

JEAN-FRANÇOIS QUESSY

Processus de Kendall sériel

Mémoire

présenté

à la Faculté des études supérieures

de l'Université Laval

pour l'obtention

du grade de maître ès sciences (M.Sc.)

Département de mathématiques et de statistique

FACULTÉ DES SCIENCES ET DE GÉNIE

UNIVERSITÉ LAVAL

Résumé

Ce mémoire examine le comportement asymptotique d'une version sérielle du processus de Kendall multivarié, défini dans le cas non sériel bivarié par Genest & Rivest (1993) et subséquentement étendu au cas multidimensionnel par Barbe et coll. (1996), qui en ont établi la convergence. L'application de résultats développés par Ghoudi & Rémillard (1998) pour les processus bâtis à partir de pseudo-observations permet de caractériser la limite du processus de Kendall sériel multivarié et de montrer qu'elle coïncide avec celle du processus non sériel dans le cas bivarié. Il est également démontré que des processus de Kendall sériels de délais différents construits à partir de la même chronique sont asymptotiquement indépendants. Ces observations garantissent la convergence de statistiques de type Kolmogorov-Smirnov et Cramér-Von Mises déduites du processus. On fournit des tables des valeurs critiques de ces statistiques sous l'hypothèse nulle d'indépendance et on compare, au moyen d'une étude de Monte-Carlo, la puissance de ces critères de tests non paramétriques à celle du tau de Kendall sériel, que Ferguson et coll. (2000) ont récemment proposé pour tester l'hypothèse de bruit blanc.

Jean-François Quessy, étudiant

Christian Genest, directeur

Bruno Rémillard, co-directeur

Remerciements

Mes premiers remerciements iront à M. Christian Genest, professeur au Département de mathématique et de statistique de l'Université Laval, et à M. Bruno Rémillard, professeur au Département de mathématique et d'informatique de l'UQTR. La confiance qu'ils m'ont témoignée, leur patience ainsi que leur grande disponibilité ont assuré la réussite de ce projet.

Je tiens également à remercier mes parents, Solange et Claude, pour leur appui inconditionnel, autant financier que moral, tout au long de ces années d'études.

Une partie de ce travail a été financée par des octrois individuels et collectifs accordés à MM. Christian Genest et Bruno Rémillard par le *Conseil de recherches en sciences naturelles et en génie du Canada*, ainsi que par le *Fonds pour la formation de chercheurs et l'aide à la recherche du Gouvernement du Québec*.

Table des matières

1	Introduction	2
2	Le processus de Kendall	5
2.1	Définition du tau de Kendall	5
2.2	Convergence du processus de Kendall	6
2.2.1	Définition du processus de Kendall	6
2.2.2	Résultat principal de Barbe et coll. (1996)	7
3	Processus bâtis à partir de pseudo-observations	10
3.1	Définition du concept de pseudo-observation	10
3.2	Convergence des processus basés sur des pseudo-observations	11
4	Processus de Kendall sériel	15

4.1	Définition du processus de Kendall sériel	
	bivarié	15
4.2	Vérification des hypothèses	17
4.2.1	Vérification de l'hypothèse I	17
4.2.2	Vérification de l'hypothèse II	19
4.2.3	Vérification de l'hypothèse III	21
4.3	Énoncé des résultats principaux	24
4.4	Processus de Kendall sériel d -varié	26
4.4.1	Vérification de l'hypothèse I	28
4.4.2	Vérification de l'hypothèse II	29
4.4.3	Vérification de l'hypothèse III	31
4.4.4	Énoncé du résultat principal	32
5	Quelques statistiques déduites du processus de Kendall sériel	34
5.1	La statistique de Kolmogorov-Smirnov	35
5.2	La statistique de Cramér-von Mises	37
5.3	Le tau de Kendall	40

5.4	Étude d'une série chaotique déterministe	42
6	Tests de puissance	46
6.1	Modèles autorégressifs	46
6.2	Autres modèles de dépendance	50
7	Conclusion	58
A	Démonstration du lemme 4.1	60
B	Calculs de limites	67
B.1	Propriétés de la fonction $q(t) = t^{1/2} \log(1/t)$	67
B.2	Propriétés de $K_d(t)$ et de $k_d(t)$	69
C	Calculs de covariances pour le cas bivarié	72
D	Programmes en langage C utilisés pour les simulations	88
	Bibliographie	89

Liste des figures

<i>Dépendance sérielle et corrélogramme pour la série déterministe définie en page 1</i>	2
--	---

Liste des tableaux

1. <i>Quantiles de la loi de la statistique de Kolmogorov-Smirnov basée sur un échantillon aléatoire de taille n, tels qu'estimés à partir de 10 000 répétitions.</i>	36
2. <i>Quantiles de la loi de la statistique de Kolmogorov-Smirnov corrigée basée sur un échantillon aléatoire de taille n, tels qu'estimés à partir de 10 000 répétitions</i>	38
3. <i>Quantiles de la loi de la statistique de Cramér-von Mises basée sur un échantillon aléatoire de taille n, tels qu'estimés à partir de 10 000 répétitions</i>	39
4. <i>Quantiles de la loi de la statistique de Kendall sérielle basée sur un échantillon aléatoire de taille n, tels qu'estimés à partir de 10 000 répétitions</i>	41
5. <i>Pourcentage de rejet de l'hypothèse nulle pour différents délais de la série chaotique de longueur 20, tel qu'estimé à partir de 10 000 répétitions</i>	43
6. <i>Pourcentage de rejet de l'hypothèse nulle pour différents délais de la série chaotique de longueur 50, tel qu'estimé à partir de 10 000 répétitions</i>	44
7. <i>Pourcentage de rejet de l'hypothèse nulle pour différents délais de la série chaotique de longueur 100, tel qu'estimé à partir de 10 000 répétitions</i>	44

8. <i>Pourcentage de rejet au seuil de 5% pour une chronique de longueur 20, tel qu'estimé à partir de 10 000 répétitions</i>	48
9. <i>Pourcentage de rejet au seuil de 5% pour une chronique de longueur 50, tel qu'estimé à partir de 10 000 répétitions</i>	49
10. <i>Pourcentage de rejet au seuil de 5% pour une chronique de longueur 100, tel qu'estimé à partir de 10 000 répétitions</i>	50
11. <i>Pourcentage de rejet au seuil de 5% pour une chronique de longueur 20, tel qu'estimé à partir de 10 000 répétitions</i>	55
12. <i>Pourcentage de rejet au seuil de 5% pour une chronique de longueur 50, tel qu'estimé à partir de 10 000 répétitions</i>	56
13. <i>Pourcentage de rejet au seuil de 5% pour une chronique de longueur 100, tel qu'estimé à partir de 10 000 répétitions</i>	57

Chapitre 1

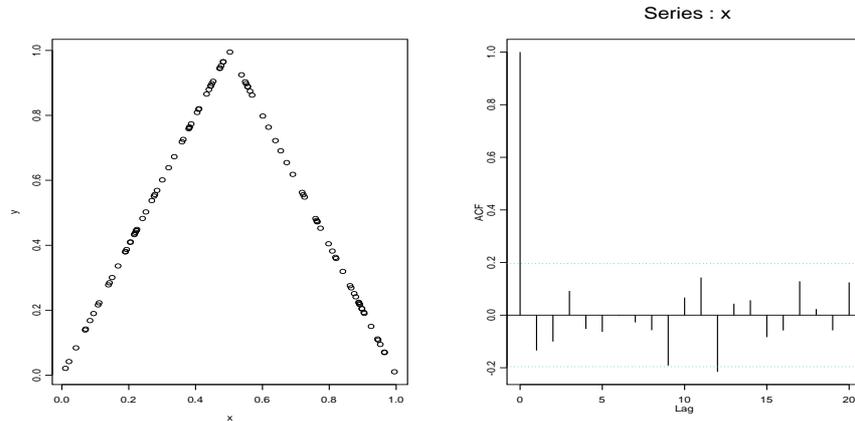
Introduction

Dans l'étude classique de la dépendance sérielle, il existe un grand nombre de mesures paramétriques, dont le corrélogramme. Ces techniques, abondamment utilisées, sont souvent inefficaces dans le cas non gaussien. À titre d'illustration, considérons la série déterministe définie par $x_{k+1} = f(x_k)$, où x_0 est choisi aléatoirement dans l'intervalle $[0, 1]$ et

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1/2 \\ 2(1-x), & 1/2 < x \leq 1. \end{cases}$$

Le graphique ci-contre (figure 1.1a) de x_{k+1} en fonction de x_k montre clairement la forte dépendance entre les termes consécutifs de cette suite. Cependant, le corrélogramme (figure 1.1b) identifierait cette série comme étant un simple bruit blanc.

Figure 1.1: Dépendance sérielle et corrélogramme pour la série déterministe définie en page 1.



En l'absence de la normalité, les statistiques non paramétriques, dont celles basées sur les rangs, peuvent constituer une solution de rechange intéressante. Par exemple, le rho de Spearman, qui est en fait une extension naturelle du coefficient de corrélation classique de Pearson, est souvent utilisé.

De nombreux auteurs, et plus particulièrement Marc Hallin et ses collaborateurs, ont étudié diverses statistiques linéaires de rangs, utilisées pour l'étude de la dépendance sérielle, c'est-à-dire de la dépendance entre les termes d'une série univariée. Un exemple récent est fourni par l'article de Ferguson et coll. (2000), qui examine les propriétés du tau de Kendall dans ce contexte. Par ailleurs, Genest & Rivest (1993) ont montré l'intérêt à des fins inférentielles d'un processus basé sur une décomposition du tau de Kendall. Comme ils l'ont montré, ce "processus de Kendall" permet d'identifier la relation de dépendance entre deux variables aléatoires. Barbe et coll. (1996)

ont ensuite prouvé la convergence de ce processus et de sa généralisation d -dimensionnelle.

Ce mémoire est consacré à l'étude de la convergence de ce même processus de Kendall dans le contexte de la dépendance sérielle.

Au chapitre 2, nous donnerons d'abord une définition du processus de Kendall et nous rappellerons le résultat principal de Barbe et coll. (1996). Nous énoncerons ensuite, au chapitre 3, un résultat obtenu par Ghoudi & Rémillard (1998) sur la convergence des processus construits à partir de pseudo-observations. Dans le chapitre 4, nous définirons le processus de Kendall sériel et démontrerons sa convergence vers un processus gaussien continu dont nous calculerons la fonction de covariance. Au chapitre 5, nous proposerons quelques statistiques déduites du processus sériel et montrerons, à l'aide de simulations, présentées au chapitre 6, qu'elles fournissent souvent de meilleurs tests d'indépendance que le tau de Kendall sériel. Finalement, après une brève conclusion, nous énumérerons au chapitre 7 quelques avenues possibles de recherches ultérieures.

Chapitre 2

Le processus de Kendall

2.1 Définition du tau de Kendall

Le tau de Kendall, noté τ , est une mesure d'association non paramétrique entre deux ou plusieurs variables. Selon Joe (1990), qui la définit dans le cas général,

$$\tau(X) = \frac{2^d}{2^{d-1} - 1} E\{H(X)\} - 1,$$

où X est un vecteur aléatoire d -dimensionnel et H est la fonction de répartition de X . Dans le cas particulier où $d = 2$, on a

$$\tau(X) = 4P\left(X_1^{(1)} \leq X_2^{(1)}, X_1^{(2)} \leq X_2^{(2)}\right) - 1,$$

où $X_i = (X_i^{(1)}, X_i^{(2)})$, $i = 1, 2$, sont des copies indépendantes de X . On note que, de façon générale, $\tau \in [-1, 1]$; de plus, $\tau = 0$ si les deux variables sont indépendantes, mais le contraire n'est pas vrai en général. La version

échantillonnale du tau de Kendall est définie par

$$\tau_n = \frac{4N}{n(n-1)} - 1,$$

où N représente le nombre de concordances dans un échantillon de taille n , c'est-à-dire le nombre de paires d'observations $X_i = (X_i^{(1)}, \dots, X_i^{(d)})$ et $X_j = (X_j^{(1)}, \dots, X_j^{(d)})$ qui satisfont soit $(X_i^{(k)} < X_j^{(k)}, \forall k)$ ou $(X_i^{(k)} > X_j^{(k)}, \forall k)$.

2.2 Convergence du processus de Kendall

2.2.1 Définition du processus de Kendall

Soit $X = (X^{(1)}, \dots, X^{(d)})$, un vecteur de variables aléatoires d -dimensionnel ayant $H(x) = P(X \leq x)$ pour fonction de répartition et H_1, \dots, H_d pour marges. Supposons que l'on dispose d'un échantillon X_1, \dots, X_n de cette loi. On définit

$$e_{i,n} = \frac{1}{n} \#\{1 \leq j \leq n : X_j \leq X_i\},$$

où $X_j \leq X_i$ signifie que $X_j^{(\ell)} \leq X_i^{(\ell)}$, $\ell = 1, \dots, d$ et $\#$ représente la cardinalité d'un ensemble. En définissant K_n comme la fonction de répartition empirique des $e_{i,n}$ et $K(t)$ comme la fonction de répartition de $\varepsilon = H(X)$, Barbe et coll. (1996) ont étudié le comportement asymptotique du processus

$$\mathbb{K}_n(t) = \sqrt{n}\{K_n(t) - K(t)\}$$

appelé "processus de Kendall." Cette appellation est justifiée du fait que τ et τ_n sont des transformations affines de $K(t)$ et $K_n(t)$ respectivement. De

façon explicite,

$$E(\varepsilon) = \int_0^1 t dK(t) = \int_0^1 \{1 - K(t)\} dt = \frac{2^{d-1} - 1}{2^d} (\tau + 1)$$

et

$$\frac{1}{n} \sum_{i=1}^n e_{i,n} = \int_0^1 t dK_n(t) = \int_0^1 \{1 - K_n(t)\} dt = \frac{2^{d-1} - 1}{2^d} (\tau_n + 1).$$

2.2.2 Résultat principal de Barbe et coll. (1996)

Hypothèse I: La fonction de répartition H est continue et la loi $K(t)$ de $\varepsilon = H(X)$ admet une densité $k(t)$ qui est également continue sur l'intervalle $(0, 1]$ en plus de vérifier la relation $k(t) = o\{t^{-1/2} \log^{-1/2-\epsilon}(1/t)\}$ quand $t \rightarrow \infty$, pour un certain $\epsilon > 0$.

Hypothèse II: Il existe une version de la loi conditionnelle du vecteur $\{H_1(X^{(1)}), \dots, H_d(X^{(d)})\}$ étant donné que $H(X) = t$ et une famille dénombrable \mathcal{P} de partitions \mathcal{C} de $[0, 1]^d$ à l'intérieur d'un nombre fini d'ensembles de Borel satisfaisant

$$\inf_{\mathcal{C} \in \mathcal{P}} \max_{C \in \mathcal{C}} \text{diam}(C) = 0,$$

et tel que l'application définie par

$$t \longmapsto \mu_t(C) = k(t) P[\{H_1(X^{(1)}), \dots, H_d(X^{(d)})\} \in C \mid H(X) = t]$$

est continue sur l'intervalle $(0, 1]$ et $\mu_1(C) = k(1) \mathbb{I}\{(1, \dots, 1) \in C\}$.

Avant d'énoncer le théorème principal, définissons, pour $s, t \in [0, 1]$,

$$Q(s, t) = P\{H(X_1) \leq s, X_1 \leq X_2 \mid H(X_2) = t\} - tK(s)$$

et

$$R(s, t) = P\{X_1 \leq (X_2 \wedge X_3) | H(X_2) = s, H(X_3) = t\} - st,$$

où $X_2 \wedge X_3$ dénote le minimum (composante par composante) entre les vecteurs X_2 et X_3 .

Théorème 2.1 (Barbe et coll. 1996) *Sous les hypothèses I et II ci-haut mentionnées, le processus empirique \mathbb{K}_n converge en loi vers un processus gaussien \mathbb{K} centré et ayant comme fonction de covariance*

$$\Gamma(s, t) = K(s \wedge t) - K(s)K(t) + k(s)k(t)R(s, t) - k(t)Q(s, t) - k(s)Q(t, s).$$

Exemple 2.1 *Considérons le cas particulier d'un vecteur bivarié $X = (X^{(1)}, X^{(2)})$ dont les composantes sont indépendantes. On a alors*

$$H(X) = H_1(X^{(1)})H_2(X^{(2)}) = UV,$$

où U et V sont des variables uniformes sur $(0, 1)$ si on suppose que H est continue (et donc que ses marges le sont également). Dans ces conditions, la fonction $K(t)$ se calcule explicitement. En effet,

$$\begin{aligned} K(t) &= P\{H(X) \leq t\} = P(UV \leq t) \\ &= \int_0^1 P\left(U \leq \frac{t}{V} \mid V = w\right) dw \\ &= \int_0^1 P\left(U \leq \frac{t}{w}\right) dw = \int_0^t dw + \int_t^1 \frac{t}{w} dw \\ &= t + t(\log 1 - \log t) = t - t \log t. \end{aligned}$$

De là, on déduit que $k(t) = \log(1/t)$. Dans ce cas particulier, on peut montrer que

$$Q(s, t) = (s \wedge t) \left\{ 1 + \log \left(\frac{t}{s \wedge t} \right) \right\} - tK(s)$$

et

$$R(s, t) = (s \wedge t) - st$$

$$+ \frac{2 \left\{ st - (s \wedge t) - (s \wedge t) \log(s \vee t) \right\} - (s \wedge t) \{ \log(s \vee t) \}^2}{\log(s) \log(t)}.$$

Aussi la fonction de covariance se réduit-elle alors à

$$\Gamma(s, t) = st - (s \wedge t) \{ 1 + \log(s \vee t) \},$$

où $s \vee t \equiv \max(s, t)$.

Chapitre 3

Processus bâtis à partir de pseudo-observations

3.1 Définition du concept de pseudo-observation

Le processus $K_n(t)$ défini au chapitre précédent est un exemple de processus construit à partir de pseudo-observations. Pour définir ce concept, dû à Ghoudi & Rémillard (1998), supposons que l'on veuille estimer la fonction de répartition d'une variable aléatoire $\theta(Q, X)$ qui n'est pas directement observable et qui dépend à la fois d'un aléa X et de sa loi Q . Si cette dernière est inconnue et estimée par Q_n , les variables aléatoires $\theta(Q_n, X_i)$ sont alors appelées des pseudo-observations. Dans le cas du processus de Kendall, $Q \equiv$

H et $Q_n \equiv H_n$. On voit donc que $e_{i,n} = H_n(X_i)$ sont des pseudo-observations.

Les résidus $\theta(X_i, Y_i) = Y_i - \hat{\beta}'X_i$ du modèle de régression linéaire constituent un autre exemple de pseudo-observations servant, entre autres, à estimer les termes d'erreur non observables.

3.2 Convergence des processus basés sur des pseudo-observations

Dans cette section, le théorème principal de Ghoudi & Rémillard (1998) concernant la convergence des processus construits à partir de pseudo-observations sera énoncé, de même que les hypothèses qui conduisent à ce résultat.

Considérons une variable aléatoire observable X prenant ses valeurs dans un ensemble quelconque \mathcal{X} et soit (X_i) , une suite d'observations stationnaire et ergodique de la variable X . Pour une fonction H pouvant dépendre de la loi de X , supposons que les variables aléatoires $\varepsilon_i = H(X_i)$ prennent leurs valeurs dans un intervalle T de \mathbb{R} . Étant donnée une estimation H_n de H , on définit les pseudo-observations $e_{i,n}$ de la manière suivante :

$$e_{i,n} = H_n(X_i), \quad 1 \leq i \leq n.$$

La fonction de répartition empirique de ces pseudo-observations est alors donnée par

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{e_{i,n} \leq t\}, \quad t \in \mathbb{R}.$$

Soit maintenant K , la fonction de répartition de la variable aléatoire $\varepsilon = H(X)$ à valeurs dans T . Les hypothèses qui assurent la convergence faible du processus $K_n(t) = \sqrt{n}\{K_n(t) - K(t)\}$ sont énoncées dans ce qui suit.

Supposons d'abord que \mathcal{X} est un espace métrique complet et séparable; dans la suite, on prendra $\mathcal{X} = \mathbb{R}^d$. Soit aussi $r : \mathcal{X} \rightarrow \mathbb{R}$, une fonction continue et positive, c'est-à-dire telle que $\inf_{x \in \mathcal{X}} r(x) > 0$. De plus, soit \mathcal{C}_r un sous-ensemble fermé de l'espace de Banach de toutes les fonctions continues $f : \mathcal{X} \rightarrow \mathbb{R}$ pour lesquelles $\|f\|_r = \sup_{x \in \mathcal{X}} |f(x)/r(x)|$ est fini.

Hypothèse I: La loi de $\varepsilon = H(X)$ admet une densité k sur T qui est bornée sur tout sous-ensemble compact de T . De plus, il existe une version de la loi conditionnelle de X étant donné $H(X) = t$, dénotée P_t , telle que pour tout $f \in \mathcal{C}_r$, pour tout ϕ continue sur \mathbb{R} et pour tout ψ continue et bornée sur \mathbb{R} , l'application

$$t \mapsto \mu\{t, (\phi \circ g)(\psi \circ r)\} = k(t)E[\{\phi \circ g(X)\}\{\psi \circ r(X)\} | H(X) = t]$$

est continue sur T , où $g = f/r$. En outre,

$$\lim_{M \rightarrow \infty} \int_M^\infty \sup_{s \in C} P_s\{r(X) > u\} du = 0$$

pour tout sous-ensemble compact C de T .

Hypothèse II: Il existe une version continue \tilde{H}_n de H_n telle que

$$\sqrt{n} \sup_{x \in \mathcal{X}} |\tilde{H}_n(x) - H_n(x)|/r(x) \xrightarrow{P} 0.$$

De plus, pour tout $f \in \mathcal{C}_r$ et tout ψ continue à valeurs dans $[0, 1]$, le processus

$$\alpha_{n,\psi \circ g}(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi\{g(X_i)\} \mathbb{I}\{\varepsilon_i \leq t + sr(X_i)\} \\ - E[\psi\{g(X_i)\} \mathbb{I}\{\varepsilon \leq t + sr(X)\}]],$$

avec $g = f/r$ est tel que pour tout sous-intervalle fermé de T et pour un certain $s \in \mathbb{R}$,

$$\sup_{t \in C} |\alpha_{n,\psi \circ g}(s/\sqrt{n}, t) - \alpha_{n,\psi \circ g}(0, t)| \xrightarrow{P} 0.$$

Enfin, si $\alpha_n(t) = \alpha_{n,1}(0, t)$ et $\mathbb{H}_n = \sqrt{n}(\tilde{H}_n - H)$, alors (α_n, \mathbb{H}_n) converge en loi sur l'espace $\mathcal{C}(\mathbb{R}) \times \mathcal{C}_r$ vers un processus (α, \mathbb{H}) . En d'autres mots, $E\{\Lambda(\alpha_n, \mathbb{H}_n)\}$ converge vers $E\{\Lambda(\alpha, \mathbb{H})\}$ pour toute fonction $\Lambda : \mathcal{C}(\mathbb{R}) \times \mathcal{C}_r \rightarrow \mathbb{R}$.

Avant d'énoncer la prochaine hypothèse, introduisons \mathcal{Q} , l'ensemble des fonctions q positives et croissantes définies dans un voisinage positif de zéro telles que $q(t)/t$ est décroissante et $q(2t)/q(t)$ est bornée supérieurement.

Hypothèse III: Si $t_* = \inf T$ est fini et n'appartient pas à T , alors il existe $q_* \in \mathcal{Q}$ et une suite (t_n) de nombres positifs qui converge vers zéro telle que $\lim_{n \rightarrow \infty} \sqrt{n}K(t_* + t_n) = 0$, $\lim_{n \rightarrow \infty} q_*(t_n)/(t_n \sqrt{n}) = 0$, $\lim_{t \downarrow 0} k(t_* + t)q_*(t) = 0$, et telle que la suite

$$\left\{ \sup_{x: H(x)-t_* > t_n} |\mathbb{H}_n(x)|/q_*\{H(x) - t_*\} \right\}_{n \geq 1}$$

est tendue.

Si $t^* = \sup T$ est fini et n'appartient pas à T , alors il existe $q^* \in \mathcal{Q}$ et une suite (s_n) de nombres positifs qui converge vers zéro telles que $\lim_{n \rightarrow \infty} \sqrt{n}K(t^* + s_n) = 0$, $\lim_{n \rightarrow \infty} q^*(s_n)/s_n\sqrt{n} = 0$, $\lim_{t \downarrow 0} k(t^* + t)q^*(t) = 0$, et telle que la suite

$$\left\{ \sup_{x: t_* - H(x) > s_n} |\mathbb{H}_n(x)|/q^*\{t_* - H(x)\} \right\}_{n \geq 1}$$

est tendue.

Théorème 3.1 (Ghoudi & Rémillard 1998) *Sous les hypothèses I et II ci-dessus, le processus empirique \mathbb{K}_n converge sur l'espace $D(T)$ vers un processus continu \mathbb{K} ayant la représentation*

$$\mathbb{K}(t) = \alpha(t) - \mu(t, \mathbb{H}).$$

Autrement dit, $E\{\phi(\mathbb{K}_n)\}$ converge vers $E\{\phi(\mathbb{K})\}$ pour toute fonction continue $\phi : D(T) \rightarrow \mathbb{R}$. Si, de plus, l'hypothèse III est vérifiée, alors \mathbb{K}_n converge sur l'espace $D(\mathbb{R})$ vers un processus continu ayant la représentation précédente sur T , et s'annulant en dehors de T .

Dans leur article, Ghoudi & Rémillard (1998) utilisent ce résultat pour redémontrer la convergence du processus de Kendall sous des conditions plus faibles que celles employées par Barbe et coll. (1996). De la même manière, ils établissent, entre autres, la convergence des processus de Spearman et de Mann-Whitney, ainsi que celle de processus bâtis à partir des résidus de modèles de régression linéaire.

Chapitre 4

Processus de Kendall sériel

4.1 Définition du processus de Kendall sériel bivarié

Soit Y_1, Y_2, \dots une suite d'observations stationnaire de loi F inconnue, que l'on suppose continue. Supposons que l'on veuille caractériser la dépendance sérielle de délai $k \geq 1$ entre les termes de cette suite à partir d'un échantillon de taille $n + k$. Pour ce faire, définissons des paires de variables aléatoires de la manière suivante :

$$X_i = (Y_i, Y_{i+k}), \quad i = 1, \dots, n.$$

Soit H la fonction de répartition commune de cette suite d'observations stationnaire et K celle des $\varepsilon_i = H(X_i)$, $1 \leq i \leq n$. Sous l'hypothèse d'indépendance, on a $\varepsilon_i = H(X_i) = F(Y_i)F(Y_{i+k}) = U_i U_{i+k}$, où U_1, \dots, U_{n+k}

sont indépendantes et uniformément distribuées sur $(0, 1)$. Dans ce qui suit, on peut donc supposer sans perte de généralité que F est uniforme sur l'intervalle $(0, 1)$.

Considérons maintenant les pseudo-observations $e_{i,n} = H_n(X_i)$, où

$$H_n(X_i) = \frac{1}{n} \# \{1 \leq j \leq n : X_j \leq X_i\}$$

et $X_j \leq X_i$ signifie que $Y_j \leq Y_i$ et $Y_{j+k} \leq Y_{i+k}$. Soit aussi

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{e_{i,n} \leq t\}$$

la version empirique de K .

On s'intéresse ici à la convergence, sous l'hypothèse d'indépendance mutuelle des Y_i , du "processus de Kendall sériel" défini par

$$\mathbb{K}_n(t) = \sqrt{n}\{K_n(t) - K(t)\}.$$

Comme ce processus est construit à partir de pseudo-observations, nous allons faire appel au théorème 2.1 de Ghoudi & Rémillard (1998), dont les hypothèses sont vérifiées à la Section 4.2. L'énoncé du résultat principal apparaît à la Section 4.3. Une généralisation d -variée du processus est ensuite considérée à la Section 4.4. Pour faciliter la lecture, nous adoptons ici les mêmes notations que Ghoudi & Rémillard (1998). Dans la suite, U et V représentent des variables aléatoires indépendantes de loi $\mathcal{U}(0, 1)$, de sorte que $X = (U, V)$ et $\varepsilon = H(X) = UV$ aient respectivement la même loi que les X_i et les ε_i sous l'hypothèse d'indépendance.

4.2 Vérification des hypothèses

Notons d'abord que sous l'hypothèse d'indépendance, la fonction $K(t)$ est la même que celle calculée dans l'exemple 2.1, à savoir

$$K(t) = t - t \log t.$$

Soit maintenant $T = (0, 1]$, $\mathcal{X} = [0, 1]^2$ et \mathcal{C}_f , l'ensemble des fonctions continues (et donc bornées) de $[0, 1]^2 \rightarrow \mathbb{R}$.

4.2.1 Vérification de l'hypothèse I

Il est clair que la densité de $\varepsilon = H(X)$, c'est-à-dire $k(t) = \log(1/t)$, est bornée sur tout compact de l'intervalle $T = (0, 1]$, puisque tout compact est inclus dans un intervalle de la forme $[a, b]$, où $0 < a < b \leq 1$.

Soit maintenant $\mu(t, f) = k(t)E\{f(X) | H(X) = t\}$, où $f \in \mathcal{C}_f$, $t \in T$. Il faut montrer que cette fonctionnelle est continue pour tout $t \in T$. Pour ce faire, posons $\Xi_1 = U$ et $\Xi_2 = UV$. On voit aisément que le jacobien de cette transformation est $1/\xi_1$. Il s'ensuit que

$$g_{\xi_1, \xi_2}(\xi_1, \xi_2) = 1/\xi_1, \quad 0 < \xi_2 < \xi_1 < 1$$

et que

$$g_{\xi_1 | \xi_2=t}(\xi_1 | \xi_2 = t) = -\frac{1}{\xi_1 \log(t)}, \quad t < \xi_1 < 1.$$

En substituant dans la définition de $\mu(t, f)$, il vient

$$\mu(t, f) = k(t)E\{f(X) | H(X) = t\}$$

$$\begin{aligned}
&= k(t)E\{f(U, V) \mid UV = t\} \\
&= k(t)E\{f(\Xi_1, t/\Xi_1) \mid \Xi_2 = t\} \\
&= k(t) \int_0^1 f(\xi_1, t/\xi_1) g_{\xi_1|\xi_2=t}(\xi_1 \mid \xi_2 = t) d\xi_1 \\
&= -k(t) \int_t^1 \frac{f(\xi_1, t/\xi_1)}{\xi_1 \log(t)} d\xi_1 \\
&= \int_t^1 \frac{1}{\xi_1} f(\xi_1, t/\xi_1) d\xi_1.
\end{aligned}$$

Il reste à montrer que $\mu(t, f)$ est continue, c'est-à-dire que

$$\lim_{n \rightarrow \infty} \mu(t_n, f) = \mu(t, f)$$

pour toute suite (t_n) dans $T = (0, 1]$ qui converge vers $t \in (0, 1]$. Pour ce faire, considérons la suite de fonctions $g_n(s) = s^{-1}f(s, t_n/s)1_{(t_n, 1]}(s)$, laquelle converge presque sûrement vers $g(s) = s^{-1}f(s, t/s)1_{(t, 1]}(s)$ puisque f est continue.

Comme f est bornée sur $[0, 1]^2$, les g_n sont uniformément bornées. On peut donc invoquer le théorème de la convergence dominée de Lebesgue pour conclure que

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mu(t_n, f) &= \lim_{n \rightarrow \infty} \int_0^1 g_n(s) ds \\
&= \int_0^1 g(s) ds \\
&= \int_t^1 s^{-1} f(s, t/s) ds \\
&= \mu(t, f).
\end{aligned}$$

Soit maintenant $r : [0, 1]^2 \rightarrow \mathbb{R}$, une fonction identiquement égale à un.

Il est évident que

$$\lim_{M \rightarrow \infty} \int_M^\infty \sup_{s \in C} P_s \{r(X) > u\} du = \lim_{M \rightarrow \infty} \int_M^\infty \sup_{s \in C} P_s(1 > u) du = 0$$

car $P_s(1 > M) = 0$ pour tout $M > 1$, et cela est vrai pour tout compact $C \subseteq T = (0, 1]$, ce qui complète la vérification de l'hypothèse I.

4.2.2 Vérification de l'hypothèse II

Soit \tilde{H}_n la fonction empirique obtenue en interpolant linéairement la fonction de répartition H_n . Il est clair que

$$|\tilde{H}_n(x) - H_n(x)| \leq \frac{1}{n}, \quad \forall x \in \mathcal{X},$$

ce qui implique que

$$\sqrt{n} \sup_{x \in \mathcal{X}} |\tilde{H}_n(x) - H_n(x)| \leq \frac{1}{\sqrt{n}}$$

et par le fait même, la première partie de l'hypothèse II est vérifiée.

Soit maintenant $\psi : [0, 1]^2 \rightarrow [0, 1]$ une fonction continue et

$$\alpha_{n, \psi}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi(X_i) \mathbb{I}\{\varepsilon_i \leq t\} - E[\psi(X) \mathbb{I}\{\varepsilon \leq t\}]], \quad t \in T.$$

Puisque $r \equiv 1$, la deuxième partie de l'hypothèse II sera vérifiée si pour tout compact $C \subset T$,

$$\sup_{t \in C} |\alpha_{n, \psi}(t + s/\sqrt{n}) - \alpha_{n, \psi}(t)| \xrightarrow{P} 0.$$

Nous allons montrer, plus généralement, que pour tout $\eta > 0$, il existe $0 < \delta < 1/2$ et n_0 tels que pour tout $n \geq n_0$, on ait

$$P\{w(\alpha_{n, \psi}, \delta) > \eta\} < \eta,$$

où

$$w(\alpha_n, \psi, \delta) = \sup_{0 \leq s, t \leq 1, |s-t| \leq \delta} |\alpha_{n, \psi}(s) - \alpha_{n, \psi}(t)|.$$

Pour ce faire, nous nous appuyerons sur le résultat suivant, dont la démonstration est donnée dans l'Annexe A.

Lemme 4.1 *La suite de processus $(\alpha_n, \psi, \mathbb{H}_n)$ est tendue et la suite (α_n, \mathbb{H}_n) converge en loi vers un processus gaussien continu et centré (α, \mathbb{H}) tel que*

$$\begin{aligned} \text{cov}\{\alpha(s), \alpha(t)\} &= 5\{(s \wedge t) - st\} - 3(s \wedge t) \log(s \wedge t) \\ &\quad + 2(s \wedge t) \log(s \vee t) + 3st\{\log(st) - \log s \log t\}, \end{aligned}$$

$$\begin{aligned} \text{cov}\{\alpha(s), \mathbb{H}(x)\} &= \{s \wedge (x^{(1)} x^{(2)})\} \left[1 + \log \left\{ \frac{x^{(1)} x^{(2)}}{s \wedge (x^{(1)} x^{(2)})} \right\} \right] \\ &\quad + x^{(2)} (s \wedge x^{(1)}) \left\{ 1 + \log \left(\frac{x^{(1)}}{s \wedge x^{(1)}} \right) \right\} \\ &\quad + x^{(1)} (s \wedge x^{(2)}) \left\{ 1 + \log \left(\frac{s \vee x^{(2)}}{s} \right) \right\} \\ &\quad - 3x^{(1)} x^{(2)} K(s), \end{aligned}$$

$$\begin{aligned} \text{cov}\{\mathbb{H}(x), \mathbb{H}(y)\} &= (x^{(1)} \wedge y^{(1)}) (x^{(2)} \wedge y^{(2)}) + x^{(1)} y^{(2)} (x^{(2)} \wedge y^{(1)}) \\ &\quad + x^{(2)} y^{(1)} (x^{(1)} \wedge y^{(2)}) - 3x^{(1)} x^{(2)} y^{(1)} y^{(2)}, \end{aligned}$$

où $x = (x^{(1)}, x^{(2)})$ et $y = (y^{(1)}, y^{(2)})$.

Ce lemme implique en particulier que la suite (α_n, ψ) est tendue dans l'espace $D[0, 1]$ des processus càdlàg munis de la topologie de Skorohod.

Or, on peut construire par interpolation linéaire un processus continu $\alpha'_{n,\psi}$ tel que $|\alpha_{n,\psi}(t) - \alpha'_{n,\psi}(t)| \leq 1/\sqrt{n}$ pour tout $t \in [0, 1]$. En vertu de l'inégalité (14.11) de Billingsley (1968), il s'ensuit que

$$\begin{aligned} w(\alpha_{n,\psi}, \delta) &\leq w(\alpha'_{n,\psi}, \delta) + 1/\sqrt{n} \\ &\leq 2w'(\alpha'_{n,\psi}, \delta) + 1/\sqrt{n} \\ &\leq 2w'(\alpha_{n,\psi}, \delta) + 3/\sqrt{n}, \end{aligned}$$

où

$$w'(\alpha_{n,\psi}, \delta) = \inf_{m \geq 2} \inf_{(t_i) \in \mathcal{P}_{m,\delta}} \max_{1 \leq i \leq m} \sup_{t_{i-1} \leq s, t \leq t_i, |s-t| \leq \delta} |\alpha_{n,\psi}(s) - \alpha_{n,\psi}(t)|$$

où $\mathcal{P}_{m,\delta}$ est l'ensemble des partitions de l'intervalle $[0, 1]$ en $m \geq 1$ points $0 = t_0 < t_1 < \dots < t_m = 1$ tels que $t_i - t_{i-1} > \delta$ pour tout $1 \leq i \leq m$.

Une application du théorème 15.2 de Billingsley (1986) permet alors de déduire que pour tout $\eta > 0$, il existe $0 < \delta < 1/2$ tels que pour tout $n \geq n_0$, on ait

$$P\{w'(\alpha_{n,\psi}, \delta) > \eta\} < \eta,$$

ce qui permet de conclure que la deuxième partie de l'hypothèse II est vérifiée. La troisième partie de cette hypothèse découle quant à elle directement du lemme 4.1.

4.2.3 Vérification de l'hypothèse III

Énonçons d'abord un corollaire au théorème d'Alexander (1987) présenté dans Barbe et coll. (1996).

Corollaire 4.1 Soit $q(t) = \sqrt{t} \log^p(1/t)$ et, pour $M > 0$, $1 < 2p < r$ et $n \geq 1$ quelconques, posons $t_n = \log^r(n)/n$ et définissons $F_{n,M}$ l'événement

$$\sup_{x:H(x)>t_n} \sqrt{n} \frac{|H_n(x) - H(x)|}{q\{H(x)\}} \leq M.$$

Si $k(t) = o\{t^{-1/2} \log^{-1/2-\epsilon}(1/t)\}$ quand $t \rightarrow \infty$, alors

$$\lim_{M \rightarrow \infty} \liminf_{n \rightarrow \infty} P(F_{n,M}) = 1$$

et

$$\sup_{x:H(x)>t_n} \frac{|H_n(x) - H(x)|}{H(x)} \leq M \log^{p-r/2}(n) \rightarrow 0$$

quand $n \rightarrow \infty$ et lorsque $F_{n,M}$ est satisfait.

L'intérêt de ce corollaire est que la suite

$$\sup_{x:H(x)>t_n} \sqrt{n} \frac{|H_n(x) - H(x)|}{q\{H(x)\}}$$

est tendue.

Pour vérifier l'hypothèse III, définissons d'abord la fonction $q(t) = -t^{1/2} \log t$ de même que la suite $t_n = (\log n)^3/n$.

Il est montré à l'Annexe B que $q(t)$ est positive et croissante dans un voisinage positif de zéro, que $q(t)/t$ est décroissante sur $(0, 1)$ et que $q(2t)/q(t)$ est bornée autour de zéro. Conformément à la notation employée par Ghoudi & Rémillard (1998), la fonction q appartient à l'ensemble \mathcal{Q} défini au Chapitre 3. Dans notre cas, $T = (0, 1]$ et $t_* = \inf T = 0$ n'appartient donc pas à T .

Dans l'Annexe B, il est montré que

$$\lim_{n \rightarrow \infty} \sqrt{n} K(t_n) = 0, \quad \lim_{n \rightarrow \infty} \frac{q(t_n)}{\sqrt{n} t_n} = 0 \quad \text{et} \quad \lim_{t \downarrow 0} k(t) q(t) = 0.$$

Avant d'aller plus loin, considérons les ensembles

$$\mathcal{A}_1 = \{1, \dots, n\} \cap \bigcup_{i=0} \{2ki + 1, \dots, (2i + 1)k\}$$

et

$$\mathcal{A}_2 = \{1, \dots, n\} \cap \bigcup_{i=0} \{(2i + 1)k + 1, \dots, 2(i + 1)k\},$$

définis pour tout délai k . Il s'agit des mêmes ensembles \mathcal{A}_ℓ utilisés à l'Annexe A pour la démonstration du lemme 4.1.

Définissons maintenant $H_{n_\ell}^{(\ell)}$ comme étant la fonction de répartition empirique construite à partir des n_ℓ observations X_i de l'ensemble \mathcal{A}_ℓ et

$$\mathbb{H}_{n_\ell}^{(\ell)} = \sqrt{n_\ell} (H_{n_\ell}^{(\ell)} - H), \quad \ell = 1, 2.$$

On voit alors que toutes les conditions d'application du corollaire 4.1 sont vérifiées, ce qui implique que les deux suites

$$\left\{ \sup_{x: H(x) > t_n} \frac{|\mathbb{H}_{n_\ell}^{(\ell)}(x)|}{q\{H(x)\}} \right\}_{n \geq 1}, \quad \ell = 1, 2$$

sont tendues.

Puisque

$$H_{n_\ell}^{(\ell)}(x) = \frac{1}{\sqrt{n_\ell}} \sum_{i \in \mathcal{A}_\ell} [\mathbb{I}_{\{X_i \leq x\}} - H(x)], \quad \ell = 1, 2,$$

alors

$$\begin{aligned} \mathbb{H}_n(x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{I}_{\{X_i \leq x\}} - H(x) \right\} \\ &= \frac{1}{\sqrt{n}} \left[\sum_{i \in \mathcal{A}_1} \left\{ \mathbb{I}_{\{X_i \leq x\}} - H(x) \right\} + \sum_{i \in \mathcal{A}_2} \left\{ \mathbb{I}_{\{X_i \leq x\}} - H(x) \right\} \right] \\ &= \frac{1}{\sqrt{n}} \left\{ \sqrt{n_1} \mathbb{H}_{n_1}^{(1)}(x) + \sqrt{n_2} \mathbb{H}_{n_2}^{(2)}(x) \right\} \\ &= \sqrt{\frac{n_1}{n_1 + n_2}} \mathbb{H}_{n_1}^{(1)}(x) + \sqrt{\frac{n_2}{n_1 + n_2}} \mathbb{H}_{n_2}^{(2)}(x). \end{aligned}$$

Ainsi,

$$\sup_{x: H(x) > t_n} \frac{|\mathbb{H}(x)|}{q\{H(x)\}} \leq \sup_{x: H(x) > t_n} \frac{|\mathbb{H}_{n_1}^{(1)}(x)|}{q\{H(x)\}} + \sup_{x: H(x) > t_n} \frac{|\mathbb{H}_{n_2}^{(2)}(x)|}{q\{H(x)\}}.$$

La suite $\sup_{x: H(x) > t_n} |\mathbb{H}(x)|/q\{H(x)\}$ étant bornée par la somme de deux suites tendues, il s'ensuit qu'elle est elle-même tendue, ce qui complète la vérification de l'hypothèse III.

4.3 Énoncé des résultats principaux

Soient \mathbb{H} la limite du processus $\mathbb{H}_n = \sqrt{n} (\tilde{H}_n - H)$ et $\alpha(t)$ la limite du processus empirique $\alpha_n(t) = \sqrt{n} [\sum_{i=1}^n \mathbb{I}\{H(X_i) \leq t\}/n - K(t)]$. Étant donné les vérifications faites à la Section 4.2, le résultat suivant est une conséquence du théorème 2.1 de Ghoudi & Rémillard (1998).

Théorème 4.1 *Le processus de Kendall sériel bivarié,*

$$\mathbb{K}_n(t) = \sqrt{n}\{K_n(t) - K(t)\},$$

tel que défini à la Section 4.1, converge vers un processus gaussien continu et centré $\mathbb{K}(t) = \alpha(t) - \mu(t, \mathbb{H})$ ayant pour fonction de covariance

$$\Gamma(s, t) = st - (s \wedge t)\{1 + \log(s \vee t)\}.$$

Les calculs permettant d'obtenir la fonction de covariance de ce processus sont détaillés à l'Annexe C.

Remarque 4.1 *Il est intéressant de remarquer que dans le cas bivarié, l'expression de la covariance est la même pour les processus de Kendall sériel et non sériel. Ces deux processus bivariés ont donc la même loi limite.*

Remarque 4.2 *Soit τ_n le tau de Kendall expérimental calculé à partir des paires $X_i = (Y_i, Y_{i+k})$, $1 \leq i \leq n$. Dans un article récent, Ferguson et coll. (2000) proposent d'utiliser cette statistique pour tester l'hypothèse d'indépendance sérielle. Sous cette hypothèse, on voit immédiatement que*

$$\sqrt{n}\tau_n = -4 \int_0^1 \mathbb{K}_n(t) dt,$$

converge vers un processus gaussien de moyenne nulle et de variance

$$16 \int_0^1 \int_0^1 \Gamma(s, t) ds dt = 4/9,$$

car

$$\begin{aligned} \int_0^1 \int_0^1 \Gamma(s, t) ds dt &= \int_0^1 \int_0^1 [st - (s \wedge t)\{1 + \log(s \vee t)\}] ds dt \\ &= \int_0^1 \int_0^t (st - s - s \log t) ds dt \\ &\quad + \int_0^1 \int_t^1 (st - t - t \log s) ds dt \\ &= \int_0^1 \left(\frac{t}{2} - \frac{t^2}{2} + \frac{t^2}{2} \log t \right) dt \\ &= \left[\frac{t^2}{4} - \frac{t^3}{6} + \frac{t^3}{6} \log t - \frac{t^3}{18} \right]_0^1 \\ &= \frac{1}{4} - \frac{1}{6} - \frac{1}{18} \\ &= \frac{1}{36}. \end{aligned}$$

Ce résultat est donc en accord avec celui de Ferguson et coll. (2000).

Jusqu'ici, nous avons toujours considéré un délai fixe. Pour énoncer le résultat suivant, définissons $\mathbb{K}_{n,i}$, un processus de Kendall sériel de délai i bâti à partir d'une série univariée, et \mathbb{K}_i , la limite de ce processus.

Théorème 4.2 $\{\mathbb{K}_{n,1}(t), \dots, \mathbb{K}_{n,d}(t)\}$ converge vers un processus gaussien centré et continu $\{\mathbb{K}_1(t), \dots, \mathbb{K}_d(t)\}$ dont la fonction de covariance limite est donnée, pour $i, j \in \{1, \dots, d\}$, par

$$\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{K}_{n,i}(s), \mathbb{K}_{n,j}(t)\} = \begin{cases} st - (s \wedge t)\{1 + \log(s \vee t)\}, & i = j \\ 0, & i \neq j, \end{cases}$$

ce qui signifie que deux processus de Kendall sériels de délais différents bâtis à partir de la même série sont indépendants.

Démonstration. Du théorème 4.1, les suites $\mathbb{K}_{n,i}(t)$, $i = 1, \dots, d$, sont tendues et convergent vers \mathbb{K}_i . Ainsi, $\{\mathbb{K}_{n,1}(t), \dots, \mathbb{K}_{n,d}(t)\}$ converge vers une loi jointe gaussienne centrée et continue notée $\{\mathbb{K}_1(t), \dots, \mathbb{K}_d(t)\}$.

La fonction de covariance limite, lorsque $i = j$, est également déduite du théorème 4.1. Les calculs qui montrent que $\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{K}_{n,i}(s), \mathbb{K}_{n,j}(t)\} = 0$, pour $i \neq j$, se trouvent à la section 5 de l'Annexe C.

4.4 Processus de Kendall sériel d -varié

Considérons à nouveau la suite Y_1, Y_2, \dots d'observations stationnaire issue d'une loi F inconnue et supposée continue. À l'aide d'un échantillon de

taille $n + d - 1$, définissons de la manière suivante les n vecteurs de variables aléatoires X_i formés en regroupant d observations consécutives de cette suite :

$$X_i = (Y_i, \dots, Y_{i+d-1}), \quad i = 1, \dots, n.$$

Ces variables nous permettront de caractériser la dépendance sérielle entre (au plus) d termes consécutifs de la suite des Y_i .

De façon similaire au cas bivarié (Section 4.1), nous allons établir la convergence du processus de Kendall sériel en vérifiant les trois hypothèses du théorème principal de Ghoudi & Rémillard (1998). Dans ce qui suit, $X = (X^{(1)}, \dots, X^{(d)})$ et H , H_n , K et K_n ont la même signification que précédemment.

Dans un premier temps, montrons comment calculer explicitement K , la fonction de répartition de $\varepsilon = H(X)$ sous l'hypothèse d'indépendance. Dans ce cas, on sait que $H(X) = \prod_{i=1}^d F(X^{(i)}) = \prod_{i=1}^d U_i$, où $U_i \sim \mathcal{U}(0, 1)$, $1 \leq i \leq d$. Puisque les variables $-\log(U_i)$ sont des observations indépendantes de la loi exponentielle d'espérance 1, leur somme est distribuée selon une loi Gamma $(d, 1)$. Par conséquent, la densité de ε est donnée par

$$k_d(u) = \frac{1}{(d-1)!} \{\log(1/u)\}^{d-1}, \quad u \in (0, 1].$$

On en déduit que

$$K_d(t) = \int_0^t k_d(u) du = \int_0^t \frac{1}{(d-1)!} \{\log(1/u)\}^{d-1} du.$$

En intégrant cette expression par partie, on s'aperçoit que

$$\int_0^t k_d(u) du = \frac{t \{\log(1/t)\}^{d-1}}{(d-1)!} + \int_0^t k_{d-1}(u) du.$$

En développant récursivement cette expression et en utilisant le fait que

$$\int_0^t k_1(u)du = t,$$

on obtient finalement

$$K(t) = t \sum_{i=1}^d \frac{\{\log(1/t)\}^{i-1}}{(i-1)!},$$

déjà rapportée par Barbe et coll. (1996).

4.4.1 Vérification de l'hypothèse I

Pour demeurer cohérent avec la notation introduite par Ghoudi & Rémillard, définissons $T = (0, 1]$, $\mathcal{X} = [0, 1]^d$ et \mathcal{C}_f , l'ensemble des fonctions continues et bornées de $[0, 1]^d \rightarrow \mathbb{R}$.

En employant le même argument qu'à la section 4.2.1, on peut voir que la densité k_d est bornée sur tout compact de $T = (0, 1]$. Considérons maintenant

$$\begin{aligned} \mu(t, f) &= k_d(t)E\{f(X) \mid H(X) = t\} \\ &= k_d(t)E \left\{ f(U_1, \dots, U_d) \mid \prod_{i=1}^d U_i = t \right\}, \quad \text{où } f \in \mathcal{C}_f, \quad t \in T. \end{aligned}$$

Pour calculer l'espérance, il faut d'abord déterminer la loi jointe de $U_1, \dots, U_{d-1}, \prod_{i=1}^d U_i$. En effectuant la transformation $\Xi_i = U_i$, $i = 1, \dots, d-1$ et $\Xi_d = \prod_{i=1}^d U_i$, dont la valeur du jacobien vaut $1/(\prod_{i=1}^{d-1} \xi_i)$, on voit que la loi jointe de Ξ_1, \dots, Ξ_d est

$$f_{\Xi_1, \dots, \Xi_d}(\xi_1, \dots, \xi_d) = \left(\prod_{i=1}^{d-1} \xi_i \right)^{-1},$$

où $0 < \xi_1, \dots, \xi_{d-1} < 1$ et $0 < \xi_d < \xi_1 < \dots < \xi_{d-1} < 1$.

De là,

$$\begin{aligned} \mu(t, f) &= k_d(t) E\{f(X) | H(X) = t\} \\ &= k_d(t) E\left\{f\left(\Xi_1, \dots, \Xi_{d-1}, t / \prod_{i=1}^{d-1} \Xi_i\right) \mid \Xi_d = t\right\} \\ &= \int_{\mathcal{D}} \frac{1}{\prod_{i=1}^{d-1} \xi_i} \cdot f\left(\xi_1, \dots, \xi_{d-1}, t / \prod_{i=1}^{d-1} \xi_i\right) d\xi_1 \cdots d\xi_{d-1}, \end{aligned}$$

où \mathcal{D} est le domaine des variables ξ_1, \dots, ξ_{d-1} .

La preuve de la continuité de $\mu(t, f)$, de même que le résultat

$$\lim_{M \rightarrow \infty} \int_M^\infty \sup_{s \in C} P_s\{r(X) > u\} du = 0,$$

s'obtiennent de la même manière que pour le cas bivarié, ce qui complète la vérification de l'hypothèse I.

4.4.2 Vérification de l'hypothèse II

La première partie de l'hypothèse II se vérifie comme précédemment en interpolant linéairement la fonction de répartition H_n par \tilde{H}_n .

Le reste de la vérification de l'hypothèse II procède de la même manière que dans le cas bivarié, en se servant du résultat suivant qui généralise le lemme 4.1 à d dimensions.

Lemme 4.2 *La suite de processus $(\alpha_{n,\psi}, \mathbb{H}_n)$, construite à partir des variables $X_i = (Y_i, \dots, Y_{i+d-1})$, est tendue et la suite (α_n, \mathbb{H}_n) , construite à*

partir des mêmes variables, converge en loi vers un processus gaussien continu (α, \mathbb{H}) dont la fonction de covariance est la limite des fonctions de covariance de (α_n, \mathbb{H}_n) .

Démonstration. Nous allons d'abord montrer que les suites $\alpha_{n,\psi}$ et \mathbb{H}_n sont tendues, ce qui impliquera la tension de la suite $(\alpha_{n,\psi}, \mathbb{H}_n)$. Pour ce faire, soient les d sous-ensembles suivants:

$$\mathcal{B}_\ell = \{1, \dots, n\} \cap \bigcup_{i=0} \{di + \ell\}, \quad \ell = 1, \dots, d.$$

Ces ensembles seront utiles car toutes les observations X_i dont les indices appartiennent à un même ensemble \mathcal{B}_ℓ sont indépendantes.

Écrivons maintenant

$$\alpha_{n,\psi}(t) = \sum_{\ell=1}^d \alpha_{n,\psi}^{(\ell)}(t)$$

où, pour $\ell \in \{1, \dots, d\}$,

$$\alpha_{n,\psi}^{(\ell)}(t) = \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{B}_\ell} \left[\psi(X_i) \mathbb{I}\{\varepsilon_i \leq t\} - E[\psi(X_i) \mathbb{I}\{\varepsilon \leq t\}] \right].$$

Le processus $\alpha_{n,\psi}(t)$ peut donc s'écrire comme la somme de d processus bâtis à partir d'observations indépendantes.

La tension des suites $\alpha_{n,\psi}^{(\ell)}(t)$ s'obtient en utilisant l'inégalité A.1 du théorème A.1. Les calculs sont exactement les mêmes que dans le cas bivarié. Il s'ensuit alors que $\alpha_{n,\psi}(t)$ est tendue, une somme finie de suites tendues étant bien entendu tendue.

Soit maintenant n_ℓ , le nombre d'éléments de l'ensemble \mathcal{B}_ℓ , où $\ell = 1, \dots, d$. Il est clair que $n = n_1 + \dots + n_d$. On pourrait facilement démontrer, par des manipulations algébriques similaires à celles présentées à la section 4.2.3, que

$$\mathbb{H}_n(x) = \sum_{\ell=1}^d \sqrt{\frac{n_\ell}{n}} \mathbb{H}_{n_\ell}^{(\ell)}(x),$$

où

$$\mathbb{H}_{n_\ell}^{(\ell)}(x) = \frac{1}{\sqrt{n_\ell}} \sum_{i \in \mathcal{B}_\ell} \left[\mathbb{I}\{X_i \leq x\} - H(X) \right].$$

Puisque $\mathbb{H}_{n_\ell}^{(\ell)}(x)$ est composée d'observations indépendantes, un résultat d'Alexander (1987) nous assure de sa convergence et donc de sa tension. La suite \mathbb{H}_n étant ainsi bornée par une somme finie de suites tendues, elle est elle-même tendue.

Finalement, pour montrer la convergence de la suite (α_n, \mathbb{H}_n) , il reste à établir que, pour tout $m \geq 1$ et toutes valeurs t_1, \dots, t_m et x_1, \dots, x_m , $\{\alpha_n(t_1), \dots, \alpha_n(t_m), \mathbb{H}_n(x_1), \dots, \mathbb{H}_n(x_m)\}$ converge vers $\{\alpha(t_1), \dots, \alpha(t_m), \mathbb{H}(x_1), \dots, \mathbb{H}(x_m)\}$. Ce résultat s'obtient en répétant *mutatis mutandis* la démonstration du lemme A.2. \diamond

4.4.3 Vérification de l'hypothèse III

Ici encore, l'argument est semblable à celui développé dans le cas bivarié. Posons $q(t) = t^{1/2} \log(1/t)$, un élément de \mathcal{Q} , et soit en outre $t_n = (\log^3 n)/n$.

Dans l'Annexe B, il est montré que $\sqrt{n}K(t_n) \rightarrow 0$, où

$$K_d(t) = t \sum_{i=1}^d \frac{\{\log(1/t)\}^{i-1}}{(i-1)!}$$

et que $k_d(t)q(t) \rightarrow 0$, où

$$k_d(t) = \frac{1}{(d-1)!} \{\log(1/t)\}^{d-1}.$$

Pour démontrer que la suite

$$\left\{ \sup_{x: H(x)-t_* > t_n} \frac{|IH_n(x)|}{q_*\{H(x) - t_*\}} \right\}_{n \geq 1}$$

est tendue, il suffit d'observer que

$$\sup_{x: H(x) > t_n} \frac{|IH_n(x)|}{q\{H(x)\}} \leq d \max_{1 \leq \ell \leq d} \left[\sup_{x: H(x) > t_n} \frac{|IH_{n_\ell}^{(\ell)}(x)|}{q\{H(x)\}} \right]$$

et que chacune des suites

$$\left\{ \sup_{x: H(x) > t_n} \frac{|IH_{n_\ell}^{(\ell)}(x)|}{q\{H(x)\}} \right\}_{n \geq 1}, \quad \ell = 1, \dots, d$$

est tendue, puisque les conditions du corollaire 4.1 sont toutes vérifiées. Ceci termine la vérification de l'hypothèse III.

4.4.4 Énoncé du résultat principal

Compte tenu de la vérifications des hypothèses faite aux Sections 4.4.1, 4.4.2 et 4.4.3, le résultat suivant est une conséquence du théorème 2.1 de Ghoudi & Rémillard (1998).

Théorème 4.3 *Le processus de Kendall sériel d -varié, $\mathbb{K}_n = \sqrt{n}(K_n - K)$, défini au début de la Section 4.4, converge vers un processus gaussien continu et centré \mathbb{K} ayant comme représentation*

$$\mathbb{K}(t) = \alpha(t) - \mu(t, \mathbb{H}), \quad 0 \leq t \leq 1,$$

où $\alpha(t)$ et \mathbb{H} sont respectivement les limites des processus α_n et \mathbb{H}_n . De plus, la fonction de covariance limite est donnée par

$$\text{cov}\{\mathbb{K}(s), \mathbb{K}(t)\} = \Gamma_{\alpha, \alpha}(s, t) - \Gamma_{\alpha, \mu}(s, t) - \Gamma_{\alpha, \mu}(t, s) + \Gamma_{\mu, \mu}(s, t),$$

où

$$\begin{aligned} \Gamma_{\alpha, \alpha}(s, t) &= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \alpha_n(t)\}, \\ \Gamma_{\alpha, \mu}(s, t) &= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \mu(t, \mathbb{H}_n)\}, \\ \Gamma_{\mu, \mu}(s, t) &= \lim_{n \rightarrow \infty} \text{cov}\{\mu(s, \mathbb{H}_n), \mu(t, \mathbb{H}_n)\}. \end{aligned}$$

Remarque 4.3 *Il ne semble pas possible d'obtenir une expression simple pour la fonction de covariance de ce processus d -varié. Nous ne croyons pas non plus que cette fonction de covariance soit généralement la même que celle du processus de Kendall non sériel défini dans Barbe et coll. (1996).*

Chapitre 5

Quelques statistiques déduites du processus de Kendall sériel

Il est possible, à partir du processus

$$\mathbb{K}_n(t) = \sqrt{n}\{K_n(t) - K(t)\},$$

de construire des statistiques permettant de tester l'hypothèse de bruit blanc dans une série chronologique univariée. Trois statistiques de ce type sont présentées dans ce chapitre, à savoir la statistique de Kolmogorov-Smirnov, la statistique de Cramér-von Mises et le tau de Kendall.

Le comportement asymptotique du processus de Kendall sériel, établi au chapitre 4 sous l'hypothèse d'indépendance, garantit la convergence en loi de ces statistiques, qui s'expriment comme fonctionnelles continues de ce processus. Comme il est toutefois difficile d'exprimer ces lois sous forme analytique, certains de leurs quantiles ont été déterminés à l'aide de simulations. Ces

résultats, nécessaires pour l'application des tests, sont donnés sous forme de tableaux dans les pages suivantes.

Au chapitre 6, l'efficacité relative de ces trois statistiques sera comparée. Comme on le verra, l'intérêt des statistiques de Kolmogorov-Smirnov et de Cramér-von Mises réside entre autres dans le fait que si la taille d'échantillon est suffisamment grande, elles permettent de rejeter l'hypothèse d'indépendance si $K(t) = P\{H(Y_1, Y_{k+1}) \leq t\} \neq t - t \log(t)$, $t \in [0, 1]$, contrairement à la statistique basée sur le tau de Kendall sériel qui n'amène le rejet que si $\int_0^1 K(t)dt \neq 3/4$.

5.1 La statistique de Kolmogorov-Smirnov

De manière générale, la statistique de Kolmogorov-Smirnov est basée sur la différence en norme sup entre deux fonctions de répartition. Dans notre cas, on peut la définir comme suit

$$S_n^{KS} = \sup_{0 \leq t \leq 1} \sqrt{n} |K_n(t) - K(t)| = \sup_{0 \leq t \leq 1} |\mathbb{K}_n(t)|,$$

puisque K_n et K coïncident en dehors de l'intervalle $[0, 1]$. La convergence du processus \mathbb{K}_n établie au chapitre 4 entraîne immédiatement celle de S_n^{KS} vers $\sup_{0 \leq t \leq 1} |K(t)|$. Les quantiles de cette loi limite étant toutefois difficiles à obtenir analytiquement, ils ont plutôt été déterminés par simulation sous l'hypothèse nulle d'indépendance sérielle.

TABLEAU 1: Quantiles de la loi de la statistique de Kolmogorov-Smirnov basée sur un échantillon aléatoire de taille n , tels qu'estimés à partir de 10 000 répétitions.

n	90	95	97.5	98	99	99.5	99.9
10	1.458	1.475	1.475	1.650	1.650	1.775	1.791
20	1.326	1.440	1.496	1.496	1.607	1.663	1.774
30	1.275	1.383	1.453	1.471	1.495	1.636	1.763
40	1.243	1.332	1.455	1.456	1.507	1.613	1.694
50	1.224	1.328	1.394	1.429	1.507	1.573	1.715
60	1.202	1.300	1.405	1.429	1.516	1.559	1.694
70	1.182	1.284	1.378	1.408	1.489	1.577	1.737
80	1.179	1.275	1.378	1.389	1.497	1.554	1.722
90	1.159	1.263	1.366	1.379	1.463	1.516	1.657
100	1.155	1.259	1.352	1.377	1.455	1.546	1.702
200	1.100	1.205	1.303	1.326	1.414	1.489	1.630
300	1.088	1.191	1.294	1.327	1.413	1.498	1.622
400	1.074	1.174	1.268	1.305	1.389	1.461	1.613
500	1.074	1.164	1.247	1.276	1.356	1.439	1.607
750	1.061	1.165	1.264	1.287	1.378	1.444	1.600
1000	1.061	1.161	1.258	1.287	1.376	1.458	1.620
2000	1.050	1.152	1.250	1.275	1.364	1.449	1.633

Le tableau 1 donne les valeurs estimées de plusieurs quantiles de la loi de S_n^{KS} sous l'hypothèse nulle pour différentes tailles d'échantillon. Ces estimations sont fondées sur 10 000 réplicats. Comme le processus est légèrement biaisé à taille finie (quoiqu'asymptotiquement sans biais), la convergence des quantiles est relativement lente. En vue de l'accélérer, nous avons donc soustrait $7/(12n)$ de chacune des pseudo-observations, comme Ferguson et coll. (2000) le font implicitement.

Le tableau 2 donne les quantiles de la statistique modifiée

$$S_n^{KS*} = \sup_{t \in \mathcal{R}} \sqrt{n} \left| K_n \left(t + \frac{7}{12n} \right) - K(t) \right|,$$

qui converge vers la même limite. Comme on peut le constater, l'approximation asymptotique devient alors valable pour de très petits échantillons.

5.2 La statistique de Cramér-von Mises

La statistique de Cramér-von Mises, définie par

$$S_n^{CVM} = n \int_0^1 \left\{ K_n \left(t + \frac{7}{12n} \right) - K(t) \right\}^2 dt,$$

mesure la distance en moyenne quadratique entre la fonction de répartition empirique corrigée $K_n\{t + 7/(12n)\}$ et sa limite $K(t)$. En se servant du théorème 4.1, on voit tout de suite que S_n^{CVM} converge vers la statistique $\int_0^1 \{K(t)\}^2 dt$.

TABLEAU 2: Quantiles de la loi de la statistique de Kolmogorov-Smirnov corrigée basée sur un échantillon aléatoire de taille n , tels qu'estimés à partir de 10 000 répétitions.

n	90	95	97.5	98	99	99.5	99.9
10	1.031	1.031	1.292	1.313	1.347	1.347	1.663
20	1.068	1.138	1.240	1.304	1.335	1.464	1.672
30	1.059	1.152	1.242	1.299	1.346	1.424	1.551
40	1.079	1.146	1.251	1.278	1.384	1.443	1.587
50	1.054	1.168	1.242	1.292	1.356	1.473	1.631
60	1.052	1.155	1.236	1.264	1.346	1.427	1.596
70	1.063	1.172	1.249	1.288	1.360	1.464	1.644
80	1.066	1.164	1.270	1.291	1.386	1.436	1.680
90	1.055	1.164	1.264	1.296	1.372	1.455	1.620
100	1.058	1.142	1.238	1.275	1.366	1.441	1.614
200	1.047	1.146	1.240	1.268	1.358	1.432	1.593
300	1.053	1.150	1.244	1.272	1.353	1.445	1.707
400	1.053	1.149	1.248	1.267	1.368	1.471	1.591
500	1.054	1.156	1.250	1.276	1.353	1.450	1.583
750	1.048	1.149	1.244	1.275	1.364	1.429	1.601
1000	1.054	1.156	1.249	1.276	1.358	1.447	1.594
2000	1.051	1.158	1.250	1.280	1.355	1.448	1.611

TABLEAU 3: Quantiles de la loi de la statistique de Cramér-von Mises basée sur un échantillon aléatoire de taille n , tels qu'estimés à partir de 10 000 répétitions.

n	90	95	97.5	98	99	99.5	99.9
10	0.154	0.192	0.233	0.250	0.289	0.339	0.435
20	0.157	0.200	0.247	0.261	0.309	0.350	0.460
30	0.161	0.202	0.247	0.261	0.309	0.358	0.455
40	0.162	0.207	0.251	0.266	0.311	0.345	0.475
50	0.160	0.200	0.251	0.267	0.311	0.361	0.462
60	0.160	0.207	0.256	0.270	0.308	0.349	0.464
70	0.159	0.206	0.247	0.266	0.318	0.360	0.483
80	0.161	0.202	0.246	0.263	0.314	0.365	0.522
90	0.161	0.206	0.254	0.268	0.315	0.370	0.468
100	0.160	0.204	0.248	0.263	0.316	0.363	0.471
200	0.159	0.200	0.244	0.259	0.303	0.354	0.435
300	0.161	0.205	0.251	0.266	0.313	0.350	0.490
400	0.159	0.202	0.250	0.262	0.305	0.368	0.498
500	0.158	0.206	0.254	0.269	0.316	0.378	0.487
750	0.160	0.204	0.249	0.268	0.314	0.361	0.493
1000	0.162	0.211	0.255	0.269	0.319	0.371	0.464
2000	0.161	0.205	0.253	0.270	0.316	0.375	0.488

Le tableau 3 illustre la convergence de plusieurs quantiles de la loi de S_n^{CVM} sous l'hypothèse nulle. Comme on peut le voir, l'approximation asymptotique est valable pour de très petits échantillons.

5.3 Le tau de Kendall

Soit τ_n , le tau de Kendall expérimental calculé au moyen des paires $X_i = (Y_i, Y_{i+k})$ formées à partir de la chronique Y_1, \dots, Y_{n+k} . Nous avons déjà signalé à la remarque 4.2 que sous l'hypothèse nulle d'indépendance sérielle,

$$\sqrt{n}\tau_n = -4 \int_0^1 \mathbb{K}_n(t) dt$$

converge vers une loi $N(0, 4/9)$.

Le tableau 4 illustre la convergence de plusieurs quantiles de la loi de la version corrigée et normalisée $1.5\sqrt{n}\{\tau_n - 7/(12n)\}$ sous la même hypothèse de bruit blanc. Comme on peut le voir, ici encore l'approximation asymptotique est valable pour de très petits échantillons.

TABLEAU 4: Quantiles de la loi de la statistique de Kendall sérielle basée sur un échantillon aléatoire de taille n , tels qu'estimés à partir de 10 000 répétitions.

n	90	95	97.5	98	99	99.5	99.9
10	1.360	1.739	1.929	2.119	2.308	2.498	2.878
20	1.297	1.632	1.968	2.035	2.303	2.571	2.907
30	1.296	1.625	1.954	2.063	2.319	2.611	3.086
40	1.273	1.629	1.937	2.032	2.316	2.530	3.075
50	1.304	1.677	1.966	2.051	2.322	2.560	2.984
60	1.291	1.652	1.949	2.027	2.285	2.569	3.227
70	1.282	1.651	1.928	2.040	2.327	2.522	3.024
80	1.319	1.671	1.965	2.074	2.334	2.552	2.971
90	1.268	1.634	1.922	1.999	2.259	2.533	3.159
100	1.294	1.654	1.978	2.074	2.362	2.584	3.226
200	1.295	1.672	1.980	2.044	2.292	2.542	3.013
300	1.313	1.682	1.993	2.091	2.380	2.575	3.210
400	1.273	1.647	1.962	2.054	2.294	2.503	3.100
500	1.283	1.658	1.976	2.105	2.375	2.605	3.119
750	1.313	1.655	1.957	2.048	2.313	2.619	3.146
1000	1.316	1.666	1.965	2.053	2.288	2.555	3.095
2000	1.277	1.634	1.949	2.065	2.324	2.572	3.171

5.4 Étude d'une série chaotique déterministe

Au chapitre 1, nous avons présenté une série déterministe que le corrélogramme identifie injustement comme un bruit blanc. Nous allons maintenant montrer que les nouveaux tests développés à partir du processus de Kendall sériel détectent cette dépendance.

Rappelons tout d'abord que les termes de cette série stationnaire sont définis récursivement par

$$Y_{i+1} = \begin{cases} 2Y_i, & 0 \leq Y_i \leq 1/2, \\ 2(1 - Y_i), & 1/2 \leq Y_i \leq 1, \end{cases}, \quad i \geq 1.$$

où $Y_1 \sim \mathcal{U}(0, 1)$. Remarquons que pour tous $0 \leq u, v \leq 1$, on a

$$\begin{aligned} H_c(u, v) &= P(Y_i \leq u, Y_{i+1} \leq v) \\ &= P(Y_i \leq u, Y_i \leq 1/2, 2Y_i \leq v) + P(Y_i \leq u, Y_i > 1/2, 2 - 2Y_i \leq v) \\ &= P\{Y_i \leq \min(u, v/2)\} + P\{1 - v/2 \leq Y_i \leq u\} \\ &= \min(u, v/2) + \max(u + v/2 - 1, 0). \end{aligned}$$

Par conséquent, $H_c(Y_i, Y_{i+1}) = Y_i$ si $Y_i \leq 1/2$ et $H_c(Y_i, Y_{i+1}) = 1 - Y_i$ lorsque $Y_i \geq 1/2$. Ainsi, pour tout $0 \leq t \leq 1$,

$$\begin{aligned} K_c(t) &= P\{H_c(Y_i, Y_{i+1}) \leq t\} \\ &= P(Y_i \leq t, Y_i \leq 1/2) + P(1 - Y_i \leq t, Y_i > 1/2) \\ &= \min(t, 1/2) + 1 - \max(1 - t, 1/2) \\ &= \min(2t, 1). \end{aligned}$$

Cette fonction étant bien différente de celle qui correspond à l'indépendance sérielle, à savoir $K(t) = t - t \log t$, les tests qui utilisent les statistiques de Kolmogorov-Smirnov et de Cramér-Von Mises, basées sur le processus $\sqrt{n}\{K_n(t) - K(t)\}$, devraient rejeter l'hypothèse nulle. Toutefois, malgré le caractère déterministe de cette série, un test effectué à partir du tau de Kendall ne devrait pas rejeter cette hypothèse d'indépendance sérielle car

$$\begin{aligned}\tau_c &= 4E\{H_c(U, V)\} - 1 = 4 \int_0^{1/2} t dK(t) - 1 \\ &= 4 \int_0^{1/2} 2tdt - 1 = 4(1/4) - 1 = 0.\end{aligned}$$

Cette observation est confirmée par les trois tableaux suivants, qui donnent, pour des séries de taille 20, 50 et 100, le pourcentage de rejet de l'hypothèse nulle au seuil de 5%, tel qu'observé à partir de 10 000 séries, et ce pour les délais 1, ..., 5. Le programme ayant servi à créer ces tables est reproduit à l'annexe D.

TABLEAU 5: Pourcentage de rejet de l'hypothèse nulle pour différents délais de la série chaotique de longueur 20, tel qu'estimé à partir de 10 000 répétitions.

Statistiques	1	2	3	4	5
tau de Kendall	22.160	7.190	5.650	5.940	5.220
Kolmogorov-Smirnov	39.720	9.120	4.050	3.560	2.690
Cramér-von Mises	51.040	11.780	5.750	5.650	4.790

TABLEAU 6: Pourcentage de rejet de l'hypothèse nulle pour différents délais de la série chaotique de longueur 50, tel qu'estimé à partir de 10 000 répétitions.

statistiques	1	2	3	4	5
tau de Kendall	24.120	8.400	5.500	5.270	5.090
Kolmogorov-Smirnov	84.350	25.060	7.990	3.850	2.880
Cramér-von Mises	95.380	20.450	6.450	4.910	4.400

TABLEAU 7: Pourcentage de rejet de l'hypothèse nulle pour différents délais de la série chaotique de longueur 100, tel qu'estimé à partir de 10 000 répétitions.

statistiques	1	2	3	4	5
tau de Kendall	24.270	9.210	5.840	5.250	4.520
Kolmogorov-Smirnov	99.940	50.440	13.270	5.040	3.360
Cramér-von Mises	100.000	37.420	7.690	4.830	3.670

Ces tableaux nous montrent que les statistiques de Kolmogorov-Smirnov et de Cramér-von Mises sont beaucoup plus efficaces que le tau de Kendall pour détecter la dépendance sérielle dans cette série déterministe, particulièrement pour des chroniques de longueur 50 et de 100. On remarque cependant que le pourcentage de rejet diminue rapidement lorsque le délai augmente. Ceci est vraisemblablement attribuable au fait que le caractère de cette série devient de plus en plus imprévisible à mesure que le délai croît. Pour de plus amples

détails concernant la nature chaotique de cette série, consulter par exemple Brock (1986).

Pour sa part, le fait que la statistique de Kendall sérielle de délai 1 rejette l'hypothèse de bruit blanc une fois sur 4 environ vient probablement de ce que la variance de cette statistique est plus grande pour la série chaotique que sous l'indépendance.

Chapitre 6

Tests de puissance

Nous avons vu à la fin de la section 5.4 que les statistiques de Kolmogorov-Smirnov et de Cramér-von Mises rejettent plus souvent l'hypothèse d'indépendance que le tau de Kendall, dans le cas particulier de la série déterministe. Dans ce chapitre, la puissance de ces trois mêmes statistiques sera étudiée sous divers scénarios de dépendance sérielle.

6.1 Modèles autorégressifs

En premier lieu, à l'instar de Ferguson et coll. (2000), nous allons considérer des contre-hypothèses markoviennes de type $X_i = \theta X_{i-1} + \varepsilon_i$, où les erreurs ε_i obéiront successivement à une loi normale, logistique, de Laplace et de Cauchy. Le coefficient d'autocorrélation θ prendra les mêmes valeurs que dans leur article, soit $1/4$, $1/8$, $1/16$, $1/32$ et 0 ($\theta = 0$ correspond à

l'indépendance). Cependant, il est à noter que contrairement à ces auteurs, les tests effectués seront bilatéraux, ce qui empêche toute comparaison directe avec leurs résultats.

Le tableau 8 rapporte le pourcentage de rejet au seuil de 5% pour une chronique de longueur 20, tel qu'estimé à partir de 10 000 répétitions. Les tableaux 9 et 10 donnent les résultats correspondants pour des séries de longueur 50 et 100 respectivement.

On constate d'abord que toutes les statistiques conservent leur seuil nominal, qui est ici de 5%. De plus, on remarque qu'à part quelques exceptions attribuables à l'erreur expérimentale, la puissance augmente lorsque θ croît, comme on pouvait s'y attendre puisque la dépendance entre les observations successives de la série chronologique devient alors plus forte.

La principale surprise vient du fait que dans presque tous les cas, le tau de Kendall sériel rejette l'hypothèse nulle plus fréquemment que les deux autres statistiques. Quelques exceptions surviennent lorsque la dépendance est très faible ($\theta = 1/16, 1/32$) et dans ces cas, c'est la statistique de Cramér-von Mises qui domine. Notons par ailleurs que cette dernière termine bon second dans presque toutes les situations considérées.

TABLEAU 8: Pourcentage de rejet au seuil de 5% pour une chronique de longueur 20, tel qu'estimé à partir de 10 000 répétitions.

Loi des erreurs	statistique	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
<i>Normale</i>	K-Smirnov	7.21	4.43	4.00	4.56	4.05
	C-V-M	14.89	7.16	5.53	5.32	5.00
	Kendall's tau	16.15	7.25	5.21	4.84	4.16
<i>Logistique</i>	K-Smirnov	7.78	4.33	3.78	4.12	4.24
	C-V-M	15.29	7.48	5.43	5.44	4.88
	Kendall's tau	16.63	7.32	5.29	4.86	4.66
<i>Laplace</i>	K-Smirnov	8.91	4.65	4.07	4.39	4.68
	C-V-M	19.33	8.83	5.81	5.38	4.98
	Kendall's tau	21.58	9.31	5.66	4.92	4.60
<i>Cauchy</i>	K-Smirnov	18.25	7.80	4.75	4.41	4.63
	C-V-M	36.87	16.12	8.84	6.59	5.28
	Kendall's tau	41.64	18.52	9.63	6.63	4.79

TABLEAU 9: Pourcentage de rejet au seuil de 5% pour une chronique de longueur 50, tel qu'estimé à partir de 10 000 répétitions.

Loi des erreurs	statistique	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
<i>Normale</i>	K-Smirnov	23.34	9.15	6.46	5.46	5.79
	C-V-M	35.03	12.56	7.01	5.85	5.24
	Kendall's tau	36.89	12.71	6.55	5.48	4.53
<i>Logistique</i>	K-Smirnov	24.37	9.16	6.50	5.55	5.45
	C-V-M	36.83	12.89	7.23	5.73	5.25
	Kendall's tau	39.33	13.06	6.56	5.26	4.75
<i>Laplace</i>	K-Smirnov	32.33	12.29	7.04	5.77	5.78
	C-V-M	47.91	17.36	8.05	6.04	5.43
	Kendall's tau	52.09	18.24	7.94	5.39	4.61
<i>Cauchy</i>	K-Smirnov	59.13	24.34	12.20	7.36	6.01
	C-V-M	77.37	35.99	16.40	8.81	5.28
	Kendall's tau	82.16	40.90	18.32	9.22	4.65

TABLEAU 10: Pourcentage de rejet au seuil de 5% pour une chronique de longueur 100, tel qu'estimé à partir de 10 000 répétitions.

Loi des erreurs	statistique	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
<i>Normale</i>	K-Smirnov	42.53	13.24	6.42	5.38	5.06
	C-V-M	60.47	20.09	8.32	6.19	5.39
	Kendall's tau	64.22	21.16	8.22	5.56	5.06
<i>Logistique</i>	K-Smirnov	45.91	14.29	7.17	5.68	5.38
	C-V-M	65.09	21.74	9.38	6.46	5.59
	Kendall's tau	69.12	22.64	9.01	5.76	5.12
<i>Laplace</i>	K-Smirnov	60.19	20.92	8.74	5.74	5.28
	C-V-M	78.79	30.42	11.77	6.84	5.26
	Kendall's tau	82.35	32.39	12.03	6.49	4.83
<i>Cauchy</i>	K-Smirnov	88.30	45.03	18.74	9.07	5.10
	C-V-M	96.85	64.11	27.94	12.34	5.03
	Kendall's tau	98.18	69.63	31.92	13.74	4.52

6.2 Autres modèles de dépendance

Dans un deuxième temps, nous avons considéré un modèle ARCH et des chroniques stationnaires dont les paires d'observations successives ont pour loi, soit une copule de Farlie-Gumbel-Morgenstern, soit une copule archimédienne de Clayton, de Frank ou de Ali-Mikhail-Haq. Les définitions de ces différents modèles, répertoriés par Nelsen (1999), sont rappelées ci-dessous.

1⁰) Modèle ARCH

Les modèles ARCH (abréviation de: autorégressifs conditionnellement hétéroscédastiques) ont été introduits dans la littérature économétrique par Engle (1982) pour mieux refléter la volatilité conditionnelle de nombreuses séries financières. Pour les fins de notre étude, un modèle ARCH(1) a été retenu, pour lequel les observations successives sont générées par

$$Y_{i+1} = \varepsilon_{i+1} \sqrt{1 + 3.5\theta Y_i^2},$$

où les ε_i sont des variables aléatoires normales centrées réduites indépendantes.

2⁰) Modèle de Clayton

La copule de ce modèle bivarié s'écrit sous la forme

$$C(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \quad 0 < u, v < 1$$

où $\alpha \geq 0$ est un paramètre de dépendance lié au tau de Kendall par la formule $\tau = \alpha/(2 + \alpha)$.

Pour générer une série d'observations ayant cette loi de dépendance, notons que

$$\begin{aligned} F_u(t) &= P(V \leq t | U = u) = \frac{\partial}{\partial u} C(u, t) \\ &= u^{-(\alpha+1)} (u^{-\alpha} + t^{-\alpha} - 1)^{-(\alpha+1)/\alpha} \\ &= \{1 + u^\alpha (t^{-\alpha} - 1)\}^{-(1+\alpha)/\alpha} \end{aligned}$$

et que par conséquent,

$$F_u^{-1}(v) = u(u^\alpha - 1 + v^{-\alpha/(1+\alpha)})^{-1/\alpha}, \quad 0 < v < 1.$$

Dans nos simulations, cependant, nous avons plutôt utilisé la paramétrisation $\theta = \alpha/(1 + \alpha) \in [0, 1)$, de telle sorte que $\tau = \theta/(2 - \theta)$ et

$$F_u^{-1}(v) = u(u^{\theta/(1-\theta)} - 1 + v^{-\theta})^{1-1/\theta}.$$

Pour simuler une chronique stationnaire dont la loi jointe de deux observations consécutives est la copule de Clayton, nous avons simplement généré des variables uniformes et indépendantes U_1, \dots, U_n et nous avons posé $Y_1 = U_1$, et $Y_{i+1} = F_{Y_i}^{-1}(U_{i+1})$ pour $i \geq 1$. Ceci a été fait pour les mêmes valeurs de θ qu'à la section précédente, à savoir $\theta = 1/4, 1/8, 1/16, 1/32$ et 0.

3⁰) Modèle de Frank

Pour ce modèle,

$$C(u, v) = \{\log(\alpha^u + \alpha^v - \alpha^{u+v} - \alpha) - \log(1 - \alpha)\} / \log \alpha, \quad 0 < u, v < 1$$

où le paramètre α est non-négatif. Un calcul simple donne

$$F_u(t) = P(V \leq t | U = u) = \frac{\partial}{\partial u} C(u, t) = \alpha^u(1 - \alpha^v) / (\alpha^u + \alpha^v - \alpha^{u+v} - \alpha)$$

et par conséquent,

$$F_u^{-1}(v) = [\log\{v\alpha + (1 - v)\alpha^u\} - \log\{v + (1 - v)\alpha^u\}] / \log \alpha.$$

Comme la copule d'indépendance est obtenue lorsque $\alpha = 1$, nous avons choisi de reparamétriser ce modèle en posant $\theta = \log(\alpha) \in (-\infty, \infty)$. Nous avons donc généré des observations de cette loi par l'entremise de

$$F_u^{-1}(v) = [\log\{v \exp(\theta) + (1 - v) \exp(u\theta)\} - \log\{v + (1 - v) \exp(u\theta)\}] / \theta$$

pour les mêmes valeurs de θ que ci-haut.

4⁰) Modèle de Ali-Mikhail-Haq

Pour ce modèle,

$$C(u, v) = \frac{uv}{1 - \theta(1 - u)(1 - v)}, \quad 0 < u, v < 1$$

et $-1 \leq \theta \leq 1$. On trouve

$$F_u(t) = \frac{v - \theta v(1 - v)}{\{1 - \theta(1 - u)(1 - v)\}^2}$$

et son inverse peut s'exprimer algébriquement en résolvant une équation du deuxième degré dont le lecteur pourra extraire la solution exacte, assez longue, du programme `modeles.c` qui se trouve à l'annexe D.

5⁰) Modèle de Farlie-Gumbel-Morgenstern

La fonction de répartition de cette copule non-archimédienne est donnée par

$$H(u, v) = uv + \theta uv(1 - u)(1 - v), \quad 0 < u, v < 1.$$

Le paramètre θ prend ici ses valeurs dans l'intervalle $[-1, 1]$. On voit facilement que

$$F_u(t) = P(V \leq t | U = u) = \frac{\partial}{\partial u} C(u, t) = t + \theta t(1 - t)(1 - 2u)$$

et de là,

$$F_u^{-1}(v) = \frac{1 + w - \sqrt{(1 + w)^2 - 4vw}}{2w}$$

où $w = \theta(1 - 2u)$.

Les résultats des simulations effectuées à l'aide de ces cinq modèles sont résumés dans les tableaux 11 à 13. Comme précédemment, nous remarquons d'abord que toutes les statistiques conservent leur seuil de 5% et que le taux de rejet s'accroît à mesure que θ augmente. Le même phénomène s'observe en outre lorsque la taille des séries croît.

Pour les modèles de Ali-Mikhail-Haq et de Farlie-Gumbel-Morgenstern, dans lesquels la dépendance entre les observations successives est très faible, aucun des trois tests considérés ne fait bonne figure lorsque $n = 20$. Bien que les résultats soient meilleurs pour des tailles de 50 et de 100, ni l'une ni l'autre des trois statistiques ne détecte facilement la contre-hypothèse.

Pour les modèles de dépendance plus forte de Clayton et de Frank, la statistique de Cramér-von Mises obtient la meilleure performance, suivie de près par le tau de Kendall. Pour des séries de taille 100, les trois statistiques performant très bien, rejetant l'hypothèse d'indépendance à plus de 90% lorsque $\theta = 1/4$.

Finalement, dans le cas du modèle ARCH, dont la structure de dépendance est de nature différente de celle des autres modèles considérés, la statistique de Kolmogorov-Smirnov, surclassée par les deux autres statistiques dans tous les autres cas, domine ces dernières lorsque la taille de la série est grande. De manière générale, les statistiques de Cramér-von Mises et de Kolmogorov-Smirnov se comportent plutôt bien dans cette situation, alors que le tau de Kendall fait piètre figure.

TABLEAU 11: Pourcentage de rejet au seuil de 5% pour une chronique de longueur 20, tel qu'estimé à partir de 10 000 répétitions.

Loi des erreurs	statistique	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
<i>Arch</i>	K-Smirnov	10.55	6.98	5.93	5.09	4.87
	C-V-M	11.01	7.59	6.97	6.31	5.39
	Kendall's tau	9.30	7.14	6.27	5.71	4.82
<i>Clayton</i>	K-Smirnov	10.44	4.91	4.49	4.34	4.66
	C-V-M	15.58	6.94	5.57	5.63	4.98
	Kendall's tau	14.99	6.24	4.92	5.02	4.32
<i>Frank</i>	K-Smirnov	29.52	11.28	6.37	5.57	4.29
	C-V-M	35.27	12.01	6.50	5.57	4.93
	Kendall's tau	34.39	10.98	5.73	5.14	4.65
<i>Ali-Mikhail-Haq</i>	K-Smirnov	4.51	3.98	4.37	4.43	4.60
	C-V-M	6.53	4.91	5.18	5.16	5.05
	Kendall's tau	6.37	4.76	4.36	4.77	4.68
<i>Farlie-G.-M.</i>	K-Smirnov	4.18	4.01	4.54	4.48	4.58
	C-V-M	6.12	5.37	5.27	4.95	5.18
	Kendall's tau	5.97	5.13	4.90	4.41	4.55

TABLEAU 12: Pourcentage de rejet au seuil de 5% pour une chronique de longueur 50, tel qu'estimé à partir de 10 000 répétitions.

Loi des erreurs	statistique	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
<i>Arch</i>	K-Smirnov	19.05	10.45	7.56	6.82	5.69
	C-V-M	15.75	9.01	6.94	6.35	5.33
	Kendall's tau	10.68	7.81	6.64	5.72	4.68
<i>Clayton</i>	K-Smirnov	30.91	10.67	6.39	6.09	5.93
	C-V-M	34.30	11.03	6.39	5.72	5.26
	Kendall's tau	31.73	9.75	5.85	5.07	4.51
<i>Frank</i>	K-Smirnov	66.07	23.61	10.47	6.93	5.75
	C-V-M	75.78	27.52	10.89	6.11	5.39
	Kendall's tau	76.57	27.84	10.39	5.66	5.00
<i>Ali-Mikhail-Haq</i>	K-Smirnov	7.54	5.43	5.48	5.35	5.48
	C-V-M	9.16	5.85	5.39	5.22	5.28
	Kendall's tau	8.58	5.47	5.09	4.65	4.76
<i>Farlie-G.-M.</i>	K-Smirnov	6.69	5.81	5.35	5.66	5.96
	C-V-M	8.60	5.81	5.30	5.66	5.47
	Kendall's tau	8.04	5.29	4.75	5.35	4.93

TABLEAU 13: Pourcentage de rejet au seuil de 5% pour une chronique de longueur 100, tel qu'estimé à partir de 10 000 répétitions.

Loi des erreurs	statistique	$\theta = 1/4$	$\theta = 1/8$	$\theta = 1/16$	$\theta = 1/32$	$\theta = 0$
<i>Arch</i>	K-Smirnov	22.70	12.18	8.08	6.88	5.17
	C-V-M	20.78	11.48	7.75	6.35	5.39
	Kendall's tau	11.15	9.17	7.03	5.84	5.05
<i>Clayton</i>	K-Smirnov	51.75	16.35	7.35	5.51	5.13
	C-V-M	57.45	19.40	8.21	6.39	5.23
	Kendall's tau	53.64	17.55	7.29	5.66	4.76
<i>Frank</i>	K-Smirnov	92.27	41.10	15.13	8.00	5.58
	C-V-M	97.05	50.64	17.66	8.39	5.97
	Kendall's tau	97.31	51.43	17.54	7.89	5.25
<i>Ali-Mikhail-Haq</i>	K-Smirnov	10.33	5.71	5.33	5.01	5.71
	C-V-M	14.35	6.98	5.77	5.58	6.04
	Kendall's tau	13.66	6.51	5.23	5.00	5.34
<i>Farlie-G.-M.</i>	K-Smirnov	8.99	5.78	5.13	5.36	5.12
	C-V-M	12.96	6.89	5.52	5.95	5.34
	Kendall's tau	12.97	6.25	5.38	5.35	4.88

Chapitre 7

Conclusion

Les simulations rapportées au chapitre 6 ont fait ressortir que dans le cas des modèles autorégressifs, dans lesquels la dépendance entre les observations successives est linéaire, la version sérielle du tau de Kendall détecte plus facilement le manque d'indépendance que les statistiques de Kolmogorov-Smirnov et de Cramér-Von Mises.

Cependant, des situations ont également été identifiées dans lesquelles les deux statistiques construites à partir du processus de Kendall s'avèrent bien meilleures. Il s'agissait en l'occurrence de modèles où la dépendance sérielle est non-linéaire. Pour les chroniques construites à partir de copules, la statistique de Cramér-von Mises semble être un bon choix. Son emploi peut également être recommandé pour les circonstances où l'on soupçonne que la dépendance sérielle est linéaire, puisque cette statistique s'est avérée presque aussi efficace que le tau de Kendall dans ces cas-là. En outre, au

moins un cas de figure a pu être trouvé où la statistique de Kolmogorov-Smirnov domine les deux autres, mais seule une analyse plus approfondie du comportement de ces trois statistiques permettrait de déterminer pour quel type de dépendance elles sont optimales.

Dans des travaux futurs, il pourrait également être intéressant de considérer des statistiques déduites de processus généralisant d'autres statistiques linéaires de rang. On pense par exemple au rho de Spearman, pour lequel Ghoudi & Rémillard (1998) ont déjà montré (dans le cas non-sériel) que les pseudo-observations seraient de la forme $e_{i,n} = H_n(Y_i)H_n(Y_{i+k})$. Dans le cas sériel, le comportement asymptotique de ce "processus de Spearman" pourrait vraisemblablement s'obtenir par une application des résultats de ces auteurs.

Finalement, plutôt que de se borner à l'étude de la convergence du processus sous l'hypothèse d'indépendance, on pourrait s'intéresser au comportement limite du processus sous diverses structures de dépendance, dont celles où la relation entre deux observations successives de la série est gouvernée par une copule archimédienne. Cette famille de lois de dépendance a été utilisée, entre autres, par Genest & Rivest (1993) pour caractériser la dépendance bivariée dans un contexte non-sériel. Une extension au contexte sériel semble possible, à condition toutefois d'émettre des hypothèses supplémentaires, notamment concernant la densité de $K(t)$.

Annexe A

Démonstration du lemme 4.1

Pour obtenir le résultat, nous allons montrer que :

1. les suites \mathbb{H}_n et $\alpha_{n,\psi}$ sont tendues, ce qui entraîne que la suite $(\alpha_{n,\psi}, \mathbb{H}_n)$ l'est également ;
2. pour tout $m \geq 1$ et toutes valeurs t_1, \dots, t_m et x_1, \dots, x_m , la suite de variables $\{\alpha_n(t_1), \mathbb{H}_n(x_1), \dots, \alpha_n(t_m), \mathbb{H}_n(x_m)\}$ converge vers la limite $\{\alpha(t_1), \mathbb{H}(x_1), \dots, \alpha(t_m), \mathbb{H}(x_m)\}$.

Lemme A.1 *Les suites de processus $\alpha_{n,\psi}$ et \mathbb{H}_n sont respectivement tendues dans les espaces $D[0, 1]$ et $D[0, 1]^2$ des processus càdlàg munis de la topologie de Skorohod.*

Démonstration. Notons d'abord que les observations X_i ne sont pas mutuellement indépendantes. À titre d'exemple, $X_1 = (Y_1, Y_{k+1})$ et $X_{k+1} =$

(Y_{k+1}, Y_{2k+1}) ont toutes les deux Y_{k+1} comme composante commune. Cependant, les deux ensembles suivants vont nous permettre de contourner les éventuels problèmes que pourrait engendrer cette dépendance dans la poursuite de notre démonstration.

Soient

$$\mathcal{A}_1 = \{1, \dots, n\} \cap \bigcup_{i=0} \{2ki + 1, \dots, (2i + 1)k\}$$

et

$$\mathcal{A}_2 = \{1, \dots, n\} \cap \bigcup_{i=0} \{(2i + 1)k + 1, \dots, 2(i + 1)k\},$$

pour un délai k donné.

Puisque toutes les observations X_i dont les indices appartiennent à un même ensemble \mathcal{A}_ℓ sont indépendantes, il est alors possible, au moyen du théorème suivant, extrait de Billingsley (1968), de démontrer la tension des suites de sous-processus

$$\alpha_{n,\psi}^{(\ell)}(t) = \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{A}_\ell} \left[\psi(X_i) \mathbb{I}\{\varepsilon_i \leq t\} - E[\psi(X_i) \mathbb{I}\{\varepsilon \leq t\}] \right], \quad \ell = 1, 2.$$

La tension de la suite $\alpha_{n,\psi} = \alpha_{n,\psi}^{(1)} + \alpha_{n,\psi}^{(2)}$ en découlera aussitôt.

Théorème A.1 (Billingsley 1968) *Supposons que $\{X_n(t_1), \dots, X_n(t_k)\}$ converge en loi vers $\{X(t_1), \dots, X(t_k)\}$ et que pour $s \leq t \leq u$ et $n \geq 1$,*

$$\begin{aligned} P\{|X_n(t) - X_n(s)| \geq \lambda, |X_n(u) - X_n(t)| \geq \lambda\} \\ \leq \frac{1}{\lambda^{2\gamma}} \{G(u) - G(s)\}^{2\alpha}, \end{aligned} \quad (\text{A.1})$$

où $\gamma \geq 0$, $\alpha > 1/2$ et G est continue et non-décroissante sur $[0, 1]$. Alors X_n converge en loi vers X .

L'inégalité A.1 est un critère pour la tension d'une suite. En vue d'appliquer ce résultat, fixons $\ell \in \{1, 2\}$ et $s \leq t \leq u$. Au vu de l'inégalité de Chebyshev, on a alors

$$\begin{aligned}
& P \left\{ |\alpha_{n,\psi}^{(\ell)}(u) - \alpha_{n,\psi}^{(\ell)}(t)| \geq \lambda, \quad |\alpha_{n,\psi}^{(\ell)}(t) - \alpha_{n,\psi}^{(\ell)}(s)| \geq \lambda \right\} \\
& \leq P \left[\left\{ \alpha_{n,\psi}^{(\ell)}(u) - \alpha_{n,\psi}^{(\ell)}(t) \right\}^2 \left\{ \alpha_{n,\psi}^{(\ell)}(t) - \alpha_{n,\psi}^{(\ell)}(s) \right\}^2 \geq \lambda^4 \right] \\
& \leq \frac{1}{\lambda^4} E \left[\left\{ \alpha_{n,\psi}^{(\ell)}(u) - \alpha_{n,\psi}^{(\ell)}(t) \right\}^2 \left\{ \alpha_{n,\psi}^{(\ell)}(t) - \alpha_{n,\psi}^{(\ell)}(s) \right\}^2 \right] \\
& = \frac{1}{\lambda^4 n^2} E \left\{ \left(\sum_{i \in \mathcal{A}_\ell} \alpha_i \right)^2 \left(\sum_{i \in \mathcal{A}_\ell} \beta_i \right)^2 \right\}
\end{aligned}$$

où

$$\alpha_i = \psi(X_i) \mathbb{I}\{t \leq \varepsilon_i \leq u\} - E[\psi(X) \mathbb{I}\{t \leq \varepsilon \leq u\}]$$

et

$$\beta_i = \psi(X_i) \mathbb{I}\{s \leq \varepsilon_i \leq t\} - E[\psi(X) \mathbb{I}\{s \leq \varepsilon \leq t\}].$$

Soit maintenant n_ℓ , le nombre d'éléments de l'ensemble \mathcal{A}_ℓ , $\ell = 1, 2$. Il est clair que $n_\ell \leq n$. Puisque les X_i sont indépendants pour $i \in \mathcal{A}_\ell$, on a

$$\begin{aligned}
& E \left\{ \left(\sum_{i \in \mathcal{A}_\ell} \alpha_i \right)^2 \left(\sum_{i \in \mathcal{A}_\ell} \beta_i \right)^2 \right\} \\
& = n_\ell E(\alpha_1^2 \beta_1^2) + n_\ell(n_\ell - 1) E(\alpha_1^2) E(\beta_1^2) + 2n_\ell(n_\ell - 1) E^2(\alpha_1 \beta_1) \\
& \leq n E(\alpha_1^2 \beta_1^2) + n(n - 1) E(\alpha_1^2) E(\beta_1^2) + 2n(n - 1) E^2(\alpha_1 \beta_1).
\end{aligned}$$

Chacun des termes $E(\alpha_1^2\beta_1^2)$, $E(\alpha_1^2)$, $E(\beta_1^2)$ et $E(\alpha_1\beta_1)$ sera traité séparément.

On a d'abord

$$\begin{aligned}
\alpha_1^2\beta_1^2 &= \psi^2(X_1)\mathbb{I}\{t \leq \varepsilon_1 \leq u\}E^2[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}] \\
&\quad + \psi^2(X_1)\mathbb{I}\{s \leq \varepsilon_1 \leq t\}E^2[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}] \\
&\quad + E^2[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}]E^2[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}] \\
&\quad - 2\psi(X_1)\mathbb{I}\{t \leq \varepsilon_1 \leq u\}E[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}]E^2[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}] \\
&\quad - 2\psi(X_1)\mathbb{I}\{s \leq \varepsilon_1 \leq t\}E[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}]E^2[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}]
\end{aligned}$$

et puisque $\psi \leq 1$, il s'ensuit que

$$\begin{aligned}
E(\alpha_1^2\beta_1^2) &\leq \{K(u) - K(t)\}\{K(t) - K(s)\}^2 + \{K(t) - K(s)\}\{K(u) - K(t)\}^2 \\
&\quad + \{K(u) - K(t)\}^2\{K(t) - K(s)\}^2 \\
&\leq 2\{K(u) - K(s)\}^3 + \{K(u) - K(s)\}^4 \\
&\leq 3\{K(u) - K(s)\}^2.
\end{aligned}$$

Ensuite,

$$\begin{aligned}
\alpha_1^2 &= \psi^2(X_1)\mathbb{I}\{t \leq \varepsilon_1 \leq u\} - 2\psi(X_1)\mathbb{I}\{t \leq \varepsilon \leq u\}E[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}] \\
&\quad + E^2[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}],
\end{aligned}$$

ce qui entraîne que

$$E(\alpha_1^2) \leq K(u) - K(t) + \{K(u) - K(t)\}^2 \leq 2\{K(u) - K(s)\}.$$

De la même façon, on trouve que

$$E(\beta_1^2) \leq K(t) - K(s) + \{K(t) - K(s)\}^2 \leq 2\{K(u) - K(s)\},$$

et par suite,

$$E(\alpha_1^2)E(\beta_1^2) \leq 4\{K(u) - K(s)\}^2.$$

Finalement,

$$\begin{aligned} \alpha_1\beta_1 &= -\psi(X_1)\mathbb{I}\{t \leq \varepsilon_1 \leq u\}E[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}] \\ &\quad -\psi(X_1)\mathbb{I}\{s \leq \varepsilon_1 \leq t\}E[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}] \\ &\quad +E[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}]E[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}] \end{aligned}$$

ce qui permet de déduire que

$$\begin{aligned} |E(\alpha_1\beta_1)| &= E[\psi(X)\mathbb{I}\{t \leq \varepsilon \leq u\}]E[\psi(X)\mathbb{I}\{s \leq \varepsilon \leq t\}] \\ &\leq \{K(u) - K(s)\}^2 \end{aligned}$$

d'où

$$E^2(\alpha_1\beta_1) \leq \{K(u) - K(s)\}^4 \leq \{K(u) - K(s)\}^2.$$

En rassemblant le tout, on voit donc que

$$\begin{aligned} &E \left\{ \left(\sum_{X_i \in \mathcal{A}_r} \alpha_i \right)^2 \left(\sum_{X_i \in \mathcal{A}_r} \beta_i \right)^2 \right\} \\ &\leq nE(\alpha_1^2\beta_1^2) + n(n-1)E(\alpha_1^2)E(\beta_1^2) + 2n(n-1)E^2(\alpha_1\beta_1) \\ &\leq 3n\{K(u) - K(s)\}^2 + 4n^2\{K(u) - K(s)\}^2 - 4n\{K(u) - K(s)\}^2 \\ &\quad + 2n^2\{K(u) - K(s)\}^2 - 2n\{K(u) - K(s)\}^2 \\ &= 6n^2\{K(u) - K(s)\}^2 - 3n\{K(u) - K(s)\}^2 \\ &\leq 6n^2\{K(u) - K(s)\}^2. \end{aligned}$$

De là,

$$\begin{aligned}
P \left\{ |\alpha_{n,\psi}^{(r)}(u) - \alpha_{n,\psi}^{(r)}(t)| \geq \lambda, \quad |\alpha_{n,\psi}^{(r)}(t) - \alpha_{n,\psi}^{(r)}(s)| \geq \lambda \right\} \\
\leq \frac{6n^2}{\lambda^4 n^2} \{K(u) - K(s)\}^2 \\
\leq \frac{1}{\lambda^4} \{G(u) - G(s)\}^2,
\end{aligned}$$

où $G(t) = \sqrt{6}K(t)$ est continue. Considérant l'inégalité A.1, il s'ensuit que $\alpha_{n,\psi}^{(\ell)}$ est tendue, pour $\ell = 1, 2$.

Pour montrer la tension de la suite \mathbb{H}_n , nous utilisons à nouveaux les ensembles \mathcal{A}_ℓ , $\ell = 1, 2$, ce qui nous permet d'écrire

$$\begin{aligned}
\mathbb{H}_n(x) &= \sqrt{\frac{n_1}{n_1 + n_2}} \mathbb{H}_{n_1}^{(1)}(x) + \sqrt{\frac{n_2}{n_1 + n_2}} \mathbb{H}_{n_2}^{(2)}(x) \\
&\leq \mathbb{H}_{n_1}^{(1)}(x) + \mathbb{H}_{n_2}^{(2)}(x),
\end{aligned}$$

où

$$\mathbb{H}_{n_\ell}^{(\ell)}(x) = \frac{1}{\sqrt{n}} \sum_{i \in \mathcal{A}_\ell} [\mathbb{H}\{X_i \leq x\} - H(x)], \quad \ell = 1, 2.$$

Comme $\mathbb{H}_{n_\ell}^{(\ell)}(x)$ est composée d'observations indépendantes, un résultat d'Alexander (1987) nous garantit sa convergence et par le fait même sa tension. Puisque \mathbb{H}_n est bornée par la somme de deux suites tendues, elle est donc elle-même tendue. \diamond

Lemme A.2 *La suite $\{\alpha_n(t_1), \dots, \alpha_n(t_m), \mathbb{H}_n(x_1), \dots, \mathbb{H}_n(x_m)\}$ converge vers $\{\alpha(t_1), \dots, \alpha(t_m), \mathbb{H}(x_1), \dots, \mathbb{H}(x_m)\}$.*

Démonstration. Pour alléger la notation, posons

$$\begin{aligned} Y_n^\top &= \{\alpha_n(t_1), \dots, \alpha_n(t_m), \mathbb{H}_n(x_1), \dots, \mathbb{H}_n(x_m)\} \\ Y^\top &= \{\alpha(t_1), \dots, \alpha(t_m), \mathbb{H}(x_1), \dots, \mathbb{H}(x_m)\}. \end{aligned}$$

Nous devons montrer que Y_n converge en loi vers Y . Cela est vrai si, pour tout vecteur $\lambda^\top = (a_1, \dots, a_m, b_1, \dots, b_m)$, $a_i, b_i \in \mathbb{R}$, $\lambda^\top Y_n$ converge en loi vers $\lambda^\top Y$. D'après le théorème central limite sur les suites de variables aléatoires k -dépendantes, $\lambda^\top Y_n$ converge vers une loi $N(0, \sigma^2)$, dont la variance σ^2 est donnée par

$$\sigma^2 = \lim_{n \rightarrow \infty} \text{var}(\lambda^\top Y_n) = \lim_{n \rightarrow \infty} \lambda^\top \cdot E(Y_n Y_n^\top) \cdot \lambda = \lambda^\top \Omega \lambda,$$

où

$$\Omega_{i,j} = \lim_{n \rightarrow \infty} \text{cov}(Y_n^{(i)}, Y_n^{(j)}) = \begin{cases} \Gamma_{\alpha, \alpha}(t_i, t_j), & 1 \leq i, j \leq m, \\ \Gamma_{\alpha, \mathbb{H}}(t_i, x_j), & 1 \leq i \leq m < j \leq 2m \\ \Gamma_{\mathbb{H}, \mathbb{H}}(x_i, x_j), & m < i, j \leq 2m, \end{cases}$$

d'après les calculs effectués en C1, C2 et C3 de l'Annexe C. Or, d'après la définition même du processus (α, \mathbb{H}) , on voit tout de suite que $\text{cov}(Y^{(i)}, Y^{(j)}) = \Omega_{i,j}$, $1 \leq i, j \leq 2m$. Donc $\sigma^2 = \lambda^\top E(Y Y^\top) \lambda$, ce qui prouve que $\lambda^\top Y_n$ converge en loi vers $\lambda^\top Y$. \diamond

Combinant le lemme A.1, en posant $\psi \equiv 1$, et le lemme A.2, on obtient la convergence de la suite de processus (α_n, \mathbb{H}_n) vers (α, \mathbb{H}) , ce qui complète la preuve du lemme 4.1.

Annexe B

Calculs de limites

B.1 Propriétés de la fonction $q(t) = t^{1/2} \log(1/t)$

Les résultats suivants concernant le comportement de la fonction $q(t) = t^{1/2} \log(1/t)$ sont nécessaires pour la vérification de l'hypothèse III, tant dans le cas d -varié que bivarié. Notons que dans tout ce qui suit, $t_n = (\log n)^3/n$.

B.1.1) Preuve que $q(t)$ est positive et croissante.

D'abord, il est clair que, pour $t \in (0, 1)$, la fonction $q(t) = -t^{1/2} \log t$ est positive. De plus,

$$\frac{\partial}{\partial t} q(t) = -\frac{(\log t + 2)}{2\sqrt{t}} > 0$$

pour tout $t < e^{-2}$, donc dans un voisinage positif de zéro.

B.1.2) Preuve que $q(t)/t$ est décroissante.

Par définition, on a

$$q(t)/t = -t^{-1/2} \log t,$$

de sorte que

$$\frac{\partial}{\partial t} \{q(t)/t\} = \frac{\log t - 2}{2t^{3/2}} < 0$$

pour $t < e^2$.

B.1.3) Preuve que $q(2t)/q(t)$ est bornée autour de zéro.

Tout d'abord, $q(2t)/q(t)$ est décroissante pour $t > 0$. En effet,

$$\frac{q(2t)}{q(t)} = \frac{\sqrt{2t} \log(2t)}{\sqrt{t} \log t} = \frac{\sqrt{2} \log 2}{\log t} + \sqrt{2},$$

ce qui fait que $q(2t)/q(t)$ sera bornée, autour de zéro, par

$$\lim_{t \downarrow 0} \frac{q(2t)}{q(t)} = \sqrt{2} \log 2 \lim_{t \downarrow 0} \frac{1}{\log t} + \sqrt{2} = \sqrt{2},$$

ce qui complète le résultat.

B.1.4) Preuve que $\lim_{n \rightarrow \infty} q(t_n)/\sqrt{nt_n} = 0$.

En faisant la transformation $n = e^u$ et en appliquant deux fois la règle de l'Hospital, on voit que

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{q(t_n)}{\sqrt{nt_n}} &= \lim_{u \rightarrow \infty} \frac{q(u^3/e^u)}{e^{u/2} u^3/e^u} \\ &= \lim_{u \rightarrow \infty} \frac{e^{u/2} u^{3/2}}{u^3 e^{u/2}} \log(e^u/u^3) \\ &= \lim_{u \rightarrow \infty} \frac{(u - 3 \log u)}{u^{3/2}} \\ &= \lim_{u \rightarrow \infty} \frac{1 - 3/u}{(3/2)u^{1/2}} \\ &= \lim_{u \rightarrow \infty} \frac{u - 3}{(3/2)u^{3/2}} \end{aligned}$$

$$= \lim_{u \rightarrow \infty} \frac{1}{(9/4)u^{1/2}} = 0.$$

B.2 Propriétés de $K_d(t)$ et de $k_d(t)$

B.2.1) Preuve que $\lim_{n \rightarrow \infty} \sqrt{n}K_d(t_n) = 0$.

Nous allons d'abord montrer le résultat pour le cas bivarié. Rappelons qu'alors, $K(t) = t - t \log t$, où $t \in (0, 1]$.

En faisant la transformation $n = e^u$ et en appliquant quatre fois la règle de l'Hospital, on trouve

$$\begin{aligned} \lim_{n \rightarrow \infty} \sqrt{n}K(t_n) &= \lim_{u \rightarrow \infty} e^{u/2}K(u^3/e^u) \\ &= \lim_{u \rightarrow \infty} \frac{u^3}{e^{u/2}} \{1 - \log(u^3/e^u)\} \\ &= \lim_{u \rightarrow \infty} \frac{u^3}{e^{u/2}} (1 - 3 \log u + u) \\ &= \lim_{u \rightarrow \infty} (u^4 + u^3 - 3u^3 \log u) / e^{u/2} \\ &= \lim_{u \rightarrow \infty} \frac{4u^3 + 3u^2 - 9u^2 \log u - 3u^2}{(1/2)e^{u/2}} \\ &= \lim_{u \rightarrow \infty} \frac{12u^2 - 18u \log u - 9u}{(1/4)e^{u/2}} \\ &= \lim_{u \rightarrow \infty} \frac{24u - 18 \log u - 27}{(1/8)e^{u/2}} \\ &= \lim_{u \rightarrow \infty} \frac{24 - 18/u}{(1/16)e^{u/2}} = 0. \end{aligned}$$

Pour étendre le résultat au cas d -varié, rappelons que

$$K_d(t) = t \sum_{i=1}^d \frac{\{\log(1/t)\}^{i-1}}{(i-1)!},$$

où $t \in (0, 1]$.

Ensuite, en posant $n = e^u$, on a

$$\begin{aligned}
\lim_{n \rightarrow \infty} \sqrt{n} K_d(t_n) &= \lim_{n \rightarrow \infty} \frac{\log^3 n}{\sqrt{n}} \sum_{i=1}^d \frac{1}{(i-1)!} \left\{ \log \left(\frac{n}{\log^3 n} \right) \right\}^{i-1} \\
&= \lim_{u \rightarrow \infty} \frac{u^3}{e^{u/2}} \sum_{i=1}^d \frac{1}{(i-1)!} \left\{ \log \left(\frac{e^u}{u^3} \right) \right\}^{i-1} \\
&= \lim_{u \rightarrow \infty} \frac{u^3}{e^{u/2}} \sum_{i=1}^d \frac{1}{(i-1)!} (u - 3 \log u)^{i-1} \\
&\leq \lim_{u \rightarrow \infty} \frac{u^3}{e^{u/2}} \sum_{i=1}^d \frac{u^{i-1}}{(i-1)!} \\
&\leq \lim_{u \rightarrow \infty} \frac{u^3}{e^{u/2}} \sum_{i=1}^d u^{d-1} \\
&= \lim_{u \rightarrow \infty} d \cdot \frac{u^{d+2}}{e^{u/2}} = 0.
\end{aligned}$$

B.1.6) Preuve que $\lim_{t \downarrow 0} k_d(t)q(t) = 0$.

Pour le cas bivarié, où $k(t) = -\log t$, deux applications successives de la règle de l'Hospital montrent que

$$\begin{aligned}
\lim_{t \downarrow 0} k(t)q(t) &= \lim_{t \downarrow 0} \sqrt{t} \log^2(t) \\
&= \lim_{t \downarrow 0} \frac{-4 \log t}{t^{-1/2}} \\
&= \lim_{t \downarrow 0} \frac{-4/t}{(-1/2)t^{-3/2}} \\
&= \lim_{t \downarrow 0} \frac{8}{t^{-1/2}} \\
&= \lim_{t \downarrow 0} 8\sqrt{t} = 0.
\end{aligned}$$

Finalement, dans le cas d -varié, où

$$k_d(t) = \frac{1}{(d-1)!} \{\log(1/t)\}^{d-1},$$

des applications successives de la règle de l'Hospital (d au total) montrent que

$$\begin{aligned} \lim_{t \downarrow 0} k_d(t)q(t) &= \lim_{t \downarrow 0} \frac{1}{(d-1)!} \{\log(1/t)\}^{d-1} t^{1/2} \log(1/t) \\ &= \lim_{t \downarrow 0} \frac{1}{(d-1)!} \frac{\{\log(1/t)\}^d}{t^{-1/2}} \\ &= \frac{d}{(d-1)!} \lim_{t \downarrow 0} \frac{\{\log(1/t)\}^{d-1}}{(1/2)t^{-1/2}} \\ &= \dots \\ &= \lim_{t \downarrow 0} \frac{d \cdot \{\log(1/t)\}^0}{(1/2^d)t^{-1/2}} = \lim_{t \downarrow 0} d \cdot 2^d \sqrt{t} = 0. \end{aligned}$$

Annexe C

Calculs de covariances pour le cas bivarié

Nous présentons ici le calcul des covariances limites suivantes, qui sont nécessaires à la caractérisation de la loi des processus gaussiens limites α , \mathbb{H} et \mathbb{K} :

$$\begin{aligned}\Gamma_{\alpha,\alpha}(s,t) &= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \alpha_n(t)\}, \\ \Gamma_{\alpha,\mathbb{H}}(s,x) &= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \mathbb{H}_n(x)\}, \\ \Gamma_{\alpha,\mu}(s,t) &= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \mu(t, \mathbb{H}_n)\}, \\ \Gamma_{\mathbb{H},\mathbb{H}}(x,y) &= \lim_{n \rightarrow \infty} \text{cov}\{\mathbb{H}_n(x), \mathbb{H}_n(y)\}, \\ \Gamma_{\mu,\mu}(s,t) &= \lim_{n \rightarrow \infty} \text{cov}\{\mu(s, \mathbb{H}_n), \mu(t, \mathbb{H}_n)\}.\end{aligned}$$

Notons d'abord que la convergence de telles suites de covariances découle directement de la proposition suivante, qui est une extension d'un résultat

de Billingsley (lemme 3, p. 172, 1968).

Proposition C.1 *Soient ξ_k et ζ_k , deux suites stationnaires et ergodiques d'espérance nulle et soient*

$$S_n = \sum_{i=1}^n \xi_i, \quad T_n = \sum_{i=1}^n \zeta_i.$$

Si les séries $\sum_{i=1}^{\infty} E(\xi_i \zeta_0)$ et $\sum_{i=1}^{\infty} E(\xi_0 \zeta_i)$ convergent absolument, alors

$$\lim_{n \rightarrow \infty} \text{cov}(S_n, T_n)/n = E(\xi_0 \zeta_0) + \sum_{i=1}^{\infty} E(\xi_0 \zeta_i) + \sum_{i=1}^{\infty} E(\xi_i \zeta_0).$$

Démonstration. Par définition, on a

$$\begin{aligned} \frac{1}{n} \text{cov}(S_n, T_n) &= \frac{1}{n} \text{cov} \left(\sum_{i=1}^n \xi_i, \sum_{i=1}^n \zeta_i \right) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E(\xi_i \zeta_j) \\ &= \frac{1}{n} \sum_{i=1}^n E(\xi_i \zeta_i) + \frac{1}{n} \sum_{i < j} E(\xi_i \zeta_j) + \frac{1}{n} \sum_{j < i} E(\xi_i \zeta_j) \\ &= E(\xi_0 \zeta_0) + \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n E(\xi_0 \zeta_{j-i}) + \frac{1}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^n E(\xi_{i-j} \zeta_0) \\ &= E(\xi_0 \zeta_0) + \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} E(\xi_0 \zeta_j) + \frac{1}{n} \sum_{j=1}^{n-1} \sum_{i=1}^{n-j} E(\xi_i \zeta_0). \end{aligned}$$

Si on pose $a_i = \sum_{j=1}^i E(\xi_0 \zeta_j)$ et $a_i^* = \sum_{j=1}^i E(\xi_j \zeta_0)$, on voit alors que

$$\frac{1}{n} \text{cov}(S_n, T_n) = E(\xi_0 \zeta_0) + \frac{1}{n} \sum_{i=1}^{n-1} a_i + \frac{1}{n} \sum_{i=1}^{n-1} a_i^*.$$

Par conséquent,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{cov}(S_n, T_n) = E(\xi_0 \zeta_0) + \sum_{i=1}^{\infty} E(\xi_0 \zeta_i) + \sum_{i=1}^{\infty} E(\xi_i \zeta_0),$$

puisque $(a_1 + \dots + a_{n-1})/n$ converge aussi vers a dès lors que a_n converge vers a . \diamond

C1: Calcul de $\Gamma_{\alpha, \alpha}(s, t) = \lim_{n \rightarrow \infty} \operatorname{cov}\{\alpha_n(s), \alpha_n(t)\}$.

Posons $\xi_i = \mathbb{I}\{\varepsilon_i \leq s\} - K(s)$ et $\zeta_i = \mathbb{I}\{\varepsilon_i \leq t\} - K(t)$. D'après la proposition C.1, on a

$$\begin{aligned} \Gamma_{\alpha, \alpha}(s, t) &= E \left[\left\{ \mathbb{I}\{\varepsilon_0 \leq s\} - K(s) \right\} \left\{ \mathbb{I}\{\varepsilon_0 \leq t\} - K(t) \right\} \right] \\ &\quad + \sum_{i=1}^{\infty} E \left[\left\{ \mathbb{I}\{\varepsilon_0 \leq s\} - K(s) \right\} \left\{ \mathbb{I}\{\varepsilon_i \leq t\} - K(t) \right\} \right] \\ &\quad + \sum_{i=1}^{\infty} E \left[\left\{ \mathbb{I}\{\varepsilon_i \leq s\} - K(s) \right\} \left\{ \mathbb{I}\{\varepsilon_0 \leq t\} - K(t) \right\} \right] \\ &= P(\varepsilon_0 \leq s, \varepsilon_0 \leq t) - K(s)K(t) \\ &\quad + 2 \sum_{i=1}^{\infty} \{P(\varepsilon_0 \leq s, \varepsilon_i \leq t) - K(s)K(t)\}. \end{aligned}$$

Puisque $P(\varepsilon_0 \leq s, \varepsilon_i \leq t) = K(s)K(t)$ sauf lorsque i est égal au délai k , on voit que

$$\begin{aligned} \Gamma_{\alpha, \alpha}(s, t) &= P(\varepsilon_0 \leq s \wedge t) + 2P(\varepsilon_0 \leq s, \varepsilon_k \leq t) - 3K(s)K(t) \\ &= K(s \wedge t) + 2P(Y_0 Y_k \leq s, Y_k Y_{2k} \leq t) - 3K(s)K(t), \end{aligned}$$

où

$$P(Y_0 Y_k \leq s, Y_k Y_{2k} \leq t) = P\left(Y_0 \leq \frac{s}{Y_k}, Y_{2k} \leq \frac{t}{Y_k}\right)$$

$$\begin{aligned}
&= \int_0^1 P\left(Y_0 \leq \frac{s}{w}\right) P\left(Y_{2k} \leq \frac{t}{w}\right) dw \\
&= \int_0^{s \wedge t} dw + \int_{s \wedge t}^{s \vee t} \frac{(s \wedge t)}{w} dw + \int_{s \vee t}^1 \frac{st}{w^2} dw \\
&= (s \wedge t) + (s \wedge t) \log\left(\frac{s \vee t}{s \wedge t}\right) + st \left(\frac{1}{s \wedge t} - 1\right) \\
&= 2(s \wedge t) + (s \wedge t) \log\left(\frac{s \vee t}{s \wedge t}\right) - st.
\end{aligned}$$

Finalement,

$$\begin{aligned}
\Gamma_{\alpha,\alpha}(s,t) &= K(s \wedge t) + 4(s \wedge t) + 2(s \wedge t) \log\left(\frac{s \vee t}{s \wedge t}\right) - 2st - 3K(s)K(t) \\
&= 5\{(s \wedge t) - st\} - 3(s \wedge t) \log(s \wedge t) + 2(s \wedge t) \log(s \vee t) \\
&\quad + 3st\{\log(st) - \log s \log t\}.
\end{aligned}$$

C2: Calcul de $\Gamma_{\alpha,\mu}(s,t) = \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \mu(t, \mathbb{H}_n)\}$.

D'après la définition de μ et en utilisant le théorème de la convergence dominée, on a

$$\begin{aligned}
\Gamma_{\alpha,\mu}(s,t) &= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \mu(t, \mathbb{H}_n)\} \\
&= \lim_{n \rightarrow \infty} \int_t^1 \frac{E\left\{\alpha_n(s) \mathbb{H}_n(u, t/u)\right\}}{u} du \\
&= \int_t^1 \lim_{n \rightarrow \infty} \frac{E\left\{\alpha_n(s) \mathbb{H}_n(u, t/u)\right\}}{u} du \\
&= \int_t^1 \lim_{n \rightarrow \infty} \frac{\text{cov}\left\{\alpha_n(s), \mathbb{H}_n(u, t/u)\right\}}{u} du.
\end{aligned}$$

Posons $\Gamma_{\alpha,\mathbb{H}}(s,u,v) = \lim_{n \rightarrow \infty} \text{cov}\{\alpha_n(s), \mathbb{H}_n(u,v)\}$. Si on prend $\xi_i = \mathbb{I}\{\varepsilon_i \leq s\} - K(s)$ et $\zeta_i = \mathbb{I}_{[0,u] \times [0,v]}(X_i) - H(u,v)$, une nouvelle application de la

proposition C.1 permet d'écrire

$$\begin{aligned}
\Gamma_{\alpha, \mathbb{H}}(s, u, v) &= E \left[\left\{ \mathbb{I}_{[0,s]}(\varepsilon_0) - K(s) \right\} \left\{ \mathbb{I}_{[0,u] \times [0,v]}(X_0) - uv \right\} \right] \\
&\quad + \sum_{i=1}^{\infty} E \left[\left\{ \mathbb{I}_{[0,s]}(\varepsilon_0) - K(s) \right\} \right. \\
&\quad \quad \quad \left. \times \left\{ \mathbb{I}_{[0,u] \times [0,v]}(X_i) - uv \right\} \right] \\
&\quad + \sum_{i=1}^{\infty} E \left[\left\{ \mathbb{I}_{[0,s]}(\varepsilon_i) - K(s) \right\} \right. \\
&\quad \quad \quad \left. \times \left\{ \mathbb{I}_{[0,u] \times [0,v]}(X_0) - uv \right\} \right] \\
&= P(Y_0 Y_k \leq s, Y_0 \leq u, Y_k \leq v) - uvK(s) \\
&\quad + \sum_{i=1}^{\infty} \{P(Y_0 Y_k \leq s, Y_i \leq u, Y_{i+k} \leq v) - uvK(s)\} \\
&\quad + \sum_{i=1}^{\infty} \{P(Y_i Y_{i+k} \leq s, Y_0 \leq u, Y_k \leq v) - uvK(s)\}
\end{aligned}$$

Or, les termes des deux sommes sont nuls sauf lorsque $i = k$. Ainsi,

$$\begin{aligned}
\Gamma_{\alpha, \mathbb{H}}(s, u, v) &= P(Y_0 Y_k \leq s, Y_0 \leq u, Y_k \leq v) \\
&\quad + P(Y_0 Y_k \leq s, Y_k \leq u, Y_{2k} \leq v) \\
&\quad + P(Y_k Y_{2k} \leq s, Y_0 \leq u, Y_k \leq v) - 3uvK(s).
\end{aligned}$$

Nous allons calculer ces trois probabilités séparément.

$$\begin{aligned}
1^\circ) P(Y_0 \leq u, Y_k \leq v, Y_0 Y_k \leq s) &= \int_0^1 P(Y_0 \leq u, w \leq v, Y_0 \leq s/w) dw \\
&= \int_0^v P\{Y_0 \leq \min(u, s/w)\} dw \\
&= \begin{cases} \int_0^{s/u} u dw + \int_{s/u}^v s/w dw, & s \leq uv \\ \int_0^v u dw, & s \geq uv \end{cases}
\end{aligned}$$

$$\begin{aligned}
&= \begin{cases} s\{1 + \log(uv/s)\}, & s \leq uv \\ uv, & s \geq uv \end{cases} \\
&= \{s \wedge (uv)\} \left\{ 1 + \log \left(\frac{uv}{s \wedge (uv)} \right) \right\}.
\end{aligned}$$

$$\begin{aligned}
2^o) P(Y_k \leq u, Y_{2k} \leq v, Y_0 Y_k \leq s) &= \int_0^1 P(Y_0 \leq s/w, w \leq u, Y_{2k} \leq v) dw \\
&= \int_0^u P(Y_0 \leq s/w) P(Y_{2k} \leq v) dw \\
&= \begin{cases} v \int_0^s dw + sv \int_s^u dw/w, & s \leq u \\ v \int_0^u dw, & s \geq u \end{cases} \\
&= \begin{cases} (sv)\{1 + \log(u/s)\}, & s \leq u \\ uv, & s \geq u \end{cases} \\
&= v(s \wedge u) \left\{ 1 + \log \left(\frac{u}{s \wedge u} \right) \right\}.
\end{aligned}$$

3°) En interchangeant les rôles de u et v ainsi que ceux de Y_0 et de Y_{2k} dans le calcul ci-haut, on trouve immédiatement

$$P(Y_k \leq v, Y_0 \leq u, Y_k Y_{2k} \leq s) = u(s \wedge v) \left\{ 1 + \log \left(\frac{v}{s \wedge v} \right) \right\}.$$

Ainsi,

$$\begin{aligned}
\Gamma_{\alpha, \mathbb{H}}(s, u, v) &= \{s \wedge (uv)\} \left\{ 1 + \log \left(\frac{uv}{s \wedge (uv)} \right) \right\} \\
&\quad + v(s \wedge u) \left\{ 1 + \log \left(\frac{u}{s \wedge u} \right) \right\} \\
&\quad + u(s \wedge v) \left\{ 1 + \log \left(\frac{v}{s \wedge v} \right) \right\} - 3uvK(s).
\end{aligned}$$

De là,

$$\Gamma_{\alpha, \mu}(s, t) = \int_t^1 \frac{\Gamma_{\alpha, \mathbb{H}}(s, u, t/u)}{u} du$$

$$\begin{aligned}
&= \int_t^1 \frac{(s \wedge t)}{u} \left\{ 1 + \log \left(\frac{t}{s \wedge t} \right) \right\} du \\
&\quad + \int_t^1 \frac{(s \wedge u)t}{u^2} \left\{ 1 + \log \left(\frac{u}{s \wedge u} \right) \right\} du \\
&\quad + \int_t^1 \left(s \wedge \frac{t}{u} \right) \left\{ 1 + \log \left(\frac{t/u}{t/u \wedge s} \right) \right\} du \\
&\quad - 3tK(s) \int_t^1 du/u \\
&= - (s \wedge t) \log t \left\{ 1 + \log \left(\frac{t}{s \wedge t} \right) \right\} \\
&\quad + 2(s \wedge t) - 2st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) + st \log s \\
&\quad + \left\{ (s \wedge t) - st \right\} \left\{ 1 + \log \left(\frac{t}{s} \right) \right\} + (s \wedge t) + st(\log t - 1) \\
&\quad + 2t \log \left(\frac{s \vee t}{t} \right) + 3st(1 - \log s) \log t.
\end{aligned}$$

De la même manière, on trouve aussi

$$\begin{aligned}
\Gamma_{\alpha, \mu}(t, s) &= - (s \wedge t) \log s \left\{ 1 + \log \left(\frac{s}{s \wedge t} \right) \right\} \\
&\quad + 2(s \wedge t) - 2st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) + st \log t \\
&\quad + \left\{ (s \wedge t) - st \right\} \left\{ 1 + \log \left(\frac{s}{t} \right) \right\} + (s \wedge t) + st(\log s - 1) \\
&\quad + 2s \log \left(\frac{s \vee t}{s} \right) + 3st \log s(1 - \log t).
\end{aligned}$$

C3: Calcul de $\Gamma_{\mu, \mu}(s, t) = \lim_{n \rightarrow \infty} \text{cov}\{\mu(s, \mathbb{H}_n), \mu(t, \mathbb{H}_n)\}$.

D'après la définition de μ et par le théorème de la convergence dominée, on obtient

$$\Gamma_{\mu, \mu}(s, t) = \lim_{n \rightarrow \infty} \text{cov}\{\mu(s, \mathbb{H}_n), \mu(t, \mathbb{H}_n)\}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} E \left\{ \int_s^1 \frac{\mathbb{H}_n(u, s/u)}{u} du \int_t^1 \frac{\mathbb{H}_n(v, t/v)}{v} dv \right\} \\
&= \int_s^1 \int_t^1 \lim_{n \rightarrow \infty} \frac{E\{\mathbb{H}_n(u, s/u) \cdot \mathbb{H}_n(v, t/v)\}}{uv} du dv \\
&= \int_s^1 \int_t^1 \lim_{n \rightarrow \infty} \frac{\text{cov}\{\mathbb{H}_n(u, s/u), \mathbb{H}_n(v, t/v)\}}{uv} du dv.
\end{aligned}$$

Posons $\Gamma_{\mathbb{H}, \mathbb{H}}(s, t, u, v) = \lim_{n \rightarrow \infty} \text{cov}\{\mathbb{H}_n(s, t), \mathbb{H}_n(u, v)\}$. En utilisant la proposition C.1 de la même manière que précédemment, on trouve

$$\begin{aligned}
\Gamma_{\mathbb{H}, \mathbb{H}}(s, t, u, v) &= P(Y_0 \leq s \wedge u, Y_k \leq t \wedge v) - stuv \\
&\quad + P(Y_0 \leq s, Y_k \leq t \wedge u, Y_{2k} \leq v) - stuv \\
&\quad + P(Y_0 \leq t, Y_k \leq s \wedge v, Y_{2k} \leq u) - stuv \\
&= (s \wedge u)(t \wedge v) + sv(t \wedge u) + tu(s \wedge v) - 3stuv.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
\Gamma_{\mu, \mu}(s, t) &= \int_s^1 \int_t^1 \frac{\Gamma_{\mathbb{H}, \mathbb{H}}(u, s/u, v, t/v)}{uv} du dv \\
&= \int_t^1 \int_s^1 \frac{(v \wedge u)}{uv} \left(\frac{s}{u} \wedge \frac{t}{v} \right) dudv \\
&\quad + \int_t^1 \int_s^1 \frac{t}{v^2} \left(\frac{s}{u} \wedge v \right) dudv \\
&\quad + \int_t^1 \int_s^1 \frac{s}{u^2} \left(u \wedge \frac{t}{v} \right) dudv \\
&\quad - 3 \int_t^1 \int_s^1 \frac{st}{uv} dudv \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

Nous calculerons ces quatre intégrales doubles séparément.

1⁰) Supposons que $s \leq t$. On a alors

$$\begin{aligned}
I_1 &= \int_t^1 \int_s^1 \frac{(v \wedge u)}{uv} \left(\frac{s}{u} \wedge \frac{t}{v} \right) dudv \\
&= \int_t^1 \int_s^{sv/t} \frac{t}{v^2} dudv + \int_t^1 \int_{sv/t}^v \frac{s}{uv} dudv \\
&\quad + \int_t^1 \int_v^1 \frac{s}{u^2} dudv \\
&= \int_t^1 \left(\frac{s}{v} - \frac{st}{v^2} + \frac{s \log t - s \log s}{v} + \frac{s}{v} - s \right) dv \\
&= -s \log t - s(1-t) + (s \log s - s \log t) \log t \\
&\quad - s \log t - s(1-t) \\
&= -2s(1-t) - s \log t \left\{ 2 + \log \left(\frac{t}{s} \right) \right\}.
\end{aligned}$$

Supposons maintenant que $s \geq t$. Dans ce cas,

$$\begin{aligned}
I_1 &= \int_t^1 \int_s^1 \frac{(v \wedge u)}{uv} \left(\frac{s}{u} \wedge \frac{t}{v} \right) dudv \\
&= \int_s^1 \int_t^1 \frac{(v \wedge u)}{uv} \left(\frac{s}{u} \wedge \frac{t}{v} \right) dvdu \\
&= -2t(1-s) - t \log s \left\{ 2 + \log \left(\frac{s}{t} \right) \right\},
\end{aligned}$$

où l'on a utilisé le résultat précédent.

En combinant ces deux résultats, on voit que

$$\begin{aligned}
I_1 &= \int_t^1 \int_s^1 \frac{(v \wedge u)}{uv} \left(\frac{s}{u} \wedge \frac{t}{v} \right) dudv \\
&= -2\{(s \wedge t) - st\} \\
&\quad - (s \wedge t) \log(s \vee t) \left\{ 2 + \log \left(\frac{s \vee t}{s \wedge t} \right) \right\}.
\end{aligned}$$

2⁰) Supposons que $s \leq t$. On a alors

$$\begin{aligned}
I_2 &= \int_t^1 \int_s^1 \frac{t}{v^2} \left(\frac{s}{u} \wedge v \right) dudv \\
&= \int_s^1 \int_t^1 \frac{t}{v^2} \left(\frac{s}{u} \wedge v \right) dvdu \\
&= \int_s^{s/t} \int_t^{s/u} \frac{t}{v} dvdu + \int_s^{s/t} \int_{s/u}^1 \frac{st}{uv^2} dvdu + \int_{s/t}^1 \int_t^1 \frac{st}{uv^2} dvdu \\
&= \int_s^{s/t} t \log \left(\frac{s}{ut} \right) du + \int_s^{s/t} \left(t - \frac{st}{u} \right) du + \int_{s/t}^1 \frac{s(1-t)}{u} du \\
&= \{st \log t - st + s\} + \{s(1-t) + st \log t\} \\
&\quad + \left\{ s(1-t) \log \left(\frac{t}{s} \right) \right\} \\
&= 2s(1-t) + s(t+1) \log t - s(1-t) \log s.
\end{aligned}$$

De la même façon, si $s \geq t$, on a

$$\begin{aligned}
I_2 &= \int_t^1 \int_s^1 \frac{t}{v^2} \left(\frac{s}{u} \wedge v \right) dudv \\
&= \int_s^1 \int_t^1 \frac{t}{v^2} \left(\frac{s}{u} \wedge v \right) dvdu \\
&= \int_s^1 \int_t^{s/u} \frac{t}{v} dvdu + \int_s^1 \int_{s/u}^1 \frac{st}{uv^2} dvdu \\
&= \int_s^1 t \log \left(\frac{s}{ut} \right) du + \int_s^1 t \left(1 - \frac{s}{u} \right) du \\
&= t(1-s) + t \log s - t(1-s) \log t \\
&\quad + t(1-s) + st \log s \\
&= 2t(1-s) + t(s+1) \log s - t(1-s) \log t.
\end{aligned}$$

Combinant les derniers résultats, on peut écrire, pour tout $0 \leq s, t \leq 1$,

$$I_2 = \int_t^1 \int_s^1 \frac{t}{v^2} \left(\frac{s}{u} \wedge v \right) dudv$$

$$\begin{aligned}
&= 2\{(s \wedge t) - st\} + \{(s \wedge t) + st\} \log(s \vee t) \\
&\quad - \{(s \wedge t) - st\} \log(s \wedge t).
\end{aligned}$$

3⁰) En utilisant le résultat précédent, nous avons, pour $0 \leq s, t \leq 1$,

$$\begin{aligned}
I_3 &= \int_t^1 \int_s^1 \frac{s}{u^2} \left(u \wedge \frac{t}{v} \right) dudv \\
&= \int_s^1 \int_t^1 \frac{s}{v^2} \left(\frac{t}{u} \wedge v \right) dudv \\
&= 2\{(s \wedge t) - st\} + \{(s \wedge t) + st\} \log(s \vee t) \\
&\quad - \{(s \wedge t) - st\} \log(s \wedge t).
\end{aligned}$$

4⁰) Finalement,

$$I_4 = -3 \int_t^1 \int_s^1 \frac{st}{uv} dudv = -3st\{\log(1/s)\}\{\log(1/t)\} = -3st \log s \log t.$$

En additionnant les expressions I_1 , I_2 , I_3 et I_4 , on obtient

$$\begin{aligned}
\Gamma_{\mu,\mu}(s, t) &= 2\{(s \wedge t) - st\}\{1 - \log(s \wedge t)\} + 2st \log(s \vee t) \\
&\quad - (s \wedge t)\{\log(s \vee t)\}^2 + (s \wedge t) \log s \log t - 3st \log s \log t.
\end{aligned}$$

C4: Covariance limite du processus de Kendall sériel.

Rappelons que $\mathbb{K}(t) = \alpha(t) - \mu(t, \mathbb{H})$. Par conséquent, d'après le lemme 4.1 et les calculs effectués en (C1), (C2) et (C3), on peut écrire

$$\begin{aligned}
\text{cov}\{\mathbb{K}(s), \mathbb{K}(t)\} &= \text{cov}\{\alpha(s), \alpha(t)\} - \text{cov}\{\alpha(s), \mu(t, \mathbb{H})\} \\
&\quad - \text{cov}\{\alpha(t), \mu(s, \mathbb{H})\} + \text{cov}\{\mu(s, \mathbb{H}), \mu(t, \mathbb{H})\} \\
&= \Gamma_{\alpha,\alpha}(s, t) - \Gamma_{\alpha,\mu}(s, t) - \Gamma_{\alpha,\mu}(t, s) + \Gamma_{\mu,\mu}(s, t)
\end{aligned}$$

$$\begin{aligned}
&= -(s \wedge t) \log t \log(s \wedge t) + 2st \log(s \vee t) + st \\
&\quad -(s \wedge t) \log s \log(s \wedge t) + (s \wedge t) \log s \log t \\
&\quad + 2st \log(s \wedge t) - (s \wedge t) - 2st \log st + (s \wedge t) \log s \\
&\quad + (s \wedge t)(\log t)^2 - 3(s \wedge t) \log(s \wedge t) - (s \wedge t)\{\log(s \vee t)\}^2 \\
&\quad + (s \wedge t)(\log s)^2 + (s \wedge t) \log t - 2t \log(s \vee t) \\
&\quad 2t \log t - 2s \log(s \vee t) + 2s \log s \\
&= st - (s \wedge t)\{1 + \log(s \vee t)\}.
\end{aligned}$$

C5: Preuve que $\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{K}_{n,i}(s), \mathbb{K}_{n,j}(t)\} = 0$.

Par le théorème 4.1, $\mathbb{K}_{n,i}(s)$ converge vers $\mathbb{K}_i(s) = \alpha_i(s) - \mu(s, \mathbb{H}_i)$, où α_i et \mathbb{H}_i sont construites en prenant i comme délai. Ainsi,

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{K}_{n,i}(s), \mathbb{K}_{n,j}(t)\} \\
&= \lim_{n \rightarrow \infty} \text{cov}\{\alpha_{n,i}(s), \alpha_{n,j}(t)\} - \lim_{n \rightarrow \infty} \text{cov}\{\alpha_{n,i}(s), \mu(t, \mathbb{H}_{n,j})\} \\
&\quad - \lim_{n \rightarrow \infty} \text{cov}\{\mu(s, \mathbb{H}_{n,i}), \alpha_{n,j}(t)\} + \lim_{n \rightarrow \infty} \text{cov}\{\mu(s, \mathbb{H}_{n,i}), \mu(t, \mathbb{H}_{n,j})\} \\
&= J_1 - J_2 - J_3 + J_4.
\end{aligned}$$

Dans ce qui suit nous supposons, sans perte de généralité, que $i < j$.

1⁰) En utilisant la proposition C.1, on a

$$\begin{aligned}
J_1 &= P(Y_0 Y_i \leq s, Y_0 Y_j \leq t) - K(s)K(t) \\
&\quad + \sum_{\ell=1}^{\infty} \{P(Y_0 Y_i \leq s, Y_\ell Y_{\ell+j} \leq t) - K(s)K(t)\} \\
&\quad + \sum_{\ell=1}^{\infty} \{P(Y_\ell Y_{\ell+i} \leq s, Y_0 Y_j \leq t) - K(s)K(t)\}
\end{aligned}$$

$$\begin{aligned}
&= P(Y_0Y_i \leq s, Y_0Y_j \leq t) + P(Y_0Y_i \leq s, Y_iY_{i+j} \leq t) \\
&\quad + P(Y_jY_{i+j} \leq s, Y_0Y_j \leq t) + P(Y_{j-i}Y_j \leq s, Y_0Y_j \leq t) \\
&\quad - 4K(s)K(t) \\
&= 4\{P(Y_0Y_1 \leq s, Y_0Y_2 \leq t) - K(s)K(t)\}.
\end{aligned}$$

Ici encore, nous avons utilisé le fait que les termes des deux séries sont tous nuls sauf lorsque $\ell = i$ dans la première série et $\ell = j - i$ ou j dans la seconde.

Or, dans un calcul précédent fait en (C1), nous avons déjà vu que

$$P(Y_0Y_1 \leq s, Y_0Y_2 \leq t) = 2(s \wedge t) - st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right),$$

ce qui fait que

$$\begin{aligned}
J_1 &= 4 \left\{ 2(s \wedge t) - st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) - K(s)K(t) \right\} \\
&= 8(s \wedge t) - 8st + 4(s \wedge t) \log(s \vee t) - 4(s \wedge t) \log(s \wedge t) \\
&\quad + 4st \log st - 4st \log s \log t.
\end{aligned}$$

2⁰)

$$J_2 = \int_t^1 \frac{\lim_{n \rightarrow \infty} \text{cov}\{\alpha_{n,i}(s), \mathbb{H}_{n,j}(u, t/u)\}}{u} du,$$

où

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \text{cov}\{\alpha_{n,i}(s), \mathbb{H}_{n,j}(u, t/u)\} \\
&= P(Y_0Y_i \leq s, Y_0 \leq u, Y_j \leq v) - uvK(s) \\
&\quad + \sum_{\ell=1}^{\infty} \{P(Y_0Y_i \leq s, Y_\ell \leq u, Y_{\ell+j} \leq v) - uvK(s)\}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{\ell=1}^{\infty} \{P(Y_{\ell}Y_{\ell+i} \leq s, Y_0 \leq u, Y_j \leq v) - uvK(s)\} \\
= & P(Y_0Y_i \leq s, Y_0 \leq u, Y_j \leq v) + P(Y_0Y_i \leq s, Y_i \leq u, Y_{i+j} \leq v) \\
& + P(Y_jY_{i+j} \leq s, Y_0 \leq u, Y_j \leq v) + P(Y_{j-i}Y_j \leq s, Y_0 \leq u, Y_j \leq v) \\
& - 5uvK(s).
\end{aligned}$$

Puisque nous avons établi en (C2) que

$$P(Y_1Y_2 \leq s, Y_1 \leq u, Y_3 \leq v) = v(s \wedge u) \left\{ 1 + \log \left(\frac{u}{s \wedge u} \right) \right\},$$

nous voyons que

$$\begin{aligned}
\lim_{n \rightarrow \infty} \text{cov} \{ \alpha_{n,i}(s), \mathbb{H}_{n,j}(u, t/u) \} & = 2v(s \wedge u) \left\{ 1 + \log \left(\frac{u}{s \wedge u} \right) \right\} \\
& + 2u(s \wedge v) \left\{ 1 + \log \left(\frac{v}{s \wedge v} \right) \right\} \\
& - 5uvK(s).
\end{aligned}$$

En se servant des calculs faits en (C2), on trouve donc

$$\begin{aligned}
J_2 & = 2 \int_t^1 \frac{(s \wedge u)t}{u^2} \left\{ 1 + \log \left(\frac{u}{s \wedge u} \right) \right\} du \\
& + 2 \int_t^1 (s \wedge t/u) \left\{ 1 + \log \left(\frac{t/u}{s \wedge t/u} \right) \right\} du \\
& + 4tK(s) \int_t^1 du/u \\
& = 2 \left\{ 2(s \wedge t) - 2st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) + st \log s \right\} \\
& + 2 \left[\{(s \wedge t) - st\} \{1 + \log(t/s)\} + (s \wedge t) - st + st \log t + 2t \log \left(\frac{s \vee t}{t} \right) \right] \\
& + 4t \log t K(s) \\
& = 8(s \wedge t) - 8st + 2(s \wedge t) \log(s \vee t) - 2(s \wedge t) \log(s \wedge t) + 4st \log st \\
& + 2(s \wedge t) \log t - 2(s \wedge t) \log s + 4t \log(s \vee t) - 4t \log t - 4st \log s \log t.
\end{aligned}$$

3⁰) Du résultat précédent, on conclut directement que

$$J_3 = 8(s \wedge t) - 8st + 2(s \wedge t) \log(s \vee t) - 2(s \wedge t) \log(s \wedge t) + 4st \log st \\ + 2(s \wedge t) \log s - 2(s \wedge t) \log t + 4s \log(s \vee t) - 4s \log s - 4st \log s \log t.$$

4⁰)

$$J_4 = \int_t^1 \int_s^1 \frac{\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{H}_{n,i}(u, s/u), \mathbb{H}_{n,j}(v, t/v)\}}{uv} dudv,$$

où

$$\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{H}_{n,i}(s, t), \mathbb{H}_{n,j}(u, v)\} \\ = P(Y_0 \leq s, Y_i \leq t, Y_0 \leq u, Y_j \leq v) - stuv \\ + \sum_{\ell=1}^{\infty} \{P(Y_0 \leq s, Y_i \leq t, Y_\ell \leq u, Y_{\ell+j} \leq v) - stuv\} \\ + \sum_{\ell=1}^{\infty} \{P(Y_\ell \leq s, Y_{i+\ell} \leq t, Y_0 \leq u, Y_j \leq v) - stuv\} \\ = P\{Y_0 \leq (s \wedge u), Y_i \leq t, Y_j \leq v\} + P\{Y_0 \leq s, Y_i \leq (t \wedge u), Y_{i+j} \leq v\} \\ + P\{Y_0 \leq u, Y_j \leq (s \wedge v), Y_{i+j} \leq v\} + P\{Y_0 \leq u, Y_j \leq (t \wedge v), Y_{j-i} \leq s\} \\ - 4stuv \\ = vt(s \wedge u) + sv(t \wedge u) + ut(s \wedge v) + us(t \wedge v) - 4stuv.$$

De là,

$$J_4 = \int_t^1 \int_s^1 \frac{st}{u^2 v^2} (u \wedge v) dudv + \int_t^1 \int_s^1 \frac{t}{v^2} (s/u \wedge v) dudv \\ + \int_t^1 \int_s^1 \frac{s}{u^2} (u \wedge t/v) dudv + \int_t^1 \int_s^1 (s/u \wedge t/v) dudv \\ - st \int_t^1 \int_s^1 \frac{dudv}{uv}$$

$$\begin{aligned}
&= st \left\{ \frac{2}{s \vee t} - 2 + \log s + \log t + \frac{1}{s \vee t} \log(s \vee t) - \frac{1}{s \vee t} \log(s \wedge t) \right\} \\
&\quad + 2(s \wedge t) - 2st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) + st \log st \\
&\quad + 2(s \wedge t) - 2st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) + st \log st \\
&\quad + 2(s \wedge t) - 2st + (s \wedge t) \log \left(\frac{s \vee t}{s \wedge t} \right) + st \log st \\
&\quad - 4st \log s \log t \\
&= 8(s \wedge t) - 8st + 4st \log st - 4st \log s \log t \\
&\quad + 4(s \wedge t) \log(s \vee t) - 4(s \wedge t) \log(s \wedge t).
\end{aligned}$$

Finalemment,

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \text{cov}\{\mathbb{K}_{n,i}(s), \mathbb{K}_{n,j}(t)\} \\
&= J_1 - J_2 - J_3 + J_4 \\
&= 8(s \wedge t) - 8st + 4(s \wedge t) \log(s \vee t) - 4(s \wedge t) \log(s \wedge t) \\
&\quad + 4st \log st - 4st \log s \log t \\
&\quad - \{8(s \wedge t) - 8st + 2(s \wedge t) \log(s \vee t) - 2(s \wedge t) \log(s \wedge t) + 4st \log st \\
&\quad + 2(s \wedge t) \log t - 2(s \wedge t) \log s + 4t \log(s \vee t) - 4t \log t - 4st \log s \log t\} \\
&\quad - \{8(s \wedge t) - 8st + 2(s \wedge t) \log(s \vee t) - 2(s \wedge t) \log(s \wedge t) + 4st \log st \\
&\quad + 2(s \wedge t) \log s - 2(s \wedge t) \log t + 4s \log(s \vee t) - 4s \log s - 4st \log s \log t\} \\
&\quad + 8(s \wedge t) - 8st + 4st \log st - 4st \log s \log t \\
&\quad + 4(s \wedge t) \log(s \vee t) - 4(s \wedge t) \log(s \wedge t) \\
&= 4(s \wedge t) \log(s \vee t) - 4t \log(s \vee t) + 4t \log t - 4s \log(s \vee t) \\
&\quad + 4s \log s - 4(s \wedge t) \log(s \wedge t) \\
&= 0.
\end{aligned}$$

Annexe D

Programmes en langage C utilisés pour les simulations

Références

- [1] Alexander, K. S. (1987). The central limit theorem for weighted empirical processes indexed by sets. *Journal of Multivariate Analysis*, **22**, 313–339.
- [2] Barbe, P., Genest, C., Ghoudi, K., & Rémillard, B. (1996). On Kendall's process. *Journal of Multivariate Analysis*, **58**, 197–229.
- [3] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.
- [4] Brock, W. A. (1986). Distinguishing random and deterministic systems: abridged version. *Journal of Economic Theory*, **40**, 168–195.
- [5] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, **50**, 987-1008.
- [6] Ferguson, T. S., Genest, C., Hallin, M. (2000). Kendall's tau for auto-correlation *La revue canadienne de statistique*, **28**, sous presse.

- [7] Genest, C. & L.-P. Rivest (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034-1043.
- [8] Ghoudi, K., & Rémillard, B. (1998). Empirical processes based on pseudo-observations. Dans *Proceedings of the International Conference on Asymptotic Methods in Statistics*, **58**, 171–197, (B. Szyskowitz ed.).
- [9] Joe, H. (1990). Multivariate concordance. *Journal of Multivariate Analysis*, **35**, 12–30.
- [10] Nelsen, Roger B. (1999) *An Introduction to Copulas*. Lecture Notes in Statistics Number 139(?), Springer-Verlag, Berlin.