
CALCUL DU TAU DE KENDALL AVEC UNE VARIABLE
ALÉATOIRE DISCRÈTE ET UNE CONTINUE.

SOUS LA SUPERVISION DE
ÉTIENNE MARCEAU

PRÉPARÉ PAR
ALEXANDRE LEPAGE,
DIAMILATOU N'DIAYE,

LE 18 JUIN 2019



UNIVERSITÉ
LAVAL

FACULTÉ DES SCIENCES ET DE GÉNIE
ÉCOLE D'ACTUARIAT
UNIVERSITÉ LAVAL

Table des matières

1	Motivation	1
2	Calcul du tau de Kendall avec une v.a. discrète et une v.a. continue.	1
3	Calcul du tau de Kendall empiriquement.	2
3.1	Calcul numérique	3
4	Enjeux des variables aléatoires discrètes par rapport aux mesures de dépendance.	4
4.1	Convergence asymptotique	5
5	Calculs numériques	6
5.1	Exemples d'inversion du tau de Kendall	7
6	Modèle collectif du risque	13
6.1	Estimation de α_0	13
6.2	Estimation de α_1	13
7	Conclusion	15

Liste des illustrations

1	Copule de Clayton	4
2	Copule de Frank	5
3	Comportement asymptotique	5
4	Comportement asymptotique de τ_n avec 1 000 simulations impliquant chacune la réalisation de 1 000 couples (N_i, X_i) , $i = 1, \dots, 1\,000$	9
5	Identification de l'intervalle d'optimisation numérique pour le scénario 1	10
6	Identification de l'intervalle d'optimisation numérique pour le scénario 2	11
7	Identification de l'intervalle d'optimisation numérique pour le scénario 3	12
8	Identification de l'intervalle d'optimisation numérique pour le scénario 4	12

Liste des tableaux

1	Réalisations du couple (X_i, Y_i) , où $X_i \sim X \sim Y_i \sim Y \sim \text{Poisson}(10)$	3
2	Comparaison des résultats obtenus avec la fonction présentée dans le code informatique 1, la fonction <code>R cor</code> et le τ théorique qui est obtenu dans le cas continu.	3
3	Résultats de l'inversion du tau de Kendall avec une copule de Frank.	10
4	Résultats de l'inversion du tau de Kendall avec une copule de AMH.	11
5	Résultats de l'inversion du tau de Kendall avec une copule de Gumbel.	12
6	Résultats de l'inversion du tau de Kendall avec une copule de Joe.	13
7	Tableau présentant les estimateurs de α_0 . L'estimateur 1 prend le tau de Kendall entre N et X_1 , l'estimateur 2 la moyenne des taus de Kendall entre N et $X_i, i = 1, 2, 3, 4$, l'estimateur 3 la moyenne des taus de Kendall entre N et tous les X_i	13
8	Tableau présentant les estimateurs de α_0 . L'estimateur 1 prend le tau de Kendall entre N et X_1 , l'estimateur 2 la moyenne des taus de Kendall entre N et $X_i, i = 1, 2, 3, 4$, l'estimateur 3 la moyenne des taus de Kendall entre N et tous les X_i	13
9	Tableau présentant les estimateurs de α_1 , les valeurs estimées de α_0 utilisées ainsi que le temps de calcul. Le tau de Kendall empirique utilisé est le tau de Kendall entre X_1 et X_2 . Les données utilisées sont issues de simulations d'échantillons de taille 10 000.	14
10	Tableau présentant les estimateurs de α_1 , les valeurs estimées de α_0 utilisées ainsi que le temps de calcul. Le tau de Kendall empirique utilisé la moyenne de tous les taus de Kendall. Les données utilisées sont issues de simulation d'échantillons de taille 10 000. . .	14

1 Motivation

Dans la littérature, on explique comment calculer les tau de Kendall avec des variables aléatoires continues ([Genest and Favre, 2007]) et avec des variables aléatoires discrètes ([Nikoloulopoulos and Karlis, 2010]). Mais qu'en est-il si on a une variable aléatoire discrète et que l'autre est continue. Le lemme 1 définit la façon de calculer ce cas particulier.

2 Calcul du tau de Kendall avec une v.a. discrète et une v.a. continue.

Soit C , une copule quelconque, et les variables aléatoires N et X définies sur \mathbb{N} et \mathbb{R}_+ respectivement. Avec le théorème de Sklar, on a que $F_{N,X}(n, x) = C(F_N(n), F_X(x))$ et la fonction de densité bivariable conjointe est donnée par

$$f_{N,X}(n, x) = \frac{\partial}{\partial x} C(F_N(n), F_X(x)) - \frac{\partial}{\partial x} C(F_N(n-1), F_X(x)).$$

Soient les couples de variables aléatoires $\{(X_1, Y_1), \dots, (X_k, Y_k)\}$. Un couple de v.a. est dit concordant si $(X_i - X_j)(Y_i - Y_j) > 0$ et discordant si $(X_i - X_j)(Y_i - Y_j) < 0$, pour $i \neq j$. Dans [Nikoloulopoulos and Karlis, 2010], on pose qu'en cas de possibilité d'égalité entre des observations d'une même variable aléatoire, la formule générale du tau de Kendall est

$$\begin{aligned} \tau(N, X) &= \mathbb{P}(\text{Concordance}) - \mathbb{P}(\text{Discordance}) \\ &= \mathbb{P}(\text{Concordance}) - [1 - \mathbb{P}(\text{Concordance}) - \mathbb{P}(\text{Égalité})] \\ &= 2\mathbb{P}(\text{Concordance}) + \mathbb{P}(\text{Égalité}) - 1. \end{aligned} \tag{1}$$

Lemme 1. Soient les couples de variables aléatoires $\{(N_1, X_1), \dots, (N_k, X_k)\}$ définis sur $\mathbb{N}^k \times \mathbb{R}_+^k$. La formule pour calculer le tau de Kendall avec une variable aléatoire discrète et une autre qui est continue est présentée en (2).

$$\tau(N, X) = 4 \sum_{n=0}^{\infty} \int_0^{\infty} F_{N,X}(n-1, x) f_{N,X}(n, x) dx + \sum_{n=0}^{\infty} (\mathbb{P}(N=n))^2 - 1. \tag{2}$$

Démonstration. Soit le couple (N_i, X_i) et (N_j, X_j) avec $i = 1, \dots, k$, $j = 1, \dots, k$ et $i \neq j$. On a que i et j sont interchangeables. Il s'ensuit que

$$\begin{aligned} \mathbb{P}(\text{Concordance}) &= \mathbb{P}(N_i < N_j, X_i < X_j) + \mathbb{P}(N_j < N_i, X_j < X_i) \\ &= 2\mathbb{P}(N_i < N_j, X_i < X_j) \end{aligned} \tag{3}$$

De (1) et (3), on a

$$\tau(N, X) = 4\mathbb{P}(N_i < N_j, X_i < X_j) + \mathbb{P}(N_i = N_j \cup X_i = X_j) - 1, \quad i \neq j. \tag{4}$$

Or

$$\begin{aligned} \mathbb{P}(N_i < N_j, X_i < X_j) &= \sum_{n=0}^{\infty} \int_{x=0}^{\infty} \mathbb{P}((N_i < n, X_i < x) \cap (N_j = n, X_j = x)) dx \\ &\stackrel{\text{ind.}}{=} \sum_{n=0}^{\infty} \int_{x=0}^{\infty} \mathbb{P}(N_i < n, X_i < x) \mathbb{P}(N_j = n, X_j = x) dx \\ &= \sum_{n=0}^{\infty} \int_{x=0}^{\infty} \mathbb{P}(N_i \leq n-1, X_i < x) f_{N,X}(n, x) dx \\ &= \sum_{n=0}^{\infty} \int_{x=0}^{\infty} F_{N,X}(n-1, x) f_{N,X}(n, x) dx. \end{aligned} \tag{5}$$

Par la suite, puisque X est une v.a. continue,

$$\mathbb{P}(N_i = N_j \cup X_i = X_j) = \mathbb{P}(N_i = N_j).$$

On obtient alors

$$\begin{aligned} \mathbb{P}(N_i = N_j \cup X_i = X_j) &= \sum_{n=0}^{\infty} \mathbb{P}(N_i = n \cap N_j = n) \\ &\stackrel{\text{ind.}}{=} \sum_{n=0}^{\infty} \mathbb{P}(N_i = n) \mathbb{P}(N_j = n) \\ &\stackrel{\text{i.d.}}{=} \sum_{n=0}^{\infty} (\mathbb{P}(N = n))^2. \end{aligned} \quad (6)$$

En insérant (5) et (6) dans (4), on obtient

$$\tau(N, X) = 4 \sum_{n=0}^{\infty} \int_0^{\infty} F_{N,X}(n-1, x) f_{N,X}(n, x) dx + \sum_{n=0}^{\infty} (\mathbb{P}(N = n))^2 - 1.$$

□

3 Calcul du tau de Kendall empiriquement.

Pour ce qui est du calcul empirique de deux variables aléatoires continues, [Genest and Favre, 2007] propose (7).

$$\begin{aligned} \tau_n(X, Y) &= \frac{P_n - Q_n}{\binom{n}{2}} \\ &= \frac{P_n - ((\binom{n}{2}) - P_n)}{\binom{n}{2}} \\ &= \frac{2P_n - \binom{n}{2}}{\binom{n}{2}} \\ &= \frac{4P_n}{n(n-1)} - 1, \end{aligned} \quad (7)$$

où P_n et Q_n représentent les nombres de couples concordants et discordants respectivement et n est le nombre d'observations.

Dans le cas où on a au moins une variable aléatoire discrète, l'équation permettant de calculer empiriquement le tau de Kendall est présentée dans le lemme 2.

Lemme 2. *Soit la séquence de couples $\{(N_1, X_1), \dots, (N_k, X_k)\}$, où N_i est une v.a. discrète et X_i est une v.a. discrète ou continue, pour $i = 1, \dots, k$. Alors (7) devient*

$$\tau_n(N, X) = \frac{4P_n + 2E_n}{n(n-1)} - 1, \quad (8)$$

où E_n est le nombre d'observations avec au moins une égalité.

Démonstration. De façon similaire à (7), dans le cas où on a au moins une variable aléatoire discrète,

on a

$$\begin{aligned}
\tau_n(N, X) &= \frac{P_n - Q_n}{\binom{n}{2}} \\
&= \frac{P_n - (\binom{n}{2} - P_n - E_n)}{\binom{n}{2}} \\
&= \frac{2P_n + E_n - \binom{n}{2}}{\binom{n}{2}} \\
&= \frac{2P_n + E_n}{\binom{n}{2}} - 1 \\
&= \frac{4P_n + 2E_n}{n(n-1)} - 1.
\end{aligned} \tag{9}$$

□

3.1 Calcul numérique

Du point de vue computationnel, (8) peut être écrit en R comme dans le code informatique 1.

Code informatique 1.

```

tau_kendall_empirique <- function(X,Y){
  # Fonction qui calcule le tau de Kendall empirique lorsqu'au moins une des
  # deux v.a. est discète.
  n <- length(X)
  concord <- outer(1:n,1:n, function(i,j) sign((X[i] - X[j]) * (Y[i] - Y[j])))
  E_n <- (length(concord[concord == 0]) - n)
  P_n <- sum(concord[concord > 0])
  tau <- (2 * P_n + E_n) / (n * (n - 1)) - 1
  return(tau)
}

```

Il se trouve que la fonction R `cor` donne des résultats similaires, mais pas exactement les mêmes comme on peut le voir dans l'exemple 1.

Exemple 1. Soient 20 réalisations du couple (X_i, Y_i) simulées à l'aide de la copule FGM dont le paramètre de dépendance est de 0.5. On a que $X_i \sim X \sim Y_i \sim Y \sim \text{Poisson}(10)$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X	9	7	7	14	9	10	11	11	8	10	6	16	10	12	7	9	12	7	6	13
Y	17	10	6	7	9	7	8	11	4	10	12	14	10	10	9	12	4	7	10	7

Tableau 1 – Réalisations du couple (X_i, Y_i) , où $X_i \sim X \sim Y_i \sim Y \sim \text{Poisson}(10)$.

Fonction empirique	-0.0579
Fonction cor de R	-0.0636
Tau théorique (continu)	0.1111

Tableau 2 – Comparaison des résultats obtenus avec la fonction présentée dans le code informatique 1, la fonction R `cor` et le τ théorique qui est obtenu dans le cas continu.

Dans le tableau 2, on voit d'abord qu'il y a une légère différence entre le tau calculé avec le code informatique 1 et la fonction R `cor`. Cependant, la raison de cet écart est inconnue pour le moment. Ensuite, on observe qu'il y a une nette différence entre le tau de Kendall théorique (continu) et celui calculé avec le code informatique 1 ou avec la fonction R `cor`. Cela s'explique par le grand nombre d'égalités dans les données simulées, du fait que les deux variables sont discrètes et que le paramètre des lois de

Poisson est relativement faible pour le nombre de simulations. Par ailleurs, dans [Nikoloulopoulos and Karlis, 2010], il est démontré que la paramétrisation des lois discrètes a une répercussion importante sur les mesures de dépendance entre les variables aléatoires.

4 Enjeux des variables aléatoires discrètes par rapport aux mesures de dépendance.

Contrairement aux variables aléatoires continues, lorsque l'on a des variables aléatoires discrètes, le choix et la paramétrisation des lois marginales influent sur le tau de Kendall. Les graphiques suivants ont pour but d'illustrer cela. L'illustration 1 représente le tau de Kendall selon le paramètre de la copule de Clayton et le paramètre de la variable aléatoire discrète. Ainsi lorsque $N \sim Pois(2)$ le tau de Kendall n'atteint pas les bornes $[-1, 1]$ en raison de valeurs égales produites. Lorsque le paramètre λ est augmenté à 20 le tau de Kendall prends des valeurs plus élevées, les égalités posant problème sont moins fréquentes étant donnée l'augmentation du paramètre.

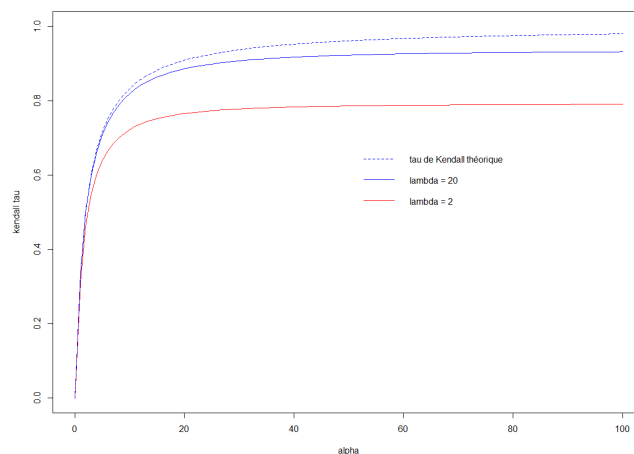


Illustration 1 – Graphique du tau de Kendall empirique selon la valeur du paramètre d'une copule de Clayton, pour un couple de variables aléatoires mixtes ou $N \sim Pois(\lambda)$, $\lambda = 2, 20$ et $X \sim Exp(1/100)$. Le tau de Kendall empirique est calculé sur des échantillons de taille 10 000.

Le même phénomène est illustré dans l'illustration 2. Une copule de Frank y est choisie afin d'observer le comportement du tau de Kendall dans le cas d'une dépendance négative. Comme précédemment plus le paramètre de la variable discrète augmente plus le tau de Kendall a une amplitude élevée. Alors, lorsque la variable aléatoire discrète possède un grand nombre de valeurs possibles son impact sur la valeur du tau de Kendall diminue.

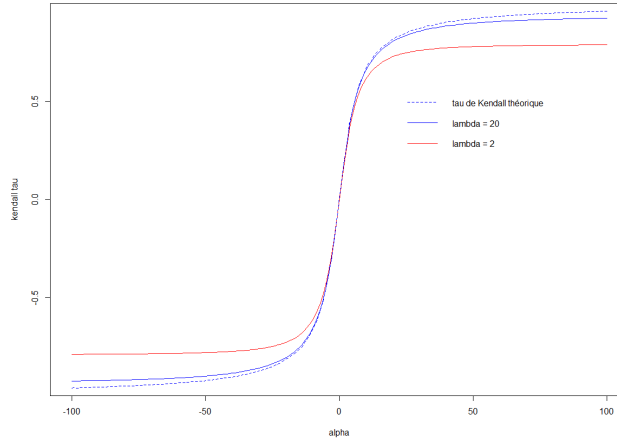


Illustration 2 – Graphique du tau de Kendall empirique selon la valeur du paramètre d’une copule de Frank, pour un couple de variables aléatoires mixtes ou $N \sim Pois(\lambda)$, $\lambda = 2, 20$ et $X \sim Exp(1/100)$. Le tau de Kendall empirique est calculé sur des échantillons de taille 10 000.

4.1 Convergence asymptotique

L’illustration 3 représente le tau de Kendall théorique, lorsque les deux lois sont continues, et le tau de Kendall empirique, lorsqu’une des deux lois est discrète, selon le paramètre d’une copule de Clayton. Le tau de Kendall est calculé par (8) avec des échantillons de taille 10 000, pour un couple de variables aléatoires mixtes ou $N \sim Pois(50)$ et $X \sim Exp(1/100)$. Ainsi asymptotiquement le tau de Kendall empirique converge vers le tau de Kendall théorique lorsque le paramètre de la variable aléatoire discrète est assez grand.

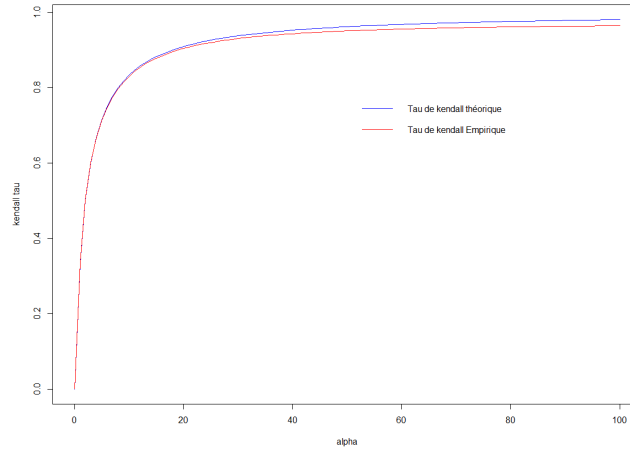


Illustration 3 – Graphique du tau de Kendall théorique et empirique selon la valeur du paramètre d’une copule de Clayton. Le tau de Kendall empirique est calculé pour un couple de variables aléatoires mixtes ou $N \sim Pois(50)$ et $X \sim Exp(1/100)$. Le tau de Kendall théorique est obtenu par la formule $\alpha/(\alpha + 2)$.

5 Calculs numériques

Lorsqu'on désire trouver le paramètre de dépendance par la méthode des moments (via le tau de Kendall ou le rho de Spearman), il faut d'abord être en mesure de calculer la valeur théorique des mesures de dépendance. À cet effet, le code informatique 2 permet de calculer le tau de Kendall théorique.

Code informatique 2.

```
library(copula)
library(rlist)

tau_kendall_theorique <- function(F_N, F_X, para, Copule, alpha, n_max, x_max){

  # Fonction qui permet de calculer le tau de kendall avec une
  # variable aléatoire discrète et une autre qui est continue.

  F_N. <- function(n) {
    # Paramétriser la fonction de répartition de N.
    do.call(F_N, list.flatten(list(q = n, para$N)))
  }
  F_X. <- function(x) {
    # Paramétriser la fonction de répartition de X.
    do.call(F_X, list.flatten(list(q = x, para$X)))
  }

  f_N <- function(n) F_N.(n) - F_N.(n-1)

  F_NX <- function(n, x) {
    # Fonction de répartition conjointe de N et X.
    f <- function(U) pCopula(U, Copule(alpha, use.indepC = "FALSE"))
    return(f(c(F_N.(n), F_X.(x))))
  }

  f_NX <- function(n, x) {
    # Fonction de densité conjointe de N et X.
    eps <- .Machine$double.eps^0.25
    f <- function(n,x) (F_NX(n, x + eps) - F_NX(n, x)) / eps
    return(f(n, x) - f(n - 1, x))
  }

  return(
    sum(sapply(0:n_max, function(n)
      4 * integrate(
        Vectorize(function(x)
          F_NX(n - 1, x) * f_NX(n, x)),
        0, x_max,
        subdivisions = 100L,
        rel.tol = .Machine$double.eps ^ 0.25
      )$value +
      f_N(n) ^ 2)) - 1
    )
  )
}
```

Les arguments pris par cette fonction sont

- **F_N** : La fonction de répartition de la variable discrète (ex : `ppois`, `pbinom`, etc.);
- **F_X** : La fonction de répartition de la variable continue (ex : `pexp`, `pgamma`, etc.);
- **para** : La liste des arguments qui servent à **F_N** et **F_X**;
(ex : pour $N \sim \text{Binom}(5, 0.4)$ et $X \sim \text{Exp}(1/100)$, on aurait
`para = list(N = list(size = 5, prob = 0.4), X = list(beta = 1/100))`)
- **Copule** : Une classe du module `Copula` (ex : `frankCopula`, `amhCopula`, etc.)¹
- **alpha** : Paramètre de dépendance inhérent à la famille de copule utilisée;
- **n_max** et **x_max** : Les valeurs maximales que peuvent prendre les variables aléatoires N et X .
(ex : pour $N \sim \text{Poisson}(1)$ et $X \sim \text{Exp}(1/100)$, on aurait
`n_max = qpois(0.99999, 1); x_max = qexp(0.99999, 1/100)`)

Maintenant que l'on peut calculer la valeur théorique du tau de Kendall pour un couple de variables aléatoires mixtes, il est possible d'inverser la fonction numériquement afin d'identifier le paramètre de dépendance d'un jeu de données à l'aide de l'estimation trouvée avec le code informatique 1. On y parvient avec le code informatique 3.

Code informatique 3.

```
inversion_tau_kendall <- function(F_N, F_X, para, Copule, bornes, tau, n_max, x_max){  
  
  # Fonction qui permet de trouver le paramètre de dépendance à l'aide du  
  # tau de Kendall empirique lorsque le modèle comprend une v.a. discrète et  
  # une v.a. continue.  
  
  alpha <- uniroot(function(alpha)  
                    tau_kendall_theorique(F_N, F_X, para,  
                    Copule, alpha, n_max, x_max) - tau,  
                    bornes)$root)  
  
  return(alpha)  
}
```

Pour cette fonction, les paramètres sont les mêmes que dans le code informatique 2 à l'exception du paramètre **alpha** qui est remplacé par **tau**. Ce dernier représente le tau empirique que l'on désire inverser pour trouver le paramètre de dépendance de la copule. De plus, l'argument **bornes** est un vecteur représentant l'intervalle des valeurs possibles du paramètre de dépendance. Afin de réduire les temps de calcul et maximiser les chances de trouver une solution à l'optimisation numérique, le mieux est de faire une analyse graphique pour identifier l'intervalle le plus petit possible dans lequel on veut appliquer la fonction. Des exemples d'application de cette procédure sont présentés dans la sous-section 5.1

5.1 Exemples d'inversion du tau de Kendall

Avec les formules proposées dans les codes informatiques 1, 2 et 3, on cherche à inverser un tau de Kendall calculé empiriquement sur une base de données afin d'identifier le paramètre qui façonne la structure de dépendance entre une variable aléatoire discrète et une qui est continue. Dans les exemples suivants, une méthodologie est proposée pour arriver à cette fin. Pour attester de la précision des codes informatiques définis dans la présente section, on utilise des données simulées dont on connaît le vrai paramètre de dépendance.

1. <https://www.rdocumentation.org/packages/copula/versions/0.999-19/topics/archmCopula-class>

Scénario 1. *Copule de Frank* :

Posons un modèle de dépendance construit sur une copule de Frank, avec comme paramètre $\alpha = 5$, et dont les deux variables aléatoires qui le composent sont $N \sim \text{Pois}(2)$ et $X \sim \text{Exp}(1/100)$. On pose les paramètre en R comme démontré dans le code informatique 4.

Code informatique 4.

```
alpha <- 5
Copule <- frankCopula
F_N <- ppois
F_X <- pexp
para <- list(N=list(lambda=2), X=list(rate=1/100))
n_max <- qpois(0.99999, para$N$lambda)
x_max <- qexp(0.99999, para$X$rate)
```

On estime un tau de Kendall empiriquement. Dans ce cas-ci, on simule 1 000 réalisations du couple (X_i, Y_i) . Puis, pour obtenir un intervalle de confiance sur l'estimateur obtenu, on répète cette simulation 1 000 fois. Comme il est illustré dans le code informatique 5. Dans une situation où on aurait un vrai jeu de données, on pourrait avoir recours à la méthode du ré-échantillonnage (*bootstrap*) afin de définir cet intervalle.

Code informatique 5.

```
tau_empirique <- numeric(nsim <- 1e+3)
set.seed(20190618)
for (i in 1:nsim) {
  UU <- rCopula(1e+3, Copule(alpha))
  X <- qpois(UU[,1], para$N$lambda)
  Y <- qexp(UU[,2], para$X$rate)
  tau_empirique[i] <- tau_kendall_empirique(X, Y)
}
mean_tau <- mean(tau_empirique)
```

Dans le contexte du scénario 1, on obtient, pour un niveau de confiance de 95%, que $\tau_n \in [0.4007240, 0.4594975]$ et la moyenne est de 0.4301108. Avec le code informatique 2, on trouve que $\tau = 0.4299595$; ce qui signifie que l'estimateur empirique est sans biais puisqu'il appartient à cet intervalle. À cet effet, dans l'illustration 4, on voit que le comportement asymptotique de l'estimateur τ_n obéit à une loi normale.

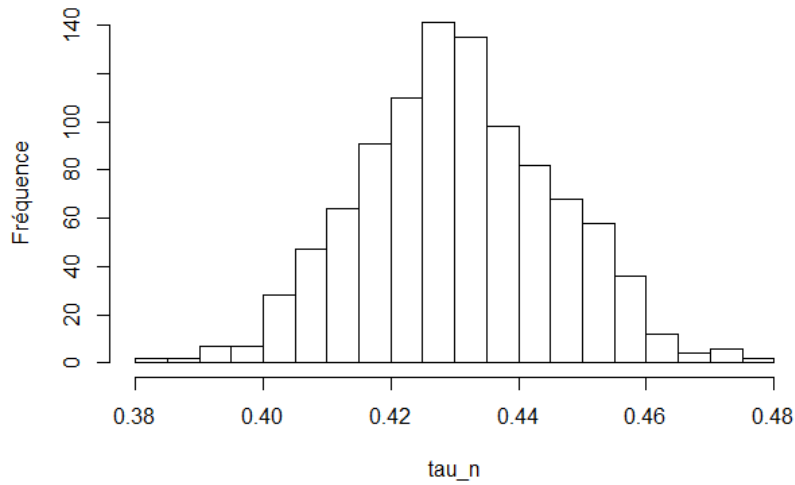


Illustration 4 – Comportement asymptotique de τ_n avec 1 000 simulations impliquant chacune la réalisation de 1 000 couples (N_i, X_i) , $i = 1, \dots, 1\,000$.

Maintenant, afin de trouver le paramètre de dépendance du jeu de données simulé, il faut identifier l'intervalle optimal sur lequel appliquer la fonction présentée dans le code informatique 3. Le code 6 permet donc de générer l'illustration 5.

Code informatique 6.

```
Tau <- sapply(c(-5:-1, 1:10), function(a)
  tau_kendall_theorique(F_N, F_X, para, Copule, a, n_max, x_max))

plot(c(-5:-1, 1:10), Tau,
     ylab="tau de kendall",
     xlab="alpha",
     type="l"
)
axis(2, tck = 1, lty = 2, col = "grey") # L'axe des ordonnées
axis(1, at=-5:10, tck=1, lty = 2, col = "grey",) # L'axe des abscisses
abline(a=tau_empirique, b=0, col="green")
```

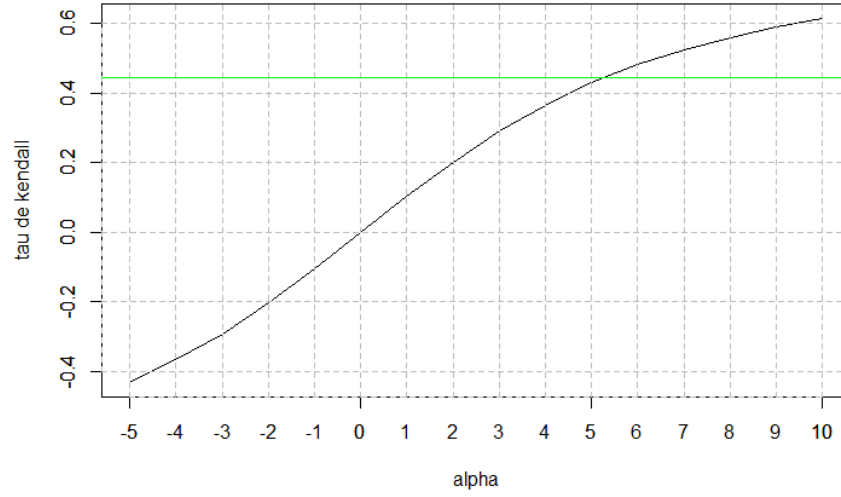


Illustration 5 – Identification de l'intervalle d'optimisation numérique pour une copule de Frank avec $\alpha = 5$.

Avec l'illustration 5, on voit qu'un intervalle optimal pourrait être $[4.5, 5.5]$. Il ne reste donc plus qu'à utiliser le code 3 pour obtenir les résultats présentés dans le tableau 3.

Vrai paramètre	5.0000
Paramètre trouvé	5.0026
Temps d'optimisation	9.7800

Tableau 3 – Résultats de l'inversion du tau de Kendall avec une copule de Frank.

Dans le tableau 3, on voit que le paramètre trouvé est très proche de la vraie valeur de α .

Scénario 2. Copule de AMH :

Posons un modèle de dépendance construit sur une copule de AMH, avec comme paramètre $\alpha = 0.5$, et dont les deux variables aléatoires demeurent inchangées par rapport au scénario 1. L'attribution des paramètres se fait de façon similaire à celle présentée dans le code informatique 4. En gardant le même encrage de simulation (*seed*), on obtient que $\tau_n \in [0.08199922, 0.16145825]$ et la moyenne est de 0.1217287, pour un τ théorique de 0.1220402. Par la suite, par analyse graphique, on voit dans l'illustration 6 que $\alpha \in [0.45, 0.55]$. Par inversion du tau empirique moyen, on obtient le tableau 4 qui présente les résultats de la simulation.

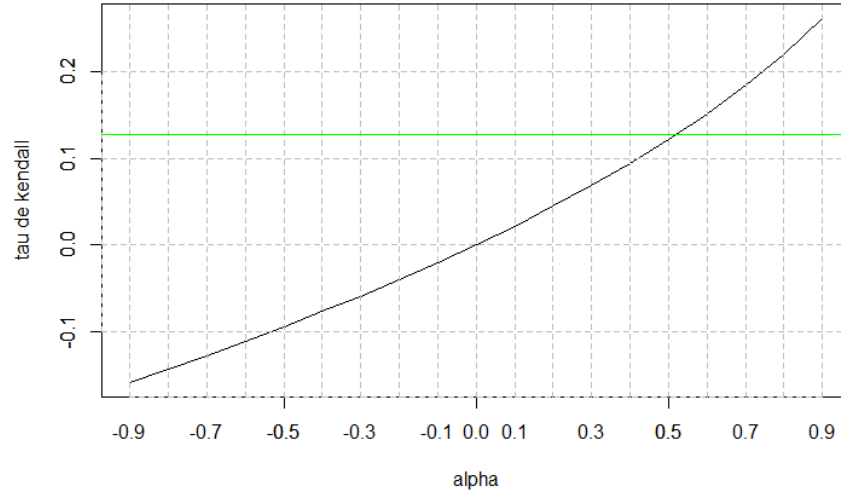


Illustration 6 – Identification de l'intervalle d'optimisation numérique pour une copule de AMH avec $\alpha = 0.5$.

Vrai paramètre	0.5000
Paramètre trouvé	0.4989
Temps d'optimisation	6.6100

Tableau 4 – Résultats de l'inversion du tau de Kendall avec une copule de AMH.

Comme pour le scénario 1, on observe un bon niveau de précision.

Scénario 3. *Copule de Gumbel* :

Posons un modèle de dépendance construit sur une copule de Gumbel, avec comme paramètre $\alpha = 5$, et dont les deux variables aléatoires demeurent inchangées par rapport aux scénarios 1 et 2. Toujours avec le même encrage de simulation, on a $\tau_n \in [0.7028951, 0.7350736]$ et une moyenne de 0.7189843 pour un τ théorique de 0.7194499. Avec l'aide de l'illustration 7, on voit que $\alpha \in [4.5, 5.5]$. On obtient alors les résultats présentés dans le tableau 5.

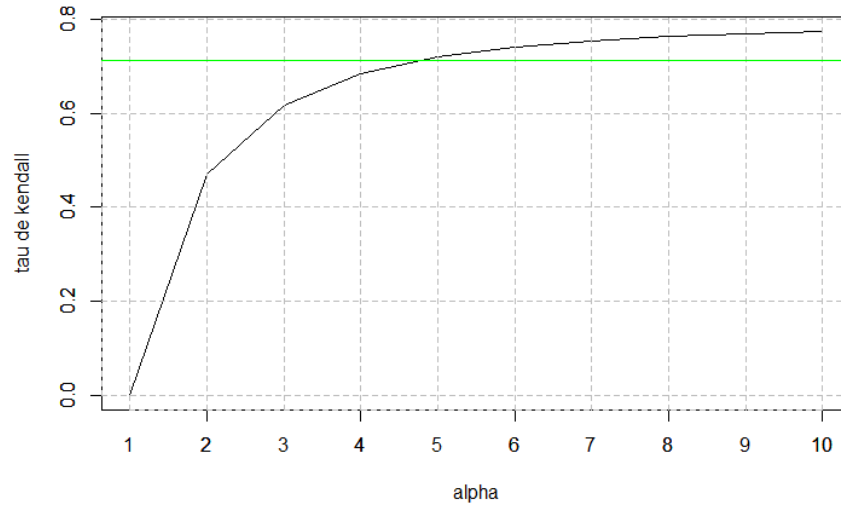


Illustration 7 – Identification de l'intervalle d'optimisation numérique pour une copule de Gumbel avec $\alpha = 5$.

Vrai paramètre	5.0000
Paramètre trouvé	4.9827
Temps d'optimisation	21.5200

Tableau 5 – Résultats de l'inversion du tau de Kendall avec une copule de Gumbel.

Scénario 4. *Copule de Joe* :

Finalement, si on pose un modèle de dépendance construit sur une copule de Joe, avec comme paramètre $\alpha = 5$, et dont les autres hypothèses restent inchangées. On a $\tau_n \in [0.6008445, 0.6502252]$ et une moyenne de 0.6255349 pour un τ théorique de 0.6260815. Avec l'aide de l'illustration 8, on voit que $\alpha \in [4.5, 6.5]$. On obtient alors les résultats présentés dans le tableau 6.

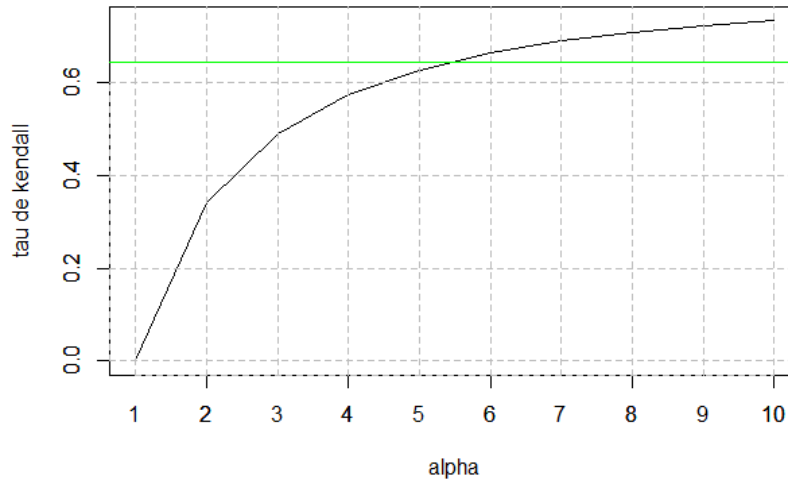


Illustration 8 – Identification de l'intervalle d'optimisation numérique pour une copule de Joe avec $\alpha = 5$.

Vrai paramètre	5.0000
Paramètre trouvé	4.9827
Temps d'optimisation	21.5200

Tableau 6 – Résultats de l'inversion du tau de Kendall avec une copule de Joe.

6 Modèle collectif du risque

6.1 Estimation de α_0

Dans cette section la méthode d'estimation précédente est appliquée au modèle collectif du risque avec $N \sim \text{Binom}(7, 0.4)$ et $X \sim \text{Exp}(1/100)$. Les exemples suivants illustrent les cas où la loi mère est logarithmique ou géométrique et la loi enfant est gamma. Les estimations ont été réalisées avec des simulations d'échantillons de taille 10 000. Trois estimateurs de α_0 sont calculés selon la méthode de calcul du tau de Kendall empirique, le premier utilise le tau de Kendall entre N et X_1 , le second la moyenne des taus de Kendall entre N et $X_i, i = 1, 2, 3, 4$ et le dernier la moyenne des taus de Kendall entre N et tous les X_i . On ne prends pas en compte les valeurs où N est nul. Pour les X_i le tau de Kendall est calculé lorsque X_i est défini.

Scénario 5. Copule hiérarchique Logarithmique-Gamma de paramètres ($\alpha_0 = 5, \alpha_1 = 4$) :

Dans ce cas la copule utilisée est la copule de Frank de paramètre α_0 . Le tableau 7 illustrent le résultat de l'estimation :

Paramètre	Estimateur 1	Estimateur 2	Estimateur 3
5.0000	5.0744	3.7114	1.9643

Tableau 7 – Tableau présentant les estimateurs de α_0 . L'estimateur 1 prend le tau de Kendall entre N et X_1 , l'estimateur 2 la moyenne des taus de Kendall entre N et $X_i, i = 1, 2, 3, 4$, l'estimateur 3 la moyenne des taus de Kendall entre N et tous les X_i .

Scénario 6. Copule hiérarchique Géométrique-Gamma de paramètres ($\alpha_0 = 0.6, \alpha_1 = 4$) :

Dans ce cas la copule utilisée est la copule AMH de paramètre α_0 . Le tableau 8 illustre le résultat de l'estimation :

Paramètre	Estimateur 1	Estimateur 2	Estimateur 3
0.6000	0.6166	0.4367	0.2963

Tableau 8 – Tableau présentant les estimateurs de α_0 . L'estimateur 1 prend le tau de Kendall entre N et X_1 , l'estimateur 2 la moyenne des taus de Kendall entre N et $X_i, i = 1, 2, 3, 4$, l'estimateur 3 la moyenne des taus de Kendall entre N et tous les X_i .

Le fait de prendre en compte les taus de Kendall entre N et tous les X_i n'améliore pas la qualité de l'estimation. La meilleure option reste de prendre le tau de Kendall entre N et X_1 .

6.2 Estimation de α_1

Le modèle collectif du risque est modélisé par une copule hiérarchique à deux paramètres. Le premier paramètre α_0 , paramètre de la loi mère modélise la dépendance entre N et X , il a été estimé dans la section précédente. Le second paramètre α_1 , paramètre de la loi enfant modélise la dépendance entre les X_i . Étant donné que les X_i sont des variables continues pour calculer le tau de Kendall empirique entre X_i et X_j on utilise la formule suivante :

$$\tau(X_i, X_j) = 4 \int_0^1 \int_0^1 C(u_1, u_2) c(u_1, u_2) - 1. \quad (10)$$

où

$$C(u_1, u_2) = \mathcal{L}_\Theta(\mathcal{L}_\Theta^{-1}(u_1) + \mathcal{L}_\Theta^{-1}(u_2))$$

avec

$$\mathcal{L}_\Theta(t) = \mathcal{L}_M(-\ln(\mathcal{L}_B(t)))$$

et

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2)$$

La relation entre les X_i est modélisée par une copule archimédienne à deux paramètres, il faut donc pour l'estimation de α_1 estimer préalablement α_0 . Les tableaux 9 et 10 présentent les résultats de l'estimation de α_1 , le paramètre α_0 utilisé pour l'estimation et le temps d'estimation comprenant l'estimation des deux paramètres de la copule archimédienne hiérarchique. Le tau de Kendall empirique entre les variables continues et la variable discrète est calculé à partir des variables N et X_1 . Pour le tableau 9 le tau de Kendall entre les variables continues est calculé par les variables X_1 et X_2 . Pour le tableau 10 le tau de Kendall empirique entre les variables continues est la moyenne des taus de Kendall entre tous les X_i . L'ajout des taus de Kendall entre toutes les variables détériore la qualité de l'estimation en raison du manque d'informations. Afin de pouvoir comparer ces deux résultats, les estimations par les deux méthodes sont effectuées à partir des mêmes échantillons.

	Paramètre α_0	Estimateur α_0	Paramètre α_1	Estimateur α_1
Géométrie-Géométrie	0.6000	0.6064	0.4000	0.4375
Géométrie-Logarithmique	0.6000	0.6037	7.0000	6.7736
Géométrie-Gamma	0.6000	0.6212	4.0000	3.8284
Logarithmique-Logarithmique	4.0000	3.9243	7.0000	5.8620
Logarithmique-Géométrie	5.0000	5.0307	0.6000	0.2811
Logarithmique-Gamma	6.0000	5.7881	4.0000	2.5337
Temps				
Géométrie-Géométrie				37.5900
Géométrie-Logarithmique				38.9700
Géométrie-Gamma				36.5300
Logarithmique-Logarithmique				66.8100
Logarithmique-Géométrie				63.2700
Logarithmique-Gamma				73.3800

Tableau 9 – Tableau présentant les estimateurs de α_1 , les valeurs estimées de α_0 utilisées ainsi que le temps de calcul. Le tau de Kendall empirique utilisé est le tau de Kendall entre X_1 et X_2 . Les données utilisées sont issues de simulations d'échantillons de taille 10 000.

	Paramètre α_0	Estimateur α_0	Paramètre α_1	Estimateur α_1
Géométrie-Géométrie	0.6000	0.6064	0.4000	0.3246
Géométrie-Logarithmique	0.6000	0.6037	7.0000	6.1878
Géométrie-Gamma	0.6000	0.6212	4.0000	2.7547
Logarithmique-Logarithmique	4.0000	3.9243	7.0000	1.0010
Logarithmique-Géométrie	5.0000	5.0307	0.6000	0.0001
Logarithmique-Gamma	6.0000	5.7881	4.0000	0.4699
Temps				
Géométrie-Géométrie				41.0400
Géométrie-Logarithmique				42.1500
Géométrie-Gamma				39.6800
Logarithmique-Logarithmique				69.7800
Logarithmique-Géométrie				66.6300
Logarithmique-Gamma				76.6200

Tableau 10 – Tableau présentant les estimateurs de α_1 , les valeurs estimées de α_0 utilisées ainsi que le temps de calcul. Le tau de Kendall empirique utilisé la moyenne de tous les taus de Kendall. Les données utilisées sont issues de simulation d'échantillons de taille 10 000.

7 Conclusion

Dans la littérature le tau de Kendall est calculé pour des variables aléatoires continues ou des variables aléatoires discrètes. Le lemme 2 propose une formule qui permet de calculer le tau de Kendall dans le cas où une des variables est discrète et l'autre est continue. Dans le cas des variables aléatoires discrètes, le choix des lois a un impact sur le tau de Kendall à cause des observations qui se répètent. Ainsi plus le nombre de valeurs prises par la variable aléatoire discrète est grand plus ce phénomène s'atténue. Le lemme 1 et le lemme 2 nous permettent de réaliser une optimisation numérique afin de déterminer le paramètre α_0 des copules archimédiennes hiérarchiques utilisées dans le modèle collectif du risque. De la même façon la formule 10 nous permet d'obtenir le paramètre α_1 de la copule archimédienne hiérarchique. Pour le modèle collectif du risque prendre un grand nombre de X_i pour le calcul du tau de Kendall empirique n'améliore pas l'estimation des paramètres α_0 et α_1 .

Références

- [Genest and Favre, 2007] Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*.
- [Nikoloulopoulos and Karlis, 2010] Nikoloulopoulos, A. K. and Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9) :1555–1568.