

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la

VIE

DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

MERZOUGUI Wafa

Titre :

Les copules et Modèle de Fragilité

Membres du Comité d'Examen :

Pr. Abdallah SAYAH UMKB Président

Pr. Fateh BENATIA UMKB Encadreur

Dr. Imane BENELMIR UMKB Examinatrice

19 Juin 2023

Dédicace

*A celle qui m'a transmis la vie, l'amour, le courage, à toi chère **maman** "M.H"
toutes mes joies, mon amour et ma reconnaissance.*

A mon **père** "N.M" pour l'éducation qu'il m'a prodigué, avec tous les moyens et
prix de toutes les sacrifices qu'il a consentis à mon égard et à mes études depuis
mon enfance.

A mon professeur **Fateh BENATIA.**

A mes chère soeurS **nouha, Mariya.**

A toute ma famille : **nawale, yasmina, karima, wahiba, nadya, Souade H.**

A mes cousins : **yasmin, imane, ranai sirine,romaysa**

A mon cousin : **zino HAMIDI.**

A mon oncle : **Brahim H, Lahmidi H, Ziyane H**

A tout ma famille : *MERZOUGUI et HAMIDI*

A mes amies : **Rehab Loubna, Masouda, Hasna, Imane h, Asma, Imane b,
Djohaina, Hayet, Salsabil, Amira.**

A toutes ces personnes, sincèrement : *MeRCi*

REMERCIEMENTS

*Je tiens à remercier **ALLAH** le tout-puissant qui m'a aidée et donnée la santé, la patience et le courage durant ces longues années d'études.*

Je tiens remercié sincèrement mon encadreur, Pr. **Fateh BENATIA** qui par ses paroles, ses écrits, ses conseils et critiques a guidé mes réflexions et a accepté de me rencontrer et de répondre à mes questions durant mes recherches.

Je remercie également tous les enseignants et employés du département de Mathématiques, avec une mention spéciale à Messieurs les membres du jury "le pr. **Abdallah SAYAH** et le Dr. **Imane BENELMIR**" qui nous ont faites l'honneur de participer à notre soutenance.

Je n'oublie pas mes parents pour leur contribution, leur soutien et leur patience.

Enfin, j'adresse mes plus sincères remerciements ma famille, mes proches, amis et tous ceux qui, de près ou de loin, m'ont soutenu et encouragé au cours de la réalisation de ce mémoire

Merci toutes et á tous.

Mer. WaFa

Table des matières

Remerciements	ii
Table des matières	ii
Table des figures	v
Liste des tables	v
Introduction	1
1 Copules bivariées et Mesures d'association	4
1.1 Théorie des copules	4
1.1.1 Définition d'une copule bivariée	4
1.1.2 Propriétés d'une copule	7
1.1.3 Familles paramétriques de copules usuelles bivariées	8
1.2 Mesure d'association	14
1.2.1 Mesure de dépendance	14
1.2.2 Mesures de concordance	17
2 Modélisation de l'analyse de survie	21
2.1 Donnée de survie	21

2.1.1	Les caractéristiques des données de survie	22
2.1.2	Distributions de la durée de survie	25
2.1.3	Les mécanismes de censure	27
2.2	Modèle de survie	28
2.2.1	Les méthodes d'analyse des données de survie	28
2.2.2	Copule de survie	32
3	Modèle de fragilité et copule	34
3.1	Modèle de fragilité	34
3.1.1	Modèles de Fragilité" UNIVARIÉS"	34
3.2	Modèle de fragilité pour copules	37
3.2.1	Copule de survie bivariée et modèle à fragilité	39
3.2.2	Copule de Clayton-Oakes et modèle à fragilité avec fragilité de	
	type gamma	41
	Conclusion	43
	Bibliographie	43

Table des figures

- 1- Densité du copule Normale [1.4](#)
- 2- Contour de la densité du Copule Normale [1.5](#)
- 3- Copule Normale [1.6](#)
- 4- Contour du Copule Normale [1.7](#)
- 5- Densité du copule de Student à 2 dl [1.8](#)
- 6- Contoure de la densité du Copule de Student [1.9](#)
- 7- Contour du Copule de Student [1.10](#)
- 8- Modèles de temps de survie [2.1](#)
- 9- les différentes dateset les durées de survies [2.2](#)
- 10- Exemple de censure à droite [2.6](#)
- 11- L'estimateur Kaplan-Meier de la fonction de survie pour les données *AML* [2.9](#)
- 12- $F_1(t_1) = a + c; F_2(t_2) = c + d; S(t_1, t_2) = b; F(t_1, t_2) = c$ [2.10](#)
- 13- Exemple de risque de décès pour un patient subissant une intervention chirurgicale [3.1](#)

Liste des tableaux

1. Tau de Kendall et Rho de Spearman de quelques copules [1.13](#)

Introduction

Le mot "copule" vient du mot latin "*copulae*" qui désigne une corrélation ou la connexion entre les parties. Ce mot a été adopté en statistique pour désigner une classe de fonction qui permet de construire et de modéliser la dépendance de manière simple. C'est par là que l'on retrouvera assez souvent les termes de corrélation et de dépendance.

Le premier chapitre se compose de deux parties. Dans la première nous introduisons les définitions de base des copules, leurs propriétés, ainsi que les principaux théorèmes liés à cette théorie, ainsi que quelques exemples de familles paramétriques de copules les plus utilisées. Dans la deuxième partie nous présentons certains aspects des mesures d'association : la mesure de concordance, la mesure de dépendance en forme de copule, telle que le rho de Spearman et le tau de Kendall. L'importance d'utiliser des copules dans la modélisation de la dépendance est aussi présentée.

Le deuxième chapitre est consacré à la modélisation de la survie, comme dans toute analyse statistique, nous présentons les fonctions de survie univariées et multivariées, des groupes auxquels ils appartiennent, la comparaison de la survie de plusieurs groupes d'individus, l'explication et la prédiction de la durée de survie présentée en fonction de certains facteurs. Approches de modélisation paramétrique, semi-paramétrique, non paramétrique ainsi que l'estimateur de Kaplan-Meier, et les mécanismes de censure de survie.

Enfin dans le troisième et dernier chapitre, nous considérons l'approche des modèles

de survie multivariée avec variable de fragilité, et présentons les copules associées à ce type de modèles, particulièrement le modèle de Clayton-Oakes, dans le cas où il y a une fragilité de type gamma.

Chapitre 1

Copules bivariées et Mesures d'association

1.1 Théorie des copules

1.1.1 Définition d'une copule bivariée

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

On appelle copule bivariée toute fonction C définie de $I^2 = [0, 1]^2$ dans I par :

$$\begin{aligned} C & : \quad I^2 \quad \rightarrow \quad I \\ (u, v) & \rightarrow C(u, v) \end{aligned}$$

vérifiant les propriétés suivantes :

i) La copule C est attachée (grounded), c'est à dire

$$\forall u, v \in I, C(u, 0) = C(0, v) = 0. \tag{1.1}$$

ii) Les marges sont uniformes, c'est à dire

$$C(u, 1) = u \text{ et } C(1, v) = v, \forall u, v \in I. \quad (1.2)$$

iii) La copule C est 2-croissante i.e

$\forall u_1, u_2, v_1, v_2 \in I$, avec $u_1 \leq u_2$ et $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0 \quad (1.3)$$

(i) et (ii) traduit, en particulier, que toute copule est une distribution dont les distributions marginales sont de loi uniforme définie sur $I = [0, 1]$. En effet en considérant le vecteur aléatoire $U = (U_1, U_2)$ où U_1 et U_2 sont deux variables aléatoires uniformes sur $I = [0, 1]$ alors on a :

$$C(u_1, u_2) = P(U_1 \leq u_1, U_2 \leq u_2)$$

(iii) est la 2-croissance ou *inégalité du rectangle* de la distribution C. Elle traduit le fait que si C admet une densité $c(u, v)$ alors elle est positive,

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \geq 0$$

Exemples de copules

Nous présentons ici quelques modèles de copules les plus utilisées.

-Copule produit :

La copule produit, noté Π , est définie par :

$$C(u, v) = \Pi(u, v) = uv, \forall u, v \in I.$$

Cette copule vérifie les propriétés [1.1](#) et [1.2](#), telle que :

$$C(0, v) = \Pi(0, v) = 0v = 0 \text{ et } C(u, 0) = \Pi(u, 0) = u0 = 0,$$

$$C(1, v) = \Pi(1, v) = 1v = v \text{ et } C(u, 1) = \Pi(u, 1) = u1 = u.$$

pour la propriété [1.3](#), on a :

$$\begin{aligned} C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) &= \Pi(u_2, v_2) - \Pi(u_2, v_1) - \Pi(u_1, v_2) + \Pi(u_1, v_1) \\ &= u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \end{aligned}$$

où

$$\text{si } u_2 \geq u_1 \implies \begin{cases} u_2v_2 \geq u_1v_2 \\ u_2v_1 \geq u_1v_1 \end{cases}$$

alors

$$u_2v_2 - u_2v_1 \geq u_1v_2 - u_1v_1 \implies u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \geq 0.$$

-Copule min :

La copule min, noté M , est définie par :

$$M(u, v) = \min(u, v), \quad \forall u, v \in I.$$

-Copule max :

La copule max, noté W , est définie par :

$$W(u, v) = \max(u + v - 1, 0), \quad \forall u, v \in I.$$

1.1.2 Propriétés d'une copule

On note tout d'abord que les copules vérifient certaines propriétés :

a- *Les marges suivent des lois uniforme:*

$$C(u, v) = 0 \text{ si } u \leq 0 \text{ ou } v \leq 0.$$

$$C(u, v) = 1 \text{ si } u \geq 1 \text{ ou } v \geq 1.$$

$$C(u, v) = u \text{ si } v \geq 1 \text{ et } C(u, v) = v \text{ si } u \geq 1.$$

b- *La continuité :* Les copules sont des fonctions continues. Plus précisément, elles vérifient la condition de Lipschitz

$$\forall u_1, u_2, v_1, v_2 \in I \text{ on a } |C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

c- *La symétrie :* On dit que C est symétrique si

$$\forall (u, v) \in I^2 : C(u, v) = C(v, u).$$

d- *L'ordre :* On dit que C_1 est plus petite que C_2 et inversement si

$$\forall (u, v) \in I^2 : C_1(u, v) \leq C_2(u, v).$$

e- *La croissance :* C est 2-croissante

$\forall (u_1, u_2), (v_1, v_2) \in I^2$ vérifiant $u_1 \leq u_2$ et $v_1 \leq v_2$ on a

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

f- *Le théorème d'invariance :* L'un des théorèmes essentiel à la théorie des copules

est celui de l'invariance par transformation strictement croissantes. Soient X et Y deux v.a continues de marginales F et G et de copule C_{XY} . Si α et β sont deux fonctions strictement croissantes, alors

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

1.1.3 Familles paramétriques de copules usuelles bivariées

Nous exposons dans la présente section les principales copules paramétriques. Ces copules présentent un intérêt particulier dans la gestion des risques(modélisation financière), de par le fait qu'elles autorisent la construction de modèles paramétriques ou semi-paramétriques.

Les copules usuelles

- la copule indépendante Π , $\Pi(u, v) = uv$ (\iff les v.a X et Y sont indépendantes).
- la copule minimale W , $\forall u, v \in I$, $W(u, v) = \max(u + v - 1, 0)$, borne inf de Fréchet
- la copule maximale M , $M(u, v) = \min(u, v)$, borne sup de Fréchet.

Pour toute copule C on a la relation :

$$W(u, v) \leq C(u, v) \leq M(u, v).$$

·la famille de Fréchet formée par les combinaison linéaire convexe de M , Π et W par exemple la copule de Mardia :

$\forall \theta \in [-1, 1], \forall u, v \in I;$

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

Les copules elliptiques

Définition 1.1.1 *On appelle copule elliptique toute copule de la forme*

$$\begin{aligned} C_\rho(u, v) &= \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{\phi_{g,1}^{-1}(u)} \int_{-\infty}^{\phi_{g',2}^{-1}(v)} g\left(\frac{\chi^2 - 2\rho\chi y + y^2}{\sqrt{1-\rho^2}}\right) d\chi dy \\ &= H_\rho(\phi_{g,1}^{-1}(u), \phi_{g',2}^{-1}(v)) \end{aligned}$$

où H_ρ

est la distribution conjointe des variables X et Y, $\phi_{g,1}^{-1}$ et $\phi_{g',2}^{-1}$ les fonctions quantiles respectives et ρ leur coefficient de corrélation.

Copules élliptiques classiques

a-La Copule normale bivariée Tout comme les distributions normale et binormale, un des types de copules très utilisé dans la modélisation financière est la copule normale standard de corrélation ρ cette copule est donnée par :

$$C_\rho(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \exp\left\{-\frac{\chi^2 - 2\rho\chi y + y^2}{2(1-\rho^2)}\right\} d\chi dy$$

où ϕ^{-1} est la fonction quantile de la loi normale standard $N(0, 1)$.

Densité du copule Normal (1.4)

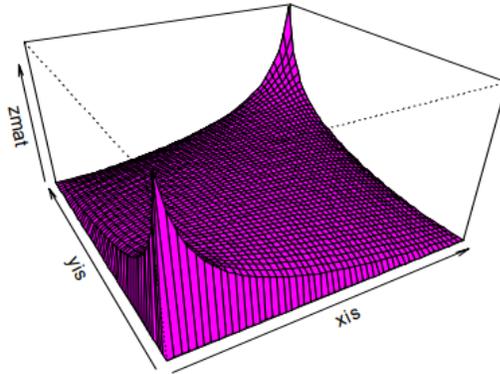


FIG.1.1-Densité du copule Normale

Contour de la densité du Copule Normal (1.5)

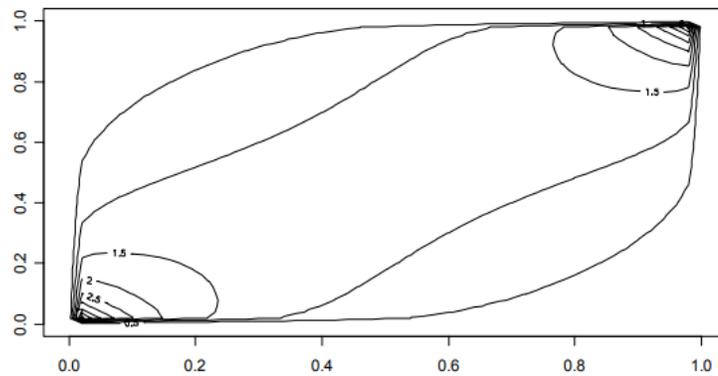


FIG.1.2-Contour de la densité du Copule Normale

Copule de la Normal

(1.6)

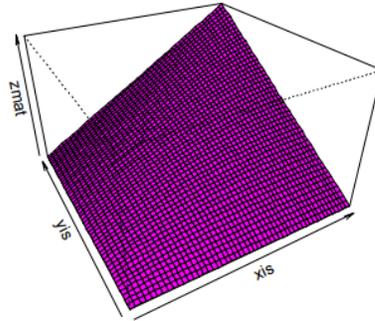


FIG.1.3-Copule Normale

Contour du Copule Normal

(1.7)

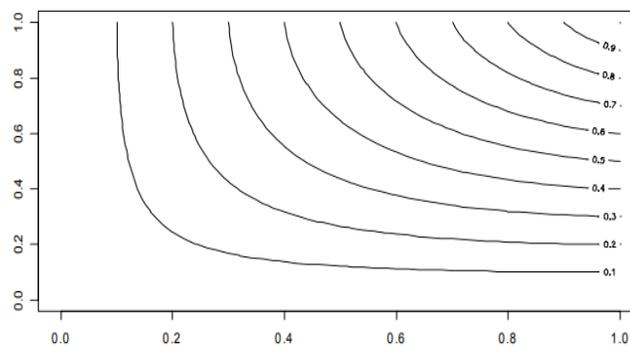


FIG.1.4-Contour du Copule Normale

b-La Copule de Student Soit ρ une matrice diagonale définie positive avec $\text{diag}(\rho) = 1$ et $t_{\rho,v}$ la distribution de student bivariée standard à v degrés de liberté et matrice de corrélation. La copule de Student associée à cette distribution est alors

définie de la façon suivante :

$$C_\rho(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_\gamma^{-1}(u)} \int_{-\infty}^{t_\gamma^{-1}(v)} \left\{ 1 + \frac{\chi^2 - 2\rho\chi y + y^2}{v(1-\rho^2)} \right\} d\chi dy$$

où t_v^{-1} est la fonction inverse de la distribution standard de Student à v degrés de liberté.

Densité du copule de Student à 2 dl (1.8)

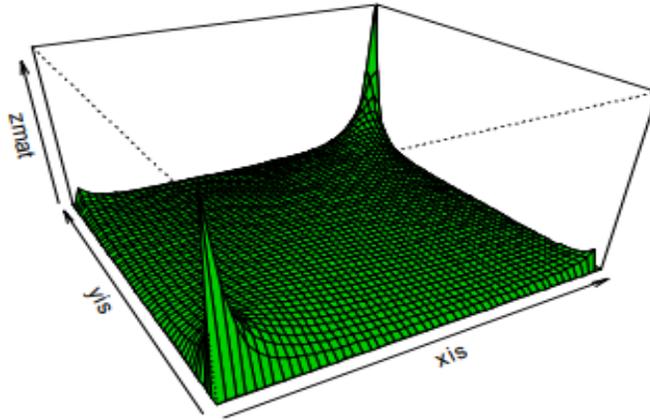


FIG.1.5-Densité du copule de Student à 2 dl

Contour de la densité du Copule de Student (1.9)

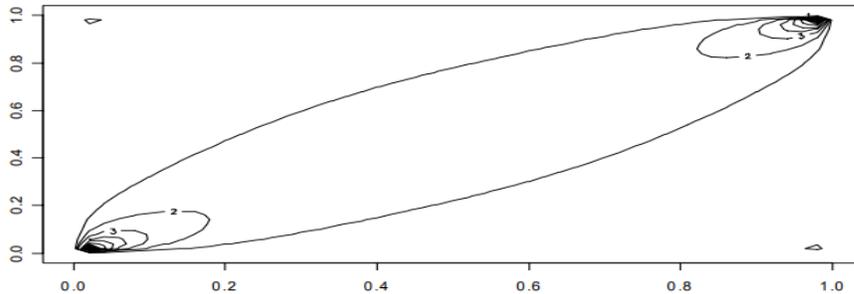


FIG.1.6-Contour de la densité du Copule de Student

Contour du Copule de Student

(1.10)

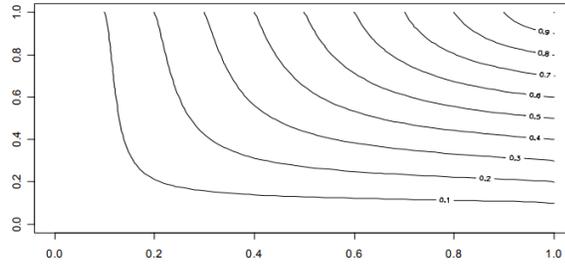


FIG.1.7-Contour du Copule de Student

Remarque 1.1.1

Les copules normale et de Student sont des copules symétriques et relativement simples d'utilisation du fait que l'on connaît bien les distributions auxquelles elles sont associées. Elles sont souvent appelées copules implicites car n'ayant pas de forme analytique explicite s'exprime par conséquent en fonction de distributions bivariées à eux associées par le théorème de Sklar.

Copule Archimédienne bivariée

Les copules Archimédiennes sont indiscutablement associées au statisticien canadien Christian Genest. Ce n'est pas lui qui les a inventées, mais il est le premier à avoir adopté une analyse statistique de ces fonctions copules. Et ses nombreuses publications ont largement contribué à les faire connaître.

Définition 1.1.2 Soit φ une fonction continue, strictement décroissante définie de I^2 dans $[0, +\infty]$ telle que $\varphi(1) = 0$, La fonction inverse généralisée de φ , notée $\varphi^{(-1)}$ avec $Dom\varphi^{(-1)} = [0, +\infty]$ et $Im\varphi^{(-1)} = I$ est définie comme suit :

$$\varphi^{(-1)} = \begin{cases} \varphi^{(-1)} & si & 0 \leq t \leq \varphi(0) \\ 0 & si & \varphi(0) \leq t \leq \infty \end{cases}$$

Remarquez que $\varphi(0)$ est continue et non-croissante sur $[0, \infty]$ et strictement décroissante sur $[0, \varphi(0)]$.

En outre $\varphi(\varphi^{(-1)}(u)) = u$ sur I et

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

Ainsi si $\varphi(0) = \infty$ alors, φ

Théorème 1.1.1 *Soient φ une fonction continue, strictement décroissante de I dans $[0, \infty]$ telle que $\varphi(1) = 0$ et $\varphi^{(-1)}$ la fonction inverse généralisée de φ .*

Soit C une fonction définie de I^2 dans I telle que :

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

Cette fonction est une Copule si et seulement si φ est convexe.

Ce type de copule est appelé Copule Archimédienne, et la fonction φ est connue sous le nom de générateur de la copule.

1.2 Mesure d'association

1.2.1 Mesure de dépendance

Dans cette partie, nous présentons la notion de mesure de la dépendance. Une mesure de dépendance donne une indication sur la mesure dans laquelle deux variables sont mutuellement liées. L'idée clé est d'évaluer une distance entre la copule associée à une paire donnée de variables aléatoires et la copule du produit.

Définition 1.2.1 (Mesures de dépendance) *Une fonction définie sur des variables aléatoires paires est une mesure de dépendance si elle satisfait les conditions dans la définition de concordance, avec les propriétés suivantes :*

1- $0 = \delta_{C_{\Pi}} \leq \delta_C \leq \delta_W = 1.$

2- $\delta_{X_1, X_2} = \delta_{C_{\Pi}} = 0$ ssi X_1 et X_2 sont indépendants.

Les mesures de dépendance les plus populaires sont définies comme suit :

Définition 1.2.2 (Phi de Hoeffding) *le Phi de Hoeffding pour un vecteur de variables aléatoires continues (X, Y) avec la copule C est donné par*

$$\phi = 3\sqrt{10 \int \int_{[0,1]^2} (C(u_1, u_2) - u_1 u_2)^2 du_1 du_2}.$$

Définition 1.2.3 (Sigma de Schweitzer-Wolff) *Sigma de Schweitzer-Wolff pour un vecteur de variables aléatoires continues (X, Y) avec la copule C est donné par :*

$$\sigma = 12\sqrt{\int \int_{[0,1]^2} |C(u_1, u_2) - u_1 u_2| du_1 du_2}.$$

Exemple de mesure de dépendance

Mesure σ de Schweizer et Wolffs : le rho de Sperman de deux v.a continue X, Y est définie par :

$$\rho_C = 12 \int_{I^2} \int (C(u, v) - uv) dudv,$$

cet intégral représente la volume entre la copule C est la copule produit Π . Si on change la différence $(C(u, v) - uv)$ par $|C(u, v) - uv|$, alors on obtient une mesure basée sur la distance L_1 entre le graphique de C et Π , cette distance représente la

mesure σ de Schweizer et Wolffs, elle est définie par :

$$\sigma_C = \sigma_{X,Y} = 12 \int_{I^2} \int |C(u, v) - uv| dudv. \quad (1.11)$$

Soit X, Y deux v.a continues d'une copule C . Alors la quantité σ_C définie dans [1.11](#) est une mesure de dépendance. Schweizer et Wolffs(1981) assure que toute distances entre les surfaces $z = C(u, v)$ et $z = uv$ représentent une mesure nonparamétrique de dépendance [89], [102]. $\forall 1 \leq p < \infty$, la distance L_p entre C et Π est définie par :

$$L_p = \left(k_p \int_{I^2} \int |C(u, v) - uv|^p dudv \right)^{\frac{1}{p}} \quad (1.12)$$

telle que k_p est une constante.

A partir la quantité dans [1.12](#) on définit les mesures de dépendances suivantes :

Mesure $\Phi_{X,Y}$: Si $p = 2$, alors nous avons

$$\Phi_{X,Y} = \Phi_C = \left(90 \int_{I^2} \int |C(u, v) - uv|^2 dudv \right)^{\frac{1}{2}},$$

tel que le carré de cette dernière, c'est à dire $\Phi_{X,Y}^2$ représente l'indice de dépendance entre les variables X et Y

Mesure $\Lambda_{X,Y}$: pour $p = \infty$, nous avons

$$\Lambda_{X,Y} = \Lambda_C = 4 \sup_{u,v \in I} |C(u, v) - uv|.$$

Après les définition de mesures dépendance σ_C , Φ_C et Λ_C on trouve que ces dernières sont basées sur le coefficient de Rho de Spearman. Il existe d'autres mesures de dépendance basées sur un autre coefficient telle que le coefficient de Gini.

1.2.2 Mesures de concordance

Caractérisation d'une mesures de concordance

Définition 1.2.4 (d'une mesures de concordance)

Une mesure d'association K entre deux variables aléatoires continues X et Y de copule C est une mesure de concordance si elle vérifie propriétés suivantes :

1. K est définie pour tout copule (X, Y) de variables aléatoires continues
2. $-1 \leq K_{X,-X} \leq K_{X,Y} \leq K_{X,X}$
3. $K_{X,Y} = K_{Y,X}$
4. Si X et Y sont indépendantes alors $K_{X,Y} = K_{Y,X}$
5. $K_{-X,Y} = K_{X,-Y} = -K_{X,Y}$
6. Si C_1 et C_2 sont deux copules telles que $C_1 < C_2$ alors on a : $K_{C_1} < K_{C_2}$
7. Si $\{(X_n, X_n)\}$ est une suite copules aléatoires dont la copule est C_n et si $\{C_n\}$ converge vers C alors K_{C_n} converge vers K_C
8. Si $\alpha(X)$ et $\beta(Y)$ sont des fonctions strictement croissantes, alors

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

Deux des mesures se concordance sont plus connues et jouent un très important rôle en statistique non paramétriques : le tau de kendall et rho de Spearman.

Tau de Kendall et rho de Spearman

Le tau de kendall Soit $(x_1, y_2); (x_2, y_2), \dots, (x_n, y_n)$ un échantillon de n observations d'un copule (X, Y) .

On définit le tau de Kendall (version échantillon) par :

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

Soient (X_1, X_2) et (Y_1, Y_2) deux observations d'un couple aléatoire (X, Y) . Si (X_1, X_2) et (Y_1, Y_2) sont i.i.d de loi H ; on définit le tau Kendall (version population) par :

$$\tau_{XY} = P[(X_1 - X_2)(X_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

Soit X et Y deux variables aléatoires continues de copule C alors le tau de Kendall X et Y est donné par :

$$\tau_{XY} = \tau_c = Q(c, c) = 4 \int \int_{I^2} C(u, v) dC(u, v) - 1$$

En fait, l'intégrale $\int \int_{I^2} C(u, v) dC(u, v)$ est la moyenne de la variable aléatoire $C(U, V)$ où U et V sont uniformément distribués sur I ; donc

$$\int \int_{I^2} C(u, v) dC(u, v) = E[C(U, V)]$$

Ainsi, on peut écrire

$$\tau_c = 4E[C(U, V)] = 4 \int_0^1 dF_C(t) - 1 = 3 - 4 \int_0^1 F_c(t) dt$$

où F_c est la distribution de la variable aléatoire $C(U, V)$

Le rho de Spearman Tout comme le tau de Kendall, le rho de Spearman est une mesure de dépendance basée sur la notion de concordance. Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois vecteurs aléatoires indépendants de même distribution H dont les distributions marginales sont F et G et dont la copule associée est C .

a) *Définition*

La version populatoire du rho de Spearman est définie comme étant proportionnelle à la différence de la probabilité de concordance et celle de discordance des couples aléatoires $(X_1, Y_1), (X_2, Y_2)$:

$$\rho = \rho_{X,Y} = \rho_c = 3([P((X_1, Y_1)(X_2, Y_2) > 0)] - P((X_1, Y_1)(X_2, Y_2) < 0))$$

b) *Propriétés*

La distribution de (X_1, Y_1) étant H et celle de (X_2, Y_2) étant Π (car les variables X_2 et Y_2 sont indépendantes) alors d'après ce qui précède on a le théorème suivant.

Théorème 1.2.1

Soit (X, Y) un couple de variables continues de copule. la version populatoire du rho de Spearman pour X et Y est donnée par :

$$\rho_{XY} = \rho_C = 3Q(C, \Pi) = 12 \int \int_{I^2} uv dC(u, v) - 3 = 12 \int \int_{I^2} C(u, v) dudv - 3$$

Car les variables U, V sont uniforme d'espérance $E(U) = E(V) = \frac{1}{2}$, et de variances $Var(U) = Var(V) = \frac{1}{12}$, alors ρ_C peut s'écrire :

$$\rho_C = \frac{E(UV) - E(U)E(V)}{\sqrt{Var(U)}\sqrt{Var(V)}}.$$

Exemple 1.2.1

Pour la copule de Marshall-Olkin, de paramètres $0 < \alpha$ et $\beta < 1$

définie par :

$$C_{\alpha,\beta}(u, v) = \begin{cases} u^{1-\alpha}v & \text{si } u^\alpha \geq v^\beta \\ uv^{1-\beta} & \text{si } u^\alpha \leq v^\beta \end{cases}$$

donc

$$\int_{I^2} \int C_{\alpha,\beta}(u,v) dudv = \frac{1}{2} \left(\frac{\alpha + \beta}{2\alpha - \alpha\beta + 2\beta} \right),$$

le rho de Spearman est

$$\rho_{\alpha,\beta} = \frac{3\alpha\beta}{2\alpha - \alpha\beta + 2\beta}.$$

Copule	Demaine de θ	Tau de Kendall	Rho de Spearman
Produit	<i>N.A</i>	0	0
FGM	$-1 \leq \theta \leq 1$	$\frac{2}{9}\theta$	$\frac{1}{3}\theta$
Clayton	$0 \leq \theta \leq \infty$	$\frac{\theta}{\theta+2}$	<i>Pas de forme fermée</i>
Frank	$-\infty \leq \theta \leq \infty$	$1 - \frac{4(1-D_1(\theta))}{\theta}$	$t - \frac{1-\exp(-\theta)}{\theta} \log \left(\frac{1-\exp(-\theta)}{1-\exp(-\theta t)} \right)$
Normale	$-1 \leq \theta \leq 1$	$\frac{2}{\pi} \arcsin(\theta)$	$\frac{6}{\pi} \arcsin\left(\frac{\theta}{2}\right)$

Tab.2.1 Tau de Kendall et RhodeSpearman de quelques copules (1.13)

Chapitre 2

Modélisation de l'analyse de survie

2.1 Donnée de survie

Les données de survie représentent le temps écoulé entre le début d'une observation et l'arrivée d'un événement. Le cas d'événement le plus simple est le décès, cependant, le terme « *donnée de survie* » couvre d'autres événements, comme l'apparition d'une maladie ou d'une épidémie. Dans l'industrie, il peut s'agir de panne d'une machine, ou en économie, du temps écoulé pour qu'une personne accepte un travail...etc.

Dans plusieurs cas l'événement est la transition d'un état à un autre. Par exemple, le décès est la transition de l'état « *vivant* » vers l'état « *mort* ». L'apparition d'une maladie est la transition de l'état « *en santé* » vers l'état « *malade* ». La figure (2.1) illustre ces deux exemples.

Selon le contexte, les termes décès, événement, échec ou transition peuvent être utilisés pour désigner l'événement constaté, et plus précisément ce qui se passe au temps de la réponse. Dans certains cas, l'aspect intéressant est la transition correspondante à l'incidence d'une maladie, et dans d'autres cas, l'aspect intéressant est l'état de la

disparition d'une maladie (Hougaard,1999).

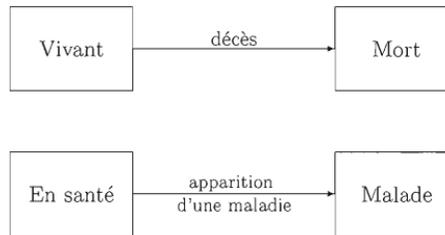


Figure .2.1 Modèles de temps de survie (2.1)

2.1.1 Les caractéristiques des données de survie

Les dates et délais principaux

Dans l'étude de survie, nous disposons d'un échantillon de n individus ou bien sujets qui possède des informations principales sur leurs dates et les différentes durées de survie.

Les principales dates

1. **Date d'origine DO** : c'est la date du début de l'observation, c'est-à-dire l'origine du début de l'analyse de survie, elle correspond au temps égal à zéro $t = 0$.
2. **Date des dernières nouvelles DDN** : pour faire l'analyse des résultats, chaque individu dispose d'une date de dernière nouvelle, qui est la date la plus récente où les informations ont été recueillies. Par exemple, dans le cas où le sujet a subi l'événement, sa date de dernière nouvelle est la date de survenue de cet événement.
3. **Date de point DP** : Nous ne pouvons pas attendre la survenue de l'événement pour tous les sujets pour faire l'analyse des observation, donc il y a ce que nous appellons la date de point qui est une date au-delà de laquelle nous arrêtons

l'observation et nous ne tenons plus compte de l'état du sujet après cet instant.

Les principaux délais

1. **Durée de surveillance** : c'est la durée entre la date initiale et la date des dernières nouvelles, c'est la période de surveillance jusqu'à l'instant de la survenue de l'événement d'un sujet i avec $T_i \in [DO, DDN]$.

2. **Recul** : noté par L_i la variable de la censure d'un individu i avec $L_i \in [DO, DP]$. c'est le délai écoulé entre la date d'origine et la date de point c'est-à-dire, le délai maximal d'observation du sujet. il est dit aussi le délai de la censure, il peut être connu ou inconnu, déterministe ou aléatoire, dépendant de l'individu ou non.

3. **Le temps de participation** : noté par t_i , c'est le délai réellement observé pour chaque individu i . noté t_i c'est une variable aléatoire qui dépend de la date de point et la date des dernières nouvelles. Nous distinguons deux cas :

a- *La date des dernières nouvelles est inférieure à la date du point :*

$$DDN < DP \iff t_i \in [DO, DDN]$$

Dans ce cas, le temps de participation est identique à la durée de surveillance ie :

$$t_i = T_i$$

Si le sujet n'a pas subi l'événement à la date des dernières nouvelles, il sera considéré comme perdu de vue, le temps de participation nous donnera une observation incomplète or que dans le cas où l'individu a subi l'événement à la date des dernières nouvelles, le temps de participation nous donnera une observation complète de son état.

b- La date des dernières nouvelles est supérieure à la date du point :

$$DDN > DP \iff t_i \in [DO, DP]$$

Dans ce cas le temps de participation est identique au temps de recul :

$$t_i = L_i$$

Le sujet est considéré comme exclu vivant, car s'il n'a pas encore subi l'événement avant la date de point, c'est son état initial qui sera pris en considération même s'il y a un changement d'état après la date de point.

Remarque 2.1.1

Les sujets perdus de vue et les sujets exclus-vivants correspondent aux observations censurés mais les deux mécanismes de censure sont de natures différentes que nous expliquerons en détail prochainement.

Nous résumons dans le schéma suivant, les différentes dates et les durées de survie vues dans cette section.

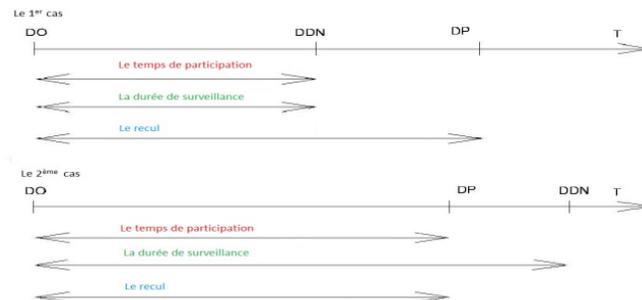


Figure 2.2 les différentes dates et les durées de survies (2.2)

2.1.2 Distributions de la durée de survie

Soit T la durée avant la survenue d'un événement. Alors T est une variable aléatoire positive dont la distribution peut être caractérisée par plusieurs fonction, dont deux ont une interprétation plus importante : la fonction de survie et la fonction de risque.

La fonction de survie

Définition 2.1.1

La fonction de survie, pour t fixé, est la probabilité de survivre jusqu'à l'instant t :

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx \quad t \geq 0, \quad (2.3)$$

où $f(x)$ est la fonction de densité de T . La fonction de survie est le complément de la fonction de répartition, c'est-à-dire

$$S(t) = 1 - F(t), \text{ où } F(t) = P(T \leq t).$$

c'est une fonction toujours non croissante à valeur dans $[0, 1]$ et c'est pourquoi on préfère parfois travailler avec la fonction de risque, qui représente comment le risque de survenue de l'événement étudié évolue dans le temps.

Propriété 2.1.1

La fonction de survie d'une v.a. T est décroissante monotone continue à gauche et vérifie

$$\lim_{t \rightarrow 0} S(t) = 1 \text{ et } \lim_{t \rightarrow \infty} S(t) = 0.$$

Remarque 2.1.2

1. F est aussi appelée fonction de distribution et $S(t)$ est la queue de distribution.

2. On peut définir $F(t) = 1 - S(t)$, pour tout $t \geq 0$,

La fonction de risque

Le risque instantané (ou taux d'incidence), pour un t fixé, est proportionnel à la probabilité que l'événement se produise juste après t sachant qu'il n'a toujours pas eu lieu à t . La fonction de risque $h(t)$ est définie par

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (2.4)$$

De [2.3](#) et [2.4](#), on peut établir les relations suivantes entre la fonction de survie et la fonction de risque instantané :

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log[S(t)]$$

et

$$S(t) = \exp \left\{ - \int_0^t h(x) dx \right\} = e^{-H(t)} \quad (2.5)$$

où

$$H(t) = \int_0^t h(x) dx$$

est définie comme la fonction de risque cumulé. Ainsi, à partir de l'équation ?? pour tout individu, plus grand est le risque de survenue d'un événement, plus petite est sa probabilité de survie avant cet événement (Tableman & kim, 2003).

Fonction de Densité

Comme toute autre v.a. continue, la durée de survie X a une fonction de densité de probabilité.

Si F admet une dérivée par rapport à la mesure de Lebesgue sur \mathbb{R}_+ , la fonction de

densité de probabilité existe, elle est définie pour tout $t \geq 0$, par :

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} = \lim_{dx \rightarrow \infty} \frac{P(t \leq X \leq t + dt)}{dt}$$

Pour t fixé, la densité de probabilité caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant t .

2.1.3 Les mécanismes de censure

La particularité des données de survie réside dans le fait que ces données sont souvent incomplètes à cause du phénomène de censure. Les méthodes statistiques pour ce genre de données doivent alors tenir compte des données censurées. Différents types de censure sont rencontrés.

★*La censure à droite* : La durée de vie est dite censurée à droite si elle n'est pas observée mais que l'on sait qu'elle est supérieure à une certaine valeur connue. On peut avoir une censure à droite due à la fin de l'étude ou due à la sortie du sujet (*sans qu'il n'ait subi l'évènement*) avant la fin de l'étude. Sur la figure 2.3, le patient 1 est suivi depuis le début de l'étude ($t = 0$) et il a subi l'évènement au temps T_1 , sa durée est donc observée. Le patient 2 est entré dans l'étude et a été suivi jusqu'à la fin sans jamais subir l'évènement, ce qui implique que sa durée T_2^+ est censurée à droite. Enfin, le patient 3, entré en cours d'étude, a été perdu de vue avant la fin, d'où sa durée T_3^+ qui est aussi censurée à droite.

★*La censure à gauche* : Elle se produit lorsqu'on commence à observer un échantillon à un point dans le temps où certains sujets pourraient déjà avoir vécu l'évènement.

★*La censure par intervalle* : Elle se rencontre, par exemple, quand on constate l'apparition d'un symptôme lors d'une visite médicale : il est apparu depuis la visite

précédente, mais on ne sait pas exactement quand.

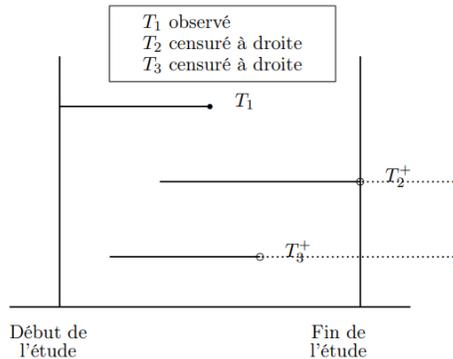


Figure 2.3 : Exemple de censure à droite (2.6)

2.2 Modèle de survie

2.2.1 Les méthodes d'analyse des données de survie

Dans cette section, nous ne considérons que la censure à droite. Soit \bar{T} la variable aléatoire qui définit la durée de survenue d'un événement. Soit C la variable de censure de \bar{T} . On définit

$$T = \min(\bar{T}, C) \text{ et } \delta = I(\bar{T} \leq C), \text{ où } \delta = 1 \text{ si } \bar{T} \leq C \text{ et } \delta = 0 \text{ si } \bar{T} > C.$$

Ainsi, pour une analyse de survie, on observe l'échantillon $(t_i, \delta_i), i = \overrightarrow{1, \dots, n}$. De plus, on peut s'intéresser aux caractéristiques de l'individu, par exemple le sexe, l'âge, le traitement reçu qui sont codées pour faciliter l'utilisation dans un vecteur de covariables Z , ce qui donne comme échantillon $(t_i, \delta_i, Z_i), i = \overrightarrow{1, \dots, n}$. Il existe plusieurs approches pour analyser les données de survie.

On distingue les méthodes non-paramétriques qui estiment la distribution de la fonction de survie sans avoir recours à un modèle, des méthodes paramétriques et des méthodes semi-paramétriques qui font appel à un modèle.

La méthode de Kaplan-Meier pour la fonction de survie

Elle est la plus utilisée des méthodes non paramétriques. Soit $t_1^* < t_2^* < \dots < t_D^*$, D temps de décès (*événement*) observés distincts et ordonnés, d_i le nombre d'évènements survenant au temps t_i^* et Y_i^* le nombre d'individus à risque, c'est-à-dire qui n'ont pas subi l'évènement et qui ne sont pas encore censurés jusqu'au temps t_i^* . L'estimateur Kaplan-Meier (1958) $\hat{S}(t)$ de la fonction de survie $S(t)$ est défini comme suit :

$$\hat{S}(t) = \begin{cases} 1, & \text{si } t < t_1^* \\ \prod_{t_i^* \leq t} (1 - \frac{d_i}{Y_i^*}), & \text{si } t \geq t_1^* \end{cases} \quad (2.7)$$

L'estimateur $\hat{S}(t)$ est une fonction en escalier non croissante et continue à droite. Il est également appelé «*Estimateur Produit-Limite*». L'estimateur de Nelson-Aalen(1972,1978) de la fonction de risque cumulé $H(t)$ est défini par :

$$\hat{H}(t) = \begin{cases} 0, & \text{si } t < t_1^* \\ \sum_{t_i^* \leq t} \frac{d_i}{Y_i^*}, & \text{si } t \geq t_1^* \end{cases} \quad (2.8)$$

La figure (2.4) montre un exemple de la courbe de la fonction de survie estimée pour les données d'un essai clinique pour évaluer l'efficacité du maintien de la chimiothérapie pour une leucémie myéloïde aigüe (Tableman & Kim, 2003). Sur cette figure, on compare la durée de survie avant rechute des malades de leucémie myéloïde aigüe qui continuent le traitement de chimiothérapie avec ceux qui ont arrêté la chimiothérapie. Elle montre que la durée avant rechute des malades qui ont continué le traitement est plus longue que pour les malades qui ont arrêté le traitement.

Les méthodes paramétriques

Dans ces méthodes, on suppose une distribution particulière pour la durée de survie. On formule ainsi le modèle en spécifiant de façon paramétrique l'une des fonctions

de la durée de vie.

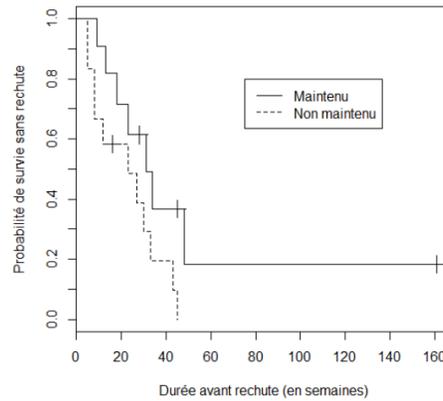


Figure 2.4: L'estimateur Kaplan–Meier de la fonction de survie pour les données AML.. (2.9)

.La ligne pleine représente la courbe de survie des malades qui ont poursuivi la chimiothérapie.

.La ligne brisée est la courbe de survie des patients qui ont arrêté la chimiothérapie.

.Les estimateurs des paramètres sont obtenus par la méthode du maximum de vraisemblance.

On peut citer comme modèles :

La loi exponentielle : C'est la seule loi qui donne un risque instantané constant. Sa fonction de densité est

$$f(x) = \lambda \exp(-\lambda x), \lambda > 0, x \geq 0.$$

La loi de Weibull : Sa densité est

$$f(x) = \frac{\alpha}{\lambda} \left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^\alpha\right\}, \quad \alpha, \lambda > 0, x \geq 0.$$

où α est le paramètre de forme et λ le paramètre d'échelle.

La loi gamma : Sa fonction de densité est :

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x), \quad \alpha, \lambda > 0, x \geq 0.$$

Autres lois : Plusieurs autres lois sont aussi utilisées en analyse de survie, par exemple la loi log-normale, la loi log-logistique, la loi gaussienne inverse ou la loi gamma généralisée (*Klein & Moeschberger, 1997*).

Dans l'approche paramétrique, les distributions peuvent dépendre de variables explicatives (covariables, comme les caractéristiques des individus) susceptibles d'influencer la survie. Considérons Z un vecteur de covariables. On suppose que la relation entre la distribution de survie et les covariables peut être étudiée par un modèle de risques proportionnels quand

$$h(t) = h_0(t) \exp(\beta^t Z),$$

où β est le vecteur des coefficients de régression et $h_0(t)$ la fonction de risque quand $Z = 0$. Par exemple, si on suppose que la durée de vie suit une loi de Weibull de paramètres α et λ alors on a

$$h_0(t) = \frac{\alpha}{\lambda} \left(\frac{t}{\lambda} \right)^{\alpha-1}$$

et le modèle à analyser sera

$$h(t) = \frac{\alpha}{\lambda} \left(\frac{t}{\lambda} \right)^{\alpha-1} \exp(\beta^t Z).$$

Les méthodes semi-paramétriques

Ces méthodes sont plus robustes que les méthodes paramétriques car elles sont moins contraignantes en termes d'hypothèses et plus efficaces que les méthodes non paramétriques du type Kaplan-Meier pour lesquelles on doit se limiter à l'estimation de la fonction de survie dans les groupes définis par les covariables (par exemple groupe traité versus groupe non traité). Elles ne supposent aucune distribution particulière de la survie et permettent de modéliser l'effet de covariables sur la distribution de la durée de vie. Le modèle de Cox, introduit par David Cox (1972), est la plus utilisée de ces méthodes en analyse des durées de vie.

2.2.2 Copule de survie

Dans de nombreuses applications, les variables aléatoires d'intérêt représentent la durée de vie des individus ou des objets dans une population. La probabilité de survie dans le cas univarié est définie par :

$$P(X > x) = 1 - F(x)$$

telle que F représente la fonction de répartition de X .

De même pour le cas bivarié, si F est la fonction jointe de survie est :

$$\bar{F}(x, y) = P(X > x, Y > y)$$

et les marginales sont \bar{F}_1 et \bar{F}_2 , alors :

$$\begin{aligned}
 \bar{F}(x, y) &= P(X > x, Y > y) \\
 &= 1 - P(X \leq x, Y \leq y) \\
 &= 1 - F_1(x) - F_2(y) + F(x, y) \\
 &= \bar{F}_1(x) + \bar{F}_2(y) - 1 + C(1 - \bar{F}_1(x), 1 - \bar{F}_2(y)).
 \end{aligned}$$

Donc la copule de survie \hat{C} de X, Y est :

$$\hat{C}(\bar{F}_1(x), \bar{F}_2(y)) = \bar{F}_1(x) + \bar{F}_2(y) - 1 + C(1 - \bar{F}_1(x), 1 - \bar{F}_2(y)).$$

De sorte que si nous définissons la fonction \hat{C} de I^2 dans I par :

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

on a :

$$\bar{F}(x, y) = \hat{C}(\bar{F}_1(x), \bar{F}_2(y))$$

a.1 Survie bivariée

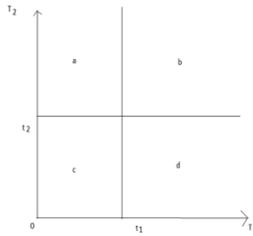


Figure 2.5 $F_1(t_1) = a + c$; $F_2(t_2) = c + d$; $S(t_1, t_2) = b$; $F(t_1, t_2) = c$ (2.10)

Chapitre 3

Modèle de fragilité et copule

3.1 Modèle de fragilité

3.1.1 Modèles de Fragilité”UNIVARIÉS”

Généralement, les modèles de survie supposent que les données de survie des patients sont de formes indépendantes et que chaque distribution des temps de survie individuels des patients est la même (temps d'évènements indépendants et identiquement distribués). Cette hypothèse de base implique une population homogène. Toutefois, dans de nombreuses situations pratiques, les patients observés diffèrent notablement. L'effet d'une drogue, d'un traitement ou l'influence de diverses variables explicatives peut être très différent entre des sous-groupes de patients. Pour rendre compte d'une telle hétérogénéité non observée dans la population étudiée, Vaupel et al.(57) ont présenté des modèles de fragilité univariés pour l'analyse de survie. L'idée maîtresse est que les individus possèdent des fragilités différentes, et que les patients les plus fragiles vont mourir plus tôt que les autres. En conséquence, il y a une sélection systématique des individus robustes, *i.e.* les patients avec une faible fragilité. Lorsque les taux de mortalité sont estimés, on peut s'intéresser à la façon dont ces taux changent

au cours du temps ou de l'âge. Très souvent, lors d'une maladie, par exemple, nécessitant une intervention chirurgicale, on observe que la fonction de risque augmente au début, atteint un maximum, puis diminue (intensité unimodale) et reprend les valeurs "normales" de risque de décès liées à l'âge (illustration figure(3.1)). Plus le patient vit après la manifestation de la maladie, plus ses chances de revenir à un risque "naturel" augmentent. Il est probable que intensités unimodales, qui retrouvent le risque "naturel", soient souvent le résultat d'un processus de sélection agissant dans une population hétérogène et ne reflètent pas la mortalité individuelle. Le taux de mortalité de la population hétérogène, sans en tenir compte, peut décliner tout simplement parce que les personnes à risque élevé (grande fragilité) sont déjà mortes. La force de mortalité d'un individu donné pourrait être sur-évaluée à cause de l'hétérogénéité de l'échantillon étudié quand cette hétérogénéité n'est pas prise en compte.

Si les facteurs de protection ou les facteurs de risque sont connus, ceux-ci peuvent être inclus dans l'analyse en utilisant le modèle à risques proportionnels, qui est de la forme de l'équation

$$\lambda(t, Z) = \lambda_0(t) \exp(\beta'Z)$$

.Le "confort mathématique" de ce modèle réside dans la séparation des effets du vieillissement dans la fonction de risque de base des effets des covariables dans la partie paramétrique.

Il existe deux principales raisons pour lesquelles il est souvent impossible d'inclure tous les facteurs importants sur le plan individuel dans l'analyse. Parfois, il y a trop de variables explicatives à prendre en compte dans le modèle, et dans d'autres cas, les chercheurs ne connaissent pas ou ne sont pas capables de mesurer toutes les covariables. Dans les deux cas, il existe deux sources de variation dans les données de survie : la *variabilité* représentée par les facteurs de risques mesurables, qui est donc théoriquement "prévisible" ; et l'hétérogénéité causée par les covariables inconnues,

qui est donc théoriquement imprévisible, à moins de connaître toutes les information pertinentes (ce qui, comme on peut le montrer, ne sera jamais le cas). Il y a des avantages à séparer ces deux sources de variation car l'hétérogénéité contrairement à la variabilité peut expliquer certains résultats ; comme par exemple dans le cas où les risques sont non proportionnels.

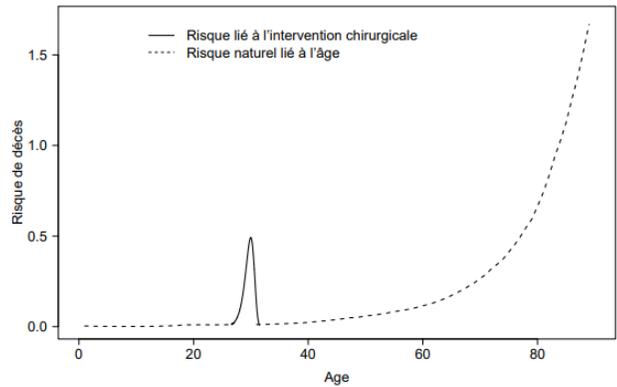


Fig.3.1—Exemple de risque de décès pour un patient subissant une intervention chirurgicale. (3.1)

Dans un modèle à risques proportionnels, la négligence d'un sous ensemble de covariables importantes conduit à des estimations biaisées des coefficients de régression et du taux de risque. La raison d'un tel biais réside dans le fait que le taux de risque temps-dépendant se traduit par des changements différents dans la composition de la population étudiée dans le temps selon les variables explicatives.

Supposons un échantillon étudié dans lequel certaines personnes "éprouvent" un plus grand risque d'évènement que d'autres, intuitivement ces personnes expérimenteront l'évènement plus tôt. Ainsi, le reste des personnes à risque a tendance à former un groupe plus ou moins sélectionné avec un risque d'évènement plus faible. Une estimation du taux de risque individuel, sans tenir compte de la fragilité non observée, devrait donc sous-estimer la véritable fonction de risque et l'ampleur de la sous-estimation devrait augmenter à mesure que le temps avance.

Le modèle de fragilité univarié est une extension du modèle de Cox dans lequel le risque d'un individu dépend en plus d'une variable aléatoire non observable Y , qui

agit multiplicativement sur la fonction de risque de base λ :

$$\lambda(t|Y, Z) = Y\lambda(t|Z) = Y\lambda_0(t) \exp(\beta Z). \quad (3.2)$$

La fragilité Y est une variable aléatoire variant avec la population et qui diminue ($Y < 1$) ou augmente ($Y > 1$) le risque individuel. Le point le plus important ici est que la fragilité est non observable. La fonction de survie S respective est donnée par

$$\begin{aligned} S(t|Y, Z) &= \exp(-\Lambda(t|Y, Z)), \\ &= \exp\left(-\int_0^t \lambda(u|Y, Z) du\right), \\ &= \exp\left(-Y \exp(\beta Z) \int_0^t \lambda_0(u) du\right), \end{aligned} \quad (3.3)$$

$S(t|Y, Z)$ peut être interprétée comme la fraction d'individus survivant au temps t après le début du suivi sachant le vecteur de covariables observables Z et la fragilité Y . Notons, que les relations [3.2](#) et [3.3](#) décrivent le même modèle en utilisant différentes notions. Jusqu'à présent, le modèle a été décrit au niveau des individus. La fonction de survie de la population totale est la moyenne des fonctions de survie sur la fonction de fragilité [3.3](#). Elle peut être considérée comme la fonction de survie d'un individu tiré aléatoirement, et correspond à ce qui est réellement observé. Ce qui peut être observé dans la population est le résultat net d'un certain nombre de personnes avec différents valeurs de Y .

3.2 Modèle de fragilité pour copules

Les modèles à effets aléatoires sont importants dans les études biologiques et épidémiologiques, car ils présentent une méthode d'hétérogénéité de modélisation.

Un modèle à effet aléatoire qui est particulièrement adapté à l'analyse de survie

multivariée est le modèle de fragilité. Pour décrire ce modèle, nous définissons la fonction de survie comme suit :

$$S(t) = P(T > t) = 1 - F(t)$$

où F est la fonction de distribution de T .

La relation entre la fonction de risque instantané de *Cox* $h(t)$ et la fonction de survie $S(t)$ est donnée par :

$$h(t) = -\frac{\partial \ln S(t)}{\partial t} = \frac{f(t)}{S(t)}$$

où $f(t)$ est la densité de probabilité de T .

Dans l'équation

$$h(t) = h(t, Z_1, Z_2, \dots, Z_P) = h_0(t) \exp \left[\sum_{j=1}^p \beta_j Z_j \right] \quad (3.4)$$

posons

$$\gamma = \exp \left[\sum_{j=1}^p \beta_j Z_j \right].$$

Alors, on peut exprimer la fonction de survie par rapport à γ comme suit :

$$S(t|\gamma) = \exp \left[- \int_0^t h(s) ds \right] = [B(t)]^\gamma \quad (3.5)$$

où

$$B(t) = \exp \left[- \int_0^t h_0(s) ds \right]$$

est la fonction de survie correspondant au risque de base.

Les modèles de fragilité se posent lorsque Z_j , et donc γ , est inobservable. Le paramètre γ s'appelle le facteur de fragilité, et quand γ augmente, la fonction de survie

diminue. Donc, on peut exprimer la fonction de survie moyenne par :

$$S(t) = E_{\gamma} [S(t|\gamma)] \quad (3.6)$$

Pour les modèles de fragilité multivariés, si nous supposons que les durées de vie T_1, T_2, \dots, T_p sont indépendantes par rapport à γ , nous aurons :

$$\begin{aligned} P [T_1 > t_1, \dots, T_p > t_p | \gamma] &= P [T_1 > t_1 | \gamma] \dots P [T_p > t_p | \gamma] \\ &= S_1(t_1 | \gamma) \dots S_p(t_p | \gamma) \\ &= [B_1(t_1)]^{\gamma} \dots [B_p(t_p)]^{\gamma} \end{aligned} \quad (3.7)$$

D'où la fonction de survie multivariée est :

$$P [T_1 > t_1, \dots, T_p > t_p] = E_{\gamma} [B_1(t_1) \dots B_p(t_p)]^{\gamma} \quad (3.8)$$

3.2.1 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

Un modèle avec fragilité est donné par :

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

* $h_{i,j}(t)$ la fonction de risque instantané au temps t pour un individu $i = \overrightarrow{1, \dots, n}$, 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 diagnostique 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 φ est l'inverse d'une transformée de Laplace,

$$C(u, v) = \varphi \{ \varphi^{-1}(u) + \varphi^{-1}(v) \}, \text{ avec } \varphi(0) = 1.$$

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 :

$$S_W(t_1, t_2) = \tilde{C}(S_{1,W}(t_1) S_{2,W}(t_2)) = \varphi_W \{ \varphi_W^{-1}(S_{1,W}(t_1)) + \varphi_W^{-1}(S_{2,W}(t_2)) \}. \quad (3.9)$$

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

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

avec

$$\Lambda_{j,w_i}(t) = \int_0^t h_{j,w_i}(s) ds, \quad j = 1, 2$$

la fonction du risque cumulé. En suite, en intégrant par rapport à la densité de la

fragilité, la fonction de survie jointe du modèle sera

$$S(t_1, t_2) = \int_0^\infty S_W(t_1, t_2) g_W(w) dw = E[\exp\{-W(\Lambda_{1,w_i}(t_1) + \Lambda_{2,w_i}(t_2))\}] \quad (3.10)$$

$$= \varphi\{\Lambda_{1,w}(t_1) + \Lambda_{2,w}(t_2)\}, \quad (3.11)$$

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

$$S_j(t) = \varphi\{\Lambda_{j,w}(t)\} \Rightarrow \Lambda_{j,w}(t) = \varphi^{-1}(S_j(t)), \quad (3.12)$$

d'où, si nous remplaçons [3.12](#) en [3.11](#), la fonction de survie jointe du modèle avec fragilité devient :

$$S(t_1, t_2) = \varphi\{\varphi^{-1}(S_1(t_1)) + \varphi^{-1}(S_2(t_2))\}. \quad (3.13)$$

De [3.9](#) et [3.13](#), nous remarquons que les deux modèles sont de nature différente car la copule utilisée dans les fonctions de survie jointes dans [3.9](#) et [3.13](#) est la même mais les fonctions marginales de survie ne sont pas les mêmes.

3.2.2 Copule de Clayton-Oakes et modèle à fragilité 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 θ est donnée par :

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

L'application directe de l'équation [3.9](#) pour la fonction de survie jointe de copule de Clayton Oakes donne

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

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 :

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

Cette expression ressemble à la forme de la copule représentée précédemment, mais

$$S_{j,m}(t) \neq S_{j,c}(t), \quad \forall j = 1, 2.$$

Conclusion

Nous avons dans ce mémoire parlé des copules et de leurs caractéristiques de base, le travail s'est concentré sur leur application aux données de survie. Le but de l'utilisation de copules est de découvrir les relations entre les composantes de survie bivariée. Un autre outil supplémentaire souvent utilisé pour modéliser les données de survie bivariées est la saisie de paramètres aléatoires individuels qui sont souvent interprétés comme des paramètres de fragilité. Pour cela nous avons détaillés les définitions et les propriétés de la famille Archimédienne, particulièrement la copule de Clayton-Oskes, ceci en les modélisant avec une fragilité de type gamma.

Le développement de la théorie des copules d'une part et celui de la fragilité permettront la production de beaucoup d'autres résultats intéressants dans ce domaine.

Bibliographie

- [1] Mémoire de *BENBRAIKA* , l'Université Mohamed khider ,BISKRA , page (22-23-30) figure(31-32-33-34-35)
- [2] *Amel CHINE* , Sur la statistique de copules ,thèse Doctorat En Science , l'Université Mohamed khider , BISKRA, page 6-7.
- [3] *BENELMIR Imane* , C (2018) , Modélisation de Dépendance par les Copules , l'Université Mohamed khider ,BISKRA , page 8-9-10.
- [4] *ABDERRAHMANE Nour El Houda* et *ZEGHMAR Radia* , C(2015-2016) , Modélisation de Dépendance par les Copules , Mémoire du Diplôme de Mastre , l'Université M'hamed Bougara Boumerdes , page 34.
- [5] *BETTEKA Samah* , C (2017) , LES VALEURS EXTREMES BIVARIEES , THESE DOCTORAT EN SCIENCES EN MATHÉMATIQUES , l'Université Mohamed khider ,BISKRA , Page 70-71-72
- [6] *Guessouri Amel* , C (2019) , Sur les Copules et applications , Mémoire Master en Mathématiques , l'Université Mohamed khider ,BISKRA , page 15-16
- [7] *YASSIR RABHI* , C (2006) , MODÉLES DE SURVIE AVEC UN POINT DE RUPTURE , MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À MONTRÉAL , page 3-4
- [8] *Lyasmine HARROUCHE* , C (2018) , Analyse statistique des modèles de survie , Diplôme de Mastre , UNIVERSITE MOULOUD MAMMERI de TIZI-OUZOU , page 12-13

- [9] *Semmari Manal* , C (2020) , Sur l'Analyse de Survie et Applications , DIPLÔME DE MASTER EN MATHÉMATIQUES , l'Université Mohamed khider ,BISKRA , Page 9-10
- [10] *VANESSA ROUSSEAU* , MODÈS FRAGILISÉS DE SURVIE RELATIVE APPLIQUÉS AUX DONNÉES CENSURÉES DANS LE CAS DE LA TRANSPLANTATION RÉNALE , THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE MONTPELLIER 1, page 39-40-41
- [11] *Houdété Odilon Loko* , Analyse de l'impact d'une intervention à grande échelle avec le modèle de risques proportionnels de Cox avec surpluse de zéros : application au projet Avahan de lutte contre le VIH\SIDA en Inde , **MÉMOIRE UNIVERSITÉ LAVAL**, page 11(FIGURE)-12-13-14
- [12] *Mohamad Ghassani* , C (2006) , Dynamiques épidémiques,risques et copules , THESE DOCTORAT DE L'UNIVERSITÉ DE GRENOBLE , page 40
- [13] *Idiou Nesrine* , C(2021) , Copulas and frailty models in multivariate survival data ,thèse Doctorat En Science l'Université Mohamed khider ,BISKRA , page 116(figure)-119-120-121
- [14] **www.bf.refer.org** , bf.refer.org/barro/chap3.html , Copules des valeurs extrêmes bivariées(19|02|2023....13 :48)

RÉSUMÉ

Dans ce mémoire nous avons présentés les copules, ainsi que leurs propriétés essentielles, particulièrement la famille Archimédienne. Le cas de la copule de Clayton-Oakes est prise comme exemple par la suite. L'analyse de survie est aussi introduite ici pour servir dans l'étude de la fragilité et son utilisation dans les modèles de fragilité pour les copules d'une façon générale.

ABSTRACT

In this thesis we present the copulas, as well as their essential properties, particularly the Archimedean family. The case of the Clayton –Oakes copula is taken as an example thereafter. Survival analysis is also introduced here for use in the study of fragility and its use in fragility models for copulas in general.

ملخص

في هذي الأطروحة نقدم الكوبيلات فضلا عن خصائصها أساسية ولا سيما عائلة الارخميدس يتم اخذ حالة كوبولا كاليتون أو كس كمثال, فيما بعد تم تقديم تحليل بقاء هنا أيضا لاستخدامه في دراسة الهشاشة واستخدامها في نماذج للكوبيلا بشكل عام