(x', y')$  d'un vecteur aléatoire continu  $(X, Y)$ , sont dites concordantes si  $(x' - x)(y' - y) > 0$

b) *Réalisations discordantes* : deux réalisations  $(x, y)$  et  $(x', y')$  d'un vecteur aléatoire continu  $(X, Y)$ , sont dites discordantes si  $(x' - x)(y' - y) < 0$ .

c) Une fonction  $\kappa$  d'un couple de variables aléatoires continues est une mesure de concordance si elle vérifie les propriétés suivantes :

- Si  $Y$  est une fonction croissante de  $X$ , alors  $\kappa(X, Y) = 1$ ;
- Si  $Y$  est une fonction décroissante de  $X$ , alors  $\kappa(X, Y) = -1$ ;
- Si  $\alpha$  et  $\beta$  sont des fonctions strictement croissantes,  $\kappa(\alpha(X), \beta(Y)) = \kappa(X, Y)$ .

Commençons par introduire le  $\tau$  de Kendall qui est une mesure de concordance, ainsi que ses importantes propriétés.

### 2.2.1. Le tau de Kendall

Développé par Kendall (1938), le tau de Kendall ([21]) mesure l'association entre deux variables  $X$  et  $Y$ .

DEFINITION 8– Soit  $(X, Y)$  et  $(X', Y')$  deux vecteurs aléatoires indépendants, et de même loi  $F$ . Le  $\tau$  de Kendall de  $(X, Y)$  est défini par :

$$\tau(X, Y) = Pr[(X - X')(Y - Y') \geq 0] - Pr[(X - X')(Y - Y') < 0]$$

Le tau de Kendall n'est autre que la différence entre la probabilité de concordance et la probabilité de discordance. C'est une mesure de concordance : en effet, il vérifie les propriétés suivantes :

- Si  $Y$  est une fonction croissante de  $X$  (et donc  $X$  une fonction croissante de  $Y$ ), alors  $\kappa(X, Y) = 1$ ;
- Si  $Y$  est une fonction décroissante de  $X$  (et donc  $X$  une fonction décroissante de  $Y$ ), alors  $\kappa(X, Y) = -1$ ;
- Si  $\alpha$  et  $\beta$  sont des fonctions strictement croissantes, alors  $\kappa(\alpha(X), \beta(Y)) = \kappa(X, Y)$ .

De plus, il vérifie aussi :

- Il est symétrique :  $\tau(X, Y) = \tau(Y, X)$ ;
- La valeur de  $\tau$  est comprise entre -1 et 1 ;
- Si  $X$  et  $Y$  sont indépendantes et continues, la valeur de tau est nulle (la réciproque n'est pas forcément vraie).

Le tau de Kendall, lui aussi, peut être déduit à partir de la copule, comme le montre la proposition suivante.



PROPOSITION 5– Soit  $(X, Y)$  un vecteur aléatoire continu de copule  $C$ . Le  $\tau$  de Kendall  $\tau(X, Y)$  a pour expression :

$$\begin{aligned}\tau(X, Y) &= 4 \int_0^1 \int_0^1 C(u, v)c(u, v)dudv - 1 \\ &= 4E(C(U, V)) - 1,\end{aligned}$$

avec  $c(u, v)$ , la densité associée à la copule  $C(u, v)$  (voir le Théorème 2.15).

Cette définition du tau de Kendall est théorique. Si l'on dispose d'un échantillon  $(x_1, y_1), \dots, (x_n, y_n)$  des variables jointes  $X$  et  $Y$ , tel que les valeurs des  $(x_i)$  et  $(y_i)$  soient distinctes, en notant  $n_c$ , le nombre de paires concordantes et  $n_d$ , le nombre de paires discordantes, l'estimation statistique de  $\tau$  est donnée par :

$$\hat{\tau} = \frac{2(n_c - n_d)}{n(n - 1)}$$

### 2.2.2. Le rho de Spearman

L'autre mesure de concordance, très utilisée en pratique, est le rho de Spearman, introduit en 1904, par Spearman ([32]). Nous donnons sa définition et quelques propriétés de cette mesure.

DEFINITION 9– *le rho de Spearman ou corrélation de Spearman, est une mesure de dépendance statistique non paramétrique entre deux variables  $X$  et  $Y$ . Il est égal à :*

$$\rho_S(X, Y) = 3(Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - Pr\{(X_1 - X_2)(Y_1 - Y_2) > 0\}),$$

où  $(X_1, Y_1), (X_2, Y_2)$  et  $(X_3, Y_3)$  sont des réalisations indépendantes du vecteur aléatoire  $(X, Y)$ .

La proposition suivante permet d'exprimer le rho de Spearman en fonction du coefficient de corrélation entre deux variables uniformes.

PROPOSITION 6– *Notons  $F$  et  $G$  les fonctions de répartition respectives des variables aléatoires  $X$  et  $Y$ . Le rho de Spearman de deux variables  $X$  et  $Y$  est égal au coefficient de corrélation entre les variables  $F(X)$  et  $G(Y)$ .*

$$\rho_S(X, Y) = \rho_P(F(X), G(Y)).$$

De plus,

- $\rho_S$  est compris entre -1 et 1 et contrairement au coefficient de corrélation de Pearson, il est toujours défini.
- Si les variables aléatoires  $X$  et  $Y$  sont indépendantes, alors  $\rho_S(X, Y) = 0$ .

Le rho de Spearman peut aussi être exprimé en fonction de la copule comme le montre la proposition suivante.

PROPOSITION 7– Soit  $(X, Y)$  un vecteur aléatoire continu de copule  $C$ . Le coefficient de spearman  $\rho_S(X, Y)$  a pour expression :

$$\rho_S(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3$$

Soit  $(x_1, y_1), \dots, (x_n, y_n)$  un ensemble d'observations des variables jointes  $X$  et  $Y$ . Si ces observations ne contiennent pas d'ex aequos (les rangs sont tous différents), l'estimation de  $\rho_S(X, Y)$  est donnée par la formule ci dessous :

$$r_S(X, Y) = 1 - \frac{6 \sum_{i=1}^n (d_i)}{n(n^2 - 1)}$$

avec  $d_i = \text{rang}(x_i) - \text{rang}(y_i)$ .

On voit par cette définition, que  $r_S$  n'est rien d'autre que le coefficient de corrélation des **rangs** des variables.

Ces trois indicateurs permettent de mesurer la dépendance dans un modèle donné.

REMARK 1– Contrairement au coefficient de corrélation de Pearson, le tau de Kendall et le rho de Spearman sont complètement caractérisés pas la copule et les distributions marginales sont inutiles.

### 3. Familles paramétriques de copules

Nous présentons dans cette section quelques exemples de copules paramétriques bivariées et multivariées.

#### 3.1. Copules usuelles

##### 3.1.1. Copule d'indépendance

Soient  $X, Y$  deux variables aléatoires continues de lois marginales respectives  $F_1$  et  $F_2$  et de fonction de répartition jointe  $F$ .

PROPOSITION 8– Si  $X_1, X_2$  sont deux variables aléatoires indépendantes, alors leur copule associée est le produit des marginales :  $C_{X_1, X_2}(x_1, x_2) = F_1(x_1).F_2(x_2)$ .

De même, dans le cas multivarié, la copule d'indépendance est donnée par :

$$C_{X_1, X_2, \dots, X_d}(x_1, x_2, \dots, x_d) = F_1(x_1).F_2(x_2) \times \dots \times F_d(x_d)$$

### 3.1.2. Copule Gaussienne

Si on note par  $\Phi_\rho$  la fonction de répartition jointe de la loi normale bivariée de marginales standards et de coefficient de corrélation linéaire  $\rho \in [0, 1]$ , et par  $\Phi$  la fonction de répartition de la loi normale standard univariée, la copule gaussienne du couple aléatoire  $X_1, X_2$  est définie par :

$$C_{\Phi_\rho}(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)).$$

$$\text{où : } \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{2x_1x_2\rho^2 - x_1^2 - x_2^2}{1-\rho^2}\right) dx_1 dx_2.$$

Cette copule est paramétrée par le coefficient de corrélation linéaire  $\rho$  et on a :

$$- C_{\Phi_\rho}(u, v) \longrightarrow W(u, v) \text{ quand } \rho \longrightarrow -1.$$

$$- C_{\Phi_\rho}(u, v) \longrightarrow M(u, v) \text{ quand } \rho \longrightarrow +1.$$

Dans le cas multivarié, en utilisant le théorème 13 (Sklar multivarié) et en notant  $\Phi_R$  la distribution normale standard multivariée, de matrice de corrélation  $R$ , de dimension  $d \times d$ , la copule gaussienne sera donnée par :

$$C_{\Phi_R}(u) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

où  $\Phi^{-1}$  est l'inverse de la fonction de répartition standard normale univariée  $\Phi$ . D'après (5) nous avons :

$$\frac{1}{(2\pi)^{\frac{d}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} X^t R^{-1} X\right) = c_{\Phi_R}(\Phi(x_1), \dots, \Phi(x_d)) \times \prod_{j=1}^d \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x_j^2\right)\right).$$

### 3.1.3. Copule de Student

La copule de Student est une copule bivariée paramétrée par le coefficient de corrélation linéaire  $\rho$  et un degré de liberté  $p$  d'une répartition de Student  $t_p$ . Cette copule est définie par :

$$\begin{aligned} C_{\rho,p}^t(u, v) &= t_{\rho,p}(t_p^{-1}(u), t_p^{-1}(v)) \\ &= \int_{-\infty}^{t_p^{-1}(u)} \int_{-\infty}^{t_p^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2+t^2-2\rho st}{p(1-\rho^2)}\right)^{-\frac{p+2}{2}} ds dt, \end{aligned}$$

avec  $t_p^{-1}$  est l'inverse généralisé, ici exacte, de la fonction de répartition de la loi de Student univariée à  $p$  degré de liberté. La densité correspondante est alors :

$$c_{\rho,p}^t(u, v) = \rho^{-\frac{1}{2}} \frac{\Gamma(\frac{p+2}{2})\Gamma(\frac{p}{2})}{\Gamma(\frac{p+1}{2})^2} \frac{\left(\frac{1+(t_p^{-1}(u))^2+(t_p^{-1}(v))^2-2\rho(t_p^{-1}(u))(t_p^{-1}(v)))}{v(1-\rho^2)}\right)^{-\frac{p+2}{2}}}{(1+(t_p^{-1}(u))^2)^{-\frac{p+2}{2}}(1+(t_p^{-1}(v))^2)^{-\frac{p+2}{2}}}.$$

Si  $T_{R,p}^d$  est la fonction de répartition de la loi de student multivariée, et  $R$  une matrice de corrélation, alors la copule de Student dans le cas multivariée est définie par :

$$C_{p,R}^{d,T}(u) = \int_{-\infty}^{t_p^{-1}(u_1)} \dots \int_{-\infty}^{t_p^{-1}(u_d)} \frac{\Gamma(\frac{p+d}{2})}{\Gamma(\frac{p}{2})(2\pi)^{\frac{d}{2}} |R|^{\frac{1}{2}}} \left(1 + \frac{(x-\mu)^t R^{-1} (x-\mu)}{v}\right)^{-\frac{p+d}{2}} dx,$$

Ainsi la densité de  $C_{p,R}^{d,T}(u)$  est donnée par :

$$c_{p,R}^{d,T}(u) = |R|^{-\frac{1}{2}} \frac{\Gamma(\frac{p+d}{2})}{\Gamma(\frac{p}{2})} \left( \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p+1}{2})} \right)^d \left( \frac{\left(1 + \frac{1}{p} \zeta^t R^{-1} \zeta\right)^{-\frac{p+d}{2}}}{\prod_{j=1}^d \left(1 + \frac{\zeta_j^2}{p}\right)^{-\frac{p+1}{2}}} \right),$$

où  $\zeta_j^2 = t_p^{-1}(u_j)$ .

REMARK 2– a) La copule gaussienne et la copule de Student appartiennent à une famille plus large de copules dites elliptiques.

b) Si le degré de liberté  $p \rightarrow \infty$ , la copule de Student converge vers la copule gaussienne et il est très difficile en pratique de différencier ces deux copules.

### 3.2. Copules archimédiennes

#### 3.2.1. Cas bivarié

DEFINITION 10– Une copule est dite archimédienne si elle est définie par :

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

où  $\varphi : [0, 1] \rightarrow \mathbb{R}_+$ , appelée générateur, est une fonction convexe et décroissante.

Cette copule possède les propriétés suivantes :

1. **Symétrie** :  $C(u, v) = C(v, u), \forall (u, v) \in [0, 1]^2$ .
2. **Associativité** :  $C(C(u, v), z) = C(u, C(v, z)), \forall (u, v, z) \in [0, 1]^3$ .
3. **Densité** :  $c(u, v) = -\frac{\varphi''(C(u, v))\varphi'(u)\varphi'(v)}{(\varphi'(C(u, v)))^3}$
4. **Invariance par homogénéité** : Pour tout  $\alpha > 0$ ,  $\alpha\varphi$  est aussi un générateur de la copule  $C$ .

De cette famille de copules on peut citer les plus connues et utilisées à ce jour, les copules de Gumbel, de Clayton et de Frank.

#### (a) Copule de Gumbel (1960) :

La copule de Gumbel, définie par le générateur  $\varphi_\alpha(t) = (-\ln t)^\alpha$ , de paramètre  $\alpha \geq 1$ , s'écrit :

$$C_\alpha^{Gum}(u, v) = \exp(-((-\ln u)^\alpha + (-\ln v)^\alpha)^{\frac{1}{\alpha}}).$$

#### (b) Copule de Clayton (1978) :

Cette copule de Cook et Johnson (1981), est définie par :

$$C_{\alpha}^{Cl}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}}$$

avec le générateur  $\varphi_{\alpha}(t) = \frac{1}{\alpha}(t^{-\alpha} - 1)$  de pseudo inverse  $\varphi_{\alpha}^{-1}(t) = (\alpha t + 1)^{-\frac{1}{\alpha}}$  et de paramètre de dépendance  $\alpha \in [-1, 0[ \cup ]0, +\infty[$ .

- Si le paramètre  $\alpha \rightarrow 0$ , alors on aura une copule d'indépendance.
- Si le paramètre  $\alpha \rightarrow \infty$ , alors on obtient la copule borne supérieur de Fréchet-Hoffding  $M$ .

**(c) Copule de Frank (1978) :**

Cette copule, de paramètre de dépendance  $\alpha \in [-\infty, 0[ \cup ]0, +\infty[$ , de générateur  $\varphi_{\alpha}(t) = -\ln\left(\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1}\right)$ , est définie par :

$$C_{\alpha}^{Fr}(u, v) = -\frac{1}{\alpha} \ln\left(1 + \frac{(e^{-u\alpha} - 1)(e^{-v\alpha} - 1)}{e^{-\alpha} - 1}\right)$$

**3.2.2. Cas multivarié**

Comme le cas bivarié, les copules archimédiennes multivariées sont définies par un générateur  $\varphi(t)$ . La forme générale de cette famille de copules est donnée par :

$$C^{d,A}(u) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)).$$

Les trois exemples de copules archimédiennes données précédemment dans le cas de dimension 2, sont généralisées au cas multivarié :

**(a) Copule de Gumbel :**

$$C_{\alpha}^{d,Gum}(u) = \exp\left\{-\left[\sum_{i=1}^d (-\log(u_i))^{\alpha}\right]^{\frac{1}{\alpha}}\right\}, \alpha \geq 1.$$

**(b) Copule de Clayton :**

$$C_{\alpha}^{d,Clay}(u) = \left(1 - d + \sum_{i=1}^d u_i^{-\alpha}\right)^{-\frac{1}{\alpha}}, \alpha > 0.$$

**(c) Copule de Frank :**

$$C_{\alpha}^{d,Fr}(u) = -\frac{1}{\alpha} \log\left\{1 + \frac{\prod_{i=1}^d (\exp(-\alpha u_i) - 1)}{(\exp(-\alpha) - 1)^{d-1}}\right\}, \alpha > 0.$$

## 4. Données de survie et copules

L'étude des données de survie est une branche de la statistique qui s'occupe de la modélisation des durées de "vie". Il s'agit en général du temps écoulé entre une date origine et une date de survenue d'évènements qui correspondent généralement à l'apparition d'une maladie, décès, rechute, etc. La probabilité qu'un individu soit vivant ou indemne au-delà du temps  $t$  est donnée par la fonction de survie. Lorsqu'il s'agit simultanément de plusieurs évènements, on parle de survie multivariée. Une distribution multivariée peut être construite grâce à l'utilisation de copules dans un cadre de survie. Nous considérons d'abord les données univariées séparément afin de caractériser les propriétés spécifiques des temps des fonctions de survie.

### 4.1. Modèles de survie

Soit  $T$  un temps de survie, c'est à dire une variable aléatoire positive ou nulle, (appelée aussi variable d'intérêt) de fonction de distribution  $F$ , et de densité  $f$ . La fonction de survie notée  $S$  à un instant  $t$ , est définie par :  $S(t) = P[T > t] = 1 - F(t)$ ,  $t \geq 0$ . Dans la modélisation de la fonction de survie, un concept fondamental est celui du taux de hasard ou de fonction de risque  $h(t)$  pour un instant  $t$  fixé, défini par

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P[t \leq T \leq t + \Delta | T \geq t]}{\Delta},$$

(Lancaster, 1990)[21]. Il est interprété comme le taux instantané de décès et s'exprime sous la forme suivante

$$h(t) = \frac{f(t)}{S(t)}.$$

La fonction du risque cumulé  $\Lambda(t)$ , (Andersen, Borgan, Gill et Keiding (1993))[3] ou fonction de hasard intégré (Hougaard [1999])[19], est l'intégrale :

$$\Lambda(t) = \int_0^t h(s) ds.$$

Le lien entre  $\Lambda$  et  $S$  est donné par :

$$S(t) = \exp(-\Lambda(t)).$$

Ce qui permet d'écrire :

$$f(t) = h(t) \exp(-\Lambda(t)).$$

Un autre concept important en modélisation de la survie est la fonction de risque de base (hasard baseline)  $h_0(t)$  (Frees et Valdez (1998),[11]). Elle intervient en particulier dans le modèle très utilisé dit de Cox (Cox [1972], [8]). Il consiste à modéliser la fonction de risque  $h(t)$ , de la façon suivante

$$h(t) = \exp(X\beta^T)h_0(t).$$

$X$  est un vecteur ligne de covariables, et  $\beta$  le transposé d'un vecteur de paramètres associés à ces covariables. Ce modèle de régression permet d'analyser les effets de covariables sur la distribution de la durée de vie. C'est un modèle à risques proportionnels. Il fait partie d'une famille plus large  $h(t) = g(X\beta^T)h_0(t)$ , avec  $g(\cdot)$ , une fonction quelconque, qui dans le modèle de Cox est la fonction exponentielle. L'un des intérêts de ce modèle est la facilité d'interprétation des paramètres. Prenons l'exemple d'une covariable  $X_j$  qui peut prendre deux valeurs : 0 si l'individu prend le traitement A et 1, s'il prend le traitement B. Le coefficient  $\beta_j$ , ou plutôt  $\exp(\beta_j)$  est le risque instantané relatif de décès du traitement B par rapport au traitement A de référence.

Dans le cas multivarié, la fonction de survie  $S(t)$  est définie par :

$$S(t_1, \dots, t_d) = Pr[T_1 > t_1, \dots, T_d > t_d],$$

où  $T_1, \dots, T_d$  sont d temps de survie. Les fonctions de survie marginales univariées, sont notées par

$$S_j(t_j) = Pr\{T_j > t_j\} = S(0, \dots, 0, t_j, 0, \dots, 0).$$

Dans cet article nous supposons que les temps de survie sont continus et prennent leur valeurs dans  $\mathbb{R}_+$ . Généralement, la fonction de répartition est définie par

$$F(t) = Pr\{T \leq t\},$$

ce qui explique la définition adoptée ici pour la fonction de survie  $S(t) = 1 - F(t) = Pr(T > t)$ . Néanmoins, nous pouvons tout aussi bien adopter la définition  $S(t) = Pr(T \geq t)$ . Notant que la relation entre la fonction de survie multivariée  $S$  et la fonction de répartition multivariée  $F$ , n'est pas aussi triviale que dans le cas univarié :

$$S(t_1, \dots, t_d) \neq 1 - F(t_1, \dots, t_d).$$

La fonction de risque et la fonction de risque cumulé multivariées sont données par :

$$\begin{aligned} h(t_1, \dots, t_d) &= \lim_{\max \Delta_j \rightarrow 0} \frac{P[t_j \leq T_j \leq t_j + \Delta_j, j = 1 \dots d | T_j \geq t_j, j = 1 \dots d]}{\Delta_1 \dots \Delta_d} \\ &= \frac{f(t_1, \dots, t_d)}{S(t_1, \dots, t_d)}, \end{aligned}$$

$$\Lambda(t_1, \dots, t_d) = \int_0^{t_1} \dots \int_0^{t_d} h(s_1, \dots, s_d) ds_1 \dots ds_d.$$

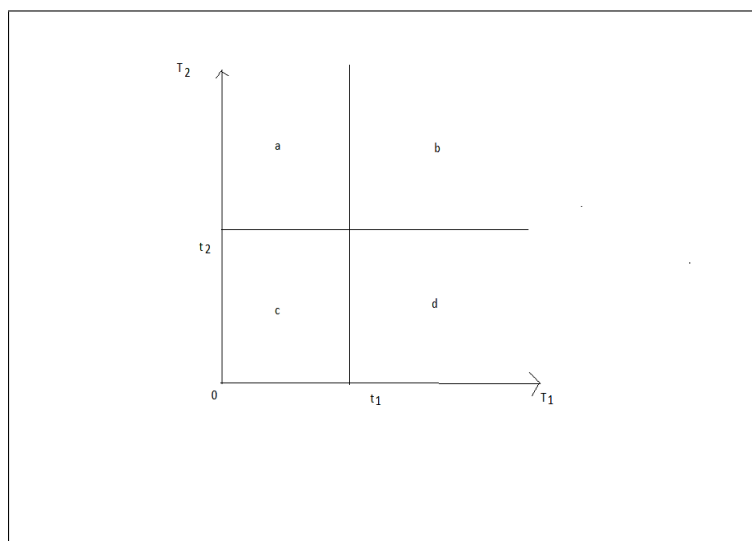
La relation entre  $S$  et  $\Lambda$  ne peut pas être formulée simplement comme dans le cas univarié. Par exemple, nous obtenons dans le cas bivarié :

$$S(t_1, t_2) = S_1(t_1)S_2(t_2)e^{-\Lambda(t_1, t_2)}$$

La construction d'une fonction de survie multivariée n'est pas commode quand on utilise des fonctions de risque directement, car elle est généralement basée sur des taux de risque conditionnels complexes (Shaked et Shanthikumar [1987])[31].

## 4.2. Les copules de survie

### a. 1 Survie bivariee



**Figure 1.** –  $F_1(t_1) = a + c$ ;  $F_2(t_2) = c + d$ ;  $S(t_1, t_2) = b$ ;  $F(t_1, t_2) = c$

Dans la majorité des applications on s'intéresse à la durée de vie d'individus statistiques d'une population donnée, ce qui donne une importance particulière à cette copule. Ici on définit une copule particulière qui est associée à cette notion de survie et nous nous intéressons principalement au cas bivarie.

Dans le cas univarié, la probabilité de survie est définie par :  $S_{X_1}(x_1) = P(X_1 > x_1) = 1 - F_1(x_1)$ , où  $F_1$  est la fonction de répartition de la variable aléatoire  $X_1$ . Si  $F = P(X_1 \leq x_1 \text{ et } X_2 \leq x_2)$  est la fonction de répartition jointe du couple aléatoire  $(X_1, X_2)$  de marginales  $F_1$  et  $F_2$  et  $C_{X_1, X_2}$  la copule de  $(X_1, X_2)$ , alors la fonction de survie du couple  $(X_1, X_2)$  sera donné par :

$$\begin{aligned} S(x_1, x_2) &= P(X_1 > x_1, X_2 > x_2) \\ &= 1 - P(X_1 \leq x_1 \text{ ou } X_2 \leq x_2) \\ &= 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2) \\ &= S_1(x_1) + S_2(x_2) - 1 + C_{X_1, X_2}(1 - S_1(x_1), 1 - S_2(x_2)), \end{aligned}$$

avec  $S_1$  et  $S_2$  les fonctions de survie marginales respectives de  $X_1$  et  $X_2$ . Donc, si nous définissons une fonction  $\tilde{C}$  de  $I^2 \rightarrow I$  :

$$\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v), \quad [7]$$

on obtient

$$S(x_1, x_2) = \tilde{C}(S_1(x_1), S_2(x_2)).$$

Remarquons que  $\tilde{C}$  est elle même une copule.

L'application du théorème de Sklar aux fonctions de survie est immédiate : Toute fonction de survie multivariée s'écrit  $S(t_1, \dots, t_d) = \tilde{C}(S_1(t_1), \dots, S_d(t_d))$ , où  $\tilde{C}$  est une copule, et  $S_1(t_1), \dots, S_d(t_d)$  les fonctions de survie marginales.



### Exemple : Modèle de Clayton.

Clayton [1978] considère un modèle d'association bivariée pour une paire ordonnée d'individus. Pour,  $(T_1$  et  $T_2$ ) les âges du premier et second membre de la paire, Clayton introduit une fonction  $\theta(t_1, t_2)$  définie ainsi

$$\theta(t_1, t_2) = \frac{h(t_1|T_2 = t_2)}{h(t_1|T_2 \geq t_2)}.$$

Cette fonction s'interprète comme le rapport du taux de risque de la distribution conditionnelle de  $T_1$ , étant donné  $T_2 = t_2$ , à celui de  $T_1$ , étant donné  $T_2 \geq t_2$ .

$$h(t_1|T_2 = t_2) = -\frac{\partial_1 S_1(t_1|T_2 = t_2)}{S_1(t_1|T_2 = t_2)} = -\frac{\partial_{1,2} S(t_1, t_2)}{\partial_2 S(t_1, t_2)},$$

et

$$h(t_1|T_2 \geq t_2) = -\frac{\partial_1 S(t_1, t_2)}{S(t_1, t_2)}.$$

D'où,

$$\theta(t_1, t_2) = -\frac{\partial_{1,2} S(t_1, t_2) \times S(t_1, t_2)}{\partial_1 S(t_1, t_2) \times \partial_2 S(t_1, t_2)} = -\frac{f(t_1, t_2) \times S(t_1, t_2)}{\partial_1 S(t_1, t_2) \times \partial_2 S(t_1, t_2)}.$$

Clayton suppose que  $\theta(t_1, t_2)$  est constante et égale à un paramètre  $\theta$ ,  $\theta \geq 0$ . Nous avons

$$\frac{\partial_{1,2} S(t_1, t_2)}{S(t_1, t_2)} - \theta \frac{\partial_1 S(t_1, t_2)}{S(t_1, t_2)} \times \frac{\partial_2 S(t_1, t_2)}{S(t_1, t_2)} = 0.$$

La fonction de survie  $S(t_1, t_2)$  est donc solution de l'équation aux dérivées partielles non linéaire du second ordre :

$$\partial_{1,2} \pi(t_1, t_2) + (\theta - 1) \partial_1 \pi(t_1, t_2) \times \partial_2 \pi(t_1, t_2) = 0,$$

avec

$$\pi(t_1, t_2) = -\ln S(t_1, t_2).$$

Clayton a montré que la solution est de la forme :

$$S(t_1, t_2) = [1 + (\theta - 1)(a_1(t_1) + a_2(t_2))]^{-\frac{1}{\theta-1}},$$

où  $a_1$  et  $a_2$  sont deux fonctions non décroissantes vérifiant  $a_1(0) = a_2(0) = 0$ .

Nous allons donner la représentation canonique de cette fonction de survie. Les marginales univariées de  $S(t_1, t_2)$  sont, respectivement :

$$S_1(t_1) = S(t_1, 0) = [1 + (\theta - 1)a_1(t_1)]^{-\frac{1}{\theta-1}},$$

et

$$S_2(t_2) = S(0, t_2) = [1 + (\theta - 1)a_2(t_2)]^{-\frac{1}{\theta-1}}.$$

En notant  $S_j(t_j) = u_j$ ,  $j = 1, 2$ , il est facile de montrer que :

$$t_j = a_j^{-1} \left( \frac{u_j^{1-\theta} - 1}{\theta - 1} \right),$$

et donc

$$S_j^{-1}(u_j) = a_j^{-1}\left(\frac{u_j^{1-\theta} - 1}{\theta - 1}\right),$$

avec  $a_j^{-1}(\cdot)$  et  $S_j^{-1}(\cdot)$  les fonctions réciproques (et non puissance inverses) respectives de  $a_j(\cdot)$  et  $S_j(\cdot)$ .

La copule de survie associée au modèle de Clayton est donc :

$$\begin{aligned} \tilde{C}(u_1, u_2) &= S(S_1^{-1}(u_1), S_2^{-1}(u_2)) \\ &= \left[1 + (\theta - 1)\left(\frac{u_1^{1-\theta} - 1}{\theta - 1} + \frac{u_2^{1-\theta} - 1}{\theta - 1}\right)\right]^{-\frac{1}{\theta-1}} \\ &= (u_1^{1-\theta} + u_2^{1-\theta} - 1)^{-\frac{1}{\theta-1}}. \end{aligned}$$

En notant  $\alpha = 1 - \theta$ , nous retrouvons la copule archimédienne de Clayton présentée en Section 3.2.

**Fin de l'exemple.**

## 5. Modèle à fragilité et copules

Le modèle à fragilité dit frailty model, est un modèle où le risque conditionnel est de type multiplicatif. Ce terme suppose, dans le contexte de la santé qu'un patient peut être plus fragile qu'un autre et donc avoir un risque de décès ou d'aggravation de sa maladie différent d'un autre. Un paramètre aléatoire (dit fragilité), potentiellement partagé par un groupe de patients (effet groupe), est introduit. La fonction de survie conjointe est obtenue en intégrant la distribution de survie conditionnelle par la densité de ce paramètre aléatoire de fragilité. Les fonctions de survie jointes pour ces modèles prennent la forme d'une copule archimédienne. Sur la base de cette constatation, il est souvent affirmé que le modèle avec fragilité correspond à un modèle particulier des copules archimédiennes (Manatunga et Oakes(1999)[25]; Viswanathan et Manatunga(2001)[34]; Andersen(2005)[2]). Il est quelque peu surprenant que ces deux approches de modélisation de nature très différente aient conduit à cette convergence.

L'idée principale consiste à introduire la dépendance entre les temps de survie  $T_1, \dots, T_d$  en utilisant une variable aléatoire  $W$  non observée, appelée fragilité. Autrement dit, il s'agit d'un modèle de survie avec variable latente. Ce modèle suppose que conditionnellement à la fragilité  $W$ , de distribution  $G$ , les temps de survie sont indépendants. Ainsi, la fonction de survie conditionnelle est donc donnée par :

$$S(t_1, \dots, t_d|w) = P[T_1 > t_1, \dots, T_d > t_d|W = w] = \prod_{j=1}^d P[T_j > t_j|W = w] = \prod_{j=1}^d S_j[t_j|W = w].$$

et la fonction de survie inconditionnelle sera donnée par :

$$S(t_1, \dots, t_n) = E(E(S(t_1, \dots, t_n|W))) = \int S(t_1, \dots, t_n|w)dG(w).$$

Afin d'avoir une représentation intéressante des modèles de fragilité, nous avons besoin du théorème de Marshall et Olkin (1988)[26] suivant :

**THEOREM 7 (MARSHALL ET OLKIN (1988))**– Soient  $F_1, \dots, F_d$  des fonctions de distribution univariées, et  $G$  une fonction de distribution à  $d$  variables telle que  $G(0, \dots, 0) = 0$ , avec des fonctions marginales univariées  $G_j, j = 1 \dots d$ . Notons les transformées de Laplace de  $G, \varphi$  et celles de  $G_j, \varphi_j$ . Soit  $C$  une fonction de distribution  $d$ -variée dont toutes les marginales univariées sont uniformes sur  $[0, 1]$ .

Si  $H_j(x) = \exp(-\varphi_j^{-1}(F_j(x)))$ , alors

$$F(x_1, \dots, x_d) = \int C[H_1(x_1)^{w_1}, \dots, H_d(x_d)^{w_d}]dG(w_1, \dots, w_d) \quad [8]$$

est une fonction de distribution à  $d$  variables avec des marginales  $F_1, \dots, F_d$ .

Marshall et Olkin ont étudié par la suite un cas particulièrement intéressant et simple de (8) (voir [1], page 15), où toutes les fonctions marginales sont identiques. L'expression (8) s'écrit alors,

$$F(x_1, \dots, x_d) = \varphi_1(\varphi_1^{-1}(F_1(x_1)) + \dots + \varphi_1^{-1}(F_d(x_d))) \quad [9]$$

Il s'agit donc d'une copule archimédienne dont le générateur est l'inverse de la transformée de Laplace. On peut maintenant énoncer la définition des fonctions de survie d'un modèle avec fragilité.

**DEFINITION 11**– Une fonction de survie est dite avec fragilité si elle s'écrit,

$$S(t_1, \dots, t_d) = \tilde{C}(S_1(t_1), \dots, S_d(t_d))$$

où  $\tilde{C}$  est une copule archimédienne dont le générateur est l'inverse de la transformée de Laplace de la distribution de la variable de fragilité  $W$ . Plus généralement, le générateur est l'inverse d'une transformée de Laplace.

### 5.1. Copule de survie bivariée et modèle à fragilité

Dans cette section, on s'intéresse à la modélisation de données de survie à deux variables. Considérons deux "temps de survie"  $(T_1, T_2)$ , correspondant par exemple à deux durées pour l'obtention d'un diagnostic effectué sur un même individu par deux techniques différentes, par exemple  $T_1$  pour la radiographie et  $T_2$  pour l'échographie. Soient  $S_1(t)$  et  $S_2(t)$  les fonctions de survie marginales pour chacune des méthodes (Goethals et al, [16]).

Un modèle avec fragilité est donné par :

$$h_{ij}(t) = w_i h_{j,w_i}(t),$$

- $h_{ij}(t)$  la fonction de risque instantané au temps  $t$  pour un individu  $i = 1, \dots, n$ , avec la technique de diagnostic  $j = 1, 2$ .
- $h_{j,w}(t)$  la fonction de risque au temps  $t$  pour un individu dont la fragilité est égale à  $w$  et la technique de diagnostic  $j$ .
- $w_i$  le terme de fragilité de l'individu  $i$ .

Pour définir les modèles de copules et les modèles avec fragilité, nous avons besoin d'une famille de copules archimédiennes particulière où le générateur  $\varphi$  est l'inverse d'une transformée de Laplace,

$C(u, v) = \varphi \{ \varphi^{-1}(u) + \varphi^{-1}(v) \}$ , avec  $\varphi(0) = 1$ . Nous n'avons donc besoin que d'une famille de fonction  $\varphi(\cdot)$ . Soit  $g_W(\cdot)$  la densité de la v.a. fragilité, définie sur le support  $[0, \infty[$  et  $\varphi_W(s)$ , sa transformée de Laplace i.e.

$$\varphi_W(s) = E \{ \exp(-sw) \} = \int_0^{\infty} \exp(-sw) g_W(w) dw.$$

Ainsi, la fonction de survie jointe conditionnelle à W, s'écrit :

$$S_W(t_1, t_2) = \tilde{C}(S_{1,W}(t_1), S_{2,W}(t_2)) = \varphi_W \{ \varphi_W^{-1}(S_{1,W}(t_1)) + \varphi_W^{-1}(S_{2,W}(t_2)) \}. \quad [10]$$

Pour le modèle avec fragilité, la fonction de survie conditionnelle est donnée par :

$$S_{W_i}(t_1, t_2) = \exp[-w_i \{ \Lambda_{1,w_i}(t_1) + \Lambda_{2,w_i}(t_2) \}], i = 1, \dots, n$$

avec  $\Lambda_{j,w_i}(t) = \int_0^t h_{j,w_i}(s) ds$ ,  $j=1,2$ , la fonction du risque cumulé. Ensuite, en intégrant par rapport à la densité de la fragilité, la fonction de survie jointe du modèle sera

$$S(t_1, t_2) = \int_0^{\infty} S_W(t_1, t_2) g_W(w) dw = E[\exp \{ -W(\Lambda_{1,w}(t_1) + \Lambda_{2,w}(t_2)) \}] \quad [11]$$

$$= \varphi \{ \Lambda_{1,w}(t_1) + \Lambda_{2,w}(t_2) \}, \quad [12]$$

Puisque la fonction de survie marginale peut s'écrire comme :

$$S_j(t) = \varphi \{ \Lambda_{j,W}(t) \} \Rightarrow \Lambda_{j,w}(t) = \varphi^{-1}(S_j(t)), \quad [13]$$

d'où, si nous remplaçons (13) en (12), la fonction de survie jointe du modèle avec fragilité devient :

$$S(t_1, t_2) = \varphi \{ \varphi^{-1}(S_1(t_1)) + \varphi^{-1}(S_2(t_2)) \}. \quad [14]$$

De (10) et (14), nous remarquons que les deux modèles sont de nature différente car la copule utilisée dans les fonctions de survie jointes dans (10) et (14) est la même mais les fonctions marginales de survie ne sont pas les mêmes.

## 5.2. Copule de Clayton-Oakes et modèle avec fragilité de type gamma

A titre d'exemple, la fonction copule pour le modèle de Clayton-Oakes, est la fonction de survie jointe d'un modèle avec fragilité dont la transformée de Laplace est celle d'une variable aléatoire de loi gamma, c'est à dire,  $\varphi_{\theta}(s) = (1 + \theta s)^{-\frac{1}{\theta}}$ . Considérons encore l'exemple des durées correspondant à deux techniques de diagnostic  $j = 1, 2$ . La transformée de Laplace inverse d'une densité de loi gamma à un seul paramètre  $\theta$  est donnée par :

$$\varphi_{\theta}^{-1}(s) = \frac{(s^{-\theta} - 1)}{\theta}, \theta \geq 0.$$

L'application directe de l'équation (10) pour la fonction de survie jointe de copule de Clayton Oakes donne

$$S_c(t_1, t_2) = C_\theta(S_{1,c}(t_1), S_{2,c}(t_2)) = [\{S_{1,c}(t_1)\}^{-\theta} + \{S_{2,c}(t_2)\}^{-\theta} - 1]^{-\frac{1}{\theta}}. \quad [15]$$

La fonction de survie jointe pour ce modèle à fragilité devient :

$$S_m(t_1, t_2) = [1 + \theta \{\Lambda_{1,u}(t_1) + \Lambda_{2,u}(t_2)\}]^{-\frac{1}{\theta}},$$

ce qui nous permet d'écrire :

$$\begin{aligned} S_m(t_1, t_2) &= \{1 + [(S_{1,m}(t_1))^{-\theta} - 1] + [(S_{2,m}(t_2))^{-\theta} - 1]\}^{-\frac{1}{\theta}} \\ &= \{(S_{1,m}(t_1))^{-\theta} + (S_{2,m}(t_2))^{-\theta} - 1\}^{-\frac{1}{\theta}} \end{aligned}$$

Cette expression ressemble à la forme de la copule représentée précédemment, mais  $S_{j,m}(t) \neq S_{j,c}(t)$ ,  $\forall j = 1, 2..$

## 6. Application à des données d'hémodialyse

Nous utilisons dans cette section, des données d'hémodialyse publiées par McGilchrist et Aisbett (1991) [27] et analysées par des modèles de fragilité. Cette étude a été motivée par l'observation d'une récurrence des infections chez les patients atteints d'insuffisance rénale et traités par hémodialyse, c'est à dire par épuration des toxines du sang produites par l'organisme et leur élimination par un filtre artificiel. Une infection peut survenir au niveau du cathéter, tube creux en plastique que le médecin place dans une veine pour l'hémodialyse. Quand l'infection est observée, le cathéter est retiré et l'infection soignée. Un cathéter est remis pour la prochaine hémodialyse, etc. Pour chaque patient, les durées entre chaque date d'insertion du cathéter et l'infection qui suit, sont observées. Seules deux observations par patient sont considérées. Il peut aussi y avoir censure, si durant la période d'étude, l'une ou les deux infections n'ont pas eu lieu.

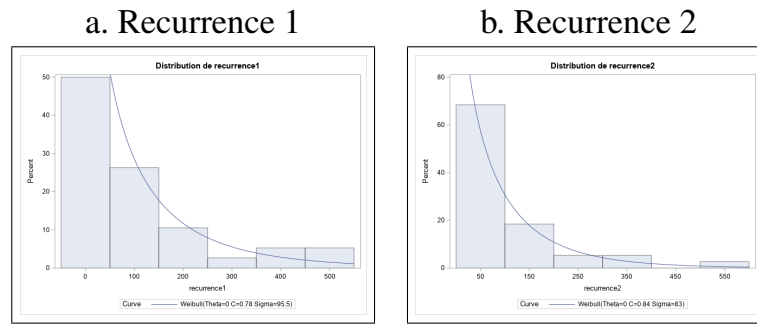
Nous ne traiterons pas le cas censuré ici, et considérerons toutes les durées comme observées. McGilchrist et Aisbett traitent le cas censuré par un modèle de fragilité.

Nous utilisons pour cela, la nouvelle procédure de SAS (Proc Copula) pour estimer le paramètre d'association et choisir ainsi la copule qui s'ajuste le mieux à ces données.

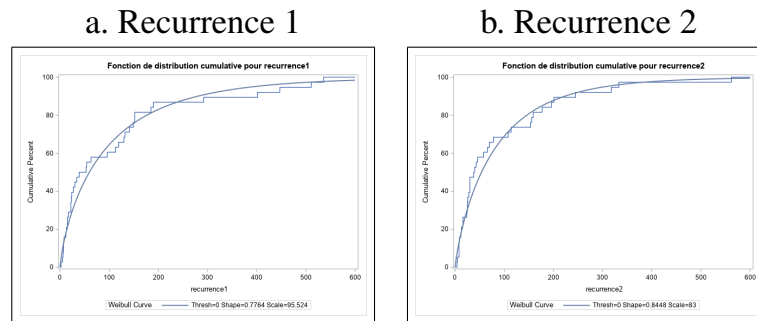
L'analyse de ces données bivariées par modèle de copule, en présence de censure, fera l'objet d'un futur travail.

Les histogrammes de la figure 2 sont ceux des durées marginales des deux temps de récurrence. Les courbes continues représentent les densités théoriques ajustées supposées être de Weibull. On constate un assez bon ajustement.

Dans la figure 3 sont représentées les fonctions de distribution marginales empiriques correspondant aux récurrences 1 et 2. Les traits continus correspondent aux fonctions de distribution de Weibull ajustées à partir des paramètres estimés.

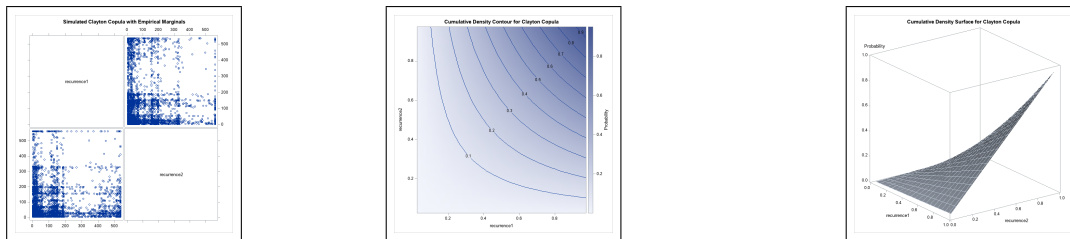


**Figure 2.** – Histogrammes empiriques et densités ajustées (Weibull) des deux temps de récurrence.



**Figure 3.** – Fonctions de répartition empiriques et leur estimées (Weibull) des deux temps de récurrence.

a. Nuage de points    b. Copule de Clayton : contours    c. Copule de Clayton : surface



**Figure 4.** – Distribution bivariable empirique et graphes associés de la copule de Clayton.

Les estimations des coefficients de corrélation de Pearson, Spearman et Kendall(tau) sont respectivement 0,07522 (avec un degré de signification  $p=0,6535$ ), 0,01040 ( $p=0,9506$ ) et 0,01004 ( $p=0,9298$ ) respectivement. Le paramètre  $\theta$  de la copule de Clayton est estimé à  $1,054 \times 10^{-7}$ . Dans la figure 4 sont représentés les nuages de points  $((t_1, t_2)$  et  $(t_2, t_1)$  (a) ainsi que les courbes de niveau de la distribution bivariable (b) et la surface correspondant à celle ci (c), estimées par une copule de Clayton. Les copules de Frank et de Gumbel peuvent être obtenues de la même manière.

La table 6.1 résume les résultats concernant le paramètre d'association  $\theta$  estimé sous différents choix de modèles de copules (Clayton, Frank et Gumbel). Le critère d'Akaike (AIC) permet de choisir le modèle qui s'ajuste le mieux aux données (valeur d'AIC la plus petite). Le choix de la copule de Gumbel semble le meilleur. Les deux temps de récurrences sont faiblement corréllés. Ces données bivariées pourraient être analysées comme des données univariées. Nous avons présenté ici une analyse succincte de l'association entre les deux temps de récurrence. Nous n'avons pas analysé les effets des covariables. Il serait intéressant de comparer l'analyse de l'effet de ces covariables en considérant les données comme univariées à celle tenant compte de l'aspect bivarié modélisé par une copule de Gumbel, retenue ici par

**Tableau 6.1.** – Estimation du paramètre d'association pour trois modèles de Copule sous critère de pénalisation d'Akaike.

Copule	$\theta$	Std. Error	AIC
Clayton	$1,054 \times 10^{-7}$	0	2,0
Frank	0,150408	1,031073	1,97873
Gumbel	1,039736	0,127503	1,89611

l'AIC. Il serait aussi intéressant de comparer les effets des covariables avec ceux obtenus par un modèle de régression de Cox avec fragilité.

Les programmes SAS utilisés sont donnés en annexe.

## Conclusion et perspectives

L'objectif de cet article, outre la présentation des copules et de leurs propriétés essentielles, s'est focalisé sur leur application aux données de survie. Les relations entre les composantes d'une survie multivariée peuvent ainsi être facilement et naturellement modélisées à l'aide de copules. Un outil additionnel souvent utilisé pour la modélisation des données de survie multivariée est l'introduction de paramètres aléatoires individuels interprétés souvent comme des paramètres de fragilité. Dans ce travail, nous avons utilisé ce modèle pour les données de survie à deux variables en considérant des copules archimédiennes. Nous nous sommes concentrés ensuite aux cas particuliers des copules de Clayton-Oakes et du modèle avec fragilité de type gamma. Pour chacun de ces deux modèles, les copules utilisées pour les fonctions de survie bivariée sont les mêmes. Toutefois les fonctions de survie marginales sont modélisées de façons différentes.

Nous nous sommes intéressés ensuite à l'applications pour des données de survie relatives à la santé. Les données de survie peuvent aussi provenir d'études de fiabilité dans le domaine industriel. Une difficulté importante dans l'analyse de ce type de données, y compris le cas univarié, est la présence de censure. Dans le cas bivarié, cette question reste en grande partie ouverte. L'analyse des données de survie bivariée et censurées par des copules fait l'objet de travaux en cours.

Dans le domaine des assurances (actuariat) et plus généralement en finance, la théorie des copules a connu beaucoup de succès, et en lien avec la théorie des valeurs extrêmes, de nombreuses copules ont été construites et utilisées. Ce travail n'aborde pas ces applications.

Ce travail ne traite pas aussi les questions d'inférence statistique qui sont souvent complexes, surtout en présence de censure. Les travaux de recherche récents dans le domaine des copules sont plus souvent de nature statistique que théorique. F. Lounas (2011)[24], Deheuvels (1979)[9], Genest (1987)[13], Genest et Rivest (1993)[14] sont d'excellentes références sur ce sujet.

Les données de survie, de type quantitatif continu en général, sont utilisés souvent dans le domaine médical dans un but diagnostic ou pronostic. On retrouve assez souvent aussi des données de nature qualitative et discrète (Fontaine, 2017)[10]. Cet article incite à plusieurs perspectives de travail futur et au développement de modèles basés sur les copules pour ce type de données.

Dans le domaine des copules, la recherche s'est appliquée aussi au développement des aspects informatiques, par la publication de programmes dans le langage R et récemment, SAS avec "Proc Copula". Nous avons utilisé "Proc Copula" pour traiter les données d'hémodialyse. Cette procédure ne permet malheureusement pas de traiter les données censurées. Un travail futur et utile sera l'écriture d'une macro SAS allant dans ce sens.

Les variables de fragilité, considérées ici, sont des variables latentes, non observées, mais unidimensionnelles. Dans l'exemple présenté, cette variable caractérisait l'effet de l'individu sur le temps de récurrence. Ces individus pourraient provenir de plusieurs centres hospitaliers. L'effet différentiel, non observé, de ces centres serait alors une variable latente. Une perspective de recherche future serait la modélisation et l'analyse par des copules de durées de vie multivariées avec variables latentes elles même multivariées. La difficulté du traitement de telles données serait surtout de type informatique.

## Remerciements

Les auteurs remercient Pr. R. Senoussi pour sa relecture attentive de cet article. Ses remarques, et les discussions multiples qui ont suivi, ont permis d'améliorer significativement une version précédente de cet article.

## Bibliographie

- Manatunga, A.K., and Oakes, D., 1999. Parametric Analysis for Matched Pair Survival Data, Springer Link
- Andersen, E.W., 2005. Two-stage estimation in copula models used in family studies, *Lifetime Data Analysis*, 11(3), pp. 333-350.
- 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
- Chaieb L., Lajmi M., 2006. Estimation de la dépendance et de lois marginales dans des modèles pour l'analyse des durées de vies multidimensionnelles. Thèse Ph. D., Université de LAVAL, QUEBEC.
- Cherbini, U., Luciano, E., Vecchiato, W., 2004. *Copula Methods in Finance*, Wiley.
- Clayton, D.G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*. 65, 141–151
- Cook, R.D., Johnson, M.E., 1981. A family of distributions for modelling nonelliptically symmetric multivariate data, *Royal Statistical Society*. 43, 210-218
- Cox, D.R., 1972, Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society -Series B*, 34, 187-220
- Deheuvels, P., 1979. La fonction de dépendance empirique et ses propriétés. *Acad. Roy. Belg. Bull. Cl. Sci.* 65, 274-292.
- Fontaine, C., 2017. Utilisation de copules paramétriques en présence de données observationnelles : cadre théorique et modélisations. *Médecine humaine et pathologie*. Université de Montpellier.
- Frees, E.W., and Valdez, E.A., 1998, Understanding relationships using copulas, *North American Actuarial Journal*, 2, 1-25
- Fréchet, M., 1957. Les tableaux de corrélation dont les marges et des bornes sont données. *Annales de l'Université de Lyon. Sciences Mathématiques et Astronomie*, 20, 13-31.
- Genest, C., 1987. Frank's family of bivariate distributions. *Biometrika*, 74, 549-555.
- Genest, C., Rivest L.P., 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88, 1034-1043.



- Georges, P., Lamy, A.G., Nicolas, E., Quibel, G., and Roncalli, T. 2001. Multivariate survival modelling : a unified approach with copulas. Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.
- Goethals, K., Janssen, P., and Duchateau, L., 2008. Frailty models and copulas : Similarities and differences. *Journal of Applied Statistics*. Vol. 35, No 9, 1071-1079.
- Hoeffding, W., 1940. Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin*, 5, 181-233.
- Hoeffding, W., 1941. Masstabinvariante Korrelationsmasse für Diskontinuierliche Verteilungen. *Archiv für Mathematische Wissenschaften*. 7, 49-70, reprinted as Scale- Invariant correlations for discontinuous distributions in *The collected works of Wassily Hoeffding*, Springer Verlag, New York(109-132).
- Hougaard, P., 1999. Fundamentals of survival data, *Biometrics*, 55, 13-22.
- Joe, H., 1997. *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.
- Kendall, M., 1938. A New Measure of Rank Correlation, *Biometrika*, 30, 81-89.
- Lancaster, T., 1979. Econometric methods for the duration of unemployment, *Econometrica* vol. 47 pp. 939-56, 1979.
- Lancaster, T., 1990. *The Econometric Analysis of Transition Data*, *Econometric Society Monographs*, 17, Cambridge University Press, Cambridge.
- Louna, F., 2011. Modélisation de la dépendance par les copules et applications. Thèse de magister. Université de Tizi-Ouzou.
- Manatunga, A.K. and Oakes, D., 1999. Parametric analysis for matched pair survival data, *Lifetime Data Analysis*, 5(4), pp. 371-387.
- Marshall, A.W. and Olkin, I., 1988. Families of multivariate distributions, *Journal of the American Statistical Association*, 83, 834-841.
- McGilchrist, C.A., and Aisbett, C.W., 1991. Regression with frailty in survival analysis. *Biometrics* 47, 461-466.
- Nelsen, R.B., 2006. *An Introduction to Copulas*, second ed. Springer, New York.
- Pearson, K., 1896. Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia, *Philosophical Transactions of the Royal Society of London*. 187 : 253-318.
- Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* 8, 229-231.
- Shaked, M. and Shanthikumar J.G., 1987. The multivariate hazard construction, *Stochastic Processes and their Applications*, 24, 241-258.
- Spearman, C., 1904. The proof and measurement of association between two things, *American Journal of Psychology*. 15(1), 72-101.
- Vaupel, J.W., Manton, K. G., and Stallard, E., 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* vol. 16 pp. 439-54.
- Viswanathan, B. and Manatunga, A.K., 2001. Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Analysis* 7, 143-155.

### **Programme SAS : Proc Copula appliquée à des données de récurrence en hémodialyse.**

```
proc univariate data=recurrences; var recurrence1 recurrence2;
histogram recurrence1 recurrence2/weibull;run;
proc univariate data=recurrences;var recurrence1 recurrence2;
cdfplot recurrence1 recurrence2/weibull;run;
proc corr data=recurrences kendall pearson spearman;
var recurrence1 recurrence2;run;
proc copula data=recurrences;var recurrence1 recurrence2;
fit clayton/marginals=empirical;
simulate /ndraws = 5000 seed = 12345678 marginals=empirical
plots = (distribution=cdf) out = fic1;run;
proc copula data=recurrences; var recurrence1 recurrence2;
```

```
fit frank/marginals=empirical;  
simulate /ndraws = 5000 seed = 12345678 marginals=empirical  
plots = (distribution=cdf) out = fic1; run;  
proc copula data=recurrences; var recurrence1 recurrence2;  
fit gumbel/marginals=empirical;  
simulate /ndraws = 5000 seed = 12345678 marginals=empirical  
plots = (distribution=cdf) out = fic1; run;
```