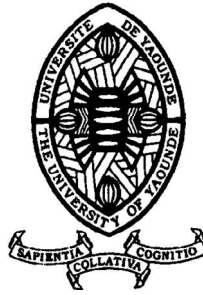


REPUBLIQUE DU CAMEROUN

Paix – Travail – Patrie

UNIVERSITE DE YAOUNDE I
ECOLE NORMALE SUPERIEURE
DEPARTEMENT DE Mathematiques



REPUBLIC OF CAMEROUN

Peace – Work – Fatherland

UNIVERSITY OF YAOUNDE I
HIGHER TEACHER TRAINING COLLEGE
DEPARTMENT OF Mathematics

Mesures de dependance de risques

Mémoire de D.I.P.E.S II de Mathematiques

Par :

THEPI NKUIMENI Franck Eric
Licencie en Mathematiques

Sous la direction
FOTSO Simeon
Charge de cours

Année Académique
2015-2016





AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire de Yaoundé I. Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : biblio.centrale.uyi@gmail.com

WARNING

This document is the fruit of an intense hard work defended and accepted before a jury and made available to the entire University of Yaounde I community. All intellectual property rights are reserved to the author. This implies proper citation and referencing when using this document.

On the other hand, any unlawful act, plagiarism, unauthorized duplication will lead to Penal pursuits.

Contact: biblio.centrale.uyi@gmail.com

◆ Dédicace ◆

À

Ma maman, MOUKAM Pauline, *mère pleine d'amour et de tendresse.*

◆ Remerciements ◆

Tous mes remerciements vont premièrement à l'endroit de l'Eternel Dieu Tout Puissant sans qui ce travail n'aurait point eut lieu. Ensuite vient mon encadreur **Dr Siméon FOTSO** pour qui j'ai beaucoup d'estime pour son dévouement dans notre encadrement et pour qui je ne saurai jamais cesser de témoigner toute ma gratitude pour son soutien et son obstination à voir ce travail parvenir à bon terme. Mes remerciements vont également à l'endroit de :

- Tous les enseignants du département de mathématiques de l'École Normale Supérieur et de l'Université de Yaoundé I et tous les autres enseignants qui de près ou de loin m'ont soutenu et enrichi par leurs modules de formations divers et variés.
- Mon père Mr **André NKUIMENI** qui n'a jamais cessé de me soutenir et de m'encadrer.
- Mes frères et sœurs **Yannick TIANI NKUIMENI** , **Stéphanie MOUKAM** , **Gisel AMASSI NKUIMENI** , **Stelle DJAPPI** et **Rosland SIEWE** pour leurs soutiens multiples en mon endroit.
- Tous mes oncles,tantes,cousins et cousines pour tous leurs encouragements.
- Mes amis ; ceux de l'amicale des anciens élèves du lycée de Nkolbisson(2AELynkol) et mes camarades **Christian SIAKAM**, **Christiana MATAGOU**, **Gladys MEFFO** qui ont été pour moi plus que des frères ou des soeurs par leurs soutiens divers dans ma lourde tâche.
- A tous ceux qui n'ont ménagé aucun effort à lire ce mémoire et en faire des critiques et suggestions importantes, je pense ici à **Billy TCHOUNKE**, **Gislène KAM**, **Marc BIMENA**, **Willian MIMKEMG** .
- Je ne saurais terminer sans remercier tous mes camarades de promotion en particulier ceux avec qui j'ai cheminé depuis les études à l'Université de Yaoundé I.

◆ Déclaration sur l'honneur ◆

Le présent travail est une oeuvre originale du candidat et n'a été soumis nulle part ailleurs, en partie ou en totalité, pour une autre évaluation académique. Les contributions externes ont été dûment mentionnées et recensées en bibliographie.

Signé,

THEPI NKUIMENI Franck Eric

◆ Table des matières ◆

Dédicace	i
Remerciements	ii
Résumé	vi
Abstract	vii
INTRODUCTION	1
1 MODELE DE DEPENDANCE	5
1.1 Dépendance parfaite : comotonie et antimonotonie	5
1.1.1 Fonction de répartition bivariée	5
1.1.2 Classe de Fréchet et borne d'une classe de Fréchet	6
1.1.3 Dépendance parfaite : comonotonie et antimonotonie	7
1.2 Dépendance parfaite et bornes de Fréchet	9
1.3 Dépendance basée sur les rangs	11
1.4 Modèle de dépendance de queue	12
2 MESURES DE DÉPENDANCE ET COEFFICIENT DE CORRÉLATION LINÉAIRE	13
2.1 Définition et propriétés d'une mesure de dépendance	13
2.2 Coefficient de corrélation de PEARSON	15
3 MESURES DE DÉPENDANCE BASÉES SUR LES RANGS	25
3.1 Coefficient de corrélation des rangs de Kendall	25
3.1.1 Définition et propriétés	25
3.1.2 Tau de kendall valeur empirique	28
3.1.3 Test d'hypothèse	29

3.2	Rho de Spearman	30
3.2.1	définition et propriété	30
3.2.2	Rho de Spearman valeur empirique	33
3.2.3	Test d'hypothèse	33
3.2.4	Liens entre le tau de Kendall et le Rho de Spearman	34
3.3	Coefficient de dépendance de queue	39
3.3.1	Interprétation du coefficient de dépendance de queue	41
3.4	Application	44
3.4.1	Présentation des données et les objectifs de l'étude	45
3.4.2	application	47
	IMPLICATION PÉDAGOGIQUE SUR LE SYSTÈME ÉDUCATIF DU SUJET	49
	CONCLUSION GÉNÉRALE ET PERSPECTIVES	50
	Bibliographie	51

◆ Résumé ◆

La prise en compte de la dépendance entre risques en assurance est un enjeu très important en vue la diversification de ceux ci. L'objectif principal de ce mémoire est de présenter les modèles de dépendance permettant, de définir la forme du lien de dépendance entre deux risques et les mesures de dépendance associées qui quantifient la force de ce lien. Nous présentons pour cela, le coefficient de corrélation linéaire qui est l'un des tout premier indice utilisé en statistique pour mesurer le degré de dépendance entre deux risques et nous montrons que celui ci n'est pas une "bonne mesure de dépendance" de risque. Par la suite, nous présentons deux mesures de dépendance de risque basées sur les notions de concordance et de discordance ; à savoir le Rho de Spearman et le Tau de Kendall. Nous montrons que ceux ci sont de " bonnes mesures de dépendance".

Mots clés : Tau de Kendall, Rho de Spearman, Modèles de dépendance, Mesure de dépendance et Coefficient de corrélation linéaire.

◆ Abstract ◆

Consideration of dependency between insurance risk is a very important issues for the diversification of those risks. The main purpose of this memoire is to present the dependency models ; to define the shape of dependence between two risks and associated dependence measures that quantify the strength of this link. To do that, we present the linear correlation coefficient which is one of the first used index in the field of statistics ; with aim of measuring the degree of dependence between two risks and to show that it is not a "good measure of risk dependance". As a result we present two dependency risk measures based on the concordance and discordance notions namely Spearman's Rho and the Tau Kendall and show that they are "good dependence measures".

Keywords : Tau Kendall, Spearman's Rho, dependency models, dependency measurement and linear correlation coefficient.

◆ INTRODUCTION ◆

Dans une panoplie de domaines de recherche comme la médecine, la biologie et la finance, des chercheurs s'évertuent constamment à comprendre les structures de dépendance entre deux ou plusieurs variables. L'idée intuitive de dépendance est que la réalisation d'une variable aléatoire a un impact sur la distribution d'autres variables. Pendant des années, les actuaires ont considéré les risques indépendants cependant la complexité et la diversification des produits d'assurance, notamment la couverture des catastrophes naturelles (incendies, inondations, tempêtes,...) ont mis en évidence la notion de dépendance entre les risques. Par exemple, en assurance d'habitation, les habitants d'une même région sont exposés aux mêmes risques en cas de survenance d'une catastrophe naturelle ; dans un couple les deux personnes sont exposés aux mêmes habitudes de vie ; en finance, les marchés boursiers réagissent similairement aux événements. Il s'est donc imposé chez l'assureur, la nécessité de prendre en compte la dépendance lorsqu'il gère plusieurs risques à la fois.

S'il n'y a qu'une seule façon pour deux risques d'être mutuellement indépendants, qui se traduit mathématiquement par le fait que leur fonction de répartition conjointe se factorise en le produit de ses marginales. Il y a bien une infinité de manières d'introduire de la dépendance dans un modèle actuariel. Le premier outil utilisé pour rendre compte de la dépendance en sciences actuarielles et en finance est la loi normale multivariée. Sa structure de dépendance est donnée par sa matrice de variance covariance ; ce modèle a été très utilisé en finance. Cette approche a été critiquée il y a quelques années et cela a conduit vers les années 1990 à développer des méthodes plus adaptées pour modéliser la dépendance entre variables aléatoires, suivant diverses lois de probabilité.

Ce mémoire a ainsi pour but de présenter d'une part les concepts de ces modèles de dépendance et les mesures de dépendance associées. Pour ce faire, notre travail s'articule au tour de trois chapitres : Le chapitre un présente quelques concepts de dépendance à savoir : la dépendance parfaite (comonotonie et l'antimonotonie) ; la dépendance basée sur les rangs (concordance et discordance) et enfin la dépendance asymptotique (dépendance de queue). Dans le chapitre deux, nous donnons la définition d'une mesure de dépendance et ses propriétés. Par la suite, nous essayons de montrer que le coefficient de corrélation linéaire, longtemps utilisé comme indice pour quantifier l'intensité de la

liaison entre deux risques n'est pas une "bonne mesure de dépendance". Dans le chapitre trois, nous présentons deux mesures de dépendance basées sur les rangs, à savoir : le tau de Kendall et le Rho de Spearman. Nous montrons par la même occasion que ceux-ci sont de "bonne mesure dépendance" et par suite nous présentons une mesure de la dépendance asymptotique entre deux risques nommée : mesure de dépendance de queue.

◆ Table des figures ◆

3.1	nuage de points deux à deux des données	48
-----	---	----

◆ Liste des tableaux ◆

3.1	Jeu de données	46
3.2	tableau des valeurs du Tau de Kendall et du Rho de Spearman sur le jeu de données .	47

MODELE DE DEPENDANCE

Si deux variables n'ont qu'une manière d'être indépendantes, la dépendance peut quant à elle revêtir de nombreuses formes. Ce sont ces formes que tentent de capter les modèles de dépendance. Ainsi on dispose de plusieurs concepts permettant de caractériser le lien de dépendance entre deux risques aléatoires. Dans ce chapitre nous présentons le cas de la dépendance parfaite (comotonie et antimonotonie), la dépendance basée sur les rangs (concordance et discordance) et la dépendance asymptotique (dépendance de queue supérieure et inférieure).

1.1 Dépendance parfaite : comotonie et antimonotonie

1.1.1 Fonction de répartition bivariée

Définition 1.1. Soit $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans \mathbb{R}^2 . la fonction de répartition bivariée de X est la fonction F définie sur \mathbb{R}^2 par :

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

Remarque 1.1.1. 1. La loi de X est caractérisée par F .

2. Les lois marginales ou marginales de X sont décrites par

$$F_1(x_1) = P(X_1 \leq x_1) = \lim_{x_2 \rightarrow +\infty} F(x_1, x_2) = F(x_1, +\infty)$$

$$F_2(x_2) = P(X_2 \leq x_2) = \lim_{x_1 \rightarrow +\infty} F(x_1, x_2) = F(+\infty, x_2)$$

Propriété 1.1. 1. F est continue à droite.

2. $\lim_{x_1 \rightarrow -\infty} F(x_1, x_2) = \lim_{x_2 \rightarrow -\infty} F(x_1, x_2) = 0.$

3. $\lim_{x_1, x_2 \rightarrow +\infty} F(x_1, x_2) = 1.$

4. F est croissante suivant chaque composante.

1.1. Dépendance parfaite : comotonie et antimonotonie

D'autre part, contrairement à la définition de fonction de répartition bivariée qui fait appel aux variables aléatoires, on peut énoncer des conditions nécessaires et suffisantes pour qu'une fonction de variables soit une fonction de répartition bivariée.

Proposition 1.1. Une fonction de 2 variables vérifiant les quatre propriétés précédentes est une fonction de répartition bivariée.

1.1.2 Classe de Fréchet et borne d'une classe de Fréchet

Connaissant la loi conjointe du couple (X_1, X_2) il est facile de déterminer les lois marginales c'est-à-dire les lois des variables X_1 et X_2 . Mais pour des lois marginales F_1 et F_2 données il existe une infinité de fonctions de répartition bivariées de marginales F_1 et F_2 . L'ensemble de toutes les fonctions de répartition bivariées de marginales F_1 et F_2 constituent la classe de Fréchet de (F_1, F_2) .

Définition 1.2. les classes de Fréchet sont des distributions bivariées avec des marges données. Ces marges peuvent être unidimensionnelles et aussi conditionnelles.

soient F_1, F_2 des distributions ; F appartient à la classe de frechet (F_1, F_2) et on note $F \in \mathcal{F}(F_1, F_2)$ si et seulement si F est une distribution de marges F_1 et F_2 .

Les classes de Fréchet constituent ainsi un cadre idéal pour analyser la dépendance puisqu'elles ne regroupent que des lois de probabilité de couples qui ne diffèrent entre elles que par leurs structures de dépendance et non par leurs marginales.

Théorème 1.1. (Denuit et al. 2004)

Si F est une fonction de répartition bivariée de marginales F_1, F_2 , alors $\forall x = (x_1, x_2) \in \mathbb{R}^2$,
 $Max(F_1(x_1) + F_2(x_2) - 1, 0) = F^-(x) \leq F(x) \leq F^+(x) = min(F_1(x_1), F_2(x_2))$

Preuve. Soit $x = (x_1, x_2) \in \mathbb{R}^2$ on a :

$$F(x) \leq \lim_{x_1 \rightarrow +\infty} F(x) = F_2(x_2)$$

$$F(x) \leq \lim_{x_2 \rightarrow +\infty} F(x) = F_1(x_1).$$

Ainsi $F(x) \leq F_1(x_1)$ et $F(x) \leq F_2(x_2)$ d'où $F(x) \leq Min(F_1(x_1), F_2(x_2))$.

De plus on a

$$\begin{aligned} F(x) &= P(X_1 \leq x_1, X_2 \leq x_2) \\ &= 1 - P\left(\bigcup_{k=1}^2 X_k > x_k\right) \\ &\geq 1 - \sum_{k=1}^2 P\{X_k > x_k\} \\ &= 1 - 2 + (P\{X_1 \leq x_1\} + P\{X_2 \leq x_2\}) \\ &= F_1(x_1) + F_2(x_2) - 1 \end{aligned}$$

1.1. Dépendance parfaite : comotonie et antimonotonie

et comme $F(x) \geq 0$ on a $F(x) \geq \text{Max}(F_1(x_1) + F_2(x_2) - 1, 0)$. ■

Définition 1.3. Les fonctions F^+ et F^- sont respectivement appelées **borne inférieure** et **borne supérieure** de la classe de Fréchet $\mathcal{F}(F_1, F_2)$.

Remarque 1.1.2. 1. Les bornes supérieure et inférieure de Fréchet sont les points extrémaux de $\mathcal{F}(F_1, F_2)$. Néanmoins, ce ne sont pas les seules. Il en existe une infinité de points extrémaux qu'il est difficile de caractériser.

2. F^+ et F^- sont des distributions bivariées.

Couples de variables aléatoires correspondant aux bornes de Fréchet

On peut se demander à quels couples de variables aléatoires correspondent les bornes de Fréchet dans $\mathcal{F}(F_1, F_2)$. Le résultat suivant répond à la question.

Proposition 1.2. Soit U une variable aléatoire de loi uniforme sur $[0; 1]$. Dans $\mathcal{F}(F_1, F_2)$,

1. F^+ est la fonction de répartition du couple $(F_1^{-1}(U), F_2^{-1}(U))$
2. F^- est la fonction de répartition du couple $(F_1^{-1}(U), F_2^{-1}(1 - U))$.

Preuve. Pour tout $(x, y) \in \mathbb{R}^2$, nous avons,

$$\begin{aligned} P(F_1^{-1}(U) \leq x, F_2^{-1}(U) \leq y) &= P(U \leq F_1(x), U \leq F_2(y)) \\ &= P(U \leq \min\{F_1(x), F_2(y)\}) \\ &= \min\{F_1(x), F_2(y)\} \\ &= F^+(x, y) \end{aligned}$$

et

$$\begin{aligned} P(F_1^{-1}(U) \leq x, F_2^{-1}(1 - U) \leq y) &= P(U \leq F_1(x), 1 - U \leq F_2(y)) \\ &= P(1 - F_2(y) \leq U \leq F_1(x)) \\ &= \max\{F_1(x) + F_2(y) - 1, 0\} \\ &= F^-(x, y). \end{aligned}$$

■

1.1.3 Dépendance parfaite : comotonie et antimonotonie

Définition 1.4. Une dépendance parfaite existe entre deux risques X et Y lorsqu'ils peuvent s'écrire comme fonction croissante ou décroissante d'une même variable aléatoire Z .

Il existe deux cas de dépendance parfaite : **la comotonie** ou dépendance parfaite positive et **l'anti-monotonie** ou dépendance parfaite négative .

1.1. Dépendance parfaite : comotonie et antimotonie

Définition 1.5. (l'antimotonie)

Le couple (X, Y) est dit **antimotonne** s'il existe une fonction croissante g_1 et une fonction décroissante g_2 et une variable aléatoire Z telles que $(X, Y) \stackrel{\text{loi}}{=} (g_1(Z), g_2(Z))$.

Définition 1.6. (monotonie)

Le couple (X, Y) est dit **comotonne** s'il existe des fonctions croissantes g_1 et g_2 et une variable aléatoire Z telles que $(X, Y) \stackrel{\text{loi}}{=} (g_1(Z), g_2(Z))$.

Remarque 1.1.3. L'égalité précédente est une égalité en loi, ce qui signifie que

$$\forall (x, y) \in \mathbb{R}^2, P(X \leq x, Y \leq y) = P(g_1(Z) \leq x, g_2(Z) \leq y).$$

Exemple 1.1.1. (Denuit et al. 2004)

Soit Z la v.a montant d'un sinistre.

Le schéma de redistribution du coût d'un sinistre entre l'assuré, l'assureur et le réassureur est le suivant.

Si le montant du sinistre est inférieur à un déductible d , l'assuré assume Z .

Si le montant du sinistre est entre le déductible et une limite L , alors l'assureur assume $Z - d$.

Si le montant du sinistre est supérieur à L , alors l'assureur se tourne vers le réassureur qui assume l'excédent. Soit X le montant du sinistre assumé par l'assureur et Y le montant assumé par le réassureur.

On a

$$X = g_1(Z) = \begin{cases} 0 & \text{si } Z \leq d \\ Z - d & \text{si } d < Z \leq L \\ L & \text{si } Z > L \end{cases}$$

et

$$Y = g_2(Z) = \begin{cases} 0 & \text{si } Z \leq L \\ Z - L & \text{si } Z > L \end{cases}$$

Alors X et Y sont deux fonctions croissantes de Z , donc X et Y sont comotonnes : Il y a une dépendance parfaite positive entre X et Y . Plus le montant du sinistre est élevé, plus les montants assumés par l'assureur et le réassureur sont élevés.

Exemple 1.1.2. (Denuit et al. 2004)

Soit Z le prix d'une action à la fin d'une certaine journée.

Le prix d'une option d'achat avec prix d'exercice K_1 sur cette action est

$$X = g_1(Z) = \text{Max}(Z - K_1, 0).$$

Le prix d'une option de vente avec prix d'exercice K_2 est $Y = g_2(Z) = \text{Max}(K_2 - Z, 0)$.

Ainsi g_1 est une fonction croissante de Z et g_2 est une fonction décroissante de Z . Par conséquent X et Y sont antimotonnes : il y'a une dépendance parfaite négative entre X et Y . Plus le prix de l'action est élevé plus X sera élevé et plus Y sera faible.

1.2 Dépendance parfaite et bornes de Fréchet

Nous donnons ici une interprétation des bornes de Fréchet.

Proposition 1.3. Soient X et Y deux variables aléatoires de fonctions de répartitions respectives F_X, F_Y , et de distribution conjointe F .

1. X et Y sont comonotones si et seulement si $F = F^+$ c'est-à-dire $F(x, y) = \min(F_X(x), F_Y(y))$, dans ce cas $X = F_X^{-1}(F_Y(Y))$.
2. X et Y sont antimonotones si et seulement si $F = F^-$ c'est-à-dire $F(x, y) = \max(F_X(x) + F_Y(y) - 1, 0)$, dans ce cas $X = F_X^{-1}(1 - F_Y(Y))$.

Preuve. Pour tout $(x, y) \in \mathbb{R}^2$, nous avons,

$$\begin{aligned}
 P(X \leq x, F_Y^{-1}(F_X(X)) \leq y) &= P(X \leq F_X^{-1}(F_X(x)), F_Y^{-1}(F_X(X)) \leq y) \\
 &= P(F_X(X) \leq F_X(x), F_X(X) \leq F_Y(y)) \\
 &= P(F_X(X) \leq \min\{F_X(x), F_Y(y)\}) \\
 &= \min\{F_X(x), F_Y(y)\} \quad \text{car } F_X(X) \sim \mathcal{U}_{[0,1]} \\
 &= F^+(x, y)
 \end{aligned}$$

$$\begin{aligned}
 P(X \leq x, F_Y^{-1}(1 - F_X(X)) \leq y) &= P(X \leq F_X^{-1}(F_X(x)), F_Y^{-1}(1 - F_X(X)) \leq y) \\
 &= P(F_X(X) \leq F_X(x), F_X(X) \geq 1 - F_Y(y)) \\
 &= P(1 - F_Y(y) \leq F_X(X) \leq F_X(x)) \\
 &= \max\{F_X(x) + F_Y(y) - 1, 0\} \quad \text{car } F_X(X) \sim \mathcal{U}_{[0,1]} \\
 &= F^-(x, y).
 \end{aligned}$$

■

Exemple 1.2.1. 1. Si $X = (X_1, X_2)$ est un vecteur aléatoire dont les marges sont **normales centrées réduites** alors

$$X_1 \text{ et } X_2 \text{ comonotones} \iff X_1 = X_2.$$

$$X_1 \text{ et } X_2 \text{ antimonotones} \iff X_1 = -X_2.$$

En effet soit $x \in \mathbb{R}$.

$$X_1 \text{ et } X_2 \text{ comonotones} \iff$$

$$\begin{aligned}
 X_1(x) &= F_1^{-1}(F_2(X_2)(x)) \\
 &= X_2(x) \text{ car } X_1 \text{ et } X_2 \text{ ont même loi}
 \end{aligned}$$

1.2. Dépendance parfaite et bornes de Fréchet

D'où $X_1 = X_2$.

X_1 et X_2 antimonotones \iff

$$\begin{aligned}X_1(x) &= F_1^{-1}(1 - F_2(X_2)(x)) \\ &= F_1^{-1}(F_2(-X_2(x))) \text{ car } F_2 \text{ est symétrique} \\ &= -X_2(x) \text{ car } X_1 \text{ et } X_2 \text{ ont même loi}\end{aligned}$$

D'où $X_1 = -X_2$.

2. Si $X = (X_1, X_2)$ est un vecteur aléatoire dont les marges sont **log-normales** alors

X_1 et X_2 comonotones $\iff X_1 = X_2$.

X_1 et X_2 antimonotones $\iff X_1 = \exp(-\ln(X_2)) = \frac{1}{X_2}$.

En effet soit $x \in \mathbb{R}$.

X_1 et X_2 comonotones \iff

$$\begin{aligned}X_1(x) &= F_1^{-1}(F_2(X_2)(x)) \\ &= X_2(x) \text{ car } X_1 \text{ et } X_2 \text{ ont même loi}\end{aligned}$$

D'où $X_1 = X_2$.

X_1 et X_2 antimonotones \iff

$$X_1(x) = F_1^{-1}(1 - F_2(X_2)(x)).$$

Or $X_i \sim LN(\sigma, \mu) \iff \ln(X_i) \sim N(\sigma, \mu)$. Ainsi comme

$$\begin{aligned}F_i(x) &= P(X_i \leq x) \\ &= P(\ln(X_i) \leq \ln(x)) \text{ car } \ln \text{ est croissante} \\ &= F_{Y_i}(x) \text{ ou } Y_i \sim N(\sigma, \mu) \\ &= F_{Y_i}(\ln(x))\end{aligned}$$

On a donc $F_{X_1}^{-1} = \exp \circ F_{Y_1}^{-1}$ et $F_{X_2} = F_{Y_2} \circ \ln$.

1.3. Dépendance basée sur les rangs

D'où

$$\begin{aligned}X_1(x) &= \exp \circ F_{Y_1}^{-1}(1 - F_{Y_2}(\ln(X_2(x)))) \\&= \exp \circ F_{Y_1}^{-1}(F_{Y_2}(-\ln(X_2(x)))) \\&= \exp(-\ln(X_2(x))) \text{ car } Y_1 \text{ et } Y_2 \text{ ont même loi} \\&= \frac{1}{X_2}\end{aligned}$$

$$\text{D'où } X_1 = \exp(-\ln(X_2)) = \frac{1}{X_2}.$$

3. Si X_1 et X_2 sont uniformes

$$X_1 \text{ et } X_2 \text{ comonotones} \iff X_1 = X_2$$

$$X_1 \text{ et } X_2 \text{ antimonotones} \iff X_1 = 1 - X_2$$

En effet, soit $x \in \mathbb{R}$

X_1 et X_2 comonotones \iff

$$\begin{aligned}X_1(x) &= F_1^{-1}(F_2(X_2(x))) \\&= X_2(x)\end{aligned}$$

$$X_1 = X_2$$

X_1 et X_2 antimonotones \iff

$$\begin{aligned}X_1(x) &= F_1^{-1}(1 - F_2(X_2(x))) \\&= F_1^{-1}(1 - x_2(x)) \\&= 1 - X_2(x)\end{aligned}$$

$$\text{d'où } X_1 = 1 - X_2$$

1.3 Dépendance basée sur les rangs

Étant données deux risques X et Y , on ne peut pas toujours les écrire comme fonction d'une même variable aléatoire Z . En d'autre terme le lien de dépendance n'est pas toujours parfait(positif ou négatif) on peut cependant définir d'autre forme de dépendance : la dépendance basée sur les rangs (concordance et discordance).

Définition 1.7. concordance

Soient X et Y deux variables aléatoires X et Y sont dites **concordantes** lorsqu'à la réalisation d'une

1.4. Modèle de dépendance de queue

petite (respectivement grande) valeur pour X est associée une petite (respectivement grande) valeur pour Y .

Notons (X_1, Y_1) et (X_2, Y_2) deux réalisations iid de (X, Y) . Les deux paires d'observations (X_1, Y_1) et (X_2, Y_2) sont **concordantes** si $X_1 < X_2$ et $Y_1 < Y_2$ ou bien $X_1 > X_2$ et $Y_1 > Y_2$. Ceci peut s'écrire (X_1, Y_1) et (X_2, Y_2) sont concordantes si $(X_1 - X_2)(Y_1 - Y_2) > 0$.

Ainsi les paires X et Y sont concordantes si elles ont tendance à générer des paires d'observations (X_i, Y_i) et (X_j, Y_j) concordantes.

Définition 1.8. discordance

C'est le concept inverse de la concordance. Soient X et Y deux variables aléatoires X et Y sont dites **discordantes** lorsqu'à la réalisation d'une petite (respectivement grande) valeur pour X est associée une grande (respectivement petite) valeur pour Y .

Notons (X_1, Y_1) et (X_2, Y_2) deux réalisations iid de (X, Y) . Les deux paires d'observations (X_1, Y_1) et (X_2, Y_2) sont **discordantes** si $X_1 < X_2$ et $Y_1 > Y_2$ ou bien $X_1 > X_2$ et $Y_1 < Y_2$. Ceci peut s'écrire (X_1, Y_1) et (X_2, Y_2) sont discordantes si $(X_1 - X_2)(Y_1 - Y_2) < 0$.

Ainsi les paires X et Y sont discordantes si elles ont tendance à générer des paires d'observations (X_i, Y_i) et (X_j, Y_j) discordantes.

1.4 Modèle de dépendance de queue

Les modèles de dépendances déjà présentés ici tiennent compte de l'ensemble de la distribution des deux variables, les modèles de dépendance de queue quant à eux tiennent compte seulement des liens entre les valeurs extrêmes des deux variables ; elles sont importantes dans l'étude de la dépendance asymptotique entre deux variables aléatoires. Le concept de dépendance de queue est utilisé en finance car les rendements des titres financiers montrent de la dépendance de queue inférieure. En effet quand un marché dégringole, les marchés autour ont aussi tendance à dégringoler.

Définition 1.9. Deux risques X et Y montrent **une dépendance de queue** si la réalisation d'une valeur extrême positive (respectivement négative) pour X entraîne la réalisation d'une valeur extrême positive (respectivement négative) pour Y .

MESURES DE DÉPENDANCE ET COEFFICIENT DE CORRÉLATION LINÉAIRE

Les modèles de dépendance définis au chapitre un, définissent le lien de dépendance entre deux risques. Les mesures de dépendances quantifient la force de ce lien de dépendance. Dans ce chapitre nous donnons la définition et les propriétés d'une mesure de dépendance et nous présentons le coefficient de corrélation linéaire qui est l'un des tout premiers indices utilisés pour quantifier le degré de dépendance entre deux risques. Par la suite nous essayerons de vérifier si ce coefficient de corrélation linéaire est une "bonne mesure de dépendance".

2.1 Définition et propriétés d'une mesure de dépendance

Définition 2.1. Une **mesure de dépendance** est une application qui à deux variables aléatoires associe un réel permettant de quantifier la force de la dépendance qui lie ces deux variables.

Cette définition n'est pas très restrictive, les probabilistes se sont rapidement demandé quelles propriétés devrait posséder une mesure de dépendance pour pouvoir être appliquée en pratique, c'est-à-dire d'être une "bonne mesure de dépendance".

Définition 2.2. (Planchet.2004)

Une mesure de dépendance $\delta(.,.)$ est une **bonne mesure de dépendance** si elle possède les bonnes propriétés suivantes.

1. **symétrie** : $\delta(X, Y) = \delta(Y, X)$
2. **Normalisation** : $-1 \leq \delta(X, Y) \leq 1$
3. **Dépendance parfaite positive** : $\delta(X, Y) = 1 \Leftrightarrow X$ et Y sont comonotones

2.1. Définition et propriétés d'une mesure de dépendance

4. **Dépendance parfaite négative** : $\delta(X, Y) = -1 \Leftrightarrow X$ et Y sont antimonotones

5. **Indépendance** : X et Y indépendants $\Rightarrow \delta(X, Y) = 0$

6. **Invariance** : quelle que soit la fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ strictement monotone

$$\delta(g(X), Y) = \begin{cases} \delta(X, Y) & \text{si } g \text{ est croissante} \\ -\delta(X, Y) & \text{si } g \text{ est décroissante} \end{cases}$$

De prime abord, on peut encore envisager d'autres bonnes propriétés, mais qui ne sont pas nécessairement compatibles avec les propriétés précédentes. Ainsi une autre propriété intéressante est

$$\delta(X, Y) = 0 \Leftrightarrow X \text{ et } Y \text{ sont indépendantes.}$$

Malheureusement, cette propriété contredit P6, comme le montre le résultat suivant.

Proposition 2.1. Il n'existe pas de mesure de dépendance δ satisfaisant les six propriétés précédente et satisfaisant

$$\delta(X, Y) = 0 \Leftrightarrow X \text{ et } Y \text{ sont indépendantes.}$$

Preuve. Considérons le couple (X_1, X_2) uniformément distribué le long du cercle unité du plan \mathbb{R}^2 , *i.e.*

$$(X_1, X_2) = (\cos Z, \sin Z) \text{ avec } Z \text{ qui suit une loi uniforme sur } [0; 2\pi].$$

Puisque $(-X_1, X_2) \stackrel{\text{loi}}{=} (X_1, X_2)$, on a : $\delta(X_1, X_2) = \delta(-X_1, X_2)$.

Par ailleurs la fonction $x \mapsto -x$ est strictement décroissante donc

$$\delta(X_1, X_2) = \delta(-X_1, X_2) = -\delta(X_1, X_2) = 0$$

alors que X_1 et X_2 sont clairement dépendantes. ■

Remarque 2.1.1. Nous justifions ici la pertinence des propriétés imposées à une mesure de dépendance de risque.

1. On impose La propriété (1) à une mesure de dépendance puisse que la dépendance entre X et Y est la même entre Y et X .
2. la propriété (2) quand t-a-elle elle permet d'avoir une mesure qui ne dépend pas des unités des variables. Cela permet ainsi de vérifier si le lien de dépendance entre deux risques est plus important que le lien de dépendance entre deux autres risques.
3. les propriétés (3) et (4) prennent leur sens, car la comonotonie et l'antimonotonie correspondent au cas de dépendance parfaite positive et négative.

2.2. Coefficient de corrélation de PEARSON

- la propriété (5) quand t-a-elle est imposée car si deux risques X et Y sont indépendants cela signifie qu'il n'existe aucun lien de dépendance entre ces risques.
- la propriété (6) prend tout son sens car le lien de dépendance entre deux risques X et Y est le même entre $g(X)$ et Y si g est une fonction croissante et contraire si g est une fonction décroissante.

Proposition 2.2. Si δ est une mesure de dépendance alors pour toute fonction g strictement monotone on a $\delta(g(X), g(Y)) = \delta(X, Y)$ pour tout couple (X, Y) de variable aléatoire considéré. C'est-à-dire qu'une mesure de dépendance est invariante par transformation strictement monotone.

Preuve. Soit δ une mesure de dépendance et soit (X, Y) un couple de variable aléatoire et soit g une fonction strictement monotone.

Si g est croissante on a : $\delta(g(X), g(Y)) = \delta(g(X), Y) = \delta(X, Y)$ d'après la propriété 5 d'une mesure de dépendance.

Si g est décroissante on a : $\delta(g(X), g(Y)) = -\delta(g(X), Y) = -(-\delta(X, Y)) = \delta(X, Y)$ d'après la propriété 5 d'une mesure de dépendance. ■

2.2 Coefficient de corrélation de PEARSON

Le coefficient de corrélation de Pearson encore appelé coefficient de corrélation linéaire est la première mesure utilisée pour quantifier la force de dépendance entre deux risques.

Il repose sur la propriété suivante $V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$. La covariance mesure le surcroît de variabilité de la somme de deux variables aléatoires par rapport à la somme de la variabilité de chacune de ces variables. La covariance permet donc d'apprécier le sens de la covariation ou variation conjointe de deux variables aléatoires. Le coefficient de corrélation de Pearson est une version normée sans dimension de la covariance.

Définition 2.3. (Coefficient de corrélation linéaire)

Soit deux variables aléatoires X et Y admettant des moments d'ordre deux. Le coefficient de corrélation de Pearson entre X et Y que l'on note $r(X, Y)$ est défini par

$$r(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

ou $V(X)$ représente la variance de X et σ_X l'écart-type de X .

Remarque 2.2.1. Si nous disposons d'un échantillon de taille n , $(x_1, y_1), \dots, (x_n, y_n)$ tiré d'une distribution conjointe (X, Y) , la quantité

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

2.2. Coefficient de corrélation de PEARSON

est une estimation de r ; c'est la valeur empirique.

Dans la suite nous donnons les propriétés du coefficient de corrélation afin de savoir si celui-ci est une "bonne mesure de dépendance".

Propriété 2.1. Quelles que soient les variables aléatoires X et Y admettant des moments d'ordre 2, on a

1. $r(X, Y) = r(Y, X)$
2. $-1 \leq r(X, Y) \leq 1$

Preuve. 1.

$$\begin{aligned} r(X, Y) &= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \\ &= \frac{Cov(Y, X)}{\sigma_Y \sigma_X} \\ &= r(Y, X) \end{aligned}$$

2. On a $V\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \geq 0$ alors

$$V\left(\frac{X}{\sigma_X}\right) + V\left(\frac{Y}{\sigma_Y}\right) + 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \geq 0$$

ainsi

$$\frac{1}{\sigma_X^2} V(x) + \frac{1}{\sigma_Y^2} V(y) + \frac{2}{\sigma_X \sigma_Y} Cov(X, Y) \geq 0$$

donc

$$2 + 2r(X, Y) \geq 0$$

d'où

$$r(X, Y) \geq -1.$$

De même $V\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \geq 0$ alors

$$V\left(\frac{X}{\sigma_X}\right) + V\left(-\frac{Y}{\sigma_Y}\right) + 2Cov\left(\frac{X}{\sigma_X}, -\frac{Y}{\sigma_Y}\right) \geq 0$$

ainsi

$$\frac{1}{\sigma_X^2} V(x) + \frac{1}{\sigma_Y^2} V(y) - \frac{2}{\sigma_X \sigma_Y} Cov(X, Y) \geq 0$$

donc

$$2 - 2r(X, Y) \geq 0$$

d'où

$$r(X, Y) \leq 1$$

2.2. Coefficient de corrélation de PEARSON

On a ainsi

$$-1 \leq r(X, Y) \leq 1.$$

■

Proposition 2.3. Soient X et Y deux variables aléatoires admettent des moments d'ordre 2.

X et Y indépendants $\Rightarrow r(X, Y) = 0$.

La preuve est immédiate car X et Y indépendants alors $Cov(X, Y) = 0$ d'où $r(X, Y) = 0$.

Remarque 2.2.2. La réciproque est fautive. En effet,

considérons $X \sim N(0, 1)$. X est symétrique alors $E(X^3) = 0$, posons $Y = X^2$ on a $E(XY) = E(X^3) = 0$ d'où $Cov(X, Y) = E(XY) - E(X)E(Y) = 0$ et donc $r(X, Y) = 0$ alors que X et Y ne sont pas indépendants.

Proposition 2.4. Si (X, Y) suit une loi normale bivariée alors $r(X, Y) = 0 \iff X$ et Y sont indépendantes

Preuve. On a déjà X et Y indépendantes $\Rightarrow r(X, Y) = 0$.

Supposons $r(X, Y) = 0$ et montrons que X et Y sont indépendantes.

$r(X, Y) = 0 \Rightarrow Cov(X, Y) = Cov(Y, X) = 0$ or $(X, Y) \sim N(\mu, \Sigma)$ où

$$\Sigma = \begin{pmatrix} Var(X) & 0 \\ 0 & Var(Y) \end{pmatrix}$$

$$\mu = (\mu_1, \mu_2).$$

Ainsi le couple (X, Y) admet pour densité f_{XY}

$$\begin{aligned} f_{XY}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right) \\ &= f_X(x_1)f_Y(x_2). \end{aligned}$$

D'où les variables aléatoires X et Y sont indépendantes. ■

Remarque 2.2.3. Le résultat précédent n'est vrai que si (X, Y) suit une loi normale bivariée. En effet si les marginales de X et Y sont normales et que la loi conjointe (X, Y) est différente de la loi normale bivariée, l'équivalence précédente est fautive.

Exemple 2.2.1. Soient X et W deux variables aléatoires indépendantes avec $X \sim N(0, 1)$ et W prend les valeurs dans $\{-1, 1\}$ avec $P(W = 1) = P(W = -1) = \frac{1}{2}$.

2.2. Coefficient de corrélation de PEARSON

Considérons $Y = XW$ on a

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(XW \leq y) \\&= P(XW \leq y \mid W = 1)P(W = 1) + P(XW \leq y \mid W = -1)P(W = -1) \\&= \frac{1}{2}P(X \leq y) + \frac{1}{2}P(-X \leq y) \\&= \frac{1}{2}P(X \leq y) + \frac{1}{2}P(X \geq -y) \\&= \frac{1}{2}P(X \leq y) + \frac{1}{2}P(X \leq y) \\&= P(X \leq y) \\&= F_X(y)\end{aligned}$$

Donc X et Y sont des variables aléatoires de loi normale, et de plus il existe un lien de dépendance entre X et Y .

Cependant

$$\begin{aligned}Cov(X, Y) &= E(XY) - E(X)E(Y) \\&= E(X^2W) - E(X)E(XW) \\&= E(X^2)E(W) - (E(X))^2E(W) \text{ car } X \text{ et } W \text{ sont indépendantes} \\&= 0 - 0 \text{ car } E(W) = 0 \\&= 0\end{aligned}$$

donc $r(X, Y) = 0$

Remarque 2.2.4. Le coefficient de corrélation de PEARSON peut donc être considéré comme un test d'indépendance lorsque le couple de variable est normale.

En effet si nous disposons d'un échantillon de taille n , extrait d'une population normale conjointe, il est naturel de vouloir tester des hypothèses concernant la véritable valeur de r .

Pour tester l'hypothèse nulle

$$H_0 : r = 0$$

contre l'hypothèse alternative

$$H_1 : r \neq 0.$$

Nous calculons la statistique t :

$$t = \rho \sqrt{\frac{n-2}{1-\rho^2}}$$

Sous H_0 la statistique t suit une loi de student avec $n - 2$ degrés de liberté. Pour un seuil de signification α donné, on rejette H_0 si $|t| \geq t_{\frac{\alpha}{2}, n-2}$; la valeur de $t_{\frac{\alpha}{2}, n-2}$ est la valeur critique du test donnée dans la table Student.

2.2. Coefficient de corrélation de PEARSON

Proposition 2.5. Soient X et Y deux variables aléatoire admettant des moment jusqu'à l'ordre 2.

$|r(X, Y)| = 1 \iff Y = aX + b$ avec $b \neq 0$. De plus le signe de r est identique au signe de a .

Preuve. Supposons $Y = aX + b$ alors

$$\begin{aligned} r(X, Y) &= \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \\ &= \frac{Cov(X, aX + b)}{\sigma_X \sigma_{aX+b}} \\ &= \frac{aCov(X, X)}{|a| \sigma_X^2} \\ &= \frac{a}{|a|} \end{aligned}$$

d'où $|r(X, Y)| = 1$.

Supposons $|r(X, Y)| = 1$ et montrons que $Y = aX + b$.

Soit (Ω, \mathbb{F}, P) un espace probabilisé posons $F = \{ \text{ensemble des variables aléatoires centrés} \}$ alors $(F, +, \cdot)$ est un espace vectoriel réel.

Cov est une forme bilinéaire symétrique sur F de plus Cov est définie positive sur F . En effet

$$\begin{aligned} Cov(X, X) &= 0 \\ &\Leftrightarrow X = cte \\ &\Leftrightarrow X = E(X) \\ &\Leftrightarrow X = 0 \end{aligned}$$

D'où Cov définit sur F un produit scalaire.

Alors d'après Inégalité de Cauchy Schwarz on a

$$\forall X, Y \in F \quad |Cov(X, Y)|^2 \leq Cov(X, X)Cov(Y, Y)$$

alors

$$\forall X, Y \in F \quad |Cov(X, Y)| \leq \sqrt{Cov(X, X)Cov(Y, Y)}$$

D'où

$$|r(X, Y)| \leq 1$$

et égalité ayant lieu si X et Y sont liées.

Ainsi si X et Y sont deux variables aléatoires quelconque on pose $X' = X - E(X)$ $Y' = Y - E(Y)$

et on a $Cov(X, Y) = Cov(X', Y')$ et $V(X) = V(X')$ et $V(Y) = V(Y')$ donc

$|r(X, Y)| = |r(X', Y')| = 1$ si et seulement si X' et Y' sont liées.

Ainsi

$$Y' = aX' \quad \text{avec} \quad a \in \mathbb{R}$$

2.2. Coefficient de corrélation de PEARSON

d'où

$$Y - E(Y) = a(X - E(X))$$

alors

$$Y = aX + E(Y) - aE(X).$$

D'où $|r(X, Y)| = 1 \Rightarrow Y = aX + b$

■

Remarque 2.2.5. 1. Une valeur de $|r(X, Y)|$ proche de 1 peut indiquer la présence d'une dépendance non-linéaire.

En effet si X suit une loi uniforme sur $[8, 10]$ et $Y = X^2$ alors $r(X, Y) \approx 0,999$, mais en aucun cas, il n'existe a et b tels que $Y = aX + b$, Dans ce cas X et Y sont comotones .

2. Le coefficient de corrélation constitue ainsi une mesure de l'intensité de la liaison linéaire entre deux variables. Plus il est proche de 1 (en valeur), plus la relation est forte.

3. La valeur ρ de l'estimateur de r n'a pas de signification. En revanche son carré *ie* ρ^2 , que l'on appelle **coefficient de détermination** s'interprète comme la proportion de variance Y (Resp X) linéairement expliqué par X (Resp Y).

Ainsi si $\rho = 0,9$ on voit que la liaison est forte, puisqu'elle est proche de 1. En revanche avec $\rho^2 = 0,81$ on peut dire que 81% de la variance de Y est expliqué par X (réciproquement).

Le lemme suivant nous sera très utile dans la suite afin d'obtenir les valeurs possibles pour le coefficient de corrélation linéaire.

Lemme 2.1. (Denuit et al. 2004)

Quels que soient les risques X_1 et X_2 de fonction de répartition jointe F_X ,

$$E[X_1 X_2] = \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \overline{F_X}(x_1, x_2) dx_1 dx_2.$$

Preuve. on a

$$\begin{aligned} \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \overline{F_X}(x_1, x_2) dx_1 dx_2 &= \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} P[X_1 > x_1, X > x_2] dx_1 dx_2 \\ &= \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \int_{y_1=x_1}^{+\infty} \int_{y_2=x_2}^{+\infty} dF_X(y_1, y_2) dx_1 dx_2 \\ &= \int_{y_1=0}^{+\infty} \int_{y_2=0}^{+\infty} \int_{x_1=0}^{y_1} \int_{x_2=0}^{y_2} dx_1 dx_2 dF_X(y_1, y_2) \quad \text{Fubini} \\ &= \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} y_1 y_2 dF_X(y_1, y_2) \\ &= E[X_1 X_2] \end{aligned}$$

■

2.2. Coefficient de corrélation de PEARSON

Corollaire 2.1. (Denuit et al. 2004) Le lemme précédent nous permet d'écrire

$$\begin{aligned} \text{cov}(X_1, X_2) &= \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \{\overline{F_X}(x_1, x_2) - \overline{F_1}(x_1)\overline{F_2}(x_2)\} dx_1 dx_2 \\ &= \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \{F_X(x_1, x_2) - F_1(x_1)F_2(x_2)\} dx_1 dx_2. \end{aligned}$$

Cette dernière représentation de la covariance la fait apparaître comme une distance entre la fonction de répartition jointe F_X de X et celle du couple de même lois marginales que X mais dont les composantes sont normales.

Propriété 2.2. Dans le cas de relation non linéaire, soient X et Y deux variables aléatoire admettant des moment jusqu'à l'ordre 2.

On a :

X et Y comonotone $\iff r(X, Y) = r_{max}$

X et Y antimonotone $\iff r(X, Y) = r_{min}$

où $r_{min} = \frac{\text{Cov}(F_X^{-1}(U), F_Y^{-1}(1-U))}{\sigma_X \sigma_Y}$ et $r_{max} = \frac{\text{Cov}(F_X^{-1}(U), F_Y^{-1}(U))}{\sigma_X \sigma_Y}$.

U suit une loi uniforme sur $[0,1]$, de plus on a

$$r_{min} \leq r(X, Y) \leq r_{max}.$$

De sorte que la valeur ± 1 n'est en général pas admissible.

Preuve. X et Y comonotones \iff le couple (X, Y) admet pour fonction de répartition F^+ qui est la fonction de répartition du couple $(F_X^{-1}(U), F_Y^{-1}(U))$ et donc $r(X, Y) = \frac{\text{Cov}(F_X^{-1}(U), F_Y^{-1}(U))}{\sigma_X \sigma_Y}$.

De même X et Y antimonotones \iff le couple (X, Y) admet pour fonction de répartition F^- qui est la fonction de répartition du couple $(F_X^{-1}(U), F_Y^{-1}(1-U))$ et donc $r(X, Y) = \frac{\text{Cov}(F_X^{-1}(U), F_Y^{-1}(1-U))}{\sigma_X \sigma_Y}$.

Pour démontrer les inégalités $r_{min} \leq r(X, Y) \leq r_{max}$ on a d'après le corollaire 2.1

$$\text{Cov}(F_X^{-1}(U), F_Y^{-1}(1-U)) = \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \{F^-(x_1, x_2) - F_1(x_1)F_2(x_2)\} dx_1 dx_2 \leq \text{cov}[X_1, X_2]$$

car $\forall (x_1, x_2) \in \mathbb{R}^2$ on a $F^-(x_1, x_2) \leq F_X(x_1, x_2)$.

De plus

$$\text{Cov}(F_X^{-1}(U), F_Y^{-1}(U)) = \int_{x_1=0}^{+\infty} \int_{x_2=0}^{+\infty} \{F^+(x_1, x_2) - F_1(x_1)F_2(x_2)\} dx_1 dx_2 \geq \text{cov}[X_1, X_2]$$

car $\forall (x_1, x_2) \in \mathbb{R}^2$ on a $F^+(x_1, x_2) \geq F_X(x_1, x_2)$.

Ainsi

$$r_{min} = \frac{\text{Cov}(F_X^{-1}(U), F_Y^{-1}(1-U))}{\sigma_X \sigma_Y} \leq \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \leq \frac{\text{Cov}(F_X^{-1}(U), F_Y^{-1}(U))}{\sigma_X \sigma_Y} = r_{max}$$

2.2. Coefficient de corrélation de PEARSON

Exemple 2.2.2. Soient $X_1 \sim \text{LN}(0, 1)$ $X_2 \sim \text{LN}(0, \sigma^2)$. Les valeurs extrêmes pour le coefficient de corrélation linéaire sont atteintes lorsque X_1 et X_2 sont en dépendance parfaite. Dès lors, on a

$$r_{max} = \frac{\exp(\sigma) - 1}{\sqrt{\exp(\sigma^2) - 1}\sqrt{e - 1}}$$

$$r_{min} = \frac{\exp(-\sigma) - 1}{\sqrt{\exp(\sigma^2) - 1}\sqrt{e - 1}}$$

En effet, on sait que $X \sim \text{LN}(\mu, \sigma) \Leftrightarrow Y = \ln(X) \sim N(\mu, \sigma)$ et on a

$$E(\ln X) = \mu \text{ et } V(\ln X) = \sigma^2$$

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \text{ et } V(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1).$$

Ainsi puisque $X_1 \sim \text{LN}(0, 1)$ et $X_2 \sim \text{LN}(0, \sigma^2)$ on a

$$E(\ln X_1) = 0 \text{ et } V(\ln X_1) = 1$$

$$E(X_1) = \exp\left(\frac{1}{2}\right) \text{ et } V(X_1) = e(e - 1)$$

$$E(X_2) = \exp\left(\frac{\sigma^2}{2}\right) \text{ et } V(X_2) = \exp(\sigma^2)(\exp(\sigma^2) - 1)$$

$$E(\ln X_2) = 0 \text{ et } V(\ln X_2) = \sigma^2.$$

Si X_1 et X_2 sont comonotones il en est de même de $\ln X_1$ et $\ln X_2$ Ainsi

$$E(\ln(X_1 X_2)) = E(\ln X_1 + \ln X_2) = E(\ln X_1) + E(\ln X_2) = 0$$

$$V(\ln X_1 X_2) = V(\ln X_1 + \ln X_2) = V(\ln X_1) + V(\ln X_2) + 2\text{cov}(\ln X_1, \ln X_2)$$

Or

$$\text{cov}(\ln X_1, \ln X_2) = r(\ln X_1, \ln X_2)\sigma_{\ln X_1}\sigma_{\ln X_2} = 1 \times 1 \times \sigma = \sigma$$

Car les variables $\ln X_1$ et $\ln X_2$ étant gaussiennes et comonotones d'où on a $r(\ln X_1, \ln X_2) = 1$.

D'où $V(\ln X_1 X_2) = 1 + 2\sigma + \sigma^2$ et donc $X_1 X_2 \sim \text{LN}(0, \sqrt{1 + 2\sigma + \sigma^2})$

Alors

$$\begin{aligned} r(X_1, X_2) &= \frac{E(X_1 X_2) - E(X_1)E(X_2)}{\sigma_{X_1}\sigma_{X_2}} \\ &= \frac{\exp\left(\frac{1+2\sigma+\sigma^2}{2}\right) - \exp\left(\frac{1}{2}\right)\exp(\sigma^2)}{\sqrt{e(e-1)}\sqrt{\exp(\sigma^2)(\exp(\sigma^2)-1)}} \\ &= \frac{\exp\left(\frac{1+\sigma^2}{2}\right)(\exp(\sigma) - 1)}{\exp\left(\frac{1+\sigma^2}{2}\right)\sqrt{\sigma-1}\sqrt{\exp(\sigma^2)-1}} \\ &= \frac{\exp(\sigma) - 1}{\sqrt{\sigma-1}\sqrt{\exp(\sigma^2)-1}} \end{aligned}$$

2.2. Coefficient de corrélation de PEARSON

D'où

$$r_{max} = \frac{\exp(\sigma) - 1}{\sqrt{\sigma - 1} \sqrt{\exp(\sigma^2 - 1)}}$$

Si X_1 et X_2 sont antimotones il en est de même de $\ln X_1$ et $\ln X_2$ Ainsi

$$E(\ln(X_1 X_2)) = E(\ln X_1 + \ln X_2) = E(\ln X_1) + E(\ln X_2) = 0$$

$$V(\ln X_1 X_2) = V(\ln X_1 + \ln X_2) = V(\ln X_1) + V(\ln X_2) + 2cov(\ln X_1, \ln X_2)$$

Or

$$cov(\ln X_1, \ln X_2) = r(\ln X_1, \ln X_2) \sigma_{\ln X_1} \sigma_{\ln X_2} = -1 \times 1 \times \sigma = -\sigma$$

Car les variables $\ln X_1$ et $\ln X_2$ étant gaussiennes et antimotones on a $r(\ln X_1, \ln X_2) = -1$.

D'où $V(\ln X_1, \ln X_2) = 1 - 2\sigma + \sigma^2$ et donc $X_1 X_2 \sim LN(0, \sqrt{1 - 2\sigma + \sigma^2})$

Alors

$$\begin{aligned} r(X_1, X_2) &= \frac{E(X_1 X_2) - E(X_1)E(X_2)}{\sigma_{X_1} \sigma_{X_2}} \\ &= \frac{\exp(\frac{1-2\sigma+\sigma^2}{2}) - \exp(\frac{1}{2})\exp(\frac{\sigma^2}{2})}{\sqrt{e(e-1)} \sqrt{\exp(\sigma^2)(\exp(\sigma^2) - 1)}} \\ &= \frac{\exp(\frac{1+\sigma^2}{2})(\exp(-\sigma) - 1)}{\exp(\frac{1+\sigma^2}{2}) \sqrt{\sigma - 1} \sqrt{\exp(\sigma^2 - 1)}} \\ &= \frac{\exp(-\sigma) - 1}{\sqrt{\sigma - 1} \sqrt{\exp(\sigma^2 - 1)}} \end{aligned}$$

D'où

$$r_{min} = \frac{\exp(-\sigma) - 1}{\sqrt{\sigma - 1} \sqrt{\exp(\sigma^2 - 1)}}$$

On constate que

$$\lim_{\sigma \rightarrow +\infty} r_{min}(\sigma) = \lim_{\sigma \rightarrow +\infty} r_{max}(\sigma) = 0$$

Par conséquent, il est possible d'avoir un couple X dont le coefficient de corrélation est quasiment nul alors que les composantes de celui-ci sont en dépendance parfaite. Ceci contredit clairement l'intuition que de faibles valeurs du coefficient de corrélation traduisent une faible dépendance.

Propriété 2.3. Le coefficient de corrélation linéaire n'est pas invariante par transformation strictement monotones.

En effet si nous considérons les données de deux variables aléatoires X et Y .

$X = (174, 175, 180, 168, 175, 170, 170, 178, 187, 178)$ et $Y = (64, 59, 64, 62, 51, 60, 68, 63, 92, 70)$
on a : $r(X, Y) = 0,66$ et $r(\ln(X), \ln(Y)) = 0,61$.

Il en ressort ainsi que le coefficient de corrélation linéaire **n'est pas une bonne mesure de dépendance** Puisque les propriétés P3, P4 et P6 de la définition 2.2 ne sont pas satisfaites.

le fait que le coefficient de corrélation linéaire n'est pas une bonne mesure de dépendance rend son

2.2. Coefficient de corrélation de PEARSON

usage problématique dans la résolution des problèmes pratiques.

De plus n'est défini que pour des couples de variables admettent des moments d'ordre 2, ce qui est un gros handicap car en finance beaucoup de risques n'admettent pas de moment d'ordre 2.

MESURES DE DÉPENDANCE BASÉES SUR LES RANGS

Dans ce chapitre, nous présentons deux mesures de dépendance de risques basées sur les rangs. **Le tau de Kendall et le Rho de Spearman** et nous montrons que contrairement au coefficient de corrélation linéaire, celles ci fournissent de bonnes mesures de la dépendance entre deux risques. En suite nous présentons une mesure de la dépendance asymptotique entre deux risques basée sur les modèles de dépendance de queue : **le coefficient de dépendance de queue**).

3.1 Coefficient de corrélation des rangs de Kendall

Dans cette partie nous présentons le coefficient de corrélation des rangs de Kendall qui contrairement au coefficient de corrélation linéaire, est une "bonne mesure de dépendance". Il utilise les concepts de concordance et discordance définis au chapitre un.

3.1.1 Définition et propriétés

Définition 3.1. Soit $X = (X_1, X_2)$ un couple de variables aléatoires réelles.

Le coefficient de corrélation des rangs de Kandell du vecteur X que l'on note $\tau(X_1, X_2)$ est défini par

$$\tau(X_1, X_2) = P[(X_1 - Y_1)(X_2 - Y_2) > 0] - P[(X_1 - Y_1)(X_2 - Y_2) < 0]$$

où (Y_1, Y_2) est un couple indépendant de (X_1, X_2) et de même loi.

Le tau de kendall compare ainsi la probabilité que deux couples indépendants et de même loi soient en concordance et la probabilité qu'ils soient en discordance.

Proposition 3.1. Pour tout couple $X = (X_1, X_2)$, on a

$$\tau(X_1, X_2) = 4E[F_X(X_1, X_2)] - 1$$

3.1. Coefficient de corrélation des rangs de Kendall

Preuve. Soit $Y = (Y_1, Y_2)$ couple indépendant et de même loi que X . On a :

$$\begin{aligned}\tau(X_1, X_2) &= P[(X_1 - Y_1)(X_2 - Y_2) > 0] - P[(X_1 - Y_1)(X_2 - Y_2) < 0] \\ &= P[(X_1 - Y_1)(X_2 - Y_2) > 0] - (1 - P[(X_1 - Y_1)(X_2 - Y_2) > 0]) \\ &= 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1 \\ &= P(X_1 > Y_1, X_2 > Y_2) + P(X_1 < Y_1, X_2 < Y_2) - 1 \\ &= \int_{\Omega} F_Y(X_1(\omega), X_2(\omega)) dP(X(\omega)) + \int_{\Omega} F_X(Y_1(\omega), Y_2(\omega)) dP(X(\omega)) - 1 \\ &= 2(E(F_Y(X_1, X_2)) + E(F_X(Y_1, Y_2))) - 1 \\ &= 4E(F_X(Y_1, Y_2)) - 1 \text{ car } X \text{ et } Y \text{ on même loi.}\end{aligned}$$

D'où $\tau(X_1, X_2) = 4E[F_X(X_1, X_2)] - 1$. ■

Proposition 3.2. Le tau de Kendall est une "bonne mesure de dépendance".

Preuve. Nous allons montrer que tau de Kendall vérifie les 6 propriétés exigées à une mesure de dépendance pour qu'elle soit "une bonne mesure de dépendance".

Soit $X = (X_1, X_2)$ un vecteur aléatoire et $Y = (Y_1, Y_2)$ un vecteur aléatoire indépendant de X et de même loi.

1. Symétrie

On a $\tau(X_1, X_2) = \tau(X_2, X_1)$. En effet,

$$\begin{aligned}\tau(X_1, X_2) &= P[(X_1 - Y_1)(X_2 - Y_2) > 0] - P[(X_1 - Y_1)(X_2 - Y_2) < 0] \\ &= P[(X_2 - Y_2)(X_1 - Y_1) > 0] - P[(X_2 - Y_2)(X_1 - Y_1) < 0] \\ &= \tau(X_2, X_1).\end{aligned}$$

2. Normalité

On a $-1 \leq \tau(X_1, X_2) \leq 1$. En effet,

$$\tau(X_1, X_2) = 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1$$

or

$$0 \leq P[(X_1 - Y_1)(X_2 - Y_2) > 0] \leq 1$$

alors

$$-1 \leq 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1 \leq 1$$

d'où

$$-1 \leq \tau(X_1, X_2) \leq 1.$$

3.1. Coefficient de corrélation des rangs de Kendall

3. Dépendance parfaite positive

X_1 et X_2 comonotones $\Leftrightarrow \tau(X_1, X_2) = 1$. En effet,

$$\begin{aligned}\tau(X_1, X_2) = 1 &\iff 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1 = 1 \\ &\iff 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] = 2 \\ &\iff P[(X_1 - Y_1)(X_2 - Y_2) > 0] = 1 \\ &\iff (X_1 - Y_1)(X_2 - Y_2) > 0 \text{ Ps} \\ &\iff X_1 \text{ et } X_2 \text{ comonotones.}\end{aligned}$$

4. Dépendance parfaite négative

X_1 et X_2 antimonotones $\Leftrightarrow \tau(X_1, X_2) = -1$. En effet,

$$\begin{aligned}\tau(X_1, X_2) = -1 &\iff 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1 = -1 \\ &\iff P[(X_1 - Y_1)(X_2 - Y_2) > 0] = 0 \\ &\iff P[(X_1 - Y_1)(X_2 - Y_2) < 0] = 1 \\ &\iff (X_1 - Y_1)(X_2 - Y_2) < 0 \text{ Ps} \\ &\iff X_1 \text{ et } X_2 \text{ antimonotones.}\end{aligned}$$

5. Dépendance

Supposons X_1 et X_2 indépendant et montrons que $\tau(X_1, X_2) = 0$.

On a d'après la propriété 3.1

$$\tau(X_1, X_2) = 4E[F_X(X_1, X_2)] - 1 = 4E[F_{X_1}(X_1)]E[F_{X_2}(X_2)] - 1$$

Car X_1 et X_2 indépendantes.

Or $F_{X_1}(X_1) \sim U[0, 1]$ et $F_{X_2}(X_2) \sim U[0, 1]$ donc $E[F_{X_1}(X_1)] = E[F_{X_2}(X_2)] = \frac{1}{2}$,
d'où

$$\tau(X_1, X_2) = 4\left(\frac{1}{4}\right) - 1 = 0$$

et donc $\tau(X_1, X_2) = 0$.

6. Invariance

Soit g une fonction strictement monotone on a

$$\tau(g(X), Y) = \begin{cases} \tau(X, Y) & \text{si } g \text{ est croissante} \\ -\tau(X, Y) & \text{si } g \text{ est décroissante.} \end{cases}$$

En effet ;

-Si g est croissante on a $g(X_1) - g(Y_1)$ et $X_1 - Y_1$ ont le même signe.

3.1. Coefficient de corrélation des rangs de Kendall

Ainsi

$$\begin{aligned}\tau(g(X), Y) &= 2P[(g(X_1) - g(Y_1))(X_2 - Y_2) > 0] - 1 \\ &= 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1 \\ &= \tau(X, Y)\end{aligned}$$

d'où $\tau(g(X), Y) = \tau(X, Y)$.

-Si g est décroissante on a $g(X_1) - g(Y_1)$ et $X_1 - Y_1$ sont de signes contraires.

Ainsi

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

d'où $\tau(g(X), Y) = -\tau(X, Y)$.

■

3.1.2 Tau de kendall valeur empirique

(Yadolah 2007)

Soient deux variables X et Y observées sur un échantillon de taille n fournissant n paires d'observations $(x_1, y_1), \dots, (x_n, y_n)$. On obtient le coefficient des rangs de Kendall en déterminant les nombres N_C et N_D des paires d'observations respectivement **concordantes** et **discordantes**.

Avec les notations introduites le tau de Kendall de X et Y est

$$\hat{\tau}(X, Y) = \frac{2(N_C - N_D)}{n(n-1)}.$$

Deux paires pour lesquelles $x_i = x_j$ ou $y_i = y_j$ ne sont ni concordantes ni discordantes et ne sont comptabilisées ni dans N_C ni dans N_D .

Il est possible de donner une formulation équivalente du tau de Kendall.

On obtient aussi le tau de Kendall en ordonnant les valeurs x_i en ordre croissant et en comptant le nombre de valeur y_i correspondantes ne satisfaisant pas cet ordre.

On note par Q le nombre d'inversions nécessaires parmi les valeurs de Y pour obtenir le même ordre(croissant) que celui des valeurs de X .

Le coefficient de corrélation des rangs de Kendall est alors défini par

$$\hat{\tau}(X, Y) = 1 - \frac{4Q}{n(n-1)}.$$

Après le calcul de la valeur empirique du taux de Kendall, il existe des tests statistiques permettant de vérifier sa significativité.

3.1.3 Test d'hypothèse

Le coefficient de corrélation des rangs de Kendall est souvent utilisé comme test statistique afin de déterminer s'il existe une relation entre deux variables aléatoires. Le test peut être un test bilatéral ou un test unilatéral.

Les hypothèses sont alors :

Cas bilatéral(A) :

$$H_0 : \tau = 0 ;$$

$$H_1 : \tau \neq 0$$

On parle de dépendance positive si les grandes valeurs de X ont tendance à être associées avec les grandes valeurs de Y et les petites valeurs de X avec les petites valeurs de Y . On parle de dépendance négative si les grandes valeurs de X ont tendance à être associées aux petites valeurs de Y et vice-versa.

Cas unilatéral(B) :

$$H_0 : \tau = 0 ;$$

$$H_1 : \tau > 0$$

C cas unilatéral :

$$H_0 : \tau = 0 ;$$

$$H_1 : \tau < 0$$

Le test statistique est défini de la façon suivante :

$$T = N_c - N_d$$

règles de décision

Les règles de décisions seront différentes selon les hypothèses posées. C'est ainsi que l'on a les règles de décision A, B, et C relatives au cas précédents.

Règles décisions A

On rejette H_0 au seuil de signification α si

$$T > t_{n, 1-\frac{\alpha}{2}} \quad \text{ou} \quad T < t_{n, \frac{\alpha}{2}}$$

3.2. Rho de Spearman

où t est la valeur critique du test donnée par la table de Kendall.

Sinon on accepte H_0 .

Règles décisions B

On rejette H_0 au seuil de signification α si

$$T > t_{n, 1-\frac{\alpha}{2}}$$

Sinon on accepte H_0 .

Règles décisions C

On rejette H_0 au seuil de signification α si

$$T < t_{n, \frac{\alpha}{2}}$$

Sinon on accepte H_0 .

3.2 Rho de Spearman

3.2.1 définition et propriété

Tout comme le tau de Kendall, le rho de Spearman est basé sur la notion de concordance et discordance. Encore appelé coefficient de corrélation de Spearman ou corrélation des rangs de Spearman, c'est une mesure de risque basée sur les rangs. Il est obtenu comme coefficient de corrélation linéaire des variables aléatoires uniformes issues des variables initiales.

Définition 3.2. Soient X et Y deux variables aléatoires de fonctions de répartition respectives F_X et F_Y .

Posons $U = F_X(X)$ et $V = F_Y(Y)$, U et V suivent des lois uniformes sur $[0; 1]$.

Le Rho de Spearman est alors défini par :

$$\rho_S(X, Y) = r(F_X(X), F_Y(Y)) = r(U, V)$$

Ainsi le coefficient de corrélation de Spearman remédie aux défauts du coefficient de corrélation linéaire de Pearson, en considérant non plus les variables de départ X et Y mais bien des versions uniformes $F_X(X)$ et $F_Y(Y)$ de celles-ci.

définitions équivalentes

Soient X et Y deux variables aléatoires et soient X^\perp et Y^\perp des variables aléatoires telles que $F_{X^\perp} = F_X$ et $F_{Y^\perp} = F_Y$ avec $F_{X^\perp Y^\perp} = F_X F_Y$.

(X^\perp, Y^\perp) est l'élément de la classe de Fréchet $\mathbf{F}(F_X, F_Y)$ correspondant au cas où X et Y sont des

3.2. Rho de Spearman

variables indépendantes. On dit encore (X^\perp, Y^\perp) sont des versions indépendantes de (X, Y) .

Le Rho de Spearman est définie par :

$$\rho_S(X, Y) = 3[P((X - X^\perp)(Y - Y^\perp) > 0) - P((X - X^\perp)(Y - Y^\perp) < 0)]$$

C'est donc le triple de la différence entre probabilité de concordance et discordance dans les couples (X, Y) et (X^\perp, Y^\perp) .

Soient $(X_1, X_2), (X'_1, X'_2)$ et (X''_1, X''_2) trois paires indépendantes de variables aléatoires continues de même loi, le Rho de Spearman se définit encore comme suit :

$$\rho_S(X_1, X_2) = 3[P((X_1 - X'_1)(X_2 - X''_2) > 0) - P((X_1 - X'_1)(X_2 - X''_2) < 0)]$$

qui se réécrit

$$\rho_S(X_1, X_2) = 6P[(X - X'_1)(X_2 - X''_2) > 0] - 3$$

Propriété 3.1. Soient X et Y deux variables aléatoires.

$$\rho_S(X, Y) = 12E(UV) - 3 \text{ où } U \text{ et } V \text{ suivent des lois uniformes sur } [0,1]$$

Preuve. Posons $U = F_X(X)$ et $V = F_Y(Y)$ alors

$$\begin{aligned} \rho_S(X, Y) &= r(F_X(X), F_Y(Y)) \\ &= r(U, V) \\ &= \frac{E(UV) - E(U)E(V)}{\sqrt{V(U)V(V)}} \\ &= \frac{E(UV) - \frac{1}{4}}{\frac{1}{12}} \\ &= 12E(UV) - 3 \end{aligned}$$

■

Propriété 3.2. Le Rho de Spearman est une mesure de dépendance.

Preuve. On va montrer que le Rho de spearman vérifie les 5 propriétés exigibles à une mesure de dépendance.

Soient X et Y deux variables aléatoires et X^\perp et Y^\perp deux variables de même loi que X (resp) Y tel que X^\perp et Y^\perp indépendantes.

1. Symétrie

$$\rho_S(X, Y) = r(F_X(X), F_Y(Y)) = r(F_Y(Y), F_X(X)) = \rho_S(Y, X)$$

donc ρ_S est symétrique car r est symétrique.

2. normalité

$\rho_S(X, Y) = r(F_X(X), F_Y(Y))$ donc ρ_S est à valeurs dans $[-1, 1]$ car r est à valeur dans $[-1, 1]$.

3. dépendance parfaite positive

Montrons que X et Y comonotones $\Leftrightarrow \rho_S(X, Y) = 1$.

On peut sans nuire à la généralité supposer que $X \sim U[0, 1]$ et $Y \sim U[0, 1]$.

Alors X et Y comonotones $\Leftrightarrow X = Y \Leftrightarrow r(X, Y) = 1 \Leftrightarrow \rho_S(X, Y) = 1$

Car X et Y sont de variables aléatoires de loi uniforme sur $[0, 1]$ et en vertu de proposition 1.3.

4. dépendance parfaite négative

Montrons que X et Y antimonotones $\Leftrightarrow \rho_S(X, Y) = -1$.

On peut sans nuire à la généralité supposer que $X \sim U[0; 1]$ et $Y \sim U[0, 1]$.

Alors X et Y antimonotones $\Leftrightarrow X = 1 - Y \Leftrightarrow r(X, Y) = -1 \Leftrightarrow \rho_S(X, Y) = -1$.

Car X et Y sont de variables aléatoires de loi uniforme sur $[0, 1]$ et en vertu de proposition 1.3.

5. indépendance

Supposons X et Y indépendants montrons que $\rho_S(X, Y) = 0$.

on a $\rho_S(X, Y) = r(F_X(X), F_Y(Y))$; et puisque X et Y sont indépendantes il en est de même de $F_X(X)$ et $F_Y(Y)$ alors $\rho_S(X, Y) = r(F_X(X), F_Y(Y)) = 0$.

Car X et Y sont de variables aléatoires de loi uniforme sur $[0, 1]$.

6. invariance

Soit g une fonction strictement monotone montrons que

$$\rho_S(g(X), Y) = \begin{cases} \rho_S(X, Y) & \text{si } g \text{ est croissante} \\ -\rho_S(X, Y) & \text{si } g \text{ est décroissante.} \end{cases}$$

On a $\rho_S(g(X), Y) = 3[P((g(X) - g(X^\perp))(Y - Y^\perp) > 0) - P((g(X) - g(X^\perp))(Y - Y^\perp) < 0)]$

alors :

- si g est croissante alors $g(X) - g(X^\perp)$ et $X - X^\perp$ ont le même signe alors

$$\begin{aligned} \rho_S(g(X), Y) &= 3[P((g(X) - g(X^\perp))(Y - Y^\perp) > 0) - P((g(X) - g(X^\perp))(Y - Y^\perp) < 0)] \\ &= 3[P((X - X^\perp)(Y - Y^\perp) > 0) - P((X - X^\perp)(Y - Y^\perp) < 0)] \\ &= \rho_S(X, Y) \end{aligned}$$

- si g est décroissante alors $g(X) - g(X^\perp)$ et $X - X^\perp$ sont de signe contraire alors

$$\begin{aligned} \rho_S(g(X), Y) &= 3[P((g(X) - g(X^\perp))(Y - Y^\perp) > 0) - P((g(X) - g(X^\perp))(Y - Y^\perp) < 0)] \\ &= 3[P((X - X^\perp)(Y - Y^\perp) < 0) - P((X - X^\perp)(Y - Y^\perp) > 0)] \\ &= -\rho_S(X, Y). \end{aligned}$$

■

3.2.2 Rho de Spearman valeur empirique

Soit $(x_1, y_1), \dots, (x_n, y_n)$ un échantillon de taille n de (X, Y) .

Pour $i = 1, \dots, n$ notons r_i^x le rang de x_i dans l'échantillon x_1, \dots, x_n et r_i^y le rang de y_i dans l'échantillon y_1, \dots, y_n . Une estimation de ρ_S est

$$\hat{\rho}_S = 1 - \frac{6 \sum_{i=1}^n (r_i^x - r_i^y)^2}{n(n^2 - 1)}$$

Si plusieurs observations ont exactement la même valeur, on attribue un rang moyen.

S'il y a plusieurs rangs moyens, il convient de faire une correction et de calculer le coefficient de corrélation classique de Pearson entre les rangs. On obtient ainsi

$$\hat{\rho}_S = \frac{n \left(\sum_{i=0}^n r_i^x r_i^y \right) - \left(\sum_{i=0}^n r_i^x \right) \left(\sum_{i=0}^n r_i^y \right)}{\sqrt{n \left(\sum_{i=0}^n (r_i^x)^2 - \left(\sum_{i=0}^n r_i^x \right)^2 \right)} \sqrt{n \left(\sum_{i=0}^n (r_i^y)^2 - \left(\sum_{i=0}^n r_i^y \right)^2 \right)}}$$

3.2.3 Test d'hypothèse

Le Rho de Spearson est souvent utilisé comme test statistique afin de déterminer s'il existe une relation entre deux variables aléatoires. Le test peut être un test bilatéral ou un test unilatéral. Les hypothèses sont alors

Cas bilatéral(A) :

$$H_0 : \rho_S = 0;$$

$$H_1 : \rho_S \neq 0$$

Cas unilatéral(B) :

$$H_0 : \rho_S = 0;$$

$$H_1 : \rho_S > 0$$

C cas unilatéral :

$$H_0 : \rho_S = 0;$$

$$H_1 : \rho_S < 0$$

règles de décision

Les règles de décision seront différentes selon les hypothèses posées. C'est ainsi que l'on a les règles

3.2. Rho de Spearman

de décision A, B, et C relatives au cas précédents.

Règle de décisions A

On rejette H_0 au seuil de signification α si

$$\rho_S > t_{n,1-\frac{\alpha}{2}} \quad \text{ou} \quad \rho_S < t_{n,\frac{\alpha}{2}}$$

ou t est la valeur critique du test donnée par la table de Spearman.

Sinon on accepte H_0 .

Règle de décisions B

On rejette H_0 au seuil de signification α si

$$\rho_S > t_{n,1-\alpha}$$

Sinon on accepte H_0 .

Règle de décisions C

On rejette H_0 au seuil de signification α si

$$\rho_S < t_{n,\alpha}$$

Sinon on accepte H_0 .

Remarque 3.2.1. La notation t que Spearman utilise n'est aucunement liée à celle de Student.

3.2.4 Liens entre le tau de Kendall et le Rho de Spearman

Il est à noter que le tau de Kendall et le rho de Spearman sont liés, au sens où il est par exemple impossible d'avoir $\tau \geq 0,4$ et $\rho_S = 0$. Dans cette section, nous établissons une série de relations entre le tau de Kendall et le rho de Spearman.

Théorème 3.1. (Denuit et al. 2004)

Les inégalités $-1 \leq 3\tau - 2\rho \leq 1$ sont vérifiées quel que soit le couple aléatoire dont les marginales sont continues.

Preuve. Soient $X = (X_1, X_2), Y = (Y_1, Y_2)$ et $Z = (Z_1, Z_2)$ trois couples indépendants et de même loi. Le tau de Kendall et rho de Spearman pour ces couples valent

$$\tau = 2P[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1$$

$$\rho = 6P[(X_1 - Y_1)(X_2 - Z_2) > 0] - 3$$

3.2. Rho de Spearman

Puisque X, Y et Z ont même loi, nous pouvons écrire

$$\tau = \frac{2}{3} \{P[(X_1 - Y_1)(X_2 - Y_2) > 0] + P[(X_1 - Z_1)(Y_2 - Z_2) > 0] + p[(Z_1 - X_1)(Z_2 - X_2) > 0]\} - 1$$

et

$$\begin{aligned} \rho &= \{P[(X_1 - Y_1)(X_2 - Z_2) > 0] + P[(X_1 - Z_1)(Y_2 - Z_2) > 0] \\ &+ p[(Y_1 - X_1)(Y_2 - Z_2) > 0] + p[(Z_1 - Y_1)(Z_2 - X_2) > 0] \\ &+ P[(Y_1 - Z_1)(Y_2 - X_2) > 0] + P[(Z_1 - X_1)(Z_2 - Y_2) > 0] \\ &- 3\} \end{aligned}$$

On peut supposer sans perte de généralité que les inégalités $X_1 < Y_1 < Z_1$ sont vérifiées, de sorte que

$$\tau = \frac{2}{3} \{P[X_2 < Y_2] + P[X_2 < Z_2] + P[X_2 < Z_2] - 1$$

et

$$\begin{aligned} \rho &= P[X_2 < Z_2] + P[X_2 > Y_2] + P[Y_2 < Z_2] \\ &+ P[Z_2 < X_2] + P[Y_2 < Z_2] + P[Z_2 > Y_2] - 3 \\ &= 2P[X_2 < Z_2] - 1 \end{aligned}$$

Car $P[Z_2 < X_2] + P[X_2 < Z_2] = 1$, $P[Y_2 < Z_2] = P[X_2 < Z_2]$ et $P[X_2 > Y_2] + P[Z_2 > Y_2] = 1$.

Soit P_{XYZ} la probabilité de $\{X_2 < Y_2 < Z_2\}$ étant donné que $\{X_1 < Y_1 < Z_1\}$. Les six probabilités P_{XYZ} (i.e $P_{XYZ}, P_{YXZ}, P_{YZX}, P_{XZY}, P_{ZYX}, P_{ZXY}$) ont une somme qui vaut 1. De ce fait, nous pouvons écrire

$$\begin{aligned} \tau &= \frac{2}{3} \{(P_{XYZ} + P_{XZY} + P_{ZXY}) + (P_{XYZ} + P_{YXZ} + P_{YZX}) + (P_{XYZ} + P_{XZY} + P_{YXZ})\} - 1 \\ &= P_{XYZ} + \frac{1}{3}(P_{XZY} + P_{YXZ}) - \frac{1}{3}(P_{YZX} + P_{ZXY}) - P_{ZYX} \end{aligned}$$

de même que

$$\begin{aligned} \rho &= 2(P_{XYZ} + P_{XZY} + P_{YXZ}) - 1 \\ &= P_{XYZ} + P_{XZY} + P_{YXZ} - P_{YZX} - P_{ZXY} - P_{ZYX} \end{aligned}$$

car on a

$$\begin{aligned} P(X_2 < Z_2) &= P(X_2 < Y_2 < Z_2) + P(X_2 < Z_2 < Y_2) + P(Y_2 < X_2 < Z_2) \\ &= P_{XYZ} + P_{XZY} + P_{YXZ} \end{aligned}$$

et

$$P_{XYZ} + P_{XZY} + P_{YXZ} + P_{YZX} + P_{ZXY} + P_{ZYX} = 1$$

Dès lors

$$3\tau - 2\rho = (P_{XYZ} + P_{YZX} + P_{ZXY}) - ((P_{XYZ} + P_{YZX} + P_{ZXY}))$$

de sorte que

$$-1 \leq 3\tau - 2\rho \leq 1.$$

3.2. Rho de Spearman

Car $0 \leq (P_{XYZ} + P_{YZX} + P_{ZXY}) \leq 1$ et $0 \leq (P_{XYZ} + P_{YZX} + P_{ZXY}) \leq 1$

Théorème 3.2. (Denuit et al. 2004)

Les inégalités

$$\frac{1 + \rho}{2} \geq \left(\frac{1 + \tau}{2}\right)^2 \quad \text{et} \quad \frac{1 - \rho}{2} \geq \left(\frac{1 - \tau}{2}\right)^2$$

sont toutes deux satisfaites quel que soit le vecteur aléatoire X dont les marginales sont continues.

Preuve. Soient X, Y et Z trois couples indépendants et de même loi. Notons par F_X leur fonction de répartition commune. définissons p la probabilité que deux ces couples soient concordants avec le troisième. Cette probabilité peut alors s'écrire

$$\begin{aligned} p &= P[Y \text{ et } Z \text{ sont concordants avec } X] \\ &= \int \int_{x \in \mathbb{R}^2} P[Y \text{ et } Z \text{ sont concordant avec } x] dF_X(x) \\ &= \int \int_{x \in \mathbb{R}^2} p[(Y_1 - x_1)(Y_2 - x_2) > 0] p[(Z_1 - x_1)(Z_2 - x_2) > 0] dF_X(x) \\ &= \int \int_{x \in \mathbb{R}^2} (p[(Y_1 - x_1)(Y_2 - x_2) > 0])^2 dF_X(x) \\ &\geq \left\{ \int \int_{x \in \mathbb{R}^2} p[(Y_1 - x_1)(Y_2 - x_2) > 0] dF_X(x) \right\}^2 \\ &= \{P[(Y_1 - X_1)(Y_2 - X_2) > 0]\}^2 \\ &= \left(\frac{1 + \tau}{2}\right)^2 \quad \text{car} \quad \tau = 2p[(X_1 - Y_1)(X_2 - Y_2) > 0] - 1 \end{aligned}$$

D'où

$$p \geq \left(\frac{1 + \tau}{2}\right)^2$$

A présent, en invoquant l'indépendance et l'identique distribution de X, Y et Z , nous pouvons encore écrire

$$\begin{aligned} p &= \frac{1}{3} \{P[Y \text{ et } Z \text{ sont concordants avec } X] \\ &\quad + P[X \text{ et } Z \text{ sont concordants avec } Y] \\ &\quad + P[X \text{ et } Y \text{ sont concordants avec } Z]\} \end{aligned}$$

En recourant aux mêmes notations que dans la preuve du théorème 3.1. nous obtenons

$$\begin{aligned} p &= \frac{1}{3} \{(P_{XYZ} + P_{XZY}) + P_{XYZ} + (P_{XYZ} + P_{YZX})\} \\ &= P_{XYZ} + \frac{1}{3} P_{XZY} + \frac{1}{3} P_{YXZ} \end{aligned}$$

Car ayant supposé que $X_1 < Y_1 < Z_1$ on a

$$P[Y \text{ et } Z \text{ sont concordants avec } X] = P[X_2 < Y_2 < Z_2] + P[X_2 < Z_2 < Y_2] = P_{XYZ} + P_{XZY}$$

$$P[X \text{ et } Z \text{ sont concordants avec } Y] = P[X_2 < Y_2 < Z_2] = P_{XYZ}$$

$$P[X \text{ et } Y \text{ sont concordants avec } Z] = P[X_2 < Y_2 < Z_2] + P[Y_2 < X_2 < Z_2] = P_{XYZ} + P_{YZX}$$

3.2. Rho de Spearman

On a montré dans la preuve du théorème 3.1. que $\rho = 2\{P_{XYZ} + P_{XZZ} + P_{YZZ}\} - 1$.

D'où

$$\frac{1+\rho}{2} = P_{XYZ} + P_{XZY} + P_{YXZ} \geq p = P_{XYZ} + \frac{1}{3}P_{XZY} + \frac{1}{3}P_{YXZ} \geq \left(\frac{1+\tau}{2}\right)^2$$

Ce qui prouve la première inégalité du théorème 3.2.

Pour la seconde inégalité on définit p comme la probabilité que deux ces couples soient discordants avec la troisième on a alors :

$$\begin{aligned} p &= P[Y \text{ et } Z \text{ sont discordants avec } X] \\ &= \int \int_{x \in \mathbb{R}^2} P[Y \text{ et } Z \text{ sont discordant avec } x] dF_X(x) \\ &= \int \int_{x \in \mathbb{R}^2} p[(Y_1 - x_1)(Y_2 - x_2) < 0] p[(Z_1 - x_1)(Z_2 - x_2) < 0] dF_X(x) \\ &= \int \int_{x \in \mathbb{R}^2} (p[(Y_1 - x_1)(Y_2 - x_2) < 0])^2 dF_X(x) \\ &\geq \left\{ \int \int_{x \in \mathbb{R}^2} p[(Y_1 - x_1)(Y_2 - x_2) < 0] dF_X(x) \right\}^2 \\ &= \{P[(Y_1 - X_1)(Y_2 - X_2) < 0]\}^2 \\ &= \left(\frac{1-\tau}{2}\right)^2 \text{ car } \tau = -2p[(X_1 - Y_1)(X_2 - Y_2) < 0] + 1 \end{aligned}$$

A présent, en invoquant l'indépendance et l'identique distribution de X , Y et Z , nous pouvons encore écrire

$$\begin{aligned} p &= \frac{1}{3} \{P[Y \text{ et } Z \text{ sont discordants avec } X] \\ &\quad + P[X \text{ et } Z \text{ sont discordants avec } Y] \\ &\quad + P[X \text{ et } Y \text{ sont discordants avec } Z]\} \end{aligned}$$

En recourant aux mêmes notations que dans la preuve du théorème 3.1. nous obtenons ayant supposé que $X_1 < Y_1 < Z_1$.

$$P[Y \text{ et } Z \text{ sont discordants avec } X] = P[Z_2 < Y_2 < X_2] + P[Y_2 < Z_2 < X_2] = P_{ZYX} + P_{YZX}$$

$$P[X \text{ et } Z \text{ sont discordants avec } Y] = P[Z_2 < Y_2 < X_2] = P_{ZYX}$$

$$P[X \text{ et } Y \text{ sont discordants avec } Z] = P[Y_2 < Z_2 < X_2] + P[Z_2 < Y_2 < X_2] = P_{YZX} + P_{ZYX}.$$

Comme $\rho = 2P[X_2 < Z_2] - 1$ alors $\frac{\rho-1}{2} = P[Z_2 < X_2]$

or $P[Z_2 < X_2] = P[Z_2 < X_2 < Y_2] + P[Y_2 < Z_2 < X_2] + P[Z_2 < Y_2 < X_2]$.

D'où

$$\frac{\rho-1}{2} = P[Z_2 < X_2 < Y_2] + P[Y_2 < Z_2 < X_2] + P[Z_2 < Y_2 < X_2] = P_{ZXY} + P_{ZXY} + P_{ZYX}$$

et donc

$$\frac{\rho-1}{2} = P_{ZXY} + P_{YZY} + P_{ZYX} \geq P_{ZXY} + \frac{1}{3}P_{YZX} + \frac{1}{3}P_{ZYX} = p \geq \left(\frac{1-\tau}{2}\right)^2.$$

3.2. Rho de Spearman

Ce qui prouve la seconde inégalité du théorème 3.2. ■

Corollaire 3.1. Quel que soit le couple aléatoire X dont les marginales sont continues, les coefficients tau de Kendall et Rho de Spearman satisfont les inégalités

$$\frac{3\tau - 1}{2} \leq \rho \leq \frac{1 + 2\tau - \tau^2}{2} \quad \text{si} \quad \tau \geq 0$$

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho \leq \frac{1 + 3\tau}{2} \quad \text{si} \quad \tau \leq 0$$

Preuve. On a

$$\frac{1 + \rho}{2} \geq \left(\frac{1 + \tau}{2}\right)^2$$

d'où

$$\rho \geq \frac{\rho^2 + 2\rho + 1}{2}$$

et

$$\frac{1 - \rho}{2} \geq \left(\frac{1 - \tau}{2}\right)^2$$

d'où

$$\rho \leq \frac{-\rho^2 + 2\rho + 1}{2}$$

de plus

$$-1 \leq 3\tau - 2\rho \leq 1$$

alors

$$\frac{3\tau - 1}{2} \leq \rho \leq \frac{3\tau + 1}{2}$$

or si $\tau \leq 0$

on a

$$\frac{3\tau + 1}{2} - \frac{-\tau^2 + 2\tau + 1}{2} = \frac{\tau(-\tau + 1)}{2} \leq 0$$

alors

$$\frac{3\tau + 1}{2} \leq \frac{-\tau^2 + 2\tau + 1}{2}$$

et

$$\frac{3\tau - 1}{2} - \frac{\tau^2 + 2\tau - 1}{2} = \frac{\tau(-\tau + 1)}{2} \leq 0$$

d'où

$$\frac{3\tau - 1}{2} \leq \frac{\tau^2 + 2\tau - 1}{2}$$

3.3. Coefficient de dépendance de queue

et donc

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho \leq \frac{1 + 3\tau}{2}.$$

De même si $\tau \leq 0$ on montre que

$$\frac{3\tau - 1}{2} \leq \rho \leq \frac{1 + 2\tau - \tau^2}{2}.$$

■

3.3 Coefficient de dépendance de queue

Jusqu'à présent les mesures de dépendance présentées servent à décrire le comportement de deux risques tout au long de leur distribution. Le coefficient de dépendance de queue quant à lui permet de décrire la dépendance des valeurs extrêmes (qui se rapportent principalement aux queues de distributions). Elle permet en effet de mesurer le niveau de la dépendance dans les valeurs extrêmes. **grande** (*upper tail dependance*) ou **petite** (*lower tail dependance*).

Ainsi, si F est une distribution bidimensionnelle, la dépendance de queue caractérise le comportement de F au point de x^+ et x^-

où

$$\begin{aligned}x^+ &= \text{Sup}\{x : F(x) \leq 1\} \\x^- &= \text{inf}\{x : F(x) > 0\}.\end{aligned}$$

Pour ce faire, nous avons besoin des inverses des fonctions marginales. Par contre, si ces dernières ne sont pas strictement croissantes, nous devons utiliser l'inverse généralisé ou le quasi-inverse, car la fonction inverse ne sera pas correctement définie.

Définition 3.3. (inverse généralisé)

Soit F une fonction de répartition.

L'**inverse généralisé** $F^{[-1]}$ de F est une fonction définie de la façon suivante

$$\begin{aligned}F^{[-1]}(t) &= \text{Sup}\{x : F(x) \leq t\} \\&= \text{inf}\{x : F(x) \geq t\}\end{aligned}$$

Soit X une variable aléatoire réelle de fonction de répartition F et d'inverse généralisé F^{-1} .

Pour $p \in]0; 1]$, la quantité $F^{-1}(p)$ s'appelle le **quatile** ou **fractile** d'ordre p de la loi de X .

Proposition 3.3. Soit F une fonction définie sur \mathbb{R} à valeurs dans $[0; 1]$, croissante et continue à droite d'inverse généralisé F^{-1} .

3.3. Coefficient de dépendance de queue

1. La fonction F^{-1} est croissante et continue à gauche et on a l'équivalence suivante

$$\forall x \in \mathbb{R}, p \in]0, 1]; F(x) \leq p \iff x \geq F^{-1}(p).$$

2. Pour tout $p \in]0, 1]$, on a $F(F^{-1}(p)) \geq p$ avec égalité si $F^{-1}(p) > -\infty$ et si F est continue en $F^{-1}(p)$.

3. Soit U une variable de loi uniforme sur $[0, 1]$. La fonction F est la fonction de répartition de la variable aléatoire $F^{-1}(U)$.

4. Soit X une variable aléatoire réelle de fonction de répartition F . Si F est continue, alors $F(X)$ suit la loi uniforme sur $[0, 1]$.

Preuve

1. Montrons tout d'abord l'équivalence, soit $x \in \mathbb{R}$ et $p \in]0, 1]$.

Si on a $F(x) \geq p$ alors $x \geq F^{-1}(p)$ par définition de F^{-1} .

Réciproquement si $x \geq F^{-1}(p)$ alors pour tout $\varepsilon > 0$ on a $x + \varepsilon > F^{-1}(p)$, ce qui implique $F(x + \varepsilon) \geq p$. Puisque F est continue à droite, en faisant tendre ε vers 0, on obtient que $F(x) \geq p$. Ceci montre l'équivalence.

Montrons que F^{-1} est continue à gauche sur $]0, 1]$.

Soit $p \in]0, 1]$,

$$F^{-1}(p - \varepsilon) > x, \forall \varepsilon \in]0, p] \iff p - \varepsilon \leq F(x), \forall \varepsilon \in]0, p] \iff p \leq F(x) \iff F^{-1}(p) \leq x.$$

Ceci assure la continuité à gauche de F^{-1} sur $]0, 1]$.

2. On suppose ici que $\lim_{x \rightarrow +\infty} F(x) = 1$.

Soit $p \in]0, 1]$ si $F^{-1}(p) \in \mathbb{R}$, on pose $x = F^{-1}(p)$, comme $F^{-1}(p) \leq x \iff p \geq F(x)$ alors $F(F^{-1}(p)) \geq p$.

Si $F^{-1}(p) = +\infty$ alors l'inégalité est vraie car $F(+\infty) = 1$.

Si $F^{-1}(p) = -\infty$ alors par définition de $F^{-1}(p)$ $F(x) \geq p \forall x \in \mathbb{R}$ et donc $F(-\infty) \geq p$. Ainsi pour tout $p \in]0, 1]$ on a donc $F(F^{-1}(p)) \geq p$.

Supposons que $F^{-1}(p) > -\infty \forall \varepsilon > 0$ $F^{-1}(p) - \varepsilon < F^{-1}(p)$ et (1) implique que $F(F^{-1}(p) - \varepsilon) < p$. Si de plus F est continue en $F^{-1}(p)$, alors en faisant tendre ε vers 0 on obtient $F(F^{-1}(p)) \leq p$ et donc $F(F^{-1}(p)) = p$.

3. Soit U une variable uniforme sur $[0, 1]$. D'après (1) on a pour tout $x \in \mathbb{R}$

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x) \text{ d'où le résultat .}$$

4. Soit U une variable aléatoire uniforme sur $[0, 1]$. Alors les variables aléatoires X et $F^{-1}(U)$ ont même loi. Par conséquent, $F(X)$ a même loi que $F(F^{-1}(U))$, variable égale à U lorsque X est finie P.S et F continue d'après (2).

3.3. Coefficient de dépendance de queue

Définition 3.4. Soient X et Y deux variables aléatoires continues de fonctions de repartions F_X et F_Y respectivement. Le **coefficient de dépendance de queue supérieure** (upper tail dependence coefficient) de X et Y , est défini par $\lambda_U = \lim_{u \rightarrow 1^-} P(Y \geq F_Y^{-1}(u) \mid X \geq F_X^{-1}(u))$ lorsque cette limite existe .

- Remarque 3.3.1.**
1. Lorsque $\lambda_U \in]0 ; 1]$, X et Y sont dites **asymptotiquement dépendantes** au niveau supérieur de la queue de distribution. Cette dépendance est d'autant plus forte lorsque λ_U est proche de 1.
 2. Si $\lambda_U = 0$ X et Y sont dites **asymptotiquement indépendantes** au niveau supérieure de la queue de distribution.

Définition 3.5. Soient X et Y deux variables aléatoires continues de fonctions de repartions F_X et F_Y respectivement. le **coefficient de dépendance de queue inférieur** (lower tail dependence coefficient) de X et Y , est défini par $\lambda_L = \lim_{u \rightarrow 0^+} P(Y \leq F_Y^{-1}(u) \mid X \leq F_X^{-1}(u))$ lorsque cette limite existe.

- Remarque 3.3.2.**
1. Lorsque $\lambda_U \in]0 ; 1]$, X et Y sont dites **asymptotiquement dépendantes** au niveau inférieur de la queue de distribution. Cette dépendance est d'autant plus forte lorsque λ_U est proche de 1.
 2. Si $\lambda_U = 0$ X et Y sont dites **asymptotiquement indépendantes** au niveau inférieur de la queue de distribution.

3.3.1 Interprétation du coefficient de dépendance de queue

Nous donnons ici une interprétation de la mesure de dépendance de queue supérieure et inférieure. pour cela considérons la mesure de dépendance **quantile-quantile** suivante

$$\lambda_U(\alpha) = P[Y > F_Y^{-1}(\alpha) \mid X > F_X^{-1}(\alpha)].$$

C'est la probabilité conditionnelle que Y soit plus grand que le quantile $F_Y^{-1}(\alpha)$ sachant que X est plus grand que le quantile $F_X^{-1}(\alpha)$. Ces deux quantiles correspondent au même seuil α . La mesure de dépendance λ_U est donc la limite de cette probabilité conditionnelle lorsque α tend vers 1 ;

$$\lambda_U = \lim_{\alpha \rightarrow 1^-} \lambda(\alpha) = \lim_{\alpha \rightarrow 1^-} P[Y > F_Y^{-1}(\alpha) \mid X > F_X^{-1}(\alpha)].$$

C'est donc la probabilité que Y est un extrême sachant que X est extrême.

Nous définissons aussi la mesure de dépendance quantile - quantile lorsque α tend vers 0 ;

$$\lambda_L(\alpha) = P[Y \leq F_Y^{-1}(\alpha) \mid X \leq F_X^{-1}(\alpha)].$$

3.3. Coefficient de dépendance de queue

La mesure de dépendance de queue inférieure est donc la limite de cette probabilité conditionnelle lorsque α tend vers 0.

$$\lambda_L = \lim_{\alpha \rightarrow 0^+} \lambda_L(\alpha) = \lim_{\alpha \rightarrow 0^+} P[Y \leq F_Y^{-1}(\alpha) \mid X \leq F_X^{-1}(\alpha)].$$

Exemple 3.3.1. Soit X et Y deux variables aléatoires. Déterminons s'ils existent λ_U et λ_L dans chacun des cas suivants :

1. X et Y comonotones

$$\begin{aligned} \lambda_U &= \lim_{u \rightarrow 1^-} P[Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)] \\ &= \lim_{u \rightarrow 1^-} \frac{P(Y > F_Y^{-1}(u), X > F_X^{-1}(u))}{P(X > F_X^{-1}(u))} \\ &= \lim_{u \rightarrow 1^-} \frac{P(F_Y(Y) > u, F_X(X) \geq u)}{P(F_X(X) > u)} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - P(F_Y(Y) \leq u) - P(F_X(X) \leq u) + P(F_Y(Y) \leq u, F_X(X) \leq u)}{1 - P(F_X(X) \leq u)} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - u - u + \text{Min}(F_Y(F_Y^{-1}(u)), F_X(F_X^{-1}(u)))}{1 - u} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + \text{Min}(u, u)}{1 - u} \\ &= \lim_{u \rightarrow 1^-} \frac{1 - u}{1 - u} \\ &= 1 \end{aligned}$$

D'où $\lambda_U = 1$.

$$\begin{aligned} \lambda_L &= \lim_{u \rightarrow 0^+} P(Y \leq F_Y^{-1}(u) \mid X \leq F_X^{-1}(u)) \\ &= \lim_{u \rightarrow 0^+} \frac{P(Y \leq F_Y^{-1}(u), X \leq F_X^{-1}(u))}{P(X \leq F_X^{-1}(u))} \\ &= \lim_{u \rightarrow 0^+} \frac{\text{Min}(F_Y(F_Y^{-1}(u)), F_X(F_X^{-1}(u)))}{P(X \leq F_X^{-1}(u))} \\ &= \lim_{u \rightarrow 0^+} \frac{\text{Min}(u, u)}{u} \\ &= \lim_{u \rightarrow 0^+} \frac{u}{u} \\ &= 1 \end{aligned}$$

D'où $\lambda_L = 1$.

Ainsi les variables comonotones sont asymptotiquement dépendantes aux niveaux supérieur et inférieur des queues de distributions.

2. X et Y antimonotones

3.3. Coefficient de dépendance de queue

$$\begin{aligned}
\lambda_U &= \lim_{u \rightarrow 1^-} P(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)) \\
&= \lim_{u \rightarrow 1^-} \frac{P(Y > F_Y^{-1}(u), X > F_X^{-1}(u))}{P(X > F_X^{-1}(u))} \\
&= \lim_{u \rightarrow 1^-} \frac{P(F_Y(Y) > u, F_X(X) \geq u)}{P(F_X(X) > u)} \\
&= \lim_{u \rightarrow 1^-} \frac{1 - P(F_Y(Y) \leq u) - P(F_X(X) \leq u) + P(F_Y(Y) \leq u, F_X(X) \leq u)}{1 - P(F_X(X) \leq u)} \\
&= \lim_{u \rightarrow 1^-} \frac{1 - u - u + \text{Max}(F_Y(F_Y^{-1}(u)) + F_X(F_X^{-1}(u)) - 1, 0)}{1 - u} \\
&= \lim_{u \rightarrow 1^-} \frac{1 - 2u + \text{Max}(u + u - 1, 0)}{1 - u} \\
&= \lim_{u \rightarrow 1^-} \frac{1 - 2u + \text{Max}(2u - 1, 0)}{1 - u} \\
&= \lim_{u \rightarrow 1^-} \frac{1 - 2u + 2u - 1}{1 - u} \\
&= 0
\end{aligned}$$

D'où $\lambda_u = 0$.

$$\begin{aligned}
\lambda_L &= \lim_{u \rightarrow 0^+} P(Y \leq F_Y^{-1}(u) \mid X \leq F_X^{-1}(u)) \\
&= \lim_{u \rightarrow 0^+} \frac{P(Y \leq F_Y^{-1}(u), X \leq F_X^{-1}(u))}{P(X \leq F_X^{-1}(u))} \\
&= \lim_{u \rightarrow 0^+} \frac{\text{Max}(F_Y(F_Y^{-1}(u)) + F_X(F_X^{-1}(u)) - 1, 0)}{P(F_X(X) \leq u)} \\
&= \lim_{u \rightarrow 0^+} \frac{\text{Max}(u + u - 1, 0)}{u} \\
&= \lim_{u \rightarrow 0^+} \frac{\text{Max}(2u - 1, 0)}{u} \\
&= 0
\end{aligned}$$

D'où $\lambda_U = 0$.

Ainsi les variables antimonotones sont asymptotiquement indépendantes aux niveaux inférieur et supérieur des queues de distributions.

3.4. Application

3. X et Y sont indépendantes

$$\begin{aligned}
 \lambda_U &= \lim_{u \rightarrow 1^-} P(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)) \\
 &= \lim_{u \rightarrow 1^-} \frac{P(Y > F_Y^{-1}(u), X > F_X^{-1}(u))}{P(X > F_X^{-1}(u))} \\
 &= \lim_{u \rightarrow 1^-} \frac{P(F_Y(Y) > u, F_X(X) \geq u)}{P(F_X(X) > u)} \\
 &= \lim_{u \rightarrow 1^-} \frac{1 - P(F_Y(Y) \leq u) - P(F_X(X) \leq u) + P(F_Y(Y) \leq u, F_X(X) \leq u)}{1 - P(F_X(X) \leq u)} \\
 &= \lim_{u \rightarrow 1^-} \frac{1 - u - u + (F_Y(F_Y^{-1}(u)))(F_X(F_X^{-1}(u)))}{1 - u} \\
 &= \lim_{u \rightarrow 1^-} \frac{1 - 2u + u^2}{1 - u} \\
 &= \lim_{u \rightarrow 1^-} \frac{1 - u}{(1 - u)^2} \\
 &= \lim_{u \rightarrow 1^-} \frac{1}{1 - u} \\
 &= \lim_{u \rightarrow 1^-} 1 - u \\
 &= 0
 \end{aligned}$$

D'où $\lambda_U = 0$.

$$\begin{aligned}
 \lambda_L &= \lim_{u \rightarrow 0^+} P(Y \leq F_Y^{-1}(u) \mid X \leq F_X^{-1}(u)) \\
 &= \lim_{u \rightarrow 0^+} \frac{P(Y \leq F_Y^{-1}(u), X \leq F_X^{-1}(u))}{P(X \leq F_X^{-1}(u))} \\
 &= \lim_{u \rightarrow 0^+} \frac{(F_Y(F_Y^{-1}(u)))(F_X(F_X^{-1}(u)))}{P(F_X(X) \leq u)} \\
 &= \lim_{u \rightarrow 0^+} \frac{u^2}{u} \\
 &= \lim_{u \rightarrow 0^+} u \\
 &= 0
 \end{aligned}$$

D'où $\lambda_L = 0$

Ainsi les variables indépendantes sont asymptotiquement indépendantes aux niveaux inférieur et supérieur des queues de distributions.

3.4 Application

Dans cette partie nous présentons les fonctions du logiciel R permettant de déterminer de façon numérique les valeurs du Tau de Kendall, Rho de Spearman et du coefficient de corrélation et leurs misent en œuvre sur une série de données.

Pour déterminer de façon numérique ces valeurs, le logiciel R dispose déjà des fonctions prédéfinies. Si on dispose d'un échantillon tiré de deux variables aléatoires X et Y , la commande permettant de déterminer la valeur numérique du tau de Kendall de X et Y est `cor(X,Y, method="Kendall")`, `cor(X,Y)` pour le coefficient de corrélation linéaire et Pour le Rho de Spearman on a deux alternatives : `cor(X , Y , method="Spearman")` ou `cor(rank(X),rank(Y))`.

3.4.1 Présentation des données et les objectifs de l'étude

Les données sont issues de (Sarporta 2006) qui est un tableau présentant différentes marques de voiture et les caractéristiques de celles-ci à savoir : le prix, la puissance, la cylindrée, le poids et la consommation des dit véhicules. L'objectif est de déterminer de façon numérique les valeurs du Tau de Kendall, Rho de Spearman et du coefficient de corrélation linéaire entre les caractéristiques de ces véhicules enfin de voir s'il existe une dépendance entre ces différentes caractéristiques prises deux à deux.

Le tableau est le suivant

TABLE 3.1 – Jeu de données

	modele	Prix	cylindre	Puissance	poids	conso
1	Daihatsu Cuore	11600	846	32	650	5.70
2	Suzuki Swift 1.0 GLS	12490	993	39	790	5.80
3	Fiat Panda Mambo L	10450	899	29	730	6.10
4	VW Polo 1.4 60	17140	1390	44	955	6.50
5	Opel Corsa 1.2i Eco	14825	1195	33	895	6.80
6	Subaru Vivio 4WD	13730	658	32	740	6.80
7	Toyota Corolla	19490	1331	55	1010	7.10
8	Ferrari 456 GT	285000	5474	325	1690	21.30
9	Mercedes S 600	183900	5987	300	2250	18.70
10	Maserati Ghibli GT	92500	2789	209	1485	14.50
11	Opel Astra 1.6i 16V	25000	1597	74	1080	7.40
12	Peugeot 306 XS 108	22350	1761	74	1100	9.00
13	Renault Safrane 2.2. V	36600	2165	101	1500	11.70
14	Seat Ibiza 2.0 GTI	22500	1983	85	1075	9.50
15	VW Golt 2.0 GTI	31580	1984	85	1155	9.50
16	Citroen ZX Volcane	28750	1998	89	1140	8.80
17	Fiat Tempra 1.6 Liberty	22600	1580	65	1080	9.30
18	Fort Escort 1.4i PT	20300	1390	54	1110	8.60
19	Honda Civic Joker 1.4	19900	1396	66	1140	7.70
20	Volvo 850 2.5	39800	2435	106	1370	10.80
21	Ford Fiesta 1.2 Zetec	19740	1242	55	940	6.60
22	Hyundai Sonata 3000	38990	2972	107	1400	11.70
23	Lancia K 3.0 LS	50800	2497	150	1550	11.90
24	Mazda Hachtback V	36200	1998	122	1330	10.80
25	Mitsubishi Galant	31990	2485	66	1300	7.60
26	Opel Omega 2.5i V6	47700	1998	125	1670	11.30
27	Peugeot 806 2.0	36950	1997	89	1560	10.80
28	Nissan Primera 2.0	26950	1984	92	1250	9.20
29	Seat Alhambra 2.0	36400	2438	85	1635	11.60
30	Toyota Previa salon	50900	2473	97	1800	12.80
31	Volvo 960 Kombi aut	49300	2473	125	1570	12.70

3.4.2 application

Pour déterminer les valeurs numériques du Tau de Kendall, Rho de Spearman et du coefficient entre les différentes caractéristiques des véhicules Prises deux à deux, les données ont été introduite dans le logiciel R sous forme d'un data frame que nous avons appelé X ayant pour colonnes le prix, la puissance, la cylindrée, le poids et la consommation des véhicules. Pour faire cela nous avons tout d'abord crée un tableur **EXcel** comportant les données puis on a exporté ce tableur dans R.

Pour déterminer Par exemple respectivement le Tau de Kendall, le Rho de Spearman et du coefficient de corrélation linéaire entre les composantes prix et puissance les commandes R sont respectivement

$$\tau(\text{prix}, \text{puissance}) = \text{cor}(\text{x}\$\text{prix}, \text{x}\$\text{puissance}, \text{method} = \text{"Kendall"})$$

$$\rho_S(\text{prix}, \text{puissance}) = \text{cor}(\text{rank}(\text{x}\$\text{prix}), \text{rank}(\text{x}\$\text{puissance}))$$

$$r(\text{prix}, \text{puissance}) = \text{cor}(\text{x}\$\text{prix}, \text{x}\$\text{puissance})$$

Il est a noté que la commande `x$prix` permet par exemple de prendre toutes les valeurs de la colonne prix dans le data frame.

Les résultats sont les suivants :

TABLE 3.2 – tableau des valeurs du Tau de Kendall et du Rho de Spearman sur le jeu de données

modalité	Tau de Kendall	Rho de Spearman	coefficient de corrélation linéaire
prix,puissance	0,83	0,94	0,93
prix,conso	0,83	0,95,	0,89
prix,cylindre	0,82	0,95	0,89
prix,poids	0,81	0,95	0,69
puissance,conso	0,78	0,92	0,95
puissance,cylindre	0,76	0,91	0,94
puissance,poids	0,72	0,88	0,77
conso,cylindre	0,77	0,90	0,92
conso,poids	0,76	0,92	0,86
poids,cylindre	0,75	0,90	0,83

3.4. Application

les résultats proches de 1 du Tau de Kendall et du Rho de Spearman entre les composantes de notre data frame, traduisent le fait qu'il y a une dépendance positive entre ces composantes. Pour le coefficient de corrélation linéaire on ne peut rien conclure car le seuil de 0,999 permettant de conclure l'existence d'une liaison linéaire entre ces composantes n'est pas atteint. On constate aussi que toutes les valeurs du Rho de Spearman sont supérieures à celles du coefficient de corrélation linéaire, confirmant ainsi le fait que Rho de Spearman est une version améliorée du coefficient de corrélation linéaire. Le nuage de points deux à deux entre les composantes du data frame obtenue grâce à la commande **Paires(x)** du logiciel R confirme cette dépendance positive entre les composantes ; prix, cylindrée, puissance, poids et consommation des véhicules.

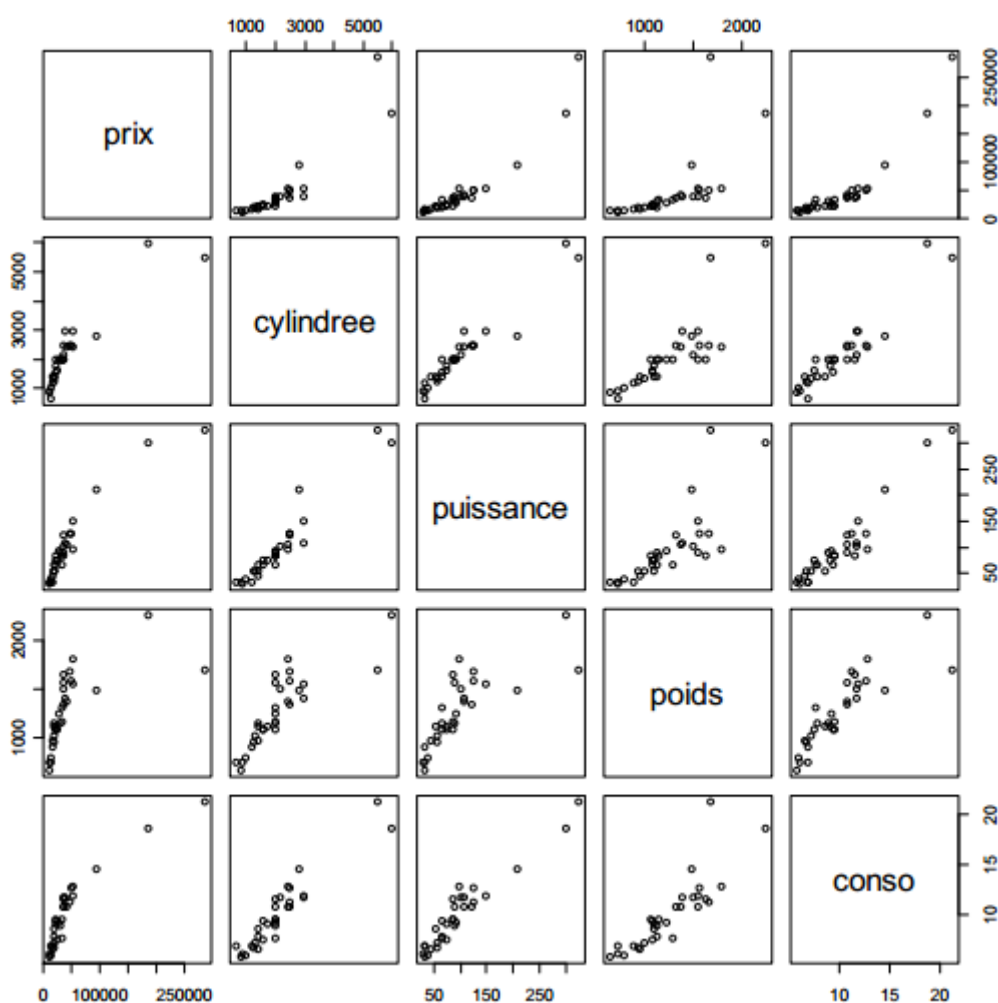


FIGURE 3.1 – nuage de points deux à deux des données

◆ IMPLICATION PÉDAGOGIQUE SUR LE SYSTÈME ÉDUCATIF DU SUJET ◆

le sujet traité dans ce mémoire entre dans le cadre des mathématiques appliquées à la finance et aborde un thème déjà présenté dans les classes de premières et terminales scientifiques.

Ce mémoire a le mérite d'avoir apporté toute la lumière sur le coefficient de corrélation linéaire enseigné dans les classes de premières et terminales scientifiques, dans l'étude des séries statistiques à double entrées. Ce coefficient de corrélation linéaire est enseigné pour l'étude de la liaison existante entre deux variables aléatoires. Cependant il est ressorti de notre travail que celui-ci a beaucoup de défauts.

Nous voulons donc ainsi recommander aux enseignants de bien spécifier aux élèves que le coefficient de corrélation linéaire ne mesure que l'intensité de la liaison linéaire entre deux variables et que celui-ci peut proposer des résultats erronés suivant la nature des variables considérées et qu'il existe d'autres fonctions mathématiques qui permettent aussi de mesurer la liaison entre deux variables aléatoires. Nous nous posons ainsi la question de savoir si l'on ne peut pas introduire au moins le Tau de Kendall ou le Rho de Spearman dans l'étude des séries statistiques à double entrées, car ceux-ci ont de bonnes propriétés par rapport au coefficient de corrélation linéaire et leur mise en œuvre est très simple.

◆ CONCLUSION GÉNÉRALE ET PERSPECTIVES ◆

Dans ce mémoire, nous avons présenté les mesures de dépendances les plus pratiques et les plus utilisées en statistique pour quantifier le degré de dépendance entre deux risques.

Pour ce faire, nous avons présenté les modèles de dépendance qui permettent de décrire la forme de la liaison entre deux risques. Ensuite, nous avons défini et donné les propriétés d'une mesure de dépendance et un accent particulier a été mis sur le coefficient de corrélation linéaire qui mais en exergue le fait que celui-ci n'est pas une "bonne mesure de dépendance". En effet, le coefficient de corrélation linéaire n'est défini que pour des couples de risques admettant des moments d'ordre deux ; il n'est pas compatible avec les bornes de Fréchet, n'est pas invariant par transformations strictement monotones et dépend des marginales des variables considérées.

Enfin, nous avons présenté deux mesures de dépendance basées sur les rangs, à savoir : le Tau de Kendall et le Rho de Spearman qui, quant à eux vérifient toutes les propriétés exigibles à une "bonne mesure de dépendance" et ne dépendent pas des marginales des variables considérées ; permettant ainsi de résoudre les problèmes posés par le coefficient de corrélation linéaire.

Cependant dans notre mémoire nous n'avons pas présenté des fonctions mathématiques permettant de modéliser la dépendance entre deux risques. Nous nous proposons ainsi dans nos prochains travaux de présenter la fonction mathématique par excellence qui permet de modéliser la dépendance entre deux risques à savoir la fonction **copule** et de voir si possible comment s'écrivent les mesures de dépendance présentées dans ce mémoire à l'aide de cette fonction copule.

◆ Bibliographie ◆

- [1] Denuit M., Charpentier A. (2004) Mathématiques de l'assurance non-vie. Tome 1 : principes fondamentaux de théorie du risque. Economica.
- [2] Gilbert Saporta. (2006) Probabilités et statistique analyse de données. France, TECHNIP, 613 Pages
- [3] Lounas Fadhila (2011) Modélisation de la dépendance par les copules et applications. Thèse Ph.D, Université de Mouloud Mammeri, Tizi-Ouzou, Faculté de science, Algérie.
- [4] Myriam Chabot concepts de dépendance et copules CaMUS4, 48-71.
- [5] Partrat CH., Besson J.L. (2004) Assurance non-vie. Modélisation, simulation. Economica.
- [6] Pierre Lafaye de micheaux, Rémy Drouilhet, Benoît Liqueur. Le logiciel R maîtriser le langage effectuer des analyses statistiques. France, Springer, 484 pages.
- [7] Planchet F. (2004) Introduction à la théorie des copules. Notes de cours, ISFA.
- [8] Saporta G. [1990] Probabilités, analyse des données et statistiques. Technip.
- [9] Ranoua Bouchouicha. Dépendance entre risques extrêmes : Application aux Hedge Funds. Working Paper GATE 2010-13. 2010.
- [10] Tom Rohmer (2014) Deux tests de détection de rupture dans la copule d'observations multivariées. Thèse Ph.D, Université de SHERBROOKE, Faculté de science, Canada.
- [11] Yadolah Dodge (2007) Statistique dictionnaire encyclopédique. Paris, Springer-Verlag, 613 pages.
- [12] YUI-KUEN TSe (2009) Nonlife Actuarial Models theory, Methods and Evaluation. Cambridge, University Press, 542 pages.