

N° d'ordre : 14/2013-M/MT

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene
Faculté de Mathématiques



MÉMOIRE

Présenté pour l'obtention du diplôme de Magister

En : Mathématiques

Spécialité : Recherche Opérationnelle (Méthodes Stochastiques)

Par Assia BAOUTA

Sujet

Etude des propriétés de mesures de dépendance sérielle

Soutenu publiquement, le 03/02/2013, devant le jury composé de :

| | | | |
|-------------------|----------------------|---------------|------------------------|
| Mr A. BERRACHEDI | Professeur | à l'U.S.T.H.B | Président. |
| Mme H. GUERBYENNE | Maître de conférence | à l'U.S.T.H.B | Directrice de mémoire. |
| Mr H. BELBACHIR | Maître de conférence | à l'U.S.T.H.B | Examineur. |
| Mr F. HAMDI | Maître de conférence | à l'U.S.T.H.B | Examineur. |

Table des matières

Table des matières

Abréviations et symboles

Introduction générale **i**

Chapitre 1 : Outils théoriques **1**

1 Généralités sur les processus aléatoires **2**

1.1 Processus stationnaires 3

1.1.1 Stationnarité stricte 3

1.1.2 Stationnarité au second ordre 4

1.2 Ergodicité 5

2 Aperçu sur quelques modèles de séries chronologiques **7**

2.1 Processus autorégressif d'ordre p 7

2.2 Processus moyenne mobile d'ordre q 8

2.3 Processus *ARCH/GARCH* 8

2.4 Processus autorégressif à seuils 11

2.5 Processus *GARCH* à seuils 12

2.6 Processus bilinéaire 14

3 Processus empiriques **16**

3.1 Définitions et résultats de base 16

3.2 Processus indexé par des fonctions 19

4 Distribution de la statistique de Cramèr-von Mises **21**

4.1 Motivation de la décomposition en composantes principales 22

| | | |
|----------|---|-----------|
| 4.1.1 | Enoncé du problème | 22 |
| 4.1.2 | Décompositions en composantes principales des vecteurs aléatoires | 22 |
| 4.1.3 | Décomposition (heuristique) en composantes principales d'un processus | 24 |
| 4.2 | Décomposition orthogonale des processus | 27 |
| 4.2.1 | Résultats fondamentaux sur les noyaux | 27 |
| 4.2.2 | Propriétés des noyaux | 27 |
| 4.2.3 | Existence de décomposition en composantes principales | 29 |
| 4.2.4 | Distribution de $\int_0^1 \mathbb{X}^2(t) dt$ via la décomposition | 31 |
| 4.3 | Décomposition en composantes principales de \mathbb{U}_n et \mathbb{U} | 33 |
| 5 | <i>U</i>-statistiques | 34 |
| 5.1 | Descriptions de la <i>U</i> -statistique | 35 |
| 5.1.1 | Définitions | 35 |
| 5.1.2 | Quelques statistiques proches : les <i>V</i> -statistiques | 37 |
| 5.1.3 | <i>U</i> -statistiques généralisées | 37 |
| 5.1.4 | Propriété d'optimalité de la <i>U</i> -statistique | 38 |
| 5.1.5 | Structure de martingales des <i>U</i> -statistiques | 39 |
| 5.1.6 | Représentation de la <i>U</i> -statistique comme une moyenne des moyennes (dépendantes) de <i>v.a.</i> i.i.d. | 42 |
| 5.1.7 | <i>U</i> -statistique pondérée | 43 |
| 5.2 | Variance et moments de la <i>U</i> -statistique | 44 |
| 5.2.1 | Variance de la <i>U</i> -statistique | 44 |
| 5.2.2 | Autres moments de la <i>U</i> -statistique | 47 |
| 5.3 | Distribution asymptotique de U_n | 48 |

| | | |
|----------|---|-----------|
| 1 | Mesure de Hoeffding | 50 |
| 1.1 | Consistance | 52 |
| 1.2 | Distribution asymptotique | 53 |
| 1.2.1 | Relation entre I_n et la statistique de von Mises | 53 |
| 1.2.2 | Distribution asymptotique de I_n dans le cas i.i.d. | 58 |
| 1.3 | Cas continu et cas discret | 59 |
| 1.4 | Test d'indépendance sérielle I_n | 63 |
| 1.5 | Extension à un retard supérieur | 63 |
| 2 | Mesure des différence pondérée de densités | 66 |
| 2.1 | Propriétés asymptotiques | 69 |
| 2.1.1 | Biais asymptotique | 71 |
| 2.1.2 | Variance asymptotique | 78 |
| 2.1.3 | Normalité asymptotique | 80 |
| 2.2 | Test basé sur $I_{n,1}$ | 84 |
| 3 | Mesure basée sur l'information conditionnelle | 84 |
| 3.1 | Définitions | 85 |
| 3.2 | Distribution asymptotique | 88 |
| 3.3 | Test d'indépendance sérielle | 88 |
| 4 | Mesure basée sur l'entropie | 89 |
| 4.1 | Distribution asymptotique | 92 |
| 4.2 | Effets de bords et noyau de Jackknife | 94 |
| 4.3 | Développement asymptotique des estimateurs d'entropie | 96 |
| 4.4 | Normalité asymptotique des estimateurs de l'entropie | 101 |
| 4.5 | Tests d'hypothèses et le bootstrap lissé | 103 |
| 4.5.1 | Test d'hypothèse | 103 |
| 4.5.2 | Test pour l'hypothèse i.i.d. $\mathcal{U} [0, 1]$ | 108 |
| 4.6 | Puissance locale asymptotique | 110 |

| | |
|---|----------------|
| <i>Chapitre 3 : Simulations et application</i> | 113 |
| 1 Mesure de Hoeffding | 113 |
| 1.1 Puissance du test basé sur I_n | 116 |
| 2 Mesure basée sur la différence pondérée des densités | 119 |
| 3 Mesure basée sur l'information conditionnelle | 124 |
| 4 Mesure basée sur l'entropie | 127 |
| 5 Application sur une série de données réelles | 131 |
| <i>Appendice</i> | i |
| <i>Référence</i> | xiv |

“ A mes chers parents, mon frère et mes deux soeurs ”

Remerciements

Je voudrais tout d'abord remercier Mme H. Guerbyenne pour m'avoir proposé le sujet et apporté le soutien nécessaire. De part ses qualités pédagogiques, ses précieux conseils, ses stimulants encouragements et sa disponibilité, j'ai pu mener à bien ce travail dans des conditions favorables. Elle m'a également permis d'approfondir mes connaissances et acquérir un savoir faire.

Mes remerciements vont à Mr A. BERRACHEDI de l'honneur qu'il m'a fait en acceptant de présider le jury et de l'intérêt qu'il a porté à ce travail.

Bien évidemment, je remercie Mr H. BELBACHIR et Mr F. HAMDI de l'honneur qu'ils me font d'être les examinateurs de ce mémoire, du temps et du soin qu'ils ont apporté à la rédaction de leurs rapports.

Un grand merci à tous le personnel de la bibliothèque de la Faculté de Mathématiques en particulier Mme Taous Oulemane pour sa disponibilité et ses qualités humaines.

Je tiens à exprimer ma reconnaissance à toute ma famille, mes amis et tous ceux qui m'ont aidé de près ou de loin.

Abréviations et symboles

| | |
|-----------------------------|---|
| $\xrightarrow{p.s.}$ | Convergence presque sûrement. |
| $\xrightarrow{\mathcal{L}}$ | Convergence en loi. |
| \xrightarrow{p} | Convergence en probabilité. |
| $v.a.$ | Variable aléatoire. |
| $X \cong Y$ | X et Y ont la même loi |
| $X \cong (\mu, \sigma^2)$ | X est de moyenne μ et de variance σ^2 |
| $X \cong F(\mu, \sigma^2)$ | X est de loi F de moyenne μ et de variance σ^2 |
| L^2 | L'espace des fonctions de carré intégrable. |

Introduction générale

" L'art de la connaissance est l'exercice des vertus,

éternelles noblesses pour le chercheur."

Lorsqu'on dispose de séries de données d'observations dans le temps, qui sont bruitées par essence, une étape préliminaire, nécessaire avant de chercher à modéliser le processus générateur des données, est de tester la dépendance sérielle. Par exemple, en finance, on teste l'hypothèse de marche aléatoire, en cherchant à vérifier si les rendements logarithmiques de prix ou de taux de change, sont indépendants et équidistribués. Une autre application consiste à vérifier la présence d'une structure de dépendance cachée dans les résidus d'un modèle estimé.

L'indépendance est l'un des concepts les plus importants en économétrie, en séries chronologiques et de façon générale, en statistique. Par conséquent, une littérature riche, proposant différents tests de dépendance a surgi.

Traditionnellement, les tests de dépendance sérielle portent sur la détection de la structure de dépendance sérielle dans les séries temporelles stationnaires en estimant la fonction d'autocorrélation $\rho_k = cov(X_{t-k}, X_t)/var(X_t)$, mais les tests basés sur cette mesure ne sont puissants que sous les hypothèses de linéarité et de gaussiannité, ce qui n'est pas, toujours, vérifié en pratique. D'où l'approche non paramétrique qui évite ces hypothèses restrictives.

Dans ce contexte, plusieurs mesures de dépendance non paramétriques ont été introduites. Un certain nombre de tests d'indépendance ont été construits sur le fait

que, l'hypothèse d'indépendance n'est vérifiée que, si et seulement si, la fonction de répartition conjointe (resp., la fonction densité de probabilité conjointe) est égale au produit des fonctions de répartition marginales (resp., au produit des fonctions de densité marginales). Donc, la statistique de test construite est une fonction de l'écart entre la fonction de répartition conjointe estimée (resp., la fonction densité conjointe estimée) et le produit des fonctions de répartition marginales estimées (resp., le produit des fonctions densité marginales estimées). Hoeffding (1948b) et Blum et al (1961) ont proposé d'estimer les fonctions de répartition conjointe et marginales par les fonctions de répartitions empiriques respectives, et introduit des statistiques basées sur les distances au sens de L_2 et L_∞ . Dans le contexte de la dépendance sérielle, Skaug et Tjøstheim (1993a) ont montré que la statistique basée sur les fonctions de répartition empiriques pour tester la dépendance sérielle de retard un (généralisé par la suite pour différents retards), a la même distribution limite que la statistique de Blum et al (1961) qui teste la dépendance entre deux variables aléatoires. Rosenblatt (1975) et Robinson (1991) ont estimé les fonctions densité par la méthode des noyaux. La statistique de Robinson se base sur l'entropie modifiée. En utilisant un fractionnement de l'échantillon, il a montré que la distribution asymptotique de cette statistique, pour la dépendance sérielle sous l'hypothèse d'indépendance, est la loi normale $\mathcal{N}(0, 1)$ et que le fractionnement de l'échantillon ne rend pas l'estimateur d'entropie modifiée incompatible avec l'entropie de la population.

Une autre classe importante qui est liée à la théorie de l'information, est celle d'estimateurs non paramétriques de l'entropie fondés sur des noyaux. Elle peut capturer toutes les dépendances deux à deux sur les retards de la série ; en outre elle est invariante sous des transformations continues monotones. Granger et Lin (1994) ont proposé une mesure d'entropie non paramétrique normalisée, estimée par la méthode des noyaux, pour identifier les retards importants dans les séries temporelles. Déterminer la distribution limite de ces estimateurs d'entropie a été jusque là difficile. Hong et White (2005) ont fourni, sous certaines hypothèses, une théorie asymptotique

tique pour une classe d'estimateurs non paramétriques d'entropie, fondés sur des noyaux, . Ils montrent également que leur théorie donne la distribution limite de la mesure d'entropie normalisée de Granger et Lin, qui était jusque là, méconnue. En outre, ils développent un test qui est, asymptotiquement, localement plus puissant que le test de Robinson. Néanmoins, la théorie de Robinson ne s'applique pas aux estimateurs d'entropie non-paramétriques fondés sur des noyaux qui n'utilisent pas un fractionnement de l'échantillon, telle que la mesure d'entropie normalisée de Granger et Lin. Par ailleurs, le fractionnement de l'échantillon nécessite certains paramètres de *tuning*. Comme Robinson (1991) l'a noté, le choix de ces paramètres reste ouvert. Une autre difficulté est que le fractionnement de l'échantillon ne marche pas lorsque la distribution marginale de la série est uniforme.

Une autre mesure importante, liée à la théorie de l'information, est la statistique introduite par Wu et *al* (1993). Elle se base sur l'information mutuelle conditionnelle. Elle est estimée par les intégrales de corrélation. Diks et Manzan (2002) ont proposé une statistique proche de celle ci. La relation entre les intégrales de corrélation généralisées et la théorie de l'information établie par Pichard et Theiler (1995), montre que la statistique de Diks et Manzan correspond à l'information mutuelle conditionnelle de second ordre. Elle est liée à la statistique de Granger et Lin (1994). Cependant, alors que leur statistique est une généralisation de la fonction d'autocorrélation, celle de Diks et Manzan est une généralisation de la fonction d'autocorrélation partielle.

Su et White (2008) ont introduit une mesure de dépendance conditionnelle basée sur une version pondérée de la distance de Hellinger entre deux densités conditionnelles $p(y|x, z)$ et $p(y|x)$. Ils ont montré que cette statistique a pour distribution limite sous l'hypothèse $H_0' Pr\{f(y|X, Z) = f(y|X)\} = 1$, une loi normale. Bien que le test basé sur cette statistique soit simple à mettre en œuvre, il présente deux limites. En premier lieu, il utilise la même fenêtre dans l'estimation non paramétrique des densités ce qui est non satisfaisant quand la dimension de (X, Y, Z) est

supérieure à trois. Deuxièmement, leur test ne peut détecter que l'écart par rapport à l'indépendance conditionnelle sur un sous ensemble compact du support de la loi de probabilité conjointe de (X, Y, Z)

Une autre mesure importante est celle basée sur les fonctions caractéristiques conditionnelles. Cette statistique suit asymptotiquement une loi normale. Su et White (2007) ont proposé un test d'indépendance conditionnelle basé sur cette statistique moins sensible au choix de suites de fenêtres

Une autre mesure liée à la théorie de l'information est celle basée sur l'entropie fondée sur les permutations. En utilisant les dynamiques symboliques, la distribution limite d'une transformation affine de cette statistique est une loi du khi-deux à $(m! - 1)$ degrés de liberté où m est la dimension d'empilement.

Dans la famille des mesures d'entropie, la mesure introduite par Tsallis (1988) est une généralisation de la mesure basée sur l'information de Kullback-Leibler et celle basée sur la distance de Hellinger. Fernandes et Néri (2010), l'ont utilisée comme une mesure de dépendance entre deux processus aléatoires satisfaisant aux conditions de α -mélange dont la dépendance sérielle est un cas particulier. Ils ont montré que, asymptotiquement, cette statistique sous H_0 , est une loi normale.

Une nouvelle mesure a été introduite, récemment, par Tjøstheim et Hufthammer (2013). Il s'agit d'une mesure de dépendance locale, la fonction de corrélation locale, basée sur une approximation d'une densité bivariée et ce, localement, par une famille de densités gaussiennes bivariées en utilisant la fonction de vraisemblance locale.

Nous nous sommes intéressées à l'étude théorique et à l'étude empirique des quatre premières statistiques de tests.

Ce mémoire intitulé, "Etude des propriétés de mesures de dépendance sérielle", est organisé comme suit

Chapitre 1 intitulé "*Outils théoriques*", présente les principaux outils mathématiques utilisés tout au long de ce mémoire

Chapitre 2 intitulé "*Tests non paramétriques de dépendance sérielle*", présente

des tests non paramétriques de dépendance sérielle basés sur des mesures non paramétriques à savoir, la mesure de Hoeffding, la mesure de différence des densités conjointe et marginales, la mesure basée sur l'information conditionnelle et la mesure d'entropie lissée. Leur intérêt est de tester l'indépendance entre les observations d'une série temporelle stationnaire. De tels tests sont puissants contre de nombreuses alternatives.

Skaug et Tjøstheim (1993a) ont montré, après l'avoir représentée comme une statistique de von Mises, que la statistique de test pour la dépendance sérielle, basée sur la mesure de Hoeffding, est asymptotiquement distribuée selon une somme pondérée infinie de loi de probabilité du khi deux indépendantes. Cette mesure est estimée en estimant les fonctions de répartition par les fonctions de répartition empiriques. Pour une fonction de répartition sous-jacente continue, Skaug et Tjøstheim (1993a) ont démontré que c'est une statistique libre ; les poids sont déterminés en résolvant une équation intégrale linéaire. Dans le cas d'une distribution discrète, il est démontré que les poids sont déterminés en cherchant les valeurs propres d'une matrice (Skaug et Tjøstheim 1993a).

Une autre mesure de différence, pondérée, a été proposée par Skaug et Tjøstheim (1993b). Elle est basée sur l'écart entre les densités conjointes et le produit des densités marginales, où les densités sont estimées par la méthode du noyau leave-one-out. Nous considérons le noyau gaussien standard et la fenêtre optimale donnée dans Silverman (1986). Ainsi le même noyau et la même fenêtre ont été utilisés pour l'estimation des densités marginales et conjointes. Skaug et Tjøstheim (1993b) ont étudié le biais et la variance asymptotique, Ils ont montré, également, que la distribution asymptotique, sous l'hypothèse d'indépendance, suit la loi normale centrée réduite.

Wu et *al* (1993), ont montré que la mesure basée sur l'information conditionnelle a pour distribution asymptotique, sous l'hypothèse d'indépendance, une loi normale de moyenne nulle et de variance V , et cette variance est estimée par les intégrales

de corrélation. Cette mesure est définie en termes de rapport des intégrales de corrélation. L'avantage de l'utiliser comme statistique de test est, qu'elle capture les dépendances tandis que le conditionnement sur les valeurs intermédiaires de la série temporelle donne également un aperçu sur l'ordre du processus sous-jacent. Il est bien connu que l'approximation normale basée sur la distribution asymptotique ne donne pas toujours de bons résultats pour des tailles modérées. Diks et Manzan (2002) ont montré dans leurs études de simulation que ces problèmes se posent également pour l'estimation de cette statistique. C'est la principale motivation pour l'utilisation d'une approche bootstrap pour déterminer la distribution de la statistique de test sous l'hypothèse d'indépendance. Diks et Manzan (2002) ont mis en œuvre la méthode de Monte Carlo dans leur test. La méthode de Monte Carlo a d'abord été suggérée par Barnard (1963) pour tester une hypothèse nulle simple. Dans les cas où l'on souhaite tester une hypothèse nulle composite, cette procédure ne peut pas être appliquée directement, puisque le processus, sous H_0 , dépend toujours de paramètres inconnus du modèle. Dans la littérature portant sur la technique du bootstrap, il est souvent souligné qu'on doit considérer des statistiques de test qui sont au moins asymptotiquement pivotales, sous H_0 . Besag et Diggle (1977) et Engen et Lillegård (1997) ont pu obtenir un niveau exact du test de Monte Carlo. L'approche est très flexible en ce sens qu'elle permet de tester les hypothèses nulles qui sont complètement non spécifiées en dehors de quelques propriétés, comme le montrent les tests de symétrie proposés par Diks et Tong (1999). Parce qu'un test de permutation peut être facilement mis en œuvre, il offre un moyen commode pour obtenir un test exact. La variance de la statistique, basée sur l'information conditionnelle, est exprimée à l'aide des fonctions de statistiques d'ordre qui sont invariantes quelque soit la permutation des données d'observation dont on dispose. Dans ce cadre, les p -valeurs (p -values, en anglais), ne sont pas affectées par une transformation translation-homothétie (ou position-échelle). En plus, les tests de permutation sont faciles à appliquer, et évitent les calculs embarrassants nécessaires pour obtenir la variance

asymptotique.

Une autre mesure importante, est la mesure d'entropie qui se base sur l'information de Kullback-Leibler. Elle peut caractériser toutes les dépendances sérielles deux à deux. De plus elle dispose d'une interprétation intéressante, dans la théorie de l'information, en ce sens qu'elle mesure la prévisibilité par la différence d'entropie entre les distributions a priori et a posteriori, et elle est invariante sous n'importe quelle transformation continue et monotone de la série en question $\{X_t\}$. La propriété d'invariance est attractive parce que $\{X_t\}$ est i.i.d. si et seulement si pour toute transformation continue monotone ψ , $\psi(\{X_t\})$ est i.i.d. Cette mesure est asymptotiquement distribuée selon une loi normale centrée réduite. Pour avoir une inférence exacte pour des échantillons de taille finie, Hong et White (2005) ont proposé et justifié une procédure consistante de bootstrap lissé pour leur test. La procédure de bootstrap naïf ne marche pas pour le test de Hong et White, en dépit de l'hypothèse nulle d'indépendance. Leur test est utile pour, par exemple, tester l'hypothèse de marche aléatoire, l'évaluation des prévisions de densité et identifier des décalages importants dans une série chronologique. Ces auteurs ont montré que ce test est asymptotiquement, localement plus puissant que le test de Robinson.

Chapitre **3** intitulé "*Simulations et application*". Il décrit les résultats obtenus pour les mesures de dépendance non paramétriques étudiées dans le chapitre **2**. Nous avons confirmé et comparé la puissance des tests basés sur ces mesures au travers des simulations intensives pour des modèles linéaires tel que *AR* et des modèles non linéaires comme *ARCH*, *GARCH*, *TAR* et bilinéaire. Une application sur une série de données réelle S&P 500 illustre et confirme ces résultats.

Chapitre

1

Outils théoriques

Dans ce chapitre, nous définissons les principaux outils mathématiques utilisés tout au long de ce mémoire. Des généralités sur les processus aléatoires telles que la stationnarité (au sens strict et large) et l'ergodicité, sont données dans la section 1. Dans la section 2, nous présentons quelques modèles de séries chronologiques linéaires tels que les modèles autorégressifs, moyenne mobile et non linéaires tel que *GARCH*, *ARCH*, *TARCH*,... qui vont servir de cadre pour les simulations. Nous introduisons brièvement quelques résultats de base de la théorie des processus empiriques dont les méthodes sont des techniques puissantes pour évaluer les propriétés asymptotiques des estimateurs issus d'approches semi-paramétriques ou non-paramétriques comme la consistance et la convergence en loi, ainsi que la distribution de la statistique de Cramér-von Mises, dans les sections 3 et 4. Nous donnerons enfin quelques résultats importants sur la *U*-statistique qui est un type spécifique d'estimateurs défini d'une manière particulière. L'utilité du concept dans la théorie statistique, est qu'il permet à un estimateur sans biais d'être de variance minimale.

1 Généralités sur les processus aléatoires

La théorie des probabilités a pour but d'étudier les phénomènes aléatoires et de trouver des modèles mathématiques adéquats. On peut classer ces phénomènes en deux classes : phénomènes statiques et phénomènes dynamiques. Pour le premier cas, il s'agit bien d'un fait qui ne dépend d'aucun d'autre fait ; par exemple la durée de vie d'un appareil, la charge électrique d'un électron, la conductivité d'un fil électrique, etc. Mathématiquement ils sont modélisés par la variable (vecteur) aléatoire. Les phénomènes ou bien les systèmes à évolution sur un domaine tel que le temps, l'espace,... par exemple, la marche au hasard d'une particule le long d'un axe, la propagation des ondes sonores, la diffusion d'un grain de pollen dans un milieu (mouvement brownien), sont modélisés par le processus aléatoire.

De nombreux phénomènes économiques réels sont donc mieux expliqués par un modèle dynamique plutôt que statique. Les systèmes dynamiques évolutifs dans le temps occupent une place majeure dans l'étude de certains phénomènes aléatoires, généralement représentés par les séries temporelles (ou chronologiques).

Une série temporelle est une réalisation d'un processus stochastique, au sens où chaque donnée de la série est la réalisation de l'une des *v.a.* qui composent le processus stochastique. Autrement dit, c'est la trajectoire du processus qui la génère.

Définition 1.0.1 *Un processus aléatoire de domaine d'évolution T , défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$ et à valeurs dans un espace d'état χ qui est muni d'une tribu \mathcal{E} , est une famille de *v.a.* chacune définie sur $(\Omega, \mathcal{F}, \mathcal{P})$ à valeur dans (χ, \mathcal{E}) . Autrement dit,*

$$X : T \times \Omega \rightarrow E;$$

$$(t, \omega) \mapsto X(t, \omega) = X_t(\omega).$$

Dans ce travail, nous considérons que le domaine d'évolution est le temps, il peut être :

1. $T = \{0, 1, 2, \dots\}$ ou $T = \mathbb{Z}$. Dans ce cas on parle d'un processus en temps

discret.

2. $T = [0, \infty[$ ou $T = \mathbb{R}$. Le processus est alors en temps continu.

Il y a deux niveaux auxquels nous pouvons observer l'évolution d'un processus :

— Fixons un $\omega \in \Omega$ (une réalisation de l'expérience aléatoire sous-jacente). L'ensemble $\{X_t(\omega), t \in T\}$ est une réalisation du processus stochastique, aussi appelée trajectoire. On peut étudier les propriétés des trajectoires.

— Fixons n instants (t_1, t_2, \dots, t_n) , avec $t_1 < t_2, \dots < t_n$ où n est un entier positif arbitraire. Considérons le vecteur aléatoire $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$, extrait du processus, et notons sa fonction de répartition conjointe :

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n)$$

Les propriétés distributionnelles du processus sont données par la collection de toutes les fonctions de répartition F_{t_1, t_2, \dots, t_n} pour tout $n \geq 1$ fini et tout n -uple (t_1, t_2, \dots, t_n) de T . Pour obtenir des processus utiles, il faut particulariser.

1.1 Processus stationnaires

Le concept fondateur de l'analyse des séries chronologiques est la stationnarité. La stationnarité signifie que les propriétés statistiques les plus importantes du processus restent constantes au cours du temps. La stationnarité est une propriété de stabilité, la distribution de X_t est identique à celle de X_{t+h} , $\forall h \in \mathbb{Z}$, et le lien entre X_t et X_{t+h} ne dépend que de l'écart h et non du temps t . La série oscille autour de sa moyenne avec une variance constante, lorsque celles-ci existent. On distingue deux types de stationnarité

1.1.1 Stationnarité stricte

Définition 1.1.1 *Un processus aléatoire $(X_t, t \in \mathbb{Z})$ est dit strictement stationnaire, si sa distribution de probabilité est invariante par translation dans le temps.*

Autrement dit, si pour tout $n \in \mathbb{N}^*$, $h, t_1, t_2, \dots, t_n \in \mathbb{Z}$ et $x_i \in \mathbb{R}$, $i = 1, \dots, n$, on a

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n) = P(X_{t_1+h} \leq x_1, X_{t_2+h} \leq x_2, \dots, X_{t_n+h} \leq x_n).$$

La stricte stationnarité implique que, si les moments existent, ils sont invariants dans le temps. Dans la pratique, il n'est pas évident de trouver la vraie loi du processus. Puisque de nombreuses propriétés probabilistes peuvent être obtenues à partir des deux premiers moments, s'ils existent, donc pour avoir la stationnarité il suffit que les deux premiers moments soient indépendants du temps, d'où la définition

1.1.2 Stationnarité au second ordre

Définition 1.1.2 *Un processus aléatoire $(X_t, t \in \mathbb{Z})$ à valeurs réelles et de domaine d'évolution \mathbb{Z} est dit stationnaire au second ordre (faiblement stationnaire, stationnaire en covariance), si*

- i) La fonction variance $\sigma^2(t) = \text{Var}(X_t) < \infty$ existe pour tout t dans le domaine d'évolution \mathbb{Z} (et donc nécessairement la fonction moyenne $\mu(t) = \mathbb{E}(X_t)$ existe et $\mathbb{E}(X_t^2) < \infty$).*
- ii) La fonction moyenne $\mu(t) = \mu$ est constante sur le domaine d'évolution et la fonction d'auto-covariance $\gamma(t, t+h) = \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)] = \gamma_h$ dépend seulement de h , différence entre les instants d'observations, ce qui entraîne que la fonction variance $\sigma^2(t)$ est constante dans le domaine d'évolution \mathbb{Z} .*

Remarque 1.1.1 *Il est clair que la stationnarité stricte implique la stationnarité au second ordre, lorsque les deux premiers moments existent, le contraire est faux, sauf dans le cas gaussien où elles sont équivalentes.*

1.2 Ergodicité

L'ergodicité exprime le fait que le processus peut prendre n'importe quelle valeur dans l'espace des états indépendamment de sa valeur initiale, et qu'il ne peut s'isoler dans un sous ensemble strictement inclus dans l'espace des états, c'est donc une propriété d'irréductibilité mais associée à des processus stationnaires. Cette propriété est très importante pour l'établissement du théorème ergodique.

Le théorème ergodique représente une généralisation de la loi des grands nombres au cas des processus non i.i.d. (indépendants identiquement distribués) mais stationnaires. Ce théorème stipule que sous certaines conditions, la moyenne empirique d'une fonction du processus converge vers une certaine *v.a.* Lorsque le processus est ergodique cette *v.a.*, si elle existe ce n'est autre que la variable dégénérée représentant la moyenne commune à tous les membres.

Avant de définir formellement la notion d'ergodicité, nous avons besoin du concept d'invariance d'un espace par rapport à une transformation

Définition 1.2.1 Soit $\mathbb{R}^{\mathbb{Z}}$ l'espace de toutes les séquences $(X_t, t \in \mathbb{Z})$ de nombres réels.

- i) L'application $\mathcal{T} : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$ est dite opérateur de translation si
- $$\mathcal{T}(\dots, X_0, X_1, X_2, \dots) = (\dots, X_1, X_2, X_3, \dots).$$
- ii) Un sous ensemble $A \subset \mathbb{R}^{\mathbb{Z}}$ est dit \mathcal{T} -invariant si $\mathcal{T}A = A$.

Il existe deux types d'ergodicité, ergodicité faible et ergodicité forte.

Définition 1.2.2 (Ergodicité forte) Un processus stochastique stationnaire $(X_t, t \in \mathbb{Z})$ est dit ergodique si $P((X_0, X_1, \dots) \in A) = 0$ ou 1 pour chaque ensemble A \mathcal{T} -invariant.

Autrement dit, le seul sous espace invariant par la translation \mathcal{T} à quelques ensembles de mesures nulles près est l'ensemble \mathbb{R}^d lui-même, ce qui exprime une certaine irréductibilité.

Lorsque le processus aléatoire $(X_t, t \in \mathbb{Z})$ est strictement stationnaire, la suite définie par $\frac{1}{n} \sum_{t=1}^n Y_t$ converge vers une *v.a.*, mais lorsque le processus $(X_t, t \in \mathbb{Z})$ est ergodique, une telle variable n'est rien d'autres que la moyenne (une constante). D'où le théorème ergodique suivant :

Théorème 1.2.1 (Karlin et Taylor, 1975) *Soit (X_t) un processus ergodique strictement stationnaire de moyenne μ finie. Alors*

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mu\right) = 1.$$

Définition 1.2.3 (Ergodicité faible) *Soit $(X_t, t \in \mathbb{Z})$ un processus stationnaire au sens faible, tel que $\mathbb{E}(X_t) = \mu < \infty$ et $\sigma^2 < \infty$ pour tout t dans \mathbb{Z} . Soit $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ la moyenne temporelle. Si \bar{X}_n converge en probabilité vers μ quand $n \rightarrow \infty$, alors $\{X_t\}$ est ergodique pour la moyenne.*

Théorème 1.2.2 (Karlin et Taylor, 1975) *Soit (X_t) un processus stationnaire au second ordre et soit $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ la moyenne empirique. Alors il existe une *v.a.* \bar{X} qui est la limite quadratique de la moyenne \bar{X}_n*

$$\lim_{n \rightarrow \infty} \|\bar{X}_n - \bar{X}\| = \lim_{n \rightarrow \infty} \left[\mathbb{E} (\bar{X}_n - \bar{X})^2 \right]^{1/2} = 0.$$

Théorème 1.2.3 (Karlin et Taylor, 1975) *Soit $(X_t, t = 0, 1, 2, \dots)$ un processus strictement stationnaire avec une moyenne finie $\mu = \mathbb{E}(X_t)$ et soit $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ la moyenne empirique. Alors la suite $\{\bar{X}_n\}$ converge vers une *v.a.* \bar{X} avec une probabilité égale à 1,*

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \bar{X}\right) = 1.$$

2 Aperçu sur quelques modèles de séries chronologiques

La spécificité de l'analyse d'une série chronologique, qui la distingue d'autres analyses statistiques, est précisément dans l'importance accordée à l'ordre dans lequel sont effectuées les observations. Les méthodes statistiques classiques demandent souvent que les variables étudiées soient stochastiquement indépendantes et observées plusieurs fois. En séries chronologiques la dépendance temporelle entre les variables constitue la source principale d'information.

Nous allons introduire quelques modèles de séries chronologiques (définition et conditions de stationnarité) dont nous avons besoin dans la partie application.

2.1 Processus autorégressif d'ordre p

Définition 2.1.1 On dit que $\{X_t, t \in \mathbb{Z}\}$ est un processus autorégressif d'ordre $p \geq 1$ s'il obéit à l'équation aux différences stochastique :

$$X_t = \varepsilon_t + \phi_0 + \sum_{k=1}^p \phi_k X_{t-k}, \quad t \in \mathbb{Z}. \quad (2.1)$$

où $\{\varepsilon_t\}$ est un bruit blanc, ϕ_0 et les ϕ_k sont des réels pour $k = \overline{1, p}$

Un bruit blanc¹ fort est un processus i.i.d. de moyenne zéro et de variance égale à une constante σ^2 :

$$\varepsilon_t \cong \mathcal{BB}(0, \sigma^2). \quad (2.2)$$

La série chronologique $\{X_t\}$ est générée d'un modèle appelé processus $AR(p)$. Le modèle (2.1) spécifie explicitement la relation entre la valeur actuelle et ses valeurs passées. Comme nous le voyons, c'est un modèle linéaire, mais il se peut que nous rencontrions des modèles autorégressifs non linéaires qui restent des cas parti-

¹dit aussi processus d'innovation, la partie non prédictible de X_t à partir de X_{t-1}, X_{t-2}, \dots

culiers, mais par convention le modèle autorégressif se réfère toujours aux modèles autorégressifs linéaires de la forme (2.1) sauf si le contraire est indiqué.

On note L l'opérateur de retard, i.e., pour un processus X_t quelconque on a $LX_t = X_{t-1}$ et plus généralement $L^i X_t = X_{t-i}$ pour tout $i \in \mathbb{N}$.

Le polynôme caractéristique associé à $AR(p)$ est défini par :

$$A(L) = 1 - \phi_1 L - \dots - \phi_p L^p.$$

Proposition 2.1.1 *Le processus $AR(p)$ est stationnaire au second ordre si et seulement si toutes les racines du polynôme caractéristique sont de module strictement supérieur à l'unité.*

2.2 Processus moyenne mobile d'ordre q

On dit que $\{X_t, t \in \mathbb{Z}\}$ est un processus moyenne mobile d'ordre q ($q \geq 1$), noté $MA(q)$ s'il est solution de l'équation récurrences stochastique suivante :

$$X_t = \theta_0 + \sum_{k=1}^q \theta_k \varepsilon_{t-k}, \quad t \in \mathbb{Z}. \quad (2.3)$$

X_t est exprimé en fonction du présent et du passé du processus d'innovation ε_t .

Un processus $MA(q)$ est toujours faiblement stationnaire, car les moments d'ordre 1 et 2 ne dépendent pas de t et la covariance ne dépend que de l'écart entre les instants considérés.

2.3 Processus $ARCH/GARCH$

Parmi les faits stylisés des séries financières rencontrés dans le domaine économétrique et financier, la volatilité instantanée², cette volatilité tend à se regrouper (*clustering volatility*) i.e., les fortes (faibles) variabilités sont fréquemment suivies

²Changement instantané du rendement d'un actif financier.

par des fortes (faibles) variabilités. Les modèles linéaires des séries chronologiques à savoir AR , MA , $ARMA$,...ne peuvent capturer ce phénomène, d'où la nécessité de développer de nouveaux modèles où d'enrichir ceux existants. Engle (1982) propose une nouvelle classe de modèles autorégressifs conditionnellement hétéroscédastiques ($ARCH$) apte à capter ce phénomène. Le modèle est formé de deux équations. La première met en relation le phénomène et certaines variables qui l'expliquent et la seconde modélise la variance conditionnelle des résidus.

Le modèle autorégressif conditionnellement hétéroscédastique $ARCH(q)$ est défini par

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2. \end{aligned} \quad (2.4)$$

où $\{\varepsilon_t\} \cong iid(0, 1)$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$.

L'introduction d'un grand nombre de retards q dans l'équation de la variance conditionnelle du modèle $ARCH(q)$ est nécessaire pour mieux représenter certaines séries monétaires et financières caractérisées par la longue mémoire de la volatilité. Ceci peut poser des problèmes d'estimation. D'où, l'extension importante au modèle autorégressif conditionnellement hétéroscédastique généralisé ($GARCH$), suggérée par Bollerslev (1986). Cette approche exige moins de paramètres à estimer que la formulation $ARCH(q)$ pour modéliser les phénomènes de persistance des chocs. La variance conditionnelle de la variable étudiée est déterminée par le carré des q termes passés du processus considéré et des p variances conditionnelles retardées.

Le modèle autorégressif conditionnellement hétéroscédastique généralisé $GARCH(p, q)$, est défini comme suit

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2. \end{aligned} \quad (2.5)$$

où $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$, $\beta_j \geq 0$, $j = 1, \dots, p$ sont les coefficients du modèle et $\{\varepsilon_t\}$ est un processus i.i.d. Si les β_j sont nuls, nous aurons un modèle $ARCH(q)$.

Le modèle $GARCH(p, q)$ a une solution stationnaire au second-ordre, non anticipative si et seulement si $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ (condition pour que la variance marginale soit constante) (Bollerslev, 1986).

Représentons X_t sous une forme matricielle

$$\underline{z}_t = \underline{b}_t + A_t \underline{z}_{t-1}. \quad (2.6)$$

où $\underline{z}_t = (X_t^2, X_{t-1}^2, \dots, X_{t-q+1}^2, \sigma_t^2, \dots, \sigma_{t-p+1}^2)' \in \mathbb{R}^{q+p}$,

$\underline{b}_t = (\alpha_0 \sigma_t^2, 0, 0, \dots, \alpha_0, 0, \dots, 0)' \in \mathbb{R}^{q+p}$,

$$A_t = \begin{pmatrix} \alpha_1 \varepsilon_t^2 & \dots & \alpha_q \varepsilon_t^2 & \beta_1 \varepsilon_t^2 & \dots & \beta_p \varepsilon_t^2 \\ I_{(q-1) \times (q-1)} & 0_{q \times 1} & & 0_{(q-1) \times p} & & \\ \alpha_1 & \dots & \alpha_q & \beta_1 & \dots & \beta_p \\ & 0_{(p-1) \times q} & & I_{(p-1) \times (p-1)} & & 0_{p \times 1} \end{pmatrix}.$$

où, pour toute suite de matrices aléatoires $A = (A_t)$, strictement stationnaire et ergodique, telle que $\mathbb{E} \log^+ \|A_t\| < \infty$ on définit

$$\begin{aligned} \gamma(A) &= \inf_{t \in \mathbb{N}} \frac{1}{t} \mathbb{E}(\log \|A_t A_{t-1} \dots A_1\|), \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\| \text{ p.s.} \end{aligned} \quad (2.7)$$

Théorème 2.3.1 (Bougerol & Picard, 1992) *Le modèle $GARCH(p, q)$ a une (unique) solution strictement stationnaire non anticipative si et seulement si*

$$\gamma(A) < 0.$$

où $\gamma(A)$ est le plus grand exposant de Lyapounov

Remarque 2.3.1 La condition de stationnarité au second ordre du modèle ARCH est $\sum_{i=1}^q \alpha_i < 1$. Pour la stationnarité stricte on applique le théorème 2.3.1 tel que

$$A_t = \begin{pmatrix} \alpha_1 \varepsilon_t^2 & \dots & \alpha_q \varepsilon_t^2 \\ I_{(q-1) \times (q-1)} & 0_{q \times 1} \end{pmatrix}$$

2.4 Processus autorégressif à seuils

Le modèle autorégressif à seuils (threshold) *TAR* (1) introduit par H. Tong (1978) présume différentes formes linéaires dans les différentes régions de l'espace d'état. La division de l'espace d'état est généralement imposée par une variable seuil, par exemple, X_{t-d} , pour $d \geq 1$. Le modèle est de la forme

$$X_t = \sum_{i=1}^k b_{i,0} + b_{i,1}X_{t-1} + \dots + b_{i,p_i}X_{t-p_i} + \sigma_i \varepsilon_t, \quad X_{t-d} \in \Omega_i. \quad (2.8)$$

où p_1, \dots, p_k sont des entiers positifs inconnus, $\sigma_i > 0$ et b_{ij} sont des paramètres inconnus. $\{\Omega_i\}$ forme une partition de la ligne droite $] -\infty, +\infty[$ dans le sens que $\bigcup_{i=1}^k \Omega_i =] -\infty, +\infty[$ et $\Omega_i \cap \Omega_j = \emptyset$ si $i \neq j$, et $\{\varepsilon_t\} \cong i.i.d.(0, \sigma^2)$. Le modèle à seuil le plus simple est à deux régimes (i.e., $k = 2$), le modèle *TAR* à deux régimes est tel que $\Omega_1 = \{X_{t-d} \leq \tau\}$ où le seuil τ est connu. A titre d'illustration, le modèle *TAR* (2) à deux régimes

$$X_t = \begin{cases} 0.62 + 1.25X_{t-1} - 0.43X_{t-2} + \varepsilon_t, & X_{t-2} \leq 3.25, \\ 2.25 + 1.52X_{t-1} - 1.24X_{t-2} + \varepsilon'_t, & X_{t-2} > 3.25. \end{cases}$$

où $\varepsilon_t \cong \mathcal{N}(0, 0.2^2)$ et $\varepsilon'_t \cong \mathcal{N}(0, 0.25^2)$

Ce modèle résulte de l'ajustement de la série de données sur le nombre de lynx canadiens pris au piège, (Tong, 1990).

Le modèle *TAR* admet une solution strictement stationnaire si
a) $\sigma_1 = \sigma_2 = \dots = \sigma_p$ b) $\max_{1 \leq i \leq k} \sum_{j=1}^p |b_j^{(i)}| < 1$ ou $\sum_{j=1}^p \max_{1 \leq i \leq k} |b_j^{(i)}| \leq 1$ avec $p = \max_{1 \leq i \leq k} p_i$

(J. Fan and Q. Yao, 2003). Notons que ces conditions sont suffisantes et non nécessaires.

2.5 Processus *GARCH* à seuils

Introduit par Zakoian (1994), il est capable de capturer l'asymétrie en spécifiant la variance conditionnelle en fonction des composantes positive et négative des innovations passées.

Notons

$$\varepsilon_t^+ = \max(\varepsilon_t, 0), \quad \varepsilon_t^- = \min(\varepsilon_t, 0)$$

ces composantes, en remarquant que $\varepsilon_t = \varepsilon_t^+ + \varepsilon_t^-$. La classe suivante introduit un effet de seuil dans la volatilité.

Soit $\{\eta_t\}$ une suite de variables i.i.d. telles que $\mathbb{E}(\eta_t) = 0$ et $Var(\eta_t) = 1$. On dit que $\{\varepsilon_t\}$ est un processus *GARCH* à seuil (Threshold *GARCH*(p, q)) s'il vérifie une équation de la forme

$$\begin{aligned} \varepsilon_t &= \sigma_t \eta_t, \\ \sigma_t &= \alpha_0 + \sum_{i=1}^q \alpha_{i,+} \varepsilon_{t-i}^+ - \alpha_{i,-} \varepsilon_{t-i}^- + \sum_{j=1}^p \beta_j \sigma_{t-j}. \end{aligned} \quad (2.9)$$

où $\alpha_0, \alpha_{i,+}, \alpha_{i,-}$ et β_j sont des réels tels que $\alpha_0 > 0, \alpha_{i,+} \geq 0, \alpha_{i,-} \geq 0, \beta_j \geq 0$.

La variable σ_t est toujours strictement positive et s'interprète comme l'écart type conditionnel de ε_t . Dans tous les cas, l'écart-type conditionnel de ε_t est $|\sigma_t|$, il n'est pas nécessaire d'imposer la positivité de σ_t (par opposition aux *GARCH* classiques fondés sur la modélisation de σ_t^2).

Il s'agit d'un modèle à seuil où la dynamique de l'écart-type conditionnel diffère selon le signe des termes d'erreur. L'équation de l'écart-type conditionnel de *TGARCH*(p, q) est une fonction linéaire par morceaux selon le signe du choc et l'écart-type conditionnel de la période précédente. Par ailleurs, dans le modèle *TGARCH*, il est possible d'observer une discontinuité de la dérivée de la variance

conditionnelle par rapport aux perturbations au voisinage de zéro de telle sorte que les problèmes d'estimation sont plus complexes (S. Avouyi-Dovi et E. Jondeau, 1999)

Ecrivant le modèle sous sa forme vectorielle

$$\underline{z}_t = \underline{b}_t + A_t \underline{z}_{t-1}.$$

où $\underline{z}_t = (\varepsilon_t^+, -\varepsilon_t^-, \dots, \varepsilon_{t-q+1}^+, -\varepsilon_{t-q+1}^-, \sigma_t, \dots, \sigma_{t-p+1})' \in \mathbb{R}^{p+2q}$,

$\underline{b}_t = \underline{b}_t(\eta_t) = (\alpha_0 \eta_t^+, \alpha_0 \eta_t^-, 0, \dots, \alpha_0, 0, \dots, 0)' \in \mathbb{R}^{p+2q}$,

et soit

$$A_t = \begin{pmatrix} \eta_t^+ \boldsymbol{\alpha}_{1:q-1} & \alpha_{q,+} \eta_t^+ & \alpha_{q,-} \eta_t^+ & \eta_t^+ \boldsymbol{\beta}_{1:p-1} & \beta_p \eta_t^+ \\ -\eta_t^- \boldsymbol{\alpha}_{1:q-1} & -\alpha_{q,+} \eta_t^- & -\alpha_{q,-} \eta_t^- & -\eta_t^- \boldsymbol{\beta}_{1:p-1} & -\beta_p \eta_t^- \\ I_{2q-2} & 0_{2q-2} & 0_{2q-2} & 0_{2q-2 \times p-1} & 0_{2q-2} \\ \boldsymbol{\alpha}_{1:q-1} & \alpha_{q,+} & \alpha_{q,-} & \boldsymbol{\beta}_{1:p-1} & \beta_p \\ 0_{p-1 \times 2q-2} & 0_{p-1} & 0_{p-1} & I_{p-1} & 0_{p-1} \end{pmatrix} \quad (2.10)$$

La matrice est de dimension $(p+2q) \times (p+2q)$, $\boldsymbol{\alpha}_{1:q-1} = (\alpha_{1,+}, \alpha_{1,-}, \dots, \alpha_{q,+}, \alpha_{q,-}) \in \mathbb{R}^{2q-2}$ et $\boldsymbol{\beta}_{1:p-1} = (\beta_1, \dots, \beta_p) \in \mathbb{R}^{p-1}$.

Le résultat suivant est analogue à celui obtenu pour la stationnarité stricte des $GARCH(p, q)$.

Théorème 2.5.1 (Zakoïan, 1994) *Une condition nécessaire et suffisante d'existence d'un processus $TGARCH(p, q)$ strictement stationnaire, solution non anticipative du modèle (2.9) est que $\gamma < 0$, où γ est le plus grand exposant de Lyapounov de la suite $\{A_t, t \in \mathbb{Z}\}$ définie par (2.10). Cette solution stationnaire non anticipative, lorsque $\gamma < 0$, est unique et ergodique.*

2.6 Processus bilinéaire

Les modèles bilinéaires ont été introduits en 1978 par Granger et Andersen, de façon à généraliser les modèles *ARMA*,

$$X_t = \alpha_0 + \sum_{i=1}^p a_i X_{t-i} + \sum_{i=0}^q b_i \varepsilon_{t-i} + \sum_{i=1}^P \sum_{j=0}^Q c_{i,j} X_{t-i-j} \varepsilon_{t-i}, \quad t \in \mathbb{Z}. \quad (2.11)$$

où $\{\varepsilon_t\}$ est une suite de *v.a* i.i.d., les paramètres $\alpha_0, a_i, b_j, c_{i,j}$ sont des réels.

Pour introduire la stationnarité stricte, nous avons besoin d'une représentation markovienne de (2.11) Par une simple extension de Pham (1985) le modèle bilinéaire peut être écrit sous la forme suivante

$$\begin{aligned} X_t &= Y_{1,t-1} + b_0 \varepsilon_t, \\ Y_t &= A_t Y_{t-1} + B_t. \end{aligned} \quad (2.12)$$

où $Y_t = (Y_{1,t}, \dots, Y_{\bar{Q},t})'$ et $Y_{i,t} = (a_i + c_{i,0} \varepsilon_t) Y_{1,t-1} + Y_{i+1,t-1} + \sum_{j=1}^{\bar{P}} c_{i,j} \varepsilon_t Y_{\bar{Q}+j,t-1} + \{(a_i b_0 + b_i) \varepsilon_t + b_0 c_{i,0} \varepsilon_t^2\}$, $Y_{\bar{Q}+i,t} = X_{t+1-i}$ pour $i = 1, \dots, \bar{P}$.

$A_t = A_0 + A_1 \varepsilon_t$ et $B_t = B_0 + B_1 \varepsilon_t + B_2 \varepsilon_t^2$, $A_i \in \mathbb{R}^{n \times n}$, $i = 1, 2$, $B_i \in \mathbb{R}^n$, $i = 0, 1, 2$ et $n = \max\{p, P + q, P + Q\}$. Avec $\bar{Q} = \max\{q, Q\}$ et $\bar{P} = n - \bar{Q} = \max\{p - \bar{Q}, P\}$, on peut partitionner les deux matrices comme suit

$$A_t = \begin{pmatrix} A_{t,11} & A_{t,12} \\ A_{t,21} & A_{t,22} \end{pmatrix}, \quad B_t = \begin{pmatrix} B_{t,1} \\ B_{t,2} \end{pmatrix},$$

où

$$A_{t,11} = \begin{pmatrix} a_1 + c_{1,0} \varepsilon_t & & & \\ \vdots & & I_{\bar{Q}-1} & \\ a_{\bar{Q}-1} + c_{\bar{Q}-1,0} \varepsilon_t & & & \\ a_{\bar{Q}} + c_{\bar{Q},0} \varepsilon_t & & 0 \dots 0 & \end{pmatrix}, \quad A_{t,12} = \begin{pmatrix} c_{1,1} \varepsilon_t & \dots & c_{1,\bar{P}} \varepsilon_t \\ \vdots & & \vdots \\ c_{\bar{Q}-1,1} \varepsilon_t & & c_{\bar{Q}-1,\bar{P}} \varepsilon_t \\ a_{\bar{Q}} + c_{\bar{Q},1} \varepsilon_t & & a_{\bar{Q}+\bar{P}} + c_{\bar{Q},\bar{P}} \varepsilon_t \end{pmatrix}$$

$$A_{t,21} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad A_{t,22} = \begin{pmatrix} 0 & 0 \\ I_{\bar{P}-1} & \vdots \\ & 0 \end{pmatrix}$$

$$B_{t,1} = \begin{pmatrix} (a_1 b_0 + b_1) \varepsilon_t + b_0 c_{1,0} \varepsilon_t^2 \\ \vdots \\ (a_{\bar{Q}-1} b_0 + b_{\bar{Q}-1}) \varepsilon_t^2 + b_0 c_{\bar{Q}-1,0} \varepsilon_t^2 \\ a_0 + (a_{\bar{Q}} b_0 + b_{\bar{Q}}) \varepsilon_t^2 + b_0 c_{\bar{Q},0} \varepsilon_t^2 \end{pmatrix}, \quad B_{t,2} = \begin{pmatrix} b_0 \varepsilon_t^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Ici dans ce qui suit nous posons : a_i , b_i et c_i égaux à zéro quand l'indice i est supérieur à p, q et Q respectivement et/ou j supérieur à P . Définissons la norme d'une matrice A comme $\|A\|_{op} = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$ et l'exposant de Lyapounov

$$\gamma = \lim_{n \rightarrow \infty} \mathbb{E} \left[\log \left(\|A_n \dots A_1\|_{op} \right) / n \right] \stackrel{p.s.}{=} \lim_{n \rightarrow \infty} \log \left(\|A_n \dots A_1\|_{op} \right) / n$$

où $\mathbb{E} \left[\log^+ \left(\|A_0\|_{op} \right) \right] < \infty$, $\log^+ x = \max(\log x, 0)$.

Hypothèse 2.6.1 $\{\varepsilon_t\}$ est un processus i.i.d. de distribution non dégénéré, et $\mathbb{E} \log^+ (\varepsilon_t^2) < \infty$.

Théorème 2.6.2 *Supposons que l'hypothèse (2.6.1) est vérifiée. Alors $\gamma < 0$ est une condition suffisante pour (2.11) d'avoir une solution unique, strictement stationnaire et ergodique donnée par*

$$\begin{aligned} X_t^* &= Y_{1,t-1}^* + b_0 \varepsilon_t, \\ Y_t^* &= B_t + \sum_{i=1}^{\infty} A_t \dots A_{t-i} B_{t-i}. \end{aligned} \tag{2.13}$$

Pour toute valeur initiale Y_0 , le processus $Y_t \xrightarrow{p.s.} Y_t^*$ quand $t \rightarrow \infty$ de telle sorte que $X_t \xrightarrow{p.s.} X_t^*$. Si en outre le modèle est irréductible $\gamma < 0$ est une condition nécessaire pour la stricte stationnarité.

3 Processus empiriques

La théorie des processus empiriques joue un rôle important en statistique puisqu'elle concerne l'ensemble des résultats limites généraux se rapportant aux échantillons aléatoires. Ces modèles mathématiques, ayant fait l'objet de recherches approfondies depuis l'origine de la statistique moderne, mentionnons, parmi d'autres, les résultats liés aux théorèmes de Glivenko-Cantelli (1933), aux théorèmes de Donsker (1951), aux lois du logarithme itéré (Chung (1949)), ou encore aux lois limites fonctionnelles, globales ou locales, (Finkelstein (1971), Deheuvels (1992, 2000), Deheuvels et Mason (1992)). Stute (1982*a*, 1982*b*, 1986*a*, 1986*b*) est l'un des premiers statisticiens à faire un usage systématique de méthodes issues de la théorie des processus empiriques dans l'étude des propriétés asymptotiques d'estimateurs fonctionnels non paramétriques. Ses travaux ont concerné principalement les estimateurs à noyau. Depuis, de nombreux auteurs, parmi lesquels nous citerons Einmahl et Mason (2000, 2005), Deheuvels et Mason (2004), ont introduit des techniques nouvelles pour aborder ces problèmes, comme celles requérant l'usage de lois limites fonctionnelles, ou encore, mettant en œuvre des variantes locales de la théorie des processus empiriques indexés par des ensembles ou par des fonctions. Dans cette perspective, les idées de Vapnik et Cervonenkis (1971, 1981), Vapnik (1982) ont conduit à des innovations majeures. Ces travaux ont permis d'établir de nouvelles propriétés des estimateurs fonctionnels classiques (concernant, par exemple, l'estimation de la densité ou de la régression).

3.1 Définitions et résultats de base

Un processus empirique est un processus aléatoire basé sur un échantillon aléatoire. Par exemple, considérons un échantillon de *v.a* i.i.d. X_1, \dots, X_n de fonction de

distribution empirique donnée par

$$F_n(x) = n^{-1} \sum_{i=1}^n 1_{[X_i \leq x]},$$

où $x \in T = \mathbb{R}$.

Plus généralement, nous pouvons considérer l'échantillon aléatoire X_1, \dots, X_n de tirages indépendants à partir d'une mesure de probabilité P sur un espace d'échantillonnage arbitraire \mathcal{X} . Nous définissons la mesure empirique par

$$P_n = n^{-1} \sum_{i=1}^n \delta_{X_i},$$

où δ_x est la mesure de Dirac au point x ,

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

pour tout ensemble mesurable A .

Le processus empirique correspondant à la fonction de répartition empirique F_n est $W_n = n^{1/2}(F_n - F)$. La distribution empirique est l'estimateur naturel de F sous-jacente si celle-ci est complètement inconnue. Car $nF_n(x)$ est distribuée selon une binomiale de moyenne $nF(x)$, cet estimateur est sans biais.

Pour tout $x \in \mathbb{R}$, il ressort de la loi des grands nombres que

$$F_n(x) \xrightarrow[n \rightarrow \infty]{p.s.} F(x).$$

De plus, le théorème central limite garantit que

$$n^{1/2}(F_n(x) - F(x)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, F(x)(1 - F(x))).$$

Deux résultats de base concernant les termes F_n et $n^{1/2}(F_n(x) - F(x))$ sont le

théorème de Glivenko-Cantelli et le théorème de Donsker. Le premier théorème étant la loi des grands nombres et il donne la convergence uniforme presque sûre.

Théorème 3.1.1 (Glivenko-Cantelli, 1933) *Soit X_1, X_2, \dots une suite de v.a. réelles i.i.d. définies sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$, de fonction de répartition F . Alors*

$$\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{p.s.} 0.$$

Le théorème de Donsker donne la convergence en loi du processus empirique $n^{1/2}(F_n(x) - F(x))$ vers un processus gaussien.

Théorème 3.1.2 (Donsker, 1951) *Soit X_1, X_2, \dots des v.a. réelles i.i.d. définies sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$, de fonction de répartition F . Alors, la suite des processus empiriques $n^{1/2}(F_n(x) - F(x))$ convergent en loi sur l'espace $D[-\infty, \infty]$ des fonctions càdlàg (continues à droite, limite à gauche) vers un processus gaussien \mathbb{G}_F de moyenne zéro et de covariance donnée par*

$$\text{Cov}(\mathbb{G}_F(s), \mathbb{G}_F(t)) = F(s \wedge t) - F(s)F(t) \quad \forall s, t \in \mathbb{R}$$

où $s \wedge t$ est le minimum entre s et t .

Le processus \mathbb{G}_F est un processus F -pont Brownien, et est un pont Brownien standard (ou uniforme) si F est la distribution uniforme sur $[0, 1]$. Il y a des généralisations pour les théorèmes 3.1.1 et 3.1.2 pour un ensemble de fonctions mesurables qui permettent d'obtenir la classe de fonctions de Glivenko-Cantelli et de Donsker. Les deux classes interviennent dans l'étude des processus indexés par un ensemble de fonctions, (pour plus de détails voir van der Vaart et Wellner, 1996, van der Vaart, 1998).

3.2 Processus indexé par des fonctions

Soit X_1, \dots, X_n un échantillon aléatoires de fonction de répartition F , sur un espace mesurable (χ, ε) . Pour une fonction f réelle mesurable définie sur (χ, ε) , écrivons

$$F_n f = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad F f = \int f dF.$$

Alors $\{F_n f : f \in \mathcal{A}\}$ est la mesure empirique indexé par \mathcal{A} , alors que $\{\mathbb{G}_n f : f \in \mathcal{A}\}$ est un processus empirique indexé par \mathcal{A} , où

$$\mathbb{G}_n f = n^{1/2}(F_n - F)f = n^{1/2} \frac{1}{n} \left(\sum_{i=1}^n f(X_i) - Ff \right)$$

Si $f \in L^1(F)$, donc $Ff = \int f dF < \infty$, alors de la loi des grands nombres, on déduit que

$$F_n(f) \xrightarrow[n \rightarrow \infty]{p.s.} F(f). \quad (3.1)$$

Supposons que \mathcal{A} est un ensemble de fonctions réelles $f : \chi \rightarrow \mathbb{R}$. Si nous avons la convergence comme dans (3.1), uniformément pour toute $f \in \mathcal{A}$

$$\|F_n f - F f\|_{\mathcal{A}} = \sup_{f \in \mathcal{A}} |F_n f - F f| \xrightarrow[n \rightarrow \infty]{p.s.*} 0,$$

³alors la classe $\mathcal{A} \subset L^1(F)$ est dite F -Glivenko-Cantelli.

Remarque 3.2.1 *La distance aléatoire $\|F_n - F\|_{\mathcal{A}}$ n'est pas forcément mesurable, d'où l'utilisation, pour contourner cette difficulté, de la convergence $p.s.*$. \mathcal{A} est aussi dite F -Glivenko-Cantelli faible lorsque la convergence a lieu en probabilité.*

Si $f \in L^2(F)$, donc $Ff^2 = \int f^2 dF < \infty$, alors en utilisant le théorème central

³On dit que X_n converge *outer* presque sûrement vers X s'il existe une suite Δ_n de variables aléatoires mesurables avec $d(X_n, X) \leq \Delta_n \forall n \in \mathbb{N}$ et avec $P\{\limsup_{n \rightarrow \infty} \Delta_n = 0\} = 1$, et on note $X_n \xrightarrow{a.s.*} X$. Bien que ce mode de convergence soit un peu différent de celui standard, il est identique lorsque tous les quantités concernées sont mesurables. Les propriétés du mode standard sont aussi généralement conservés dans ce nouveau mode.

limite, on obtient

$$n^{1/2}(F_n f - F f) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, F(f - F f)^2). \quad (3.2)$$

Supposons que \mathcal{A} est un ensemble de fonctions réelles $f : \chi \rightarrow \mathbb{R}$. Si nous avons la convergence comme dans (3.2), uniformément pour toute $f \in \mathcal{A}$ alors

$$n^{1/2}(F_n - F)(f) \Longrightarrow \mathbb{G}(f) \quad \text{dans } \ell^\infty(\mathcal{A})$$

⁴et \mathbb{G} est un processus pont Brownien de moyenne zéro de fonction covariance

$$\text{Cov}(\mathbb{G}(f), \mathbb{G}(g)) = Ffg - FfFg.$$

Alors \mathcal{A} est dite classe de Donsker.

$\ell^\infty(\mathcal{A})$ est l'espace défini comme l'ensemble de toutes les fonctions réelles uniformément bornées sur \mathcal{A} i.e., de toutes les fonctions $z : \mathcal{A} \rightarrow \mathbb{R}$ tel que $\|z\|_{\mathcal{A}} = \sup_{f \in \mathcal{A}} |z(f)| < \infty$.

⁴Pour des éléments aléatoires $X_n, n \geq 0$ sur $(M, \mathcal{M}_\delta^B)$, on dit que X_n converge faiblement vers X_0 ou P_n (mesure de probabilité induite dans $(M, \mathcal{M}_\delta^B)$ par X_n) converge faiblement vers P_0 à condition $\int_\chi f dP_n = \mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X_0) = \int_\chi f dP_0$ quand $n \rightarrow \infty$ pour toute fonction réelles sur M , bornée, δ -uniformément continue et δ -mesurable, notée

$$\begin{aligned} X_n &\Rightarrow X_0 \text{ sur } (\chi, \varepsilon, \delta) \text{ quand } n \rightarrow \infty \\ \text{ou } P_n &\Rightarrow P_0 \text{ sur } (\chi, \varepsilon, \delta) \text{ quand } n \rightarrow \infty \end{aligned}$$

4 Distribution de la statistique de Cramèr-von Mises

Le but est de trouver la distribution de la statistique de Cramèr-von Mises

$$\begin{aligned} W_n^2 &= \int_{-\infty}^{+\infty} n [F_n(x) - F(x)]^2 dF(x) \\ &= \int_{-\infty}^{+\infty} \mathbb{U}_n^2(F(x)) dF(x) \\ &= \int_0^1 \mathbb{U}_n^2(t) dt. \end{aligned}$$

Tout comme une matrice de covariance Σ qui peut être représentée comme $\Sigma = \sum_{j=1}^n \lambda_j \gamma_j \gamma_j'$ où les λ_j sont les valeurs propres et γ_j les vecteurs propres orthonormés de Σ , de même la fonction covariance K de plusieurs processus peut être représentée comme

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \quad (4.1)$$

Pour des fonction f_j orthonormées par rapport à la métrique de L^2 . Soient Z_1^*, Z_2^*, \dots des *v.a* i.i.d. $\mathcal{N}(0, 1)$. Comme $\sum_{j=1}^n \sqrt{\lambda_j} Z_j^* \gamma_j$ est une *v.a.* $\mathcal{N}(0, \Sigma)$, de même pour le processus \mathbb{X} défini par

$$\mathbb{X}(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j^* f_j(t) \quad (4.2)$$

est un processus normal de fonction moyenne nulle et de fonction de covariance K .

En intégrant (4.2), nous voyons que

$$\int_0^1 \mathbb{X}^2(t) dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \quad (4.3)$$

est distribué comme une somme infinie de *v.a.* indépendantes de khi deux pondérées.

4.1 Motivation de la décomposition en composantes principales

4.1.1 Enoncé du problème

Sous l'hypothèse nulle d'indépendance, la statistique de Cramèr-von Mises se réduit à la forme

$$W_n^2 = \int_0^1 [U_n(t)]^2 dt \xrightarrow[n \rightarrow \infty]{\mathcal{L}} W^2 = \int_0^1 [U(t)]^2 dt \quad (4.4)$$

Soit Y un vecteur $m \times 1$ dont la $i^{\text{ème}}$ composante est $\mathbb{U}(i/(m+1))/\sqrt{m}$. Alors $Y'Y \xrightarrow{p.s} W^2$ quand $n \rightarrow \infty$. Dans la prochaine section nous décomposons un vecteur aléatoire Y arbitraire de loi $(0, \Sigma)$ en composantes principales et puis nous exprimons $Y'Y$ en terme de ses composantes principales. Après cela, nous donnons un traitement heuristique (qui va être traité rigoureusement) pour une telle décomposition du processus \mathbb{X} en composantes principales sur $[0, 1]$, et ensuite on exprime $\int_0^1 \mathbb{X}^2(t) dt$ en terme de ses composantes principales.

4.1.2 Décompositions en composantes principales des vecteurs aléatoires

Soit le vecteur aléatoire $n \times 1$

$$Y \cong (0, \Sigma) \quad (4.5)$$

Nous savons par le théorème des axes principaux ou théorème spectral (in Lay 2003, p 458) qu'il existe une matrice orthogonale Γ (de lignes $\gamma'_1, \dots, \gamma'_n$ qui forment une base orthonormée pour l'espace de dimension n) pour la quelle

$$Z = \Gamma Y \cong (0, \Lambda) \quad (4.6)$$

avec $\Lambda = \Gamma \Sigma \Gamma'$, une matrice diagonale dont les éléments diagonaux sont $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ qui sont toutes ≥ 0 (> 0 pour Σ non singulière). Les coordonnées

$$Z_j = \gamma'_j Y = \langle \gamma_j, Y \rangle \text{ pour } 1 \leq j \leq n \quad (4.7)$$

sont les composante principales de Y , notons que les λ_j et les γ_j sont obtenus comme des solutions de l'équation

$$\gamma' \Sigma = \lambda \gamma' \quad (4.8)$$

Nous avons

$$\Sigma = \Gamma' \Lambda \Gamma = \sum_{j=1}^n \lambda_j \begin{bmatrix} \gamma_j \gamma'_j \end{bmatrix} \quad (4.9)$$

Notons que $\begin{bmatrix} \gamma_j \gamma'_j \end{bmatrix} y = \gamma_j (\gamma'_j y) = \langle y, \gamma'_j \rangle \gamma_j$ est la projection de y sur le vecteur unité dans la direction de γ_j , ainsi chacune des matrices

$$\begin{bmatrix} \gamma_j \gamma'_j \end{bmatrix} \text{ défini la projection sur } \gamma_j. \quad (4.10)$$

Aussi

$$Y = \sum_{j=1}^n \langle Y, \gamma_j \rangle \gamma_j = \sum_{j=1}^n Z_j \gamma_j \quad (4.11)$$

$$= \sum_{j=1}^n \sqrt{\lambda_j} Z_j^* \gamma_j \quad (4.12)$$

où $Z_j^* = Z_j / \sqrt{\lambda_j}$ des *v.a.* $(0, 1)$ non corrélées.

Présentons Y en termes des composantes principales Z_j en (4.11) et en termes des composantes principales normalisées Z_j^* en (4.12). De (4.11) ou de (4.6), nous avons

$$Y' Y = \sum_{j=1}^n \langle \gamma_j, Y \rangle^2 = \sum_{j=1}^n Z_j^2 = \sum_{j=1}^n \lambda_j Z_j^{*2} \quad (4.13)$$

Finalement, notons que

$$Z_1^*, \dots, Z_n^* \text{ sont indépendantes } \mathcal{N}(0, 1) \text{ si } Y \cong \mathcal{N}(0, \Sigma). \quad (4.14)$$

Dans ce cas, $Y'Y$ est distribué comme une somme pondérée des khi deux indépendantes, où les poids sont les valeurs propres de la matrice covariance Σ et les Z_j^* sont les projections de Y sur le vecteur propre associé.

4.1.3 Décomposition (heuristique) en composantes principales d'un processus

Les processus présentant une propriété de régularité ont des trajectoires complètement spécifiées si on connaît les trajectoires en tout point d'un ensemble dénombrable dense. C'est ainsi qu'un nombre dénombrable de variables devrait être suffisant pour définir un processus comme le pont Brownien. Il semble raisonnable que la décomposition harmonique d'un processus de moyenne nulle $\{\mathbb{X}(t), 0 \leq t \leq 1\}$ pourrait prendre la forme

$$\mathbb{X} = \sum_{j=1}^{\infty} Z_j f_j \quad (4.15)$$

où les Z_j sont des *v.a.* non Corrélées de moyenne zéro et les f_j sont des fonctions orthonormées sur $[0, 1]$. Nous utilisons la notation

$$\langle g_1, g_2 \rangle = \int_0^1 g_1(t) g_2(t) dt \quad (4.16)$$

et

$$\|g\| = \langle g, g \rangle^{1/2} = \left[\int_0^1 g^2(t) dt \right]^{1/2}$$

si la représentation de (4.15) est vraie, des calculs purement formels suggèrent

$$\langle \mathbb{X}, f_k \rangle = \left\langle \sum_{j=1}^{\infty} Z_j f_j, f_k \right\rangle \quad (4.17)$$

$$= \sum_{j=1}^{\infty} Z_j \langle f_j, f_k \rangle = Z_k \quad (4.18)$$

et aussi

$$\begin{aligned} K(s, t) &= \text{cov} [\mathbb{X}(s), \mathbb{X}(t)] = \mathbb{E} \left[\sum_{j=1}^{\infty} Z_j f_j(s) \sum_{k=1}^{\infty} Z_k f_k(t) \right] \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \text{cov}(Z_j, Z_k) f_j(s) f_k(t) \\ &= \sum_{j=1}^{\infty} \text{var}(Z_j) f_j(s) f_j(t) \\ &= \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \end{aligned} \quad (4.19)$$

où $\lambda_j = \text{var}(Z_j)$ et $\text{cov}(Z_j, Z_k) = 0$ si $j \neq k$. L'équation (4.19) suggère comment les f_j peuvent être déterminées, nous avons

$$\begin{aligned} \int_0^1 f_k(s) K(s, t) ds &= \int_0^1 f_k(s) \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) ds \\ &= \sum_{j=1}^{\infty} \lambda_j f_j(t) \langle f_k, f_j \rangle \\ &= \lambda_k f_k(t) \end{aligned}$$

Nous cherchons donc toutes les solutions (valeurs propres λ et fonctions propres f) de l'équation intégrale

$$\int_0^1 f(s) K(s, t) ds = \lambda f(t) \quad 0 \leq t \leq 1 \quad (4.20)$$

Notons que (4.19) porte en elle l'implication que

$$\begin{aligned}
 \text{cov}(Z_j, Z_k) &= \mathbb{E} \left[\int_0^1 \mathbb{X}(s) f_j(s) ds \int_0^1 \mathbb{X}(t) f_k(t) dt \right] \\
 &= \int_0^1 \left[\int_0^1 f_j(s) K(s, t) ds \right] f_k(t) dt \\
 &= \int_0^1 \lambda_j f_j(t) f_k(t) dt \\
 &= \lambda_j \delta_{jk}
 \end{aligned}$$

où δ_{jk} est le symbole de Kronecker.

Les variables de (4.17) sont les composantes principales empiriques de \mathbb{X} , et (4.15) est dite la décomposition en composantes principales de \mathbb{X} . Soit $Z_j^* = Z_j / \sqrt{\lambda_j}$ les composantes principales normalisées, comme $Z_j^* \cong \mathcal{N}(0, 1)$, notons que l'équation (4.15) peut être exprimée comme

$$\mathbb{X} = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j^* f_j \quad (4.21)$$

Finalement de (4.21) nous avons

$$\int_0^1 \mathbb{X}^2(t) dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \quad (4.22)$$

pour Z_j^* non corrélées $(0, 1)$.

Ainsi, $\int_0^1 \mathbb{X}^2(t) dt$ est distribuée comme une somme infinie des khi deux indépendantes pondérées (4.4), dans le cas où \mathbb{X} est un processus gaussien, les poids sont les valeurs propres du noyau de covariance et Z_j sont les projections de \mathbb{X} sur les fonctions propres associées.

4.2 Décomposition orthogonale des processus

4.2.1 Résultats fondamentaux sur les noyaux

Soit L_2 l'ensemble de toutes les fonctions réelles mesurables g sur $[0, 1]$ ayant $\int_0^1 g^2(t) dt < \infty$; ici dt est la mesure de Lebesgue. Soit $\langle g_1, g_2 \rangle = \int_0^1 g_1(t) g_2(t) dt$ le produit scalaire sur L_2 , et soit $\|g\| = \langle g, g \rangle^{1/2} = \left[\int_0^1 g^2(t) dt \right]^{1/2}$ la norme usuelle sur L_2 . Rappelons que L_2 est un espace linéaire normé complet ou espace de Hilbert. Un noyau désigne une fonction non nulle, symétrique⁵ et de carré intégrable $K(s, t)$ définie pour $0 \leq s, t \leq 1$ i.e.,

$$K \neq 0, K(s, t) = K(t, s) \text{ et } \int_0^1 \int_0^1 K^2(s, t) ds dt < \infty. \quad (4.23)$$

Le noyau est dit défini positif si

$$\int_0^1 \int_0^1 g(s) K(s, t) g(t) ds dt > 0 \quad (4.24)$$

pour toute $g \in L^2$ avec $g \neq 0$; et est semi défini positif si (4.24) ≥ 0 .

Si pour un nombre réel λ il existe une fonction f dans L^2 tel que

$$\int_0^1 f(s) K(s, t) ds = \lambda f(t) \quad (4.25)$$

alors λ est la valeur propre du noyau K et f est la fonction propre associée.

4.2.2 Propriétés des noyaux

Les points suivants donnent quelques propriétés importantes des noyaux symétriques.

1. Les valeurs propres du noyau sont au plus dénombrables.
2. Les fonctions propres correspondantes à des valeurs propres distinctes sont

⁵C'est la symétrie considérée tout au long de ce mémoire e.i., la fonction est invariante en permutant ses arguments, $K(s, t) = K(t, s)$.

- orthogonales (i.e., $\langle f_1, f_2 \rangle = 0$ pour les fonctions propres associées aux valeurs propres $\lambda_1 \neq \lambda_2$)
3. Correspondant à une valeur propre non nulle, il y a au plus un nombre fini de fonctions linéairement indépendantes associées; ce nombre est appelé la multiplicité de la valeur propre.
 4. Si $\lambda_1, \lambda_2, \dots$ des valeurs propres non nulles, chacune apparaissant autant de fois que sa multiplicité, alors les fonctions propres correspondantes f_1, f_2, \dots devraient être orthonormées (i.e., $\langle f_i, f_j \rangle = 0$ et $\|f_i\| = 1 \forall i, j$)
 5. Nous avons nécessairement $\sum_i \lambda_i^2 \leq \int_0^1 \int_0^1 K^2(s, t) ds dt$ ce qui implique que, s'il y a un nombre infini de valeurs propres λ_j , alors $\lambda_j \rightarrow 0$ quand $j \rightarrow \infty$.
 6. La plus grande valeur propre λ non nulle en valeur absolue satisfait $|\lambda| = [\sup \int_0^1 \int_0^1 g(s) K(s, t) g(t) ds dt; g \in L^2$ telle que $\|g\| = 1]$ (Kanwal, 1971).
 7. Le noyau K est semi défini positif si et seulement si toutes les valeurs propres non nulles λ_j sont strictement positives.
 8. Le noyau K est défini positif si et seulement si toutes les valeurs propres λ_j sont strictement positives et les fonctions propres correspondantes f_j forment un système orthonormé complet.

Proposition 4.2.1 (in Shorack et Wellner, 2009, p 208) *L'ensemble orthonormé f_1, f_2, \dots dans L_2 est complet, signifie que les conditions équivalentes suivantes sont vérifiées*

- i) f_1, f_2, \dots est un ensemble orthonormé maximal.
- ii) L'ensemble de toute les combinaisons linéaires, finies des f_j est $\|[\cdot]\|$ -dense dans L_2 .
- iii) $\forall g \in L_2$, nous avons $\|g\|^2 = \sum_{j=1}^{\infty} \langle g, f_j \rangle^2$.
- iv) $\forall g, h \in L_2$, nous avons $\langle g, h \rangle = \sum_{j=1}^{\infty} \langle g, f_j \rangle \langle h, f_j \rangle$.

Remarquons que, pour chaque m fixé le choix $a_j = \langle g, f_j \rangle$ minimise $\left\| \left[\sum_{j=1}^m a_j f_j - g \right] \right\|^2$; ainsi (ii) implique que $\sum_{j=1}^m a_j f_j$ converge vers g au sens de $[\cdot]$ p.s. quand $m \rightarrow \infty$ et on note $\sum_{j=1}^m a_j f_j \rightarrow_{[\cdot]} g$.

Théorème 4.2.1 (Mercer, 1951) *Si K est continu pour $0 \leq s, t \leq 1$ et est un noyau semi défini positif (qui vérifie 4.23), alors*

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \quad 0 \leq s, t \leq 1 \quad (4.26)$$

où la série converge uniformément et absolument. Aussi, $\sum_{j=1}^{\infty} \lambda_j < \infty$ (ici les λ_j et les f_j sont respectivement les valeurs et les vecteurs propres associés à (4.25), les f_j sont orthonormées).

4.2.3 Existence de décomposition en composantes principales

De ce qui précède, nous commençons par l'appliquer aux processus sur $[0, 1]$.

Proposition 4.2.2 (in Shorack et Wellner, 2009) *Soit K la fonction covariance d'un processus non trivial $\{\mathbb{X}(t), 0 \leq t \leq 1\}$ de moyenne nulle dont les trajectoires sont des fonctions de L_2 , p.s.*

Si K vérifie

$$\int_0^1 K(t, t) dt < \infty$$

alors K est un noyau semi défini positif. Si λ_1, f_1 et λ_2, f_2 sont deux solutions de (4.25) pour lesquelles f_1 et f_2 sont orthonormées, alors

$$Z_j = \int_0^1 \mathbb{X}(t) f_j(t) dt \quad \text{pour } j = 1, 2$$

sont des v.a. non corrélées $(0, \lambda_j)$.

Hypothèse 4.2.2 Supposons que K est la fonction de covariance du processus non trivial $\{\mathbb{X}(t), 0 \leq t \leq 1\}$ de moyenne nulle dont les trajectoires sont dans le sous

espace \mathcal{L}^6 de L_2 , *p.s.*, et supposons que K satisfait

$$\int_0^1 K(t, t) dt < \infty \quad (4.27)$$

Supposons aussi que $\lambda_1 \geq \lambda_2 \geq \dots > 0$ comprend la totalité du spectre des valeurs propres de K , que les fonctions propres orthonormées associées f_1, f_2, \dots de l'ensemble complet de sous espace \mathcal{L} , et que nous avons la décomposition de Kac et Siegert (1947)

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \quad 0 \leq s, t \leq 1 \quad (4.28)$$

de la proposition 4.2.2

$$Z_j = \int_0^1 \mathbb{X}(t) f_j(t) dt \cong (0, \lambda_j) \quad j \geq 1 \quad (4.29)$$

sont des *v.a.* non corrélées.

Remarque 4.2.1 *Si on suppose que K de la proposition 4.2.2 est continu, le théorème de Mercer garantit que (4.28) est toujours vérifiée ; en fait, la convergence dans (4.28) est uniforme et absolue dans ce cas, et $\sum_{j=1}^{\infty} \lambda_j < \infty$. Cependant, il y a un exemple intéressant où (4.28) est vérifiée sans que K soit continu. (théorème 5.4.1 de Anderson et Darling, 1952).*

Théorème 4.2.3 (Kac et Siegert, 1947) *Supposons que \mathbb{X} et K vérifient l'hypothèse 4.2.2. Alors*

$$\sum_{j=1}^{\infty} \lambda_j < \infty \quad (4.30)$$

$$\sum_{j=1}^m Z_j f_j \xrightarrow{||\cdot||} \mathbb{X} \quad m \rightarrow \infty \quad p.s \quad (4.31)$$

où

$$Z_j = \langle \mathbb{X}, f_j \rangle \cong (0, \lambda_j) \quad \text{non corrélées} \quad (4.32)$$

⁶Si F et F_0 sont continues, alors $L = \{t \in (-\infty, +\infty) : F(t) - F_0(t) = \|F - F_0\|\}$.

$$\int_0^1 \mathbb{X}^2(t) dt \stackrel{p.s.}{=} \sum_{j=1}^{\infty} Z_j^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \quad (4.33)$$

où $Z_j^* = Z_j/\sqrt{\lambda_j} \cong (0, 1)$.

$$\mathbb{E} \left(\mathbb{X}(t) - \sum_{j=1}^m Z_j f_j(t) \right)^2 \rightarrow 0, \quad \forall t \text{ quand } m \rightarrow \infty \quad (4.34)$$

$$\mathbb{X} = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j^* f_j$$

avec $Z_j^* = Z_j/\sqrt{\lambda_j}$ des variables non corrélées $(0, 1)$.

4.2.4 Distribution de $\int_0^1 \mathbb{X}^2(t) dt$ via la décomposition

Supposons que

$$T = \int_0^1 \mathbb{X}^2(t) dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2} \quad (4.35)$$

où Z_j^* sont des *v.a.* i.i.d. $\mathcal{N}(0, 1)$. (c'est le cas si nous ajoutons l'hypothèse de normalité au théorème 4.2.3).

Comment utiliser cette représentation pour obtenir la distribution ?

Méthode 1 La *v.a.* T de (4.35) a pour fonction caractéristique

$$\phi_T(\lambda) = \mathbb{E}(e^{i\lambda T}) = \prod_{j=1}^{\infty} (1 - 2i\lambda_j \lambda)^{-1/2} \quad (4.36)$$

Il est possible d'inverser cette fonction caractéristique (ou la fonction génératrice des moments). Des exemples de cette technique sont donnés dans Anderson et Darling (1952) pour le traitement de W^2 et Watsons (1961) pour le traitement de U^2 .

Méthode 2 Si on tronque la série infinie dans (4.35), alors la somme des *v.a.* indépendantes de khi deux pondérées qui reste, peut être traitée par des méthodes numériques de Imhof (1961) ou Slepian (1968).

Méthode 3 Il est possible de calculer les premiers moments de T , puis les relier

avec la courbe de Pearson⁷ (voir Stephens, 1976).

Méthode 4 On peut utiliser la méthode de Monte Carlo (voir Stephens 1976).

Calculons maintenant les moments de T (pour plus de détails voir Anderson et Darling, 1952).

k_m : le $m^{\text{ème}}$ cumulant⁸ de T qui est donnée par les coefficients de $(i\lambda)^m / m!$ dans le développement en série de $\log \phi_T(\lambda)$ de (4.36). Ainsi

$$k_m = 2^{m-1} (m-1)! \sum_{j=1}^{\infty} \lambda_j^m \quad (4.37)$$

alors

$$\mathbb{E}(T) = \sum_{j=1}^{\infty} \lambda_j \text{ et } Var(T) = 2 \sum_{j=1}^{\infty} \lambda_j^2 \quad (4.38)$$

ce développement nécessite la connaissance de toutes les λ_j . Notons que, quand l'expression (4.29) est vérifiée, avec $\int_0^1 K(t, t) dt < \infty$, nous avons

$$\int_0^1 K(t, t) dt = \int_0^1 \sum_{j=1}^{\infty} \lambda_j f_j(t)^2 dt = \sum_{j=1}^{\infty} \lambda_j = \mathbb{E}(T) \quad (4.39)$$

⁷Un système de distributions proposé par Karl Pearson est constitué de sept solutions à l'équation différentielle qui donne une approximation d'une vaste gamme de distributions de différentes formes. Gruska, Mirkhani & Lamberson (1989) décrivent en détails la façon dont les différentes courbes de Pearson peuvent être ajustées par une distribution empirique.

⁸Pour une variable aléatoire X de moyenne zéro et de fonction caractéristique donnée par :

$$\varphi_X(\lambda) = \mathbb{E}(e^{i\lambda X})$$

Développons en série de Taylor le logarithme de la fonction caractéristique, nous obtenons

$$\ln \varphi_X(\lambda) = k_1(it) + k_2(it)^2/2 + \dots + k_r(it)^r/r! + \dots$$

où k_r sont des constantes appelées cumulants de la distribution de X .

et

$$\begin{aligned}
& 2 \int_0^1 \int_0^1 K^2(s, t) dt ds \\
&= 2 \int_0^1 \int_0^1 \sum_{j=1}^{\infty} \lambda_j \lambda_{j'} f_j(s) f_{j'}(s) f_j(t) f_{j'}(t) ds dt \\
&= 2 \sum_{j=1}^{\infty} \lambda_j^2 = \text{Var}(T) \text{ car le système } \{f_j(\cdot), j \in \mathbb{N}\} \text{ est orthonormal} \quad (4.40)
\end{aligned}$$

(le théorème de Fubini permet la permutation entre les opérateurs). Plus généralement,

$$k_m = 2^{m-1} (m-1)! \int_0^1 K_m(t, t) dt$$

$$\begin{aligned}
K_m(s, t) &= \int_0^1 K_{m-1}(s, r) K(r, t) dr \\
&= \sum_{j=1}^{\infty} \lambda_j^m f_j(s) f_j(t).
\end{aligned}$$

4.3 Décomposition en composantes principales de \mathbb{U}_n et \mathbb{U}

Proposition 4.3.1 (in Shorack et Wellner, 2009) *La fonction de covariance d'un processus gaussien \mathbb{U}*

$$K_{\mathbb{U}}(s, t) = s \wedge t - st \quad 0 \leq s, t \leq 1$$

$$s \wedge t = \min(s, t)$$

peut être décomposée comme suit :

$$K_{\mathbb{U}}(s, t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \quad 0 \leq s, t \leq 1 \quad (4.41)$$

où

$$\lambda_j = (j\pi)^{-2} \text{ et } f_j(t) = \sqrt{2} \sin(j\pi t) \quad \text{pour } 0 \leq t \leq 1. \quad (4.42)$$

$\{f_j(\cdot), j \in \mathbb{N}\}$ formant un système orthonormé. La série dans (4.41) converge uniformément et absolument sur $[0, 1]$.

Après avoir décomposé $K_{\mathbb{U}}$ (ce qui équivaut à $K_{\mathbb{U}_n}$), nous avons les résultats suivants

Théorème 4.3.1 (in Shorack et Wellner, 2009) *Nous avons*

$$\mathbb{U} \cong \sum_{j=1}^{\infty} \frac{Z_j^* f_j}{j\pi} \quad \text{sur } (C, \mathfrak{C}),$$

où $f_j(t) = \sqrt{2} \sin(j\pi t)$ comme dans la proposition 4.3.1 et où Z_1^*, Z_2^*, \dots sont des v.a i.i.d. $\mathcal{N}(0, 1)$.

De plus,

$$W^2 = \int_0^1 \mathbb{U}^2(t) dt = \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{(j\pi)^2}.$$

5 U -statistiques

D'un point de vue purement mathématique, il peut être souhaitable et approprié de voir n'importe quelle statistique comme un membre d'une classe générale de statistiques qui possèdent certaines caractéristiques communes et importantes. Ainsi, plusieurs statistiques intéressantes ont été formulées comme une généralisation de statistiques particulières.

La classe des statistiques qui va être considérée dans cette partie a été introduite dans l'article fondamental de Hoeffding (1948a). Le développement repose sur les travaux de Halmos (1946). La Classe se présente comme une généralisation de la moyenne empirique. Typiquement, les membres de la classe sont des statistiques asymptotiquement normales. Elles ont aussi de bonnes propriétés de convergence.

5.1 Descriptions de la U -statistique

Les U -Statistiques sont étroitement, liées à la classe des statistiques introduite par von Mises (1947). De nombreuses statistiques d'intérêt appartiennent à ces deux classes et plusieurs autres peuvent être approximées par un membre de ses classes.

Des définitions de bases sont données dans 5.1.1, et une classe de statistiques plus générale est présentée dans 5.1.2. Ces considérations s'appliquent à la U -statistique fondée sur un échantillon. Une généralisation à plusieurs échantillons est donnée dans 5.1.3, et à des versions pondérées dans 5.1.7. Une propriété importante d'optimalité des U -statistiques, dans l'estimation sans biais, est montrée dans 5.1.4. La représentation des U -statistiques comme une martingale fait l'objet de 5.1.5, et comme une moyenne de moyennes i.i.d. dans 5.1.6.

5.1.1 Définitions

Soit \mathfrak{F} une famille de lois de probabilité sur un espace quelconque mesurable. L'approche adoptée ici, est non paramétrique. Ceci signifie que \mathfrak{F} est considérée comme une grande famille de distributions sous des restrictions telles que la continuité ou l'existence des moments. Soit $\theta(F)$ une fonctionnelle à valeur réelles définie sur \mathfrak{F} .

Définition 5.1.1 *On dit que $\theta(F)$ est un paramètre estimable dans \mathfrak{F} si pour $m \in \mathbb{N}^*$, il existe un estimateur sans biais de $\theta(F)$ basé sur m v.a. i.i.d. de fonction de répartition F c'est-à-dire, une fonction mesurable $h(x_1, \dots, x_m)$, à valeurs réelles, telle que*

$$\theta(F) = \mathbb{E}_F(h(X_1, \dots, X_m)) = \int \dots \int h(x_1, \dots, x_m) dF(x_1) \dots dF(x_m) \quad \text{pour tout } F \in \mathfrak{F}.$$

Le plus petit entier m tel que $\theta(F)$ est estimable est le degré de $\theta(F)$. Il convient de noter que la fonction h peut être supposée symétrique en ses arguments i.e., invariante par permutation de ses arguments. Ceci parce que si f est un estimateur

sans biais de $\theta(F)$, alors la moyenne de f appliquée à toutes les permutations des variables est toujours sans biais et est, en outre, symétrique.

$$h(x_1, \dots, x_m) = \frac{1}{m!} \sum_{\pi_i \in S_m} f(x_{\pi_1}, \dots, x_{\pi_m})$$

où $\sum_{\pi_i \in S_m}$ est la somme sur toutes les $m!$ permutations (π_1, \dots, π_m) de $(1, \dots, m)$ et S_m est le groupe des permutations de l'ensemble $\{1, 2, \dots, m\}$.

Définition 5.1.2 Soit X_1, \dots, X_n un n -échantillon de fonction de répartition F où l'on suppose que $n \geq m$ et soit h , une fonction réelle mesurable définie sur \mathbb{R}^m , une U -statistique de noyau h est définie comme suit

$$U_n = \left(\frac{n!}{(n-m)!} \right)^{-1} \sum_p h(X_{i_1}, \dots, X_{i_m}), \quad (5.1)$$

où \sum_p est la somme sur les $n!/(n-m)!$ arrangements de m entiers choisis parmi $(1, \dots, n)$. Si le noyau h est symétrique, U_n a la forme équivalente

$$\left(\frac{n!}{m!(n-m)!} \right)^{-1} \sum_c h(X_{i_1}, \dots, X_{i_m}). \quad (5.2)$$

\sum_c est la somme sur les $n!/m!(n-m)!$ combinaisons de m entiers $i_1 < \dots < i_m$ choisis parmi $(1, \dots, n)$.

Si $\theta(F) = \mathbb{E}_F(h(X_1, \dots, X_m))$ existe pour toute $F \in \mathfrak{F}$, alors U_n est l'estimateur sans biais de $\theta(F)$. De plus il a la propriété d'être le meilleur estimateur sans biais de $\theta(F)$ (au sens de la variance minimale) si \mathfrak{F} est suffisamment grande. Par exemple si elle contient toutes les distributions F pour lesquelles $\theta(F)$ est finie, alors les statistiques d'ordre forment une statistique exhaustive complète de $F \in \mathfrak{F}$.

5.1.2 Quelques statistiques proches : les V -statistiques

Soit U_n , une U -statistique

$$U_n = \binom{n}{m}^{-1} \sum_c h(X_{i_1}, \dots, X_{i_m}).$$

La statistique de von Mises correspondante est définie par

$$\begin{aligned} V_n &= \frac{1}{n^m} \sum_{i_1=1}^n \dots \sum_{i_m=1}^n h(X_{i_1}, \dots, X_{i_m}), \\ &= \theta(F_n), \end{aligned} \quad (5.3)$$

où F_n est la fonction de répartition empirique.

5.1.3 U -statistiques généralisées

L'extension au cas de plusieurs échantillons est directe. Considérons les groupes k -dépendants⁹ d'observations indépendantes $\{X_1^{(1)}, X_2^{(1)}, \dots\}, \dots, \{X_1^{(k)}, X_2^{(k)}, \dots\}$ de distributions $F^{(1)}, \dots, F^{(k)}$ respectivement. Soit $\theta = \theta(F^{(1)}, \dots, F^{(k)})$, la fonction paramétrique pour laquelle il existe un estimateur sans biais.

$$\theta = \mathbb{E} \left\{ h \left(X_1^{(1)}, \dots, X_{m_1}^{(1)}, \dots, X_1^{(k)}, \dots, X_{m_k}^{(k)} \right) \right\}, \quad (5.4)$$

où h est définie comme ci-dessus, symétrique à l'intérieur de chacun des k blocs. pour un noyau h , $n_1 \geq m_1, \dots, n_k \geq m_k$, la U -statistique est définie par

$$U_n = \left[\prod_{j=1}^k \binom{n_j}{m_j} \right]^{-1} \sum_c h(X_{i_{11}}^{(1)}, \dots, X_{i_{1m_1}}^{(1)}, \dots, X_{i_{k1}}^{(k)}, \dots, X_{i_{km_k}}^{(k)}). \quad (5.5)$$

Ici $\{i_{j1}, \dots, i_{jm_j}\}$ est l'ensemble de m_j éléments distincts de l'ensemble $\{1, 2, \dots, n_j\}$,

⁹Soit $\{X_t\}$ une suite de variables aléatoires strictement stationnaire, $\{X_t\}$ est dite α -dépendant $\alpha \in \mathbb{N}$, si pour tout t les ensembles des variables aléatoires $\{X_j, j \leq t\}$ et $\{X_j, j \geq t + \alpha + 1\}$ sont indépendants.

$1 \leq j \leq k$, et \sum_c est la somme sur toutes les $n_j!/m_j!(n_j - m_j)!$ combinaisons.

5.1.4 Propriété d'optimalité de la U -statistique

La U -statistique peut être considérée comme la projection du noyau sur la statistique d'ordre. Pour un noyau $h(x_1, \dots, x_m)$ et un échantillon X_1, \dots, X_n , $n \geq m$, la U -statistique correspondante est donc

$$U_n = \mathbb{E} \{h(X_1, \dots, X_m | X_{(n)})\}, \quad (5.6)$$

où $X_{(n)}$ est la statistique d'ordre (X_{n1}, \dots, X_{nn}) .

Une implication de cette représentation est que, toute statistique $S = S(X_1, \dots, X_n)$ pour un estimateur sans biais de $\theta = \theta(F)$ peut être amélioré par la U -statistique correspondante, d'où le théorème suivant

Théorème 5.1.1 (in Serfling, 1980) *Soit $S = S(X_1, \dots, X_n)$, un estimateur sans biais de $\theta(F)$ basé sur l'échantillon X_1, \dots, X_n de distribution F . Alors la U -statistique correspondante est aussi sans biais et*

$$\text{Var}_F(U) \leq \text{Var}_F(S).$$

avec l'égalité si et seulement si $P_F(U = S) = 1$.

La statistique d'ordre $X_{(n)}$ est exhaustive pour toute famille \mathfrak{F} de distributions contenant F . La U -statistique est la projection sur la statistique exhaustive. Ainsi, le résultat précédent est simplement un cas particulier du théorème de Rao-Blackwell. Si $X_{(n)}$ est exhaustive, et complète, alors U_n est l'estimateur sans biais à variance minimale de $\theta(F)$.

5.1.5 Structure de martingales des U -statistiques

Certaines propriétés importantes des U -statistiques telles que la représentation comme une moyenne de $v.a.$ i.i.d. et le comportement presque sûr $(U_n \xrightarrow{p.s.} \theta, U_n - \xrightarrow{p.s.} 0 \text{ quand } n \rightarrow \infty)^{10}$, provient de sa structure de martingale.

Définition 5.1.3 Soit l'espace de probabilité (Ω, \mathcal{A}, P) , une suite de $v.a.$ $\{Y_n\}$ et une suite de σ -algèbre $\{\mathcal{F}_n\} \subset \mathcal{A}$ telle que Y_n est \mathcal{F}_n -mesurable et $\mathbb{E}|Y_n| < \infty$. Alors la suite $\{Y_n, \mathcal{F}_n\}$ est une martingale prospective si

$$a) \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$$

$$b) \mathbb{E}\{Y_{n+1} | \mathcal{F}_n\} = Y_n \text{ avec une probabilité égale à 1 pour tout } n.$$

et une martingale rétrospective si

$$a) \mathcal{F}_1 \supset \mathcal{F}_2 \supset \dots$$

$$b) \mathbb{E}\{Y_n | \mathcal{F}_{n+1}\} = Y_{n+1} \text{ avec une probabilité égale à 1 pour tout } n.$$

Les Lemmes suivants dûs à Hoeffding (1961) et Berk (1966), respectivement, donnent les caractérisations de la martingale prospective et la martingale rétrospective pour les U -statistiques. Pour le premier Lemme, nous avons besoin de quelques notations. Considérons le noyau symétrique $h(x_1, \dots, x_m)$ qui satisfait $\mathbb{E}_F \{h(X_1, \dots, X_m)\} < \infty$. Définissons les fonctions associées

$$h_c(x_1, \dots, x_c) = \mathbb{E}_F \{h(x_1, \dots, x_c, X_{c+1}, \dots, X_m)\}. \quad (5.7)$$

pour tout $c = 1, \dots, m - 1$ et posons $h_m \equiv h$. Comme

$$\int_A h_c(x_1, \dots, x_c) dF(x_1) \dots dF(x_c) = \int_{A \times R^{m-c}} h(x_1, \dots, x_m) dF(x_1) \dots dF(x_m). \quad (5.8)$$

¹⁰ \widehat{U}_n est la projection de la U -statistique U_n définie pour $\mathbb{E}_F |h| < \infty$, par $\widehat{U}_n = \sum_{i=1}^n \mathbb{E}_F (U_n | X_i) - (n-1)\theta$

pour tout ensemble de Borel A dans \mathbb{R}^c , h_c est l'espérance conditionnelle de $h(X_1, \dots, X_m)$ sachant X_1, \dots, X_c

$$h_c(x_1, \dots, x_c) = \mathbb{E}_F \{h(X_1, \dots, X_m | X_1 = x_1, \dots, X_c = x_c)\}. \quad (5.9)$$

De plus, notons que pour $1 \leq c \leq m - 1$

$$h_c(x_1, \dots, x_c) = \mathbb{E}_F \{h_{c+1}(x_1, \dots, x_c, X_{c+1})\}. \quad (5.10)$$

Il est commode de centrer, en définissant

$$\theta(F) = \mathbb{E}_F \{h(X_1, \dots, X_m)\}.$$

$$\tilde{h} = h - \theta(F). \quad (5.11)$$

et

$$\tilde{h}_c = h_c - \theta(F), \quad 1 \leq c \leq m. \quad (5.12)$$

Définissons maintenant

$$\begin{aligned} g_1(x_1) &= \tilde{h}_1(x_1), \\ g_2(x_1, x_2) &= \tilde{h}_2(x_1, x_2) - g_1(x_1) - g_1(x_2), \\ g_3(x_1, x_2, x_3) &= \tilde{h}_3(x_1, x_2, x_3) - \sum_{i=1}^3 g_1(x_i) - \sum_{1 \leq i < j \leq 3} g_2(x_i, x_j), \\ &\vdots \\ g_m(x_1, \dots, x_m) &= \tilde{h}_m(x_1, \dots, x_m) - \sum_{i=1}^m g_1(x_i) - \sum_{1 \leq i_1 < i_2 \leq m} g_2(x_{i_1}, x_{i_2}) \\ &\quad - \dots - \sum_{1 \leq i_1 < \dots < i_{m-1} \leq m} g_{m-1}(x_{i_1}, \dots, x_{i_{m-1}}). \end{aligned} \quad (5.13)$$

Les g_c sont symétriques si \tilde{h} est symétrique et l'on a

$$\begin{aligned}\mathbb{E}_F \{g_1(x_1)\} &= 0, \\ \mathbb{E}_F \{g_2(x_1, x_2)\} &= 0, \\ &\vdots \\ \mathbb{E}_F \{g_m(x_1, \dots, x_{m-1}, X_m)\} &= 0.\end{aligned}$$

puisque \tilde{h} est centrée et donc \tilde{h}_c est centrée.

Maintenant, considérons un échantillon X_1, \dots, X_n ($n \geq m$) et notons que la U -statistique U_n correspondant au noyau h satisfait

$$U_n - \theta(F) = \binom{n}{m}^{-1} S_n,$$

où

$$S_n = \sum_{1 \leq i_1 < \dots < i_m \leq n} \tilde{h}(X_{i_1}, \dots, X_{i_m}) \quad (5.14)$$

Finalement, pour $1 \leq c \leq m$, soit

$$S_{cn} = \sum_{1 \leq i_1 < \dots < i_c \leq n} g_c(X_{i_1}, \dots, X_{i_c})$$

Le lemme de Hoeffding que nous allons citer souligne les propriétés de martingales pour la suite $\{S_{cn}\}_{n \geq c}$ pour tout $c = 1, \dots, m$, et donne la représentation de U_n en terme de S_{1n}, \dots, S_{mn} .

Lemme 5.1.1 (Hoeffding) *Soit $h = h(x_1, \dots, x_m)$ un noyau symétrique pour $\theta = \theta(F)$, avec $\mathbb{E}_F |h| < \infty$. Alors*

$$U_n - \theta = \sum_{c=1}^m \binom{m}{n} \binom{n}{c}^{-1} S_{cn} \quad (5.15)$$

De plus, pour tout $c = 1, \dots, m$,

$$E_F \{S_{cn} | X_1, \dots, X_k\} = S_{ck}, \quad c \leq k \leq n. \quad (5.16)$$

Donc, avec $\mathcal{F}_k = \sigma \{X_1, \dots, X_k\}$, la suite $\{S_{cn}, \mathcal{F}_n\}_{n \geq c}$ est une martingale prospective.¹¹

5.1.6 Représentation de la U -statistique comme une moyenne des moyennes (dépendantes) de $v.a.$ i.i.d.

Considérons le noyau symétrique $h(x_1, \dots, x_m)$ et un échantillon X_1, \dots, X_n de taille $n \geq m$, définissons $k = \lfloor n/m \rfloor$, le plus grand entier inférieur ou égal à n/m , et

$$W(x_1, \dots, x_n) = \frac{h(x_1, \dots, x_m) + h(x_{m+1}, \dots, x_{2m}) + \dots + h(x_{k(m-1)+1}, \dots, x_{km})}{k}$$

\sum_p est la somme sur toutes les $n!$ permutations (i_1, \dots, i_n) de $(1, \dots, n)$ et \sum_c est la somme sur toutes les $\binom{n}{m}$ combinaisons (i_1, \dots, i_m) de $(1, \dots, n)$, nous avons

$$k \sum_p W(x_{i_1}, \dots, x_{i_n}) = km!(n-m)! \sum_c h(x_{i_1}, \dots, x_{i_m}),$$

et alors

$$\sum_p W(X_{i_1}, \dots, X_{i_n}) = m!(n-m)! \binom{n}{m} U_n, \quad (5.17)$$

or

$$U_n = \frac{1}{n!} \sum_p W(x_{i_1}, \dots, x_{i_n}).$$

Ceci exprime U_n comme une moyenne de $n!$ termes, dont chacun est lui même une moyenne de k $v.a.$ i.i.d. Ce type de représentation été introduit par Hoeffding

¹¹Preuve : Voir Serfling (1980), pp 181.

(1963).

5.1.7 U -statistique pondérée

Considérons maintenant, un noyau quelconque $h(x_1, \dots, x_m)$ pas nécessairement symétrique pour les observations X_1, \dots, X_n . Supposons aussi que chaque terme $h(X_{i_1}, \dots, X_{i_m})$ est pondéré par le facteur $w(x_{i_1}, \dots, x_{i_m})$ dépendant seulement des indices i_1, \dots, i_m . Dans ce cas la U -statistique prend la forme plus générale

$$T_n = \sum_c w(i_1, \dots, i_m) h(X_{i_1}, \dots, X_{i_m}).$$

\sum_c étant, comme ci-dessus, la somme sur les $\binom{n}{m}$ combinaisons de m éléments parmi n .

Dans le cas où h est symétrique et les poids $w(i_1, \dots, i_m)$ prennent seulement les valeurs 0 ou 1, la statistique T_n est une U -statistique incomplète ou bien réduite. De telles statistiques ont été étudiées par Blom (1976) et Brown et Kildea (1978).

Certaines statistiques de permutation survenant dans l'inférence non paramétrique sont asymptotiquement équivalentes aux statistiques définies ci-dessus, avec un poids qui ne prend pas nécessairement les valeurs 0 et 1. Pour cette application et d'autres, les statistiques de la forme T_n avec un noyau h symétrique et $m = 2$ ont été étudiées par Shapiro et Hubert (1979).

Finalement, certaines statistiques de rang pondéré pour une régression linéaire simple, prennent la forme de T_n . D'après Sievers (1978), considérons le modèle de la régression linéaire simple

$$y_i = \alpha + \beta x_i + e_i, \quad 1 \leq i \leq n$$

où α et β sont des paramètres inconnus, x_1, \dots, x_n sont les observations de la variable exogène et e_1, \dots, e_n sont des *v.a* i.i.d. de distribution F . Sievers (1978) considère des

inférences pour β basées sur les v .asuivantes :

$$T_\beta = \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{ij} \phi(Y_i - \alpha - \beta x_i, Y_j - \alpha - \beta x_j)$$

où $\phi(u, v) = 1_{[u \leq v]}$, les poids $a_{ij} \geq 0$ sont arbitraires, et $x_1 \leq \dots \leq x_n$ avec au moins une inégalité stricte. Par exemple, le test de $H_0 : \beta = 0$ contre $H_1 : \beta > 0$ peut être basé sur la statistique T_{β_0} . Sous l'hypothèse nulle, la distribution de T_{β_0} est la même que celle de T_0 pour $\beta = 0$. Ceci est équivalent à

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{ij} \phi(e_i, e_j)$$

qui est de la même forme que T_n définie ci-dessus. Les a_{ij} ici sont choisis de sorte que l'efficacité asymptotique soit la plus élevée, dans ce cas $a_{ij} = x_j - x_i$ (Serfling, 1980).

5.2 Variance et moments de la U -statistique

Les formules exactes de la variance de la U -statistique sont données dans 5.2.1. Les moments d'ordres supérieurs sont difficiles à obtenir, mais les limites utiles sont données dans 5.2.2.

5.2.1 Variance de la U -statistique

Considérons le noyau symétrique $h(x_1, \dots, x_m)$ tel que

$$\mathbb{E}_F \{h(X_1, \dots, X_m)\} < \infty$$

Nous utilisons de nouveau les fonctions h_c et \tilde{h}_c introduites dans 5.1.5. rappelons que $h_m = h$ et pour $1 \leq c \leq m - 1$,

$$h_c(x_1, \dots, x_c) = \mathbb{E}_F \{h(x_1, \dots, x_c, X_{c+1}, \dots, X_m)\},$$

$\tilde{h} = h - \theta$, $\tilde{h}_c = h_c - \theta$, $1 \leq c \leq m - 1$, où

$$\theta = \theta(F) = \mathbb{E}_F \{h(X_1, \dots, X_m)\}$$

Notons que

$$\mathbb{E}_F \left\{ \tilde{h}_c(X_1, \dots, X_c) \right\} = 0, \quad 1 \leq c \leq m$$

Définissons $\zeta_0 = 0$ et pour $1 \leq c \leq m$,

$$\zeta_c = \text{Var}_F \{h_c(X_1, \dots, X_c)\} = \mathbb{E}_F \left\{ \tilde{h}_c^2(X_1, \dots, X_c) \right\}$$

Nous aurons

$$0 = \zeta_0 \leq \zeta_1 \leq \dots \leq \zeta_m = \text{Var}_F \{h\} < \infty \quad (5.18)$$

Remarque 5.2.1 (5.18) est montrée en utilisant (5.13).

Ensuite, considérons deux ensembles $\{a_1, \dots, a_m\}$ et $\{b_1, \dots, b_m\}$ de m entiers distincts de $\{1, \dots, n\}$ et soit c le nombre des entiers communs entre les deux ensembles.

Il s'ensuit de la symétrie de \tilde{h} et de l'indépendance de $\{X_1, \dots, X_n\}$ que

$$\mathbb{E}_F \left\{ \tilde{h}(X_{a_1}, \dots, X_{a_m}) \tilde{h}(X_{b_1}, \dots, X_{b_m}) \right\} = \zeta_c$$

Notons également que le nombre de choix possibles de ces deux ensembles qui contiennent c éléments communs est $\binom{n}{m} \binom{m}{c} \binom{n-m}{m-c}$.

Avec ces préliminaires, nous pouvons obtenir la variance de la U -statistique. Soit

$$U_n - \theta = \binom{n}{m}^{-1} \sum_c \tilde{h}(X_{i_1}, \dots, X_{i_m})$$

nous aurons

$$\begin{aligned}
\text{Var}_F \{U_n\} &= \mathbb{E}_F \{(U_n - \theta)^2\} \\
&= \binom{n}{m}^{-2} \sum_c \sum_c \mathbb{E}_F \left\{ \tilde{h}(X_{a_1}, \dots, X_{a_m}) \tilde{h}(X_{b_1}, \dots, X_{b_m}) \right\} \\
&= \binom{n}{m}^{-2} \sum_{c=0}^m \binom{n}{m} \binom{m}{c} \binom{n-m}{m-c} \zeta_c.
\end{aligned}$$

Ce résultat et d'autres relations utiles de Hoeffding (1948a) sont comme suit

Lemme 5.2.1 (in Serfling, 1980) *La variance de U_n est donnée par*

$$\text{Var}_F \{U_n\} = \binom{n}{m}^{-1} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \zeta_c \quad (5.19)$$

et satisfait

- i) $\frac{m^2}{n} \zeta_1 \leq \text{Var}_F \{U_n\} \leq \frac{m}{n} \zeta_m$,
- ii) $(n+1) \text{Var}_F \{U_{n+1}\} \leq \text{Var}_F \{U_n\}$,
- iii) $\text{Var}_F \{U_n\} = \frac{m^2 \zeta_1}{n} + O(n^{-2})$, $n \rightarrow \infty$.

Une formule alternative de $\text{Var}_F \{U_n\}$ est obtenue en utilisant la fonction g_c introduite dans 5.1.5 au lieu de h_c et \tilde{h}_c , la représentation est donnée dans le lemme 5.1.1.

Considérons les ensemble $\{i_1, \dots, i_c\}$ et $\{j_1, \dots, j_d\}$ d'entiers distincts de $\{1, \dots, n\}$, où $1 \leq c, d \leq m$. Il est clair que, si un des ensembles de $\{i_1, \dots, i_c\}$ n'est pas contenu dans $\{j_1, \dots, j_d\}$, alors

$$\mathbb{E}_F \{g_c(X_{i_1}, \dots, X_{i_c} | X_{j_1}, \dots, X_{j_d})\} = 0$$

A partir de là,, il ressort que $\mathbb{E}_F \{g_c(X_{i_1}, \dots, X_{i_c}) g_c(X_{j_1}, \dots, X_{j_d})\} = 0$ sauf pour

$c = d$ et $\{i_1, \dots, i_c\} = \{j_1, \dots, j_d\}$. Par conséquent

$$S_{nc} = \sum_{1 \leq i_1 < \dots < i_c \leq n} g_c(X_{i_1} \dots X_{i_c})$$

nous avons

$$\mathbb{E}\{S_c, S_d\} = \begin{cases} \binom{n}{c} \mathbb{E}\{g_c^2\}, & c = d, \\ 0 & c \neq d. \end{cases}$$

D'où

Lemme 5.2.2 (in Serfling, 1980) *La variance de la U-statistique est donnée par*

$$\text{Var}_F\{U_n\} = \sum_{c=1}^m \binom{m}{c}^{-2} \binom{n}{c}^{-1} \mathbb{E}\{g_c^2\}.$$

Le taux de convergence de $\text{Var}_F\{U_n\}$ vers 0 dépend du plus petit c pour le quel ζ_c est non nulle. De l'un des deux lemmes 5.1.1 ou 5.2.2 nous obtenons

Corollaire 5.2.1 (in Serfling, 1980) *Soit $c \geq 1$ et supposons que $\zeta_0 = \dots = \zeta_{c-1} = 0 < \zeta_c$, alors*

$$\mathbb{E}(U_n - \theta)^2 = O(n^{-c}), \quad n \rightarrow \infty$$

Notons que la condition $\zeta_d = 0$, $d < c$ est équivalente à la condition $\mathbb{E}\{\tilde{h}_d^2\} = 0$, $d < c$ et aussi à la condition $\mathbb{E}\{\tilde{g}_d^2\} = 0$, $d < c$.

5.2.2 Autres moments de la U-statistique

Les généralisations exactes des lemmes 5.1.1 et 5.2.2, pour des moments d'ordres supérieurs sont difficiles à élaborer. Cependant, pour la plupart des applications, des limites convenables suffisent.

Lemme 5.2.3 (in Serfling, 1980) Soit r un nombre réel supérieur ou égal à deux, Supposons que $\mathbb{E}|h|^r < \infty$. Alors

$$\mathbb{E}|U_n - \theta|^r = O(n^{-(1/2)^r}), \quad n \rightarrow \infty$$

Lemme 5.2.4 (in Serfling, 1980) Soit $c_1 \geq 1$ et supposons que $\zeta_0 = \dots = \zeta_{c-1} = 0 < \zeta_c$. Soit r un entier supérieur ou égal à deux et supposons que $\mathbb{E}|h|^r < \infty$. Alors

$$\mathbb{E}|U_n - \theta|^r = O(n^{-[(1/2)(rc+1)]}), \quad n \rightarrow \infty$$

où $[\cdot]$ désigne la partie entière.

5.3 Distribution asymptotique de U_n

Selon la valeur de ζ_1 ($\zeta_1 > 0$ ou $\zeta_1 = 0$), nous distinguons deux cas pour la distribution asymptotique de U_n .

Cas 5.3.1 ($\zeta_1 > 0$)

Théorème 5.3.2 (Hoeffding, 1948) Si $\mathbb{E}_F h^2 < \infty$ et $\zeta_1 > 0$, alors $n^{1/2}(U_n - \theta) \xrightarrow{L} \mathcal{N}(0, m^2 \zeta_1)$ Ainsi

$$U_n \text{ est } AN\left(\theta, \frac{m^2 \zeta_1}{n}\right).$$

où AN signifie asymptotiquement normale.

Cas 5.3.3 ($\zeta_1 = 0 < \zeta_2$) Pour la fonction $\tilde{h}_2(x_1, x_2)$ associée au noyau $h = h(x_1, \dots, x_m)$ $m \geq 2$, définissons l'opérateur A sur l'espace fonctionnel $L_2(\mathbb{R}, F)$ par

$$Ag(x) = \int \tilde{h}_2(x, y) g(y) dF(y), \quad x \in \mathbb{R}, g \in L^2$$

Pour tout opérateur A , les valeurs propres $\lambda_1, \lambda_2, \dots$, des nombres réels (pas forcément distincts) correspondent aux solutions g_1, g_2, \dots de l'équation

$$Ag - \lambda g = 0$$

nous avons

Théorème 5.3.4 (in Serfling, 1980) Si $\mathbb{E}_F h^2 < \infty$ et $\zeta_1 < 0 < \zeta_1$, alors

$$n(U_n - \theta) \xrightarrow{\mathcal{L}} \frac{m(m-1)}{2} Y,$$

où Y est une v.a. de la forme

$$Y = \sum_{j=1}^{\infty} \lambda_j (\chi_{1j}^2 - 1),$$

où $\chi_{11}^2, \chi_{12}^2, \dots$ sont des khi deux indépendantes à un degré de liberté, Y a pour fonction caractéristique

$$\mathbb{E}_F \{e^{itY}\} = \prod_{j=1}^{\infty} (1 - 2it\lambda_j)^{-1/2}.$$

Chapitre

2

Tests non paramétriques d'indépendance sérielle

Mesurer la dépendance et tester l'indépendance sont des problèmes importants dans tous les domaines de la statistique lorsqu'on est confronté à des données d'observations. Dans ce chapitre, nous allons voir les aspects de ces problèmes dans le cadre de l'analyse des séries chronologiques. Pour une série temporelle stationnaire donnée $\{X_t\}$ nous nous intéressons à tester l'indépendance entre les X_t .

De tels tests ont un intérêt dans la recherche d'un modèle mathématique susceptible de représenter nos données. Les tests de type corrélation comme, par exemple le test de Box-Ljung, sont souvent appliqués à la classe des modèles *ARMA* (voir par exemple Brockwell and Davis, 1987, Ch. 9.4). Cependant il est connu que tels tests ont une puissance faible contre diverses alternatives non linéaires.

1 Mesure de Hoeffding

Une mesure de la dépendance, qui s'impose intuitivement, est celle introduite par Hoeffding en 1948b, définie en termes de fonctions de répartition conjointe et marginales. Skaug et Tjøstheim (1993a), ont prouvé que l'estimateur fondé sur les

fonctions de répartition empiriques est consistant si le processus sous-jacent est ergodique. Il est de plus, asymptotiquement distribué selon une somme pondérée infinie de loi du khi-deux indépendantes. Ces auteurs ont proposé un test d'indépendance de deux variables de retards différents (et pas un ensemble de $j + 1$ variables simultanées).

Soit $\{X_t, t \geq 0\}$ un processus stationnaire de fonction de répartition marginale $G(x) = P(X_t \leq x)$ et de fonction de répartition bi-variée $F(x, y) = P(X_{t-1} \leq x; X_t \leq y)$.

Soit

$$I = \int \{F(x, y) - G(x)G(y)\}^2 dF(x, y). \quad (1.1)$$

Un estimateur de I est obtenu par la substitution des fonctions de répartition empiriques dans (1.1). Soit le processus $\{Z_t, t \geq 1\}$ défini par $Z_t = (Z_t^{(1)}, Z_t^{(2)}) = (X_{t-1}, X_t)$. Pour des observations données $\{X_t, 0 \leq t \leq n\}$, la fonction de répartition empirique bi-variée est définie par

$$F_n(z) = F_n(z^{(1)}, z^{(2)}) = \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(1)} \leq z^{(1)}]} \mathbf{1}_{[Z_t^{(2)} \leq z^{(2)}]},$$

où $\mathbf{1}_{[\cdot]}$ désigne la fonction indicatrice. L'estimateur de I est donné par

$$\begin{aligned} I_n &= \int \{F_n(z^{(1)}, z^{(2)}) - F_n(z^{(1)}, \infty)F_n(\infty, z^{(2)})\}^2 dF_n(z^{(1)}, z^{(2)}), \\ &= \frac{1}{n} \sum_{t=1}^n \{F_n(X_{t-1}, X_t) - F_n(X_{t-1}, \infty)F_n(\infty, X_t)\}^2. \end{aligned} \quad (1.2)$$

Notons que $F_n(\cdot, \infty)$ et $F_n(\infty, \cdot)$ sont des estimateurs de $G(\cdot)$.

La fonctionnelle empirique I_n est la statistique considérée par Blum et *al* (1961), Deheuvels (1981) et Carlstein (1988) pour tester l'indépendance entre deux processus aléatoires $\{X_t, t \geq 0\}$ et $\{Y_t, t \geq 0\}$. Scargle (1981) a introduit la mesure de

dépendance sérielle suivante

$$D = \sum_{j=1}^m \int |F^j(x, y) - G(x)G(y)|^2 dx dy.$$

où $F^j(x, y) = P(X_t \leq x, X_{t-j} \leq y)$.

Elle est similaire à (1.1) pour le retard $j = 1$, la différence est que le facteur poids $dF(x, y)$ est remplacé par $dx dy$.

1.1 Consistance

Skaug et Tjøstheim (1993a), ont montré que I_n est un estimateur fortement consistant de I .

Théorème 1.1.1 (Skaug et Tjøstheim, 1993a) *Soit $\{X_t\}$, un processus ergodique, Alors $I_n \rightarrow I$ presque sûrement quand $n \rightarrow \infty$.*

Nous reprenons la démonstration de Skaug et Tjøstheim (1993a) pour laquelle nous avons clarifié certains passages.

Preuve. Posons

$A_n(z) = F_n(z) - F_n(z^{(1)}, \infty)F_n(\infty, z^{(2)})$, et $A(z) = F(z) - G(z^{(1)})G(z^{(2)})$ où $z = (z^{(1)}, z^{(2)})$. Donc

$$\begin{aligned} I_n &= \int A_n^2(z) dF_n(z) \\ &= \int A^2(z) dF(z) + \int A^2(z) d\{F_n(z) - F(z)\} + \int \{A_n^2(z) - A^2(z)\} dF_n(z). \end{aligned}$$

Le premier terme sur la droite de l'équation ci-dessus est I . Le deuxième terme $\int A^2(z) d\{F_n(z) - F(z)\} = n^{-1} \sum_{t=1}^n A^2(Z_t) - \mathbb{E}\{A^2(Z_t)\}$ tend vers zéro presque sûrement quand $n \rightarrow \infty$, puisque $\{A^2(z_t)\}$ est ergodique. Pour le troisième terme

nous avons

$$\left| \int \{A_n^2(z) - A^2(z)\} dF_n(z) \right| \leq \sup_z |A_n^2(z) - A^2(z)| \leq 2 \sup_z |A_n(z) - A(z)|. \quad (1.3)$$

puisque $A_n^2 - A^2 = (A_n - A)(A_n + A)$. De plus, $|A_n + A| = |F_n(z) - F_n(z^{(1)}, \infty)F_n(\infty, z^{(2)}) + F(z) - G(z^{(1)})G(z^{(2)})| \leq 2$ car $0 \leq F_n(z) + F(z) \leq 2$ et $-2 \leq -F_n(z^{(1)}, \infty)F_n(\infty, z^{(2)}) - G(z^{(1)})G(z^{(2)}) \leq 0$. Stute et Schumann (1980) ont montré que

$$\sup_z |F_n(z) - F(z)| \xrightarrow{n \rightarrow \infty} 0,$$

presque sûrement. Par conséquent le développement

$$\begin{aligned} A_n(z) - A(z) &= \{F_n(z) - F(z)\} \\ &\quad - \{F_n(z^{(1)}, \infty) - G(z^{(1)}) + G(z^{(1)})\} \{F_n(\infty, z^{(2)}) - G(z^{(2)}) + G(z^{(2)})\} \\ &\quad + G(z^{(1)})G(z^{(2)}) \\ &= \{F_n(z) - F(z)\} \\ &\quad - G(z^{(2)}) \{F_n(z^{(1)}, \infty) - G(z^{(1)})\} \\ &\quad - G(z^{(1)}) \{F_n(\infty, z^{(2)}) - G(z^{(2)})\} \\ &\quad - \{F_n(z^{(1)}, \infty) - G(z^{(1)})\} \{F_n(\infty, z^{(2)}) - G(z^{(2)})\}, \end{aligned}$$

montre que $\sup_z |A_n(z) - A(z)| \rightarrow 0$, quand $n \rightarrow \infty$, presque sûrement. Combinons ceci avec (1.3) nous aurons $\int \{A_n^2(z) - A^2(z)\} dF_n(z) \rightarrow 0$. ■

1.2 Distribution asymptotique

1.2.1 Relation entre I_n et la statistique de von Mises

Sous l'hypothèse d'indépendance Skaug et Tjøstheim (1993a) montrent que I_n est, à un $Op(n^{-3/2})$ près, asymptotiquement égale à la statistique de Von Mises

définie par

$$V_n = \frac{1}{n^2} \sum_{s,t}^n h(Z_s, Z_t),$$

où $h : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ est une fonction symétrique. Cette statistique est étroitement liée à la U -statistique

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq s < t \leq n} h(Z_s, Z_t).$$

La distribution limite de I_n peut, alors, être déterminée à partir de la théorie déjà existante sur les U -statistiques. C'est l'approche utilisée par Carlstein (1988) pour déterminer la distribution limite de sa statistique de test.

Lemme 1.2.1 (Skaug et Tjøstheim, 1993a) *Soit $\{X_t, t \geq 0\}$, un processus aléatoire i.i.d., alors*

$$I_n = \frac{1}{n^2} \sum_{s,t}^n h(Z_s, Z_t) + Op(n^{-3/2}), \quad (1.4)$$

où

$$h(z_1, z_2) = \int_{\mathbb{R}} q(u, z_1) q(u, z_2) dF(u), \quad (1.5)$$

avec

$$q(u, z) = \left\{ \mathbf{1}_{[z^{(1)} \leq u^{(1)}]} - G(u^{(1)}) \right\} \left\{ \mathbf{1}_{[z^{(2)} \leq u^{(2)}]} - G(u^{(2)}) \right\} \quad (1.6)$$

où $u = (u^{(1)}, u^{(2)})$ et $z = (z^{(1)}, z^{(2)})$.

Avant d'établir la preuve de ce lemme, on a besoin du lemme suivant

Soit $\{Z_t, t \geq 1\}$, un processus stationnaire bi-dimensionnel de fonction de répartition $F(z)$. Pour les observations $\{Z_t, 1 \leq t \leq n\}$, soit le processus empirique défini par

$$U_n(z) = \sqrt{n} \{F_n(z) - F(z)\}.$$

Lemme 1.2.2 (Skaug et Tjøstheim, 1993c) *Soit $a(z_1, \dots, z_m)$, une fonction mesurable définie de \mathbb{R}^{2m} dans \mathbb{R} . Alors si $\{Z_t\}$ est α -dépendant*

$$\int \dots \int a(z_1, \dots, z_m) \prod_{i=1}^m dU_n(z_i) = O_p(1),$$

Skaug Tjøstheim ont fait une extension de la preuve de Lemme B de Serfling (1980), p. 223 au cas d'un processus α -dépendant, alors

$$\mathbb{E} \left(\int \dots \int a(z_1, \dots, z_m) \prod_{i=1}^m dU_n(z_i) \right)^2 = O(1). \quad (1.7)$$

Preuve. Notons d'abord que

$$\begin{aligned} I_n &= \int_{\mathbb{R}^2} \{F_n(u) - F_n(u^{(1)}, \infty)F_n(\infty, u^{(2)})\}^2 dF(u) \\ &\quad + n^{-1/2} \int_{\mathbb{R}^2} \{F_n(u) - F_n(u^{(1)}, \infty)F_n(\infty, u^{(2)})\}^2 dU_n(u) \\ &= A_n + B_n. \end{aligned}$$

où $u = (u^{(1)}, u^{(2)})$.

Utilisant l'égalité $a - bc = (a_0 - b_0c_0) + (a - a_0) - b_0(c - c_0) - c_0(b - b_0) - (b - b_0)(c - c_0)$, nous trouvons du fait que $F(z) = G(z^{(1)})G(z^{(2)})$

$$\begin{aligned} &F_n(u) - F_n(u^{(1)}, \infty)F_n(\infty, u^{(2)}) \\ &= \{F_n(u) - F(u)\} \\ &\quad - G(u^{(2)}) \{F_n(u^{(1)}, \infty) - G(u^{(1)})\} \\ &\quad - G(u^{(1)}) \{F_n(\infty, u^{(2)}) - G(u^{(2)})\} \\ &\quad - \{F_n(u^{(1)}, \infty) - G(u^{(1)})\} \{F_n(\infty, u^{(2)}) - G(u^{(2)})\} \\ &= \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(1)} \leq u^{(1)}]} \mathbf{1}_{[Z_t^{(2)} \leq u^{(2)}]} - G(z^{(1)})G(z^{(2)}) \\ &\quad - G(u^{(2)}) \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(1)} \leq u^{(1)}]} - G(z^{(1)}) \right) \\ &\quad - G(u^{(1)}) \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(2)} \leq u^{(2)}]} - G(z^{(2)}) \right) \end{aligned}$$

$$\begin{aligned}
& - \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(1)} \leq u^{(1)}]} - G(z^{(1)}) \right) \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(2)} \leq u^{(2)}]} - G(z^{(2)}) \right) \\
= & \frac{1}{n} \sum_{t=1}^n \left(\mathbf{1}_{[Z_t^{(1)} \leq u^{(1)}]} - G(u^{(1)}) + G(u^{(1)}) \right) \\
& \times \left(\mathbf{1}_{[Z_t^{(2)} \leq u^{(2)}]} - G(z^{(2)}) + G(z^{(2)}) \right) - G(z^{(1)})G(z^{(2)}) \\
& - G(u^{(2)}) \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(1)} \leq u^{(1)}]} - G(z^{(1)}) \right) \\
& - G(u^{(1)}) \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(2)} \leq u^{(2)}]} - G(z^{(2)}) \right) \\
& - \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(1)} \leq u^{(1)}]} - G(z^{(1)}) \right) \times \left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{[Z_t^{(2)} \leq u^{(2)}]} - G(z^{(2)}) \right) \\
= & \int_{\mathbb{R}^2} M_1(u, z) M_2(u, z) dF_n(z) - \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} M_1(u, z_1) M_2(u, z_2) dF_n(z_1) dF_n(z_2) \\
= & n^{-1/2} \int_{\mathbb{R}^2} M_1(u, z) M_2(u, z) dU_n(z) \\
& - n^{-1} \int_{\mathbb{R}^4} M_1(u, z_1) M_2(u, z_2) dU_n(z_1) dU_n(z_2), \quad (1.8)
\end{aligned}$$

où $M_i(u, z) = \mathbf{1}_{[z^{(i)} \leq u^{(i)}]} - G(u^{(i)})$, $i = 1, 2$. La dernière égalité dans (1.8) est due au fait que $\int_{\mathbb{R}^2} M_i(u, z) dF(z) = 0$, $i = 1, 2$. Ainsi, d'après le Lemme 1.2.2, nous avons

$$\begin{aligned}
A_n & = \int_{\mathbb{R}^2} \{ F_n(u) - F_n(u^{(1)}, \infty) F_n(\infty, u^{(2)}) \}^2 dF(u) \\
& = \int_{\mathbb{R}^2} \left\{ n^{-1/2} \int_{\mathbb{R}^2} M_1(u, z) M_2(u, z) dU_n(z) \right. \\
& \quad \left. - n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} M_1(u, z_1) M_2(u, z_2) dU_n(z_1) dU_n(z_2) \right\}^2 dF(u) \\
& = n^{-1} \int_{\mathbb{R}^2} \left\{ \int_{\mathbb{R}^2} M_1(u, z) M_2(u, z) dU_n(z) \right\}^2 dF(u) \\
& \quad + n^{-2} \int_{\mathbb{R}^2} \left\{ \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} M_1(u, z_1) M_2(u, z_2) dU_n(z_1) dU_n(z_2) \right\}^2 dF(u) \\
& \quad - n^{-3/2} \int_{\mathbb{R}^2} \left\{ \int_{\mathbb{R}^2} M_1(u, z) M_2(u, z) dU_n(z) \right\} \\
& \quad \times \left\{ \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} M_1(u, z_1) M_2(u, z_2) dU_n(z_1) dU_n(z_2) \right\} dF(u)
\end{aligned}$$

Notons les termes multipliés par n^{-1} , n^{-2} et $n^{-3/2}$ par *I*, *II* et *III* respectivement,

commençons par

$$\begin{aligned}
 I &= n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int M_i(u, z_j) dF(u) \right\} dU_n(z_1) dU_n(z_2) \\
 &= n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int M_i(u, z_j) dF(u) \right\} d[\{F_n(z_1) - F(z_1)\} \{F_n(z_1) - F(z_1)\}] \\
 &= n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int M_i(u, z_j) dF(u) \right\} dF_n(z_1) dF_n(z_2) \\
 &\quad - n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int M_i(u, z_j) dF(u) \right\} dF_n(z_1) dF(z_2) \\
 &\quad - n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int M_i(u, z_j) dF(u) \right\} dF(z_1) dF_n(z_2) \\
 &\quad + n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int M_i(u, z_j) dF(u) \right\} dF(z_1) dF(z_2) \\
 &= n^{-1} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \prod_{i,j=1,2} \left\{ \int_{\mathbb{R}^2} M_i(u, z_j) dF(u) \right\} dF_n(z_1) dF_n(z_2) \\
 &= \frac{1}{n^2} \sum_{s,t=1}^m h(Z_s, Z_t).
 \end{aligned}$$

où $h(z_1, z_2) = \int_{\mathbb{R}^2} q(u, z_1) q(u, z_2) dF(u)$ et $q(u, z_i) = \prod_{j=1,2} M_j(u, z_i)$ pour $i = 1, 2$.

pour II , d'après l'extension du Skaug et Tjøstheim donnée dans (1.7), nous avons

$$\begin{aligned}
 II &= n^{-2} \mathbb{E} \left\{ \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} M_1(u, z_1) M_2(u, z_2) dU_n(z_1) dU_n(z_2) \right\}^2 \\
 &= O(n^{-2}).
 \end{aligned}$$

et pour III , d'après le Lemme 1.2.2, nous avons

$$\begin{aligned}
 III &= n^{-3/2} \int_{\mathbb{R}^2} \left\{ \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} M_1(u, z) M_2(u, z) M_1(u, z_1) M_2(u, z_2) \right. \\
 &\quad \left. dU_n(z) dU_n(z_1) dU_n(z_2) \right\} dF(u) \\
 &= O_p(n^{-3/2}).
 \end{aligned}$$

En combinant les trois termes, nous trouvons

$$A_n = \frac{1}{n^2} \sum_{s,t=1}^m h(Z_s, Z_t) + O_p(n^{-3/2}). \quad (1.9)$$

De la même manière pour $B_n = n^{-1/2} \int_{\mathbb{R}^2} \{F_n(u) - F_n(u^{(1)}, \infty)F_n(\infty, u^{(2)})\}^2 dU_n(u)$ en utilisant la relation (1.8) et le Lemme 1.2.2, nous trouvons $B_n = O_p(n^{-3/2})$. Combinons ceci avec (1.9), nous trouvons (1.4). ■

1.2.2 Distribution asymptotique de I_n dans le cas i.i.d.

De (1.4) nous avons

$$nI_n = nV_n + O_p(n^{-1/2}),$$

où $V_n = \frac{1}{n^2} \sum_{s,t}^n h(Z_s, Z_t)$. Sous l'hypothèse d'indépendance, le noyau donné dans (1.5) peut se factoriser sous la forme suivante

$$h(z_1, z_2) = g(z_1^{(1)}, z_2^{(1)})g(z_1^{(2)}, z_2^{(2)}),$$

où

$$g(x, y) = \int \{\mathbf{1}_{[x \leq w]} - G(w)\} \{\mathbf{1}_{[y \leq w]} - G(w)\} dG(w).$$

définissons l'opérateur¹² $B : L_2(G) \rightarrow L_2(G)$

$$Bf(x) = \int g(x, y)f(y)dG(y),$$

qui est de type Hilbert-Schmidt. L'ensemble $\{\eta_m, \pi_m, m \geq 1\}$, des valeurs propres et des fonctions propres de B , ont les propriétés suivantes

¹² $L^2(G)$ espace des fonctions mesurables de carré G -intégrable.

$$\begin{aligned}
i) \quad & \mathbb{E}\{\pi_i(X_0)\pi_j(X_0)\} = \mathbf{1}_{[i=j]} \\
ii) \quad & \mathbb{E}\{\pi_m(X_0)\} = 0 \quad m = 1, 2, \dots \\
iii) \quad & \mathbb{E}\{g(X_0, X_0)\} = \sum_{m=1}^{\infty} \eta_m < \infty
\end{aligned}$$

et vérifient d'après le Théorème de Mercer généralisé, 1902

$$Bf(x) - \eta f(x) = 0 \quad (G \text{ presque partout}), \quad (1.10)$$

pour une fonction propre f . Cet ensemble des valeurs propres détermine la distribution de I_n sous l'hypothèse nulle.

Théorème 1.2.1 (Skaug et Tjøstheim, 1993a) *Soit $\{X_t, 0 \leq t \leq n\}$, une réalisation d'un processus aléatoire constitué de variables indépendantes identiquement distribuées, alors I_n converge en loi vers*

$$nI_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \sum_{k,j} \lambda_{kj} W_{kj}^2$$

où $\lambda_{kj} = \eta_k \eta_j$, $\{\eta_m, m \geq 1\}$ est l'ensemble des valeurs propres de l'opérateur B et $\{W_{kj}, k, j \geq 1\}$ sont des v.a indépendantes de loi $\mathcal{N}(0, 1)$.

1.3 Cas continu et cas discret

Le cas continu et le cas discret sont assez différents l'un de l'autre. Ils vont être traités séparément.

Cas continu Sous l'hypothèse d'indépendance, I_n est libre relativement à la loi sous-jacente du processus stochastique. Pour voir cela, considérons le processus $\{U_t, t \geq 0\}$ défini par $U_t = G(X_t)$, en supposant que la fonction de répartition marginale G admet un inverse G^{-1} , U_t est uniformément distribuée sur $[0, 1]$. De

plus, soit le processus $\{W_t, t \geq 1\}$ défini par $W_t = (U_{t-1}, U_t)$. Alors

$$I_n = \frac{1}{n} \sum_{t=1}^n \{F_{W,n}(U_{t-1}, U_t) - F_{W,n}(U_{t-1}, \infty)F_{W,n}(\infty, U_t)\}^2,$$

où $F_{W,n}$ est la fonction de répartition empirique de $\{W_t, 1 \leq t \leq n\}$. Ainsi si $\{X_t, t \geq 0\}$ sont des *v.a* indépendantes, I_n peut être construite à partir de *v.a* indépendantes de loi uniforme sur $[0, 1]$. Nous notons que la statistique de test de Scargle (1981),

$$D_n = \sum_{j=1}^m \sum_{i=1}^{N-1} \sum_{k=1}^{N-1} |F^j(X_i, X_k) - F(X_i, \infty)G(\infty, X_k)|^2 \Delta X_i \Delta X_j,$$

où $X_1 \leq X_2 \leq \dots \leq X_n$ et $\Delta X_i = X_{i+1} - X_i$, $F^j(X_i, X_k) = \frac{1}{n-j-1} \sum_{i=1} \sum_{k=1} \mathbf{1}_{[X_i \leq X_{i-j}]} \mathbf{1}_{[X_k \leq X_{k-j}]}$, n'est pas libre de la loi sous-jacente du processus stochastique.

Soit X_0 une *v.a* de loi uniforme sur $[0, 1]$. L'opérateur B devient

$$Bf(x) = \int_0^1 g(x, y) f(y) dy,$$

où

$$g(x, y) = \int \{\mathbf{1}_{[x \leq w]} - w\} \{\mathbf{1}_{[y \leq w]} - w\} dw = -\max(x, y) + \frac{1}{2}(x^2 + y^2) + \frac{1}{3}.$$

Les valeurs propres de B peuvent être trouvées comme dans la preuve de la proposition 1 de Shorack et Wellner (1986, pp. 213). Nous avons alors, $\eta_m = (m\pi)^2$ et les coefficients dans le Théorème 1.2.1 sont donnés par $\lambda_{ij} = (ij\pi^2)^{-2}$ qui montre que I_n a la même distribution limite que la statistique de test de Blum et al (1961) ($B_n = \int \{F_n(x, y) - F_n(x, \infty)F_n(\infty, y)\}^2 dF_n(x, y)$).

Cas discret Dans ce cas I_n n'est pas une distribution libre de la loi sous-jacente du processus stochastique. Or $I = \int \{F(xy) - G(x)G(y)\}^2 dF(x, y)$ n'est

pas absolument une mesure de dépendance, du fait que I peut être nulle en présence de la dépendance (Joag-Dev, 1991). Pourtant, le test de I_n à une puissance contre plusieurs alternatives discrètes.

Pour $x_1 < x_2 \dots$, soit $P(X_0 = x_i) = p_i$, tel que $\sum_{i=1}^{\infty} p_i = 1$. Alors

$$G(x) = \sum_{i=1}^{\infty} p_i \mathbf{1}_{[x_i \leq x]}.$$

Skaug (1993c) a montré que le problème des valeurs propres dans (1.10) se réduit à la recherche des valeurs propres de la matrice symétrique $\Sigma = \{\sigma_{ij}, 1 \leq i, j \leq m-1\}$, où

$$\sigma_{ij} = (p_i p_j)^{1/2} \{G(x_{\min(i,j)}) - G(x_i)G(x_j)\},$$

et $G(x_{\min(i,j)}) = \sum_{k \geq 1} p_k \mathbf{1}_{[x_k \leq x_i]} \mathbf{1}_{[x_k \leq x_j]}$.

Les valeurs propres de Σ doivent être recalculées pour chaque m -uplets p_1, \dots, p_m .

En effet,

$$g(x, y) = \sum_{i=1}^{\infty} p_i [\mathbf{1}_{[x_i \leq x]} - G(x_i)] [\mathbf{1}_{[y_i \leq y]} - G(x_i)]$$

qui donne

$$\begin{aligned} Bf(x) &= \sum_{j=1}^{\infty} p_j \sum_{i=1}^{\infty} p_i [\mathbf{1}_{[x_i \leq x]} - G(x_i)] [\mathbf{1}_{[x_j \leq x_i]} - G(x_i)] f(x_j) \\ &= \sum_{i=1}^{\infty} p_i [\mathbf{1}_{[x_i \leq x]} - G(x_i)] \sum_{j=1}^{\infty} p_j [\mathbf{1}_{[x_j \leq x_i]} - G(x_i)] f(x_j) \quad (1.11) \\ &= \sum_{i=1}^{\infty} \sigma_i(f) f_i(x) \end{aligned}$$

où $f_i(x) = \sqrt{p_i} \{\mathbf{1}_{[x_i \leq x]} - G(x_i)\}$ et

$$\begin{aligned} \sigma_i(f) &= \sqrt{p_i} \sum_{j=1}^{\infty} p_j [\mathbf{1}_{[x_j \leq x_i]} - G(x_i)] f(x_j) \\ &= \sqrt{p_i} \left[\sum_{j=1}^i p_j f(x_j) - G(x_i) \mathbb{E}\{f(X)\} \right] \end{aligned}$$

de (1.11) on voit que la fonction propre de B est G presque sûrement de la forme

$\sum_k \alpha_k f_k(x)$ où $\alpha_1, \alpha_2, \dots$ sont des constantes. Définissons $\sigma_{ik} = \sigma_i(f_k)$ pour $i, k \geq 1$ et comme $\mathbb{E}\{f_k(X)\} = 0$, nous avons

$$\sigma_{ik} = \sqrt{p_i p_k} \{G(x_{\min(i,k)}) - G(x_i)G(x_k)\} \quad (1.12)$$

Alors nous avons

$$\begin{aligned} B \left(\sum_k \alpha_k f_k(x) \right) &= \sum_i p_i [\mathbf{1}_{[x_i \leq x]} - G(x_i)] \sum_j p_j [\mathbf{1}_{[x_j \leq x_i]} - G(x_i)] \sum_k \alpha_k f_k(x_j) \\ &= \sum_k \alpha_k \sum_i \sigma_{ik} f_i = \sum_i f_i \sum_k \alpha_k \sigma_{ik} \end{aligned}$$

et les conditions sur les valeurs propres deviennent

$$\eta \sum_i \alpha_i f_i(x) = \sum_i f_i(x) \sum_k \alpha_k \sigma_{ik} \quad x = x_1, x_2, \dots \quad (1.13)$$

Si on restreint les conditions pour une taille d'échantillon finie, on peut aller plus loin. Par conséquent, supposons que x_1, \dots, x_m est le support de G . Alors, notons que de (1.12), $\sigma_{im} = \sigma_{mi} = 0$, $i = 1, \dots, m$ et que $f_m(x) = 0$ (G p.s). Ces observations peuvent être utilisées pour montrer que dans le cas fini, le système d'équations (62) se réduit à un système linéaire de dimension $(m - 1)$.

$$\eta \alpha = \Sigma \alpha$$

où $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_{m-1})$ et Σ est la matrice de dimension $(m - 1) \times (m - 1)$ avec σ_{ik} est l'élément (i, k) . Donc les valeurs propres de B sont les valeurs propres de la matrice symétrique Σ . Le cas le plus simple pour $m = 2$, dans le cas où $P(X_0 = x_1) = p$ et $P(X_0 = x_2) = 1 - p$ pour $0 < p < 1$. Alors $\sigma_{11} = p(p - p^2) = p^2(1 - p)$ et $\eta = p^2(1 - p)$. Du théorème 1.2.1, nI_n converge en loi vers $p^4(1 - p)W^2$, où $W \equiv \mathcal{N}(0, 1)$. Ce résultat a été testé en faisant des simulations pour $n = 100$ et

différentes valeurs de $p \in (0, 1)$. La distribution asymptotique donne une description exacte de la distribution de la loi simulée de nI_n .

1.4 Test d'indépendance sérielle I_n

L'hypothèse nulle H_0 , est que les *v.a* $\{X_t, t \geq 0\}$ sont indépendantes. D'après le Théorème 1.1.1, il est raisonnable de rejeter H_0 si les valeurs de I_n sont grandes. Alors un test pour un seuil approximatif ε consiste à : rejeter H_0 si $nI_n > u_\varepsilon$, où u_ε est le quantile d'ordre $1 - \varepsilon$ de la distribution limite de nI_n sous H_0 , donnée dans le Théorème 1.2.1. Bien que nous ayons choisi que notre hypothèse soit l'indépendance dans $\{X_t, t \geq 0\}$, le Théorème 1.1.1 montre seulement que ce test est consistant contre d'autres alternatives, où X_t et X_{t-1} sont dépendantes. Dans 1.5, nous présentons une extension du test qui est sensible au retard supérieur.

1.5 Extension à un retard supérieur

Le test pour l'hypothèse que $\{X_t\}$ est constitué de *v.a* indépendantes, basé sur I_n est sensible aux alternatives où X_{t-1} et X_t sont dépendantes. Pour certaines alternatives il est nécessaire de vérifier l'indépendance des variables dans l'ensemble (X_{t-j}, \dots, X_t) où j est une constante fixée. Skaug et Tjøstheim (1993a), montrent que la meilleure façon de le faire n'est pas de considérer un ensemble de $j + 1$ variables simultanées mais seulement de considérer deux variables de retard différent. Cette approche est quand même incomplète mais Skaug et Tjøstheim (1993a), montrent qu'elle a de bonnes propriétés contre certains types usuels de dépendance. Soit le processus de vecteurs $\{Z_{j,t}, t \geq k\}$ pour $1 \leq j \leq m$, défini par $Z_{j,t} = (X_{t-j}, X_t)$. Alors $Z_{1,t}$ est égale à Z_t défini auparavant. Soit la fonction de répartition empirique bi-variée de retard j

$$F_{j,n}(z) = F_{j,n}(z^{(1)}, z^{(2)}) = \frac{1}{n-j+1} \sum_{t=j}^n \mathbf{1}_{[Z_{j,t}^{(1)} \leq z^{(1)}]} \mathbf{1}_{[Z_{j,t}^{(2)} \leq z^{(2)}]}.$$

De plus, soit

$$I_n(j) = \int \{F_{j,n}(z^{(1)}, z^{(2)}) - F_{j,n}(z^{(1)}, \infty) F_{j,n}(\infty, z^{(2)})\}^2 dF_{j,n}(z^{(1)}, z^{(2)}).$$

Alors, la statistique pour tester l'indépendance des composantes du vecteur (X_{t-j}, \dots, X_t) est

$$G_n(m) = \sum_{j=1}^m I_n(j).$$

Le théorème suivant donne la loi de $G_n(m)$ sous H_0 .

Théorème 1.5.1 (Skaug et Tjøstheim, 1993a) *Soit $\{X_t, t \geq 0\}$ des v.a i.i.d. Alors*

$$nG_n(m) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \sum_{i,j=1}^{\infty} \lambda_{ij} W_{ij}(m),$$

où λ_{ij} sont comme dans le théorème 1.2.1 et $\{W_{ij}(m), i, j \geq 1\}$ sont i.i.d. de loi du khi deux à m degrés de liberté.

Preuve. La statistique $G_n(m)$ est asymptotiquement une statistique de Von Mises dégénérée, pour laquelle les résultats de Carlstein (1988), sont appliqués.

Avec q défini en (1.6), nous avons par analogie avec (1.4)

$$I_n(j) = \frac{1}{n^2} \sum_{s,t \geq m}^n \int q(u, Z_{j,s}) q(u, Z_{j,t}) dF(u) + O_p(n^{-3/2}) \quad (1.14)$$

Soit $Z = (Z^{(m)}, \dots, Z^{(0)}) = (X_{t-m}, \dots, X_t)$, un vecteur de dimension $m+1$ et

$$q_j(u, z) = \left\{ \mathbf{1}_{[z^{(j)} \leq u^{(1)}]} - G(u^{(1)}) \right\} \left\{ \mathbf{1}_{[z^{(0)} \leq u^{(2)}]} - G(u^{(2)}) \right\}$$

Alors

$$G_n(m) = \frac{1}{n^2} \sum_{s,t \geq m}^n \sum_{j=1}^m h_j(Z_s, Z_t) + O_p(n^{-3/2})$$

où $h_j(z_1, z_2) = \int q_j(u, z_1) q_j(u, z_2) dF(u)$ et $Z_t = (Z_t^{(m)}, \dots, Z_t^{(0)})$. Ainsi, comme

$\int h_j(z_1, z_2) \prod_{i=1}^m dG(z_2^{(i)}) = 0$, il s'ensuit qu'à un terme asymptotiquement négligeable près, $G_n(m)$ est une statistique de Von Mises dégénérée de noyau $h = \sum_{j=1}^m h_j$.

Soit $A_m : L^2 \left(\underbrace{G \times G \times \dots \times G}_{m \text{ fois}} \right) \rightarrow L^2 \left(\underbrace{G \times G \times \dots \times G}_{m \text{ fois}} \right)$ l'opérateur intégral

$$\begin{aligned} A_m f(z_1) &= \int_{\mathbb{R}} h(z_1, z_2) f(z_2) \prod_{i=1}^m dG(z_2^{(i)}) \\ &= \sum_{j=1}^m \int_{\mathbb{R}} h_j(z_1, z_2) f(z_2) dG(z_2^{(0)}) dG(z_2^{(j)}) \end{aligned}$$

Soit $\{(\eta_r, \pi_r), r \geq 1\}$ les valeurs et les fonctions propres respectivement. Comme $\mathbb{E}\{\pi_r(X_0)\} = 0$, il ressort que

$$\begin{aligned} &\int h_j(z_1, z_2) f(z_2) \pi_i(z_2^{(0)}) \pi_k(z_2^{(r)}) \prod_{l=1}^m dG(z_2^{(l)}) \\ &= \begin{cases} \eta_i \eta_k \pi_i(z_2^{(0)}) \pi_k(z_2^{(r)}) & j = r \\ 0 & j \neq r \end{cases} \end{aligned}$$

Ceci montre que $\{\phi_{ikj}(z) = \pi_i(z^{(0)}) \pi_k(z^{(j)}), i, k \geq 1, 1 \leq j \leq m\}$, les fonctions propres de A_m avec $\lambda_{ik} = \eta_i \eta_k$ étant les valeurs propres corréspondantes à $\{\phi_{ikj}(z), 1 \leq j \leq m\}$. Ainsi, λ_{ik} est répété m fois dans l'ensemble des valeurs propres de A_m . L'ensemble $\{\phi_{ikj}(z), 1 \leq j \leq m\}$ à la propriété : $\mathbb{E}\{\phi_{i_1 k_1 j_1}(Z_s) \phi_{i_2 k_2 j_2}(Z_t)\} = \mathbf{1}_{[s=t]} \mathbf{1}_{[i_1=i_2]} \mathbf{1}_{[k_1=k_2]} \mathbf{1}_{[j_1=j_2]}$. Donc W_{ikj} dans le théorème 1 de Carlstein (1988) sont des variables indépendantes de loi $\mathcal{N}(0, 1)$, nous avons la convergence en loi

$$nG_n(m) \xrightarrow{n \rightarrow \infty} \sum_{\substack{i, k \geq 1 \\ 1 \leq j \leq m}} \lambda_{ik} W_{ikj}^2 = \sum_{i, k=1}^{\infty} \lambda_{ik} W_{ik}(m)$$

où $W_{ik}(m) = \sum_{j=1}^m W_{ikj}^2$ qui est la distribution de khi deux à m degrés de liberté ■

2 Mesure des différence pondérée de densités

Cette mesure se base essentiellement sur la différence des densités conjointmarginale $p_2(x, y) - p(x)p(y)$, où p_2 est la densité bi-variée et les p les densités marginales de (X_{t-1}, X_t) pour des séries chronologiques stationnaires $\{X_t\}$. Nous supposons que ces densités existent. Sous l'hypothèse d'indépendance, on a $p_2 = p^2$, ceci peut être exploité pour construire des fonctionnelles pour mesurer la dépendance et tester l'indépendance. La même procédure s'applique pour (X_{t-k}, X_t) .

Les densités conjointes d'ordre supérieur pour $(X_t, X_{t-1}, \dots, X_{t-k})$ peuvent être traitées de la même façon, mais lorsque k croît des problèmes peuvent surgir à cause de la dimension dans l'estimation des densités conjointes.

Cette mesure est définie par

$$I_1 = \int \{p_2(x, y) - p(x)p(y)\} p_2(x, y) dx dy. \quad (2.1)$$

Cette mesure n'est pas positive. Quand $p_2(x, y) < p(x)p(y)$, le poids $p_2(x, y)$ est petit, tandis que quand $p_2(x, y) > p(x)p(y)$ il est grand ; ainsi les contributions positives à l'intégrale dominent les négatives. Dans le cas gaussien il est possible de montrer le résultat formel.

Proposition 2.0.1 (Skaug et Tjøstheim, 1993b) *Soit $\{X_t\}$ un processus gaussien stationnaire. Alors $I_1 \geq 0$. Elle est nulle si et seulement si X_t et X_{t-1} sont indépendantes.*

Preuve. Soit $\sigma^2 = \text{var}(X_t)$ et $\rho = \text{corr}(X_t, X_{t-1})$. Alors

$$\int p_2^2(x, y) dx dy - \int p^2(x)p^2(y) dx dy = (4\pi\sigma^2)^{-1} \left\{ (1 - \rho^2)^{-1/2} - 1 \right\} \geq 0. \quad (2.2)$$

Utilisant l'inégalité de Schwarz

$$\begin{aligned} I_1 &= \int p_2^2(x, y) dx dy - \int p(x) p(y) p_2(x, y) dx dy \\ &\geq \int p_2^2(x, y) dx dy - \left\{ \int p_2^2(x, y) dx dy \right\}^{1/2} \left\{ \int p^2(x) p^2(y) dx dy \right\}^{1/2} \end{aligned} \quad (2.3)$$

Il s'ensuit de (2.2) que $I_1 \geq 0$. De sa définition, nous avons $I_1 = 0$ si X_t et X_{t-1} sont indépendantes. D'autre part si $I_1 = 0$ alors de (2.3)

$$\int p_2^2(x, y) dx dy \leq \int p^2(x) p^2(y) dx dy$$

et il s'ensuit de (2.2) que nous devons avoir $\rho = 0$ i.e., X_t et X_{t-1} doivent être indépendantes sous cette hypothèse gaussienne. ■

et donc, I_1 n'est pas toujours positive, le contre exemple le montre.

Soit $p_2(x, y)$ donnée par

$$p_2(x, y) = \begin{cases} \alpha & -1 < x \leq 0, 0 < y \leq 1 \\ 1 - 2\alpha & 0 < x, y \leq 1 \\ \alpha & 0 < x \leq 1, -1 < y \leq 0 \\ 0 & \text{sinon} \end{cases}$$

Le calcul de I_1 donne $\alpha^2(4\alpha - 1)$, et donc, I_1 est négative si $0 < \alpha < \frac{1}{4}$.

Il y a au moins deux façons de déterminer la mesure I_1 . Pour les deux méthodes nous avons besoin d'estimer \hat{p}_2 et \hat{p} par la méthode du noyau. D'autres type d'estimations sont possibles. Chan et Tran (1992) ont utilisé l'estimation basée sur les histogrammes.

La plupart des méthodes d'estimation de I_1 consistent à utiliser l'intégration numérique,

$$I'_{n,1} = \int \{\hat{p}_2(x, y) - \hat{p}(x)\hat{p}(y)\} \hat{p}_2(x, y) w_2(x, y) dx dy. \quad (2.4)$$

où $w_2(x, y)$ est la fonction poids. On peut travailler avec les moyennes empiriques, pour des observations données $\{X_0, X_1, \dots, X_n\}$ (Skaug et Tjøstheim, 1993b),

$$I_{n,1} = \frac{1}{n} \sum_t \{\widehat{p}_2(X_t, X_{t-1}) - \widehat{p}(X_t) \widehat{p}(X_{t-1})\} w_2(X_t, X_{t-1}) \quad (2.5)$$

Sous les hypothèses du théorème ergodique faible

$$I_{n,1} \xrightarrow{p} \int \{p_2(x, y) - p(x)p(y)\} p_2(x, y) w_2(x, y) dx dy$$

En enlevant le facteur poids nous aurons I_1 . L'introduction de w_2 dans (2.5) signifie que les observations extrêmes peuvent être cachées. En fait dans toutes les simulations de $I_{n,1}$ nous prenons $w_2(x, y) = 1$.

Pour des raisons de calculs, nous préférons d'utiliser (2.5) plutôt que (2.4).

Considérons les estimateurs des densités donnés par la méthode des noyaux

$$\widehat{p}_2(x, y) = \frac{1}{n} \sum_s k_h(X_s - x) k_h(X_{s-1} - y) \quad \text{et} \quad \widehat{p}(x) = \frac{1}{n} \sum_s k_h(X_s - x) \quad (2.6)$$

où $h = h(n)$ est la fenêtre, $k_h(x) = h^{-1}k(h^{-1}x)$ avec k est la fonction noyau satisfaisant les hypothèses standard telles que $\int k(x) dx = 1$, $k(-x) = k(x)$, $k(x)$ est bornée et

$\int x^2 k(x) dx < \infty$. Nous considérons la même fenêtre pour les estimateurs densité dans les deux cas, uni et bidimensionnelles tout à fait comme Rosenblatt (1975) et Wahlen (1991). Ceci ne s'accorde pas avec l'optimalité des estimateurs des densités mais simplifie les calculs. Utilisons (2.6), nous obtenons

$$\begin{aligned} & \int \{\widehat{p}_2(x, y)\}^2 dx dy \\ &= \frac{1}{n^2} \sum_s \sum_t k_h(X_s - x) k_h(X_{s-1} - y) k_h(X_t - x) k_h(X_{t-1} - y) \\ &= \frac{1}{n} \sum_s \frac{1}{n} \sum_t G_h(X_s - X_t) G_h(X_{s-1} - X_{t-1}) \end{aligned}$$

$$\triangleq \frac{1}{n} \sum_t \widehat{p}_{2,G}(X_t, X_{t-1}),$$

où G set la fonction noyau obtenue en convoluant k avec lui même, i.e.,

$$G(z) = \int k(x) k(z-x) dx$$

De la même manière,

$$\begin{aligned} & \int \widehat{p}(x) \widehat{p}(y) \widehat{p}_2(x, y) dx dy \\ &= \frac{1}{n^3} \sum_s \sum_t \sum_u k_h(X_s - x) k_h(X_t - y) k_h(X_u - x) k_h(X_{u-1} - y) dx dy \\ &= \frac{1}{n} \sum_s \frac{1}{n} \sum_t G_h(X_s - X_t) \frac{1}{n} \sum_u G_h(X_u - X_{t-1}) \\ &\triangleq \frac{1}{n} \sum_t \widehat{p}_G(X_t) \widehat{p}_G(X_{t-1}), \end{aligned}$$

Ainsi, $I_{n,1}$ devient

$$\begin{aligned} & \int \{\widehat{p}_2(x, y) - \widehat{p}(x) \widehat{p}(y)\} \widehat{p}_2(x, y) dx dy \quad (2.7) \\ &= \frac{1}{n} \sum_t \widehat{p}_{2,G}(X_t, X_{t-1}) - \widehat{p}_G(X_t) \widehat{p}_G(X_{t-1}) \end{aligned}$$

Pour un noyau gaussien avec $k(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2)$, G est aussi gaussienne avec $G(x) = (2\pi)^{-1/2} 2^{-1/2} \exp(-\frac{1}{2}(x/2^{1/2})^2)$. De plus la fonction poids w peut être intégrée dans (2.7). Alors il y a des termes additionnels d'ordre $O(h^2)$ qui peuvent être négligés asymptotiquement.

2.1 Propriétés asymptotiques

Pour l'analyse asymptotique de $I_{n,1}$, nous considérons l'estimation par la méthode des noyaux et le développement de Taylor. Le développement de Taylor pour une fonctionnelle entropie été utilisée par Joe (1989b) dans le cas i.i.d, nous utilisons

les mêmes notations. Ainsi, soit

$$F_n(x, y) = F_{n,2}(x, y) = \frac{1}{n} \sum_t \mathbf{1}_{[X_t \leq x, X_{t-1} \leq y]}.$$

la fonction de répartition empirique bivariée générée par $\{X_0, X_1, \dots, X_n\}$.

Alors $I_{n,1}$ peut s'écrire

$$I_{n,1} = \int \{\widehat{p}_2(x, y) - \widehat{p}(x)\widehat{p}(y)\} w(x, y) dF_n(x, y).$$

De plus, soit

$$F(x, y) = P(X_t \leq x, X_{t-1} \leq y) = \int_{-\infty}^x \int_{-\infty}^y p(u, v) dudv$$

et

$$U_n(x, y) = n^{-1/2} \{F_n(x, y) - F(x, y)\}$$

où $F_n(x)$, $F(x)$ et $U_n(x)$ sont les quantités univariées correspondantes. Nous définissons

$$p_h(x, y) = p_{h(n),2}(x, y) \triangleq \mathbb{E}[\widehat{p}(x, y)]$$

et notons

$$p_h(x, y) = \int k_h(x-u) k_h(y-v) p(u, v) dudv$$

et de même pour $p_h(x)$. $\widehat{A}(x, y) = \widehat{p}(x, y) - \widehat{p}(x)\widehat{p}(y)$ et $A_h(x, y) = p_h(x, y) - p_h(x)p_h(y)$. $I_{n,1}$ peut s'écrire

$$\begin{aligned} I_{n,1} &= \int A_h(x, y) w(x, y) dF(x, y) + \int \{\widehat{A}(x, y) - A_h(x, y)\} w(x, y) dF(x, y) \\ &\quad + n^{-1/2} \int A_h(x, y) w(x, y) dU_n(x, y) \\ &\quad + n^{-1/2} \int \{\widehat{A}(x, y) - A_h(x, y)\} w(x, y) dU_n(x, y). \end{aligned}$$

Faisons un développement de Taylor pour $\widehat{A}(x, y) = \widehat{p}(x, y) - \widehat{p}(x)\widehat{p}(y)$ au voisinage de $\{p_h(x, y), p_h(x), p_h(y)\}$. Utilisant le fait que $n^{1/2} \{\widehat{p}(x, y) - p_h(x, y)\} = \int k_h(x-u)k_h(y-v)dU_n(u, v)$ et $n^{1/2} \{\widehat{p}(x) - p_h(x)\} = \int k_h(x-u)dU_n(u)$, nous obtenons l'identité suivante

$$\begin{aligned}
I_{n,1} &= \int \{p_h(x, y) - p_h(x)p_h(y)\} w(x, y) dF(x, y) \\
&\quad + n^{-1/2} \left\{ \int k_h(x-u)k_h(y-v)w(x, y)dU_n(u, v)dF(x, y) \right. \\
&\quad - \int k_h(x-u)p_h(y)w(x, y)dU_n(u)dF(x, y) \\
&\quad - \int k_h(y-v)p_h(x)w(x, y)dU_n(v)dF(x, y) \\
&\quad \left. + \int \{p_h(x, y) - p_h(x)p_h(y)\} w(x, y)dU_n(x, y) \right\} \\
&\quad + n^{-1} \left\{ - \int k_h(x-u)k_h(y-v)w(x, y)dU_n(u)dU_n(v)dF(x, y) \right. \\
&\quad + \int k_h(x-u)k_h(y-v)w(x, y)dU_n(u, v)dU_n(x, y) \\
&\quad - \int k_h(x-u)p_h(y)w(x, y)dU_n(u)dU_n(x, y) \\
&\quad \left. - \int k_h(x-v)p_h(x)w(x, y)dU_n(v)dU_n(x, y) \right\} \\
&\quad - n^{-3/2} \int k_h(x-u)k_h(y-v)w(x, y)dU_n(u)dU_n(v)dU_n(x, y) \quad (2.8)
\end{aligned}$$

Sous l'hypothèse d'indépendance, nous avons $p_h(x, y) \equiv p_h(x)p_h(y)$, alors le premier terme du développement (2.8) s'annule.

2.1.1 Biais asymptotique

Le but est de trouver un test d'indépendance et il est nécessaire d'évaluer $\mathbb{E}(I_{n,1})$ sous H_0 . L'indépendance des X_t n'implique pas l'indépendance entre les paires (X_{t+1}, X_t) , (X_t, X_{t-1}) , et (X_{t-1}, X_{t-2}) et nous avons besoin d'une extension du théorème 2.1 de Joe (1989b).

Lemme 2.1.1 (Skaug and Tjøstheim 1993b) Soit X_t une suite de variables aléatoire i.i.d. et soit $F(x, y, z) = F_3(x, y, z) = F(x)F(y)F(z)$ la fonction de répartition jointe de (X_{t+1}, X_t, X_{t-1}) . Alors pour des fonctions quelconques $a_n(x)$, $a_n(x, y)$, $a_n(x, y; u)$, $a_n(x, y; u, v)$, et $b(x, y; u, v)$ pour lesquelles les intégrales ci-dessous existent, nous avons

$$\mathbb{E} \left\{ \int a_n(x) dU_n(x) \right\} = \mathbb{E} \left\{ \int a_n(x, y) dU_n(x, y) \right\} = 0 \quad (2.9)$$

$$\mathbb{E} \left\{ \int a_n(x, y) dU_n(x) dU_n(y) \right\} = \int a_n(x, x) dF(x) - \int a_n(x, y) dF(x, y) \quad (2.10)$$

$$\begin{aligned} & \mathbb{E} \left\{ \int a_n(x, y; u, v) dU_n(x, y) dU_n(u, v) \right\} \\ &= \int a_n(x, y; x, y) dF(x, y) + \int \{a_n(x, y; y, u) + a_n(y, u; x, y)\} dF(x, y, u) \\ & \quad - 3 \int a_n(x, y; u, v) dF(x, y) dF(u, v) + O(n^{-1}) \end{aligned} \quad (2.11)$$

$$\begin{aligned} \mathbb{E} \left\{ \int a_n(x, y; u) dU_n(x, y) dU_n(u) \right\} &= \int a_n(x, y; x) + a_n(x, y; y) dF(x, y) \\ & \quad - 2 \int a_n(x, y; u) dF(x, y, u) + O(n^{-1}) \end{aligned} \quad (2.12)$$

$$\mathbb{E} \left\{ \int b_n(x, y; u, v) dU_n(x, y) dU_n(u) dU_n(v) \right\} = O(n^{-1/2}) \quad (2.13)$$

Preuve. Pour la première identité de (2.9), nous avons

$$\begin{aligned} \mathbb{E} \left\{ \int a_n(x) dU_n(x) \right\} &= n^{1/2} \mathbb{E} \left\{ \int a_n(x) d(F_n(x) - F(x)) \right\} \\ &= n^{1/2} \mathbb{E} \left\{ \int a_n(x) dF_n(x) - \int a_n dF(x) \right\} \\ &= n^{1/2} \mathbb{E} \left\{ \frac{1}{n} \sum_{t=1}^n a_n(X_t) - \mathbb{E} a_n(X_1) \right\} \\ &= n^{1/2} \left\{ \frac{1}{n} \sum_{t=1}^n \mathbb{E} \{a_n(X_t)\} - \mathbb{E} \{a_n(X_1)\} \right\} = 0 \end{aligned}$$

De même pour la deuxième identité dans (2.9). Pour (2.10), en utilisant la définition de U_n

$$\begin{aligned} & E \left\{ \int a_n(x, y) dU_n(x) dU_n(y) \right\} \\ &= n^{-1} \mathbb{E} \left\{ \sum_t \sum_s a_n(X_t, X_s) \right\} - \mathbb{E} \left\{ \sum_t \int a_n(X_t, y) dF(y) \right\} \\ & \quad - \mathbb{E} \left\{ \sum_s \int a_n(x, X_s) dF(x) \right\} + n \int a_n(x, y) dF(x) dF(y). \end{aligned} \quad (2.14)$$

de l'équation (2.14), pour la double somme nous trouvons $\int a_n(x, x) dF(x) + (n-1) \int a_n(x, y) dF(x, y)$, pour le deuxième terme nous trouvons $n \int a_n(x, y) dF(x, y)$ de même pour le troisième; en combinant ceux-ci avec le quatrième terme, le résultat est établi.

Nous allons donner la preuve de (2.11), celles de (2.12) et (2.13) se font de la même manière. Utilisant la définition de U_n , (2.11) peut être écrite

$$\begin{aligned} & n^{-1} \mathbb{E} \left\{ \sum_t \sum_s a_n(X_t, X_{t-1}, X_s, X_{s-1}) \right\} - \mathbb{E} \left\{ \sum_t \int a_n(X_t, X_{t-1}, u, v) dF(u, v) \right\} \\ & - \mathbb{E} \left\{ \sum_s \int a_n(x, y, X_s, X_{s-1}) dF(x, y) \right\} + n \int a_n(x, y; u, v) dF(x, y) dF(u, v). \end{aligned} \quad (2.15)$$

Les termes au milieu et la première double somme de (2.15) pour $|s-t| > 1$ donne lieu à $-2n + n(n-3)/n$ termes de même type que le quatrième terme. Combinons les, y compris le dernier terme à droite de (2.11). Pour $s=t$ et $|s-t|=1$ de la double somme, nous obtenons le premier terme et $(n-1)/n$ fois le deuxième terme sur la droite de (2.11). ■

Le lemme 2.1.1 est utilisé pour obtenir le biais de $I_{n,1}$. Nous faisons cela pour les estimateurs des densités donnés par la méthode du noyau leave-one-out (voir

Silverman 1980, pp.49)

$$\begin{aligned}\hat{p}_2(X_t, X_{t-1}) &= \frac{1}{n-1} \sum_{s \neq t} k_h(X_s - X_t) k_h(X_{s-1} - X_{t-1}) \\ \hat{p}(X_t) &= \frac{1}{n-1} \sum_{s \neq t} k_h(X_s - X_t)\end{aligned}\quad (2.16)$$

et pour les estimateurs

$$\begin{aligned}\hat{p}(X_t, X_{t-1}) &= \frac{1}{n} \sum_s k_h(X_s - X_t) k_h(X_{s-1} - X_{t-1}) \\ \hat{p}(X_t) &= \frac{1}{n} \sum_s k_h(X_s - X_t)\end{aligned}\quad (2.17)$$

De plus, nous utilisons le produit des fonctions pondérées avec $w_2(x, y) = \mathbf{1}_{[x \in S]} \mathbf{1}_{[y \in S]}$, où $\mathbf{1}_{[z \in S]}$ est une fonction indicatrice du support S .

Proposition 2.1.1 (Skaug et Tjøstheim 1993b) *Soit X_t une suite de variables aléatoire i.i.d. et soient $p(x, y)$ et $p(x)$ deux densités de classe C^2 (soient deux fois continûment différentiables) Supposons que les conditions citées auparavant sur k (la fonction noyau) sont vérifiées et que $h = h(n) \rightarrow 0$ et $nh^2 \rightarrow \infty$ quand $n \rightarrow \infty$. Alors, en utilisant les estimateurs de (2.17), nous avons*

$$\begin{aligned}\mathbb{E}(I_{n,1}) &= n^{-1} h^{-2} k^2(0) \left\{ \int p(x) w(x) dx \right\}^2 \\ &\quad - 2n^{-1} h^{-1} k(0) \int p^2(x) w(x) dx \int p(x) w(x) dx \\ &\quad + n^{-1} \left[2 \left\{ \int p^2(x) w(x) dx \right\}^2 - \int p^3(x) w(x) dx \right] \\ &\quad + O(n^{-1}h) + O(n^{-2}h^{-2})\end{aligned}\quad (2.18)$$

Si les estimateurs leave-one-out (2.16) sont utilisées alors les deux termes sur la droite de (2.18) s'annulent.

Preuve. Prenons l'espérance de (2.8), remarquons de (2.9) que les termes multipliés par $n^{-1/2}$ à partir de (2.8) sont nuls. Notons les termes multipliés par n^{-1} par $n^{-1}(-S_1 + S_2 - S_3 - S_4)$; de (2.10) et sous H_0 , nous avons

$$\begin{aligned}\mathbb{E}(S_1) &= \int k_h(x-u) k_h(y-u) p(x) w(x) p(y) w(y) p(u) dx dy du \\ &\quad - \int k_h(x-u) k_h(y-v) p(x) w(x) p(y) w(y) p(u) p(v) dx dy du dv.\end{aligned}\tag{2.19}$$

Utilisons le développement de Taylor au voisinage de $h = 0$ pour

$$p_h(x) = \mathbb{E}[\widehat{p}(x)] = p(x) + \frac{1}{2}h^2 p''(x) + o(h^2)\tag{2.20}$$

Faisons un développement de Taylor à l'ordre deux pour (2.19)

$$\mathbb{E}(S_1) = \mathbb{E}(S_1)|_{h=0} + \frac{1}{2}h^2 \mathbb{E}''(S_1) + o(h^2)$$

nous trouvons

$$\begin{aligned}\mathbb{E}(S_1) &= \int h^{-1}k((x-u)/h) h^{-1}k((y-u)/h) p(x) w(x) p(y) w(y) p(u) dx dy du \\ &\quad - \int h^{-1}k((x-u)/h) h^{-1}k((y-v)/h) p(x) w(x) \\ &\hspace{20em} p(y) w(y) p(u) p(v) dx dy du dv \\ &= \int k(z) k(t) p(hz+u) w(hz+u) p(ht+u) w(ht+u) p(u) dz dt du \\ &\quad - \int k(z) k(t) p(hz+u) w(hz+u) p(ht+v) w(ht+v) \\ &\hspace{20em} p(u) p(v) dz dt du dv\end{aligned}$$

$$\begin{aligned}\mathbb{E}(S_1)|_{h=0} &= \int k(z) k(t) p(u) w(u) p(u) w(u) p(u) dz dt du \\ &\quad - \int k(z) k(t) p(u) w(u) p(v) w(v) p(u) p(v) dz dt du dv\end{aligned}$$

$$\begin{aligned}
&= \underbrace{\int k(z) dz}_{=1} \underbrace{\int k(t) dt}_{=1} \int p^3(u) w(u) du \\
&\quad - \underbrace{\int k(z) dz}_{=1} \underbrace{\int k(t) dt}_{=1} \int p(u) w(u) p(v) w(v) dudv \\
&= \int p^3(u) w(u) du - \left\{ \int p^2(u) w(u) du \right\}^2
\end{aligned}$$

pour $\mathbb{E}'(S_1)$, elle est nulle.

$$\begin{aligned}
\mathbb{E}'(S_1) &= \int zk(z)k(t)p'(hz+u)w(hz+u)p(ht+u)w(ht+u)p(u)dzdtdu \\
&\quad + \int zk(z)k(t)p(hz+u)w'(hz+u)p(ht+u)w(ht+u)p(u)dzdtdu \\
&\quad + \int tk(z)k(t)p(hz+u)w(hz+u)p'(ht+u)w(ht+u)p(u)dzdtdu \\
&\quad + \int tk(z)k(t)p(hz+u)w(hz+u)p(ht+u)w'(ht+u)p(u)dzdtdu \\
&\quad - \int zk(z)k(t)p'(hz+u)w(hz+u)p(ht+v)w(ht+v)p(u)p(v)dzdtdu \\
&\quad - \int zk(z)k(t)p(hz+u)w'(hz+u)p(ht+v)w(ht+v)p(u)p(v)dzdtdu \\
&\quad - \int tk(z)k(t)p(hz+u)w(hz+u)p'(ht+v)w(ht+v)p(u)p(v)dzdtdu \\
&\quad - \int tk(z)k(t)p(hz+u)w(hz+u)p(hz+v)w'(hz+v)p(u)p(v)dzdtdu \\
&= \underbrace{\int zk(z)dz}_{=0} \int k(t)dt \int p'(hz+u)w(hz+u)p(ht+u)w(ht+u)p(u)tdu \\
&\quad + \underbrace{\int zk(z)dz}_{=0} \int k(t)dt \int p(hz+u)w'(hz+u)p(ht+u)w(ht+u)p(u)tdu \\
&\quad + \int k(z)dz \underbrace{\int tk(t)dt}_{=0} \int p(hz+u)w(hz+u)p'(ht+u)w(ht+u)p(u)tdu \\
&\quad + \int k(z)dz \underbrace{\int tk(t)dt}_{=0} \int p(hz+u)w(hz+u)p(ht+u)w'(ht+u)p(u)tdu \\
&\quad - \underbrace{\int zk(z)dz}_{=0} \int k(t)dt \int p'(hz+u)w(hz+u)p(ht+v)w(ht+v)p(u)p(v)
\end{aligned}$$

$p(v) dudv$

$$\begin{aligned}
& - \underbrace{\int zk(z) dz}_{=0} \int k(t) dt \int p(hz+u) w'(hz+u) p(ht+v) w(ht+v) p(u) p(v) dudv \\
& - \int k(z) dz \underbrace{\int tk(t) dt}_{=0} \int p(hz+u) w(hz+u) p'(ht+v) w(ht+v) p(u) p(v) dudv \\
& - \int k(z) dz \underbrace{\int tk(t) dt}_{=0} \int p(hz+u) w(hz+u) p(ht+v) w'(ht+v) p(u) p(v) dudv \\
& = 0
\end{aligned}$$

De la même manière pour l'ordre 2

$$\begin{aligned}
\mathbb{E}'(S_1)|_{h=0} &= \int p''(u) w^2(u) p(u) du + \int p'(u) w' p(u) w(u) du \\
&+ \int p'(u) w'(u) p(u) w(u) du + \int p^2(u) w''(u) w(u) du \\
&+ \int p''(u) w^2(u) p(u) du + \int p'(u) w' p(u) w(u) du \\
&+ \int p'(u) w'(u) p(u) w(u) du + \int p^2(u) w''(u) w(u) du \\
&- \int p''(u) w(u) p(v) w(v) dudv - \int p'(u) w'(u) p(v) w(v) dudv \\
&- \int p'(u) w'(u) p(v) w(v) dudv - \int p(u) w''(u) p(v) w(v) dudv \\
&- \int p(u) w(u) p''(v) w(v) dudv - \int p(u) w(u) p'(v) w'(v) dudv \\
&- \int p(u) w(u) p''(v) w(v) dudv - \int p(u) w(u) p'(v) w'(v) dudv \\
&- \int p(u) w(u) p'(v) w'(v) dudv - \int p(u) w(u) p'(v) w''(v) dudv
\end{aligned}$$

qui sont des quantités finies, donc

$$\mathbb{E}(S_1) = \int p^3(x) w(x) dx - \left\{ \int p^2(x) w(x) dx \right\}^2 + O(h^2).$$

De la même manière, en utilisant (2.11) et (2.12) nous concluons que

$$\begin{aligned} \mathbb{E}(S_2) &= h^{-2}k^2(0) \left\{ \int p(x)w(x)dx \right\}^2 + 2 \int p^3(x)w(x)dx \\ &\quad - 3 \left\{ \int p^2(x)w(x)dx \right\}^2 + O(h^2) + O(n^{-1}) \end{aligned}$$

et

$$\begin{aligned} \mathbb{E}(S_3) &= \mathbb{E}(S_4) = h^{-1}k(0) \int p^2(x)w(x)dx \int p(x)w(x)dx + \int p^3(x)w(x)dx \\ &\quad - 2 \left\{ \int p^2(x)w(x)dx \right\}^2 + O(h). \end{aligned}$$

De plus, en utilisant (2.13) le terme de l'ordre le plus élevé de (2.8) (i.e., celui multiplié par $n^{-3/2}$) est donné par $n^{-2}h^{-2}k^2(0) \int w(x)dx$. En rassemblant les termes, nous obtenons (2.18) ■

Si nous utilisons la méthode de leave-one-out, ceci signifie que l'on suppose $k(0) = 0$ et il en est de même dans la deuxième partie de la proposition.

Dans le cas gaussien i.e., $X_t \sim N(0, 1)$ et $w(x) \equiv 1$,

$$2 \left\{ \int p^2(x)dx \right\}^2 - \int p^3(x)dx = \frac{1}{2\pi} (1 - 3^{-1/2}) \approx 0.067.$$

Dans la pratique, nous constatons que le biais de $I_{n,1}$ peut être négligé quand nous utilisons les estimations de la méthode leave-one-out (2.16) dans la construction de la statistique correspondante pour $n \geq 100$ i.e., nous remplaçons $\mathbb{E}(I_{n,1})$ par zéro. Ce n'est pas le cas si les estimations de (2.17) sont utilisées.

2.1.2 Variance asymptotique

Nous utilisons la même méthode que dans 2.1.1 et posons que $w(x, y) = \mathbf{1}_{[x \in S]} \mathbf{1}_{[y \in S]}$ où $S \subset \mathbb{R}$ est un ensemble Lebesgue mesurable.

Proposition 2.1.2 (Skaug et Tjøstheim 1993b) *Sous les mêmes hypothèses que la proposition 2.1.1, nous avons pour les deux estimations (2.16) et (2.17)*

$$\begin{aligned} \text{var}(I_{n,1}) &= n^{-1} \left(\int p^3(x) w(x) dx - \left\{ \int p^2(x) w(x) dx \right\}^2 \right) \\ &\quad + O(n^{-1}h^2) + O(n^{-2}h^{-2}). \end{aligned} \quad (2.21)$$

Preuve. Nous commençons par trouver la variance du terme en $n^{-1/2}$ de (2.8). Sous H_0 , ce terme peut être écrit comme

$$B_n = n^{-1/2} \left\{ \int p_{h,w}(x, y) dU_n(x, y) - 2 \int p_h(x) p(x) w(x) d(x) \int p_{h,w}(x) dU_n(x) \right\}, \quad (2.22)$$

où

$$p_{h,w}(x, y) = \int k_h(x-u) k_h(y-v) p(u, v) w(x, y) d(u, v)$$

et de même pour $p_{h,w}(x)$. Par suite de (2.9), nous avons $\text{var}(B_n) = \mathbb{E}(B_n^2)$, et nous pouvons utiliser (2.10)-(2.12) pour calculer les espérances des intégrales. Nous obtenons, en utilisant $p_h(x, y) = p_h(x) p_h(y)$ et $p_{h,w}(x, y) = p_{h,w}(x) p_{h,w}(y)$

$$\begin{aligned} &\mathbb{E} \left[\left\{ \int p_{h,w}(x, y) dU_n(x, y) \right\}^2 \right] \\ &= \left\{ \int p_{h,w}^2(x) p(x) d(x) \right\}^2 + 2 \frac{n-1}{n} \int p_{h,w}^2(x) p(x) d(x) \left\{ \int p_{h,w}(x) p(x) d(x) \right\}^2 \\ &\quad - 3 \left\{ \int p_{h,w}(x) p(x) d(x) \right\}^4 \end{aligned}$$

$$\begin{aligned} &\mathbb{E} \left[\int p_{h,w}(x, y) dU_n(x, y) \int p_h(x) p(x) w(x) d(x) \int p_{h,w}(x) dU_n(x) \right] \\ &= \int p_h(x) p(x) w(x) d(x) \left[2 \int p_{h,w}^2(x) p(x) d(x) \int p_{h,w}(x) p(x) d(x) \right. \\ &\quad \left. - 2 \left\{ \int p_{h,w}(x) p(x) d(x) \right\}^3 \right], \end{aligned}$$

$$\begin{aligned}
& \mathbb{E} \left[\left\{ \int p_h(x) p(x) d(x) \int p_{h,w}(x) dU_n(x) \right\}^2 \right] \\
&= \left\{ \int p_h(x) p(x) w(x) d(x) \right\}^2 \left[\int p_{h,w}^2(x) p(x) d(x) - \left\{ \int p_{h,w}(x) p(x) d(x) \right\}^2 \right].
\end{aligned} \tag{2.23}$$

De (2.20)

$$\int p_{h,w}(x) p(x) d(x) = \int p^2(x) w(x) d(x) + O(h^2)$$

avec des expressions similaires pour des intégrales impliquant $p_h(x)$ et $p_{h,w}(x)$. Par un calcul direct

$$\text{var}(B_n) = n^{-1} \left(\int p^3(x) w(x) d(x) - \left\{ \int p^2(x) w(x) d(x) \right\}^2 \right)^2 + O(n^{-1}h^2).$$

Les termes multipliés par n^{-1} et par $n^{-3/2}$ de (2.8) peuvent être traités de la même façon et en faisant une extension de 2.1.1 (et du lemme 2.1 de Joe, 1989b). La variance du terme multiplié par n^{-1} est donnée par

$$2n^{-2}h^{-2} \left\{ \int k^2(x) dx \right\} \left\{ \int p^2(x) w(x) dx \right\}^2 \tag{2.24}$$

Celle du terme multiplié par $n^{-3/2}$ est de l'ordre $O(n^{-3}h^{-2})$. Les covariances des différents termes de (2.8) sont de l'ordre $o(n^{-2}h^{-2})$, ainsi (2.21) est obtenue. ■

2.1.3 Normalité asymptotique

Nous travaillerons sous l'hypothèse nulle bien que la normalité asymptotique peut être prouvé dans le cas de dépendance, à l'aide du développement de Taylor tronqué et du théorème de mélange. Commençons par (2.8) et notons les termes multipliés par $n^{-1/2}$, n^{-1} et $n^{-3/2}$ par I , II et III respectivement, de sorte que

$$I_{n,1} = n^{-1/2}I + n^{-1}II + n^{-3/2}III$$

Ici $n^{-1/2}I = B_n$ où B_n est donné dans (2.22). De ce qui précède, nous avons $\mathbb{E} \left[\{n^{-1/2}I\}^2 \right] = O(n^{-1})$, $n^{-2}\mathbb{E} \left[\{II - \mathbb{E}(II)\}^2 \right] = O(n^{-2}h^{-2})$, $\mathbb{E} \left[\{n^{-3/2}III\} \right] = O(n^{-2}h^{-2})$ et $\text{var} (n^{-3/2}III) = O(n^{-3}h^{-2})$. Par conséquent, $n^{-1/2}I = O_p(n^{-1/2})$ et sous les hypothèses de la proposition 2.1.2, $n^{-1} \{II - \mathbb{E}(II)\} + n^{-3/2}III = o_p(n^{-1/2})$. Il ressort de résultats standards (voir Brockwell and Davis 1987 pp 198) que $n^{-1/2} \{I_{n,1} - n^{-1}\mathbb{E}(II)\}$ et $I = n^{1/2}B_n$ ont la même distribution asymptotique.

Théorème 2.1.1 (Skaug et Tjøstheim 1993b) *Sous les hypothèses de la proposition 2.1.2 et pour la fonction poids $w(x, y) = \mathbf{1}_{[x \in S]} \mathbf{1}_{[y \in S]}$ de support compact S et $\int p^3(x) w(x) dx - \left\{ \int p^2(x) w(x) dx \right\}^2 \neq 0$, alors*

$$n^{1/2} \{I_{n,1} - n^{-1}\mathbb{E}(II)\} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, \left\{ \int p^3(x) w(x) dx - \left\{ \int p^2(x) w(x) dx \right\}^2 \right\} \right),$$

où $\mathbb{E}(II)$ est donnée dans la proposition 2.1.2.

Preuve. Il suffit de prouver la normalité asymptotique pour $n^{1/2}B_n$ où B_n est donnée dans (2.22). Utilisant la définition de U_n

$$\begin{aligned} n^{1/2}B_n &= n^{-1/2} \sum_t \left\{ p_{h,w}(X_t, X_{t-1}) - \int p_{h,w}(x, y) p(x, y) dx dy \right\} \\ &\quad - 2n^{-1/2} \int p_h(x) p(x) w(x) dx \sum_t \left\{ p_{h,w}(X_t) - \int p_{h,w}(x) p(x) dx \right\} \\ &= n^{-1/2} \sum_t \{C_{h,w}(X_t, X_{t-1}) - 2D_{h,w}(X_t)\} = n^{-1/2} \sum_t G_{h,w}(X_t, X_{t-1}), \end{aligned}$$

où

$$\begin{aligned} C_{h,w}(u, v) &= p_{h,w}(u, v) - \int p_{h,w}(x, y) p(x, y) dx dy, \\ D_{h,w}(u) &= \int p_h(x) p(x) w(x) dx \left\{ p_{h,w}(u) - \int p_{h,w}(x) p(x) dx \right\} \end{aligned}$$

et

$$G_{h,w}(u, v) = C_{h,w}(u, v) - 2G_{h,w}(u).$$

Nous pouvons l'écrire comme

$$\begin{aligned} n^{-1/2} \sum_t G_{h,w}(X_t, X_{t-1}) &= n^{-1/2} \sum_t G_w(X_t, X_{t-1}) \\ &\quad + n^{-1/2} \sum_t \{G_{h,w}(X_t, X_{t-1}) - G_w(X_t, X_{t-1})\} \end{aligned}$$

où

$$\begin{aligned} n^{-1/2} \sum_t G_w(X_t, X_{t-1}) &= \int p(x, y) w(x, y) dU_n(x, y) \\ &\quad - 2 \int p(x) w(x) d(x) \int p(x) w(x) dU_n(x). \end{aligned}$$

En utilisant le même raisonnement que 2.1.2 et les hypothèses du théorème, il s'ensuit que

$$\mathbb{E} \left[\left\{ n^{-1/2} \sum_{t=1}^n G_w(X_t, X_{t-1}) \right\}^2 \right] \rightarrow \left[\int p^3(x) w(x) dx - \left\{ \int p^2(x) w(x) dx \right\}^2 \right]^2 > 0$$

quand $n \rightarrow \infty$. Puisque $\{X_t\}$ est i.i.d. et G_w est une fonction mesurable sur \mathbb{R}^2 , le processus $\{Y_t\} = \{G_w(X_t, X_{t-1})\}$ est strictement stationnaire et 1-dépendant. A partir du théorème central limite pour un processus α -dépendant (Brockwell and Davis 1987, théorème 6.4.2)

$$n^{-1/2} \sum_t G_w(X_t, X_{t-1}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N \left(0, \left[\int p^3(x) w(x) dx - \left\{ \int p(x) w(x) dx \right\}^2 \right]^2 \right).$$

Le théorème sera démontré si l'on peut démontrer que

$$n^{-1/2} \sum_t \{G_{h,w}(X_t, X_{t-1}) - G_w(X_t, X_{t-1})\} = o_p(1),$$

ou bien

$$n^{-1} \mathbb{E} \left[\left(\sum_t \{G_{h,w}(X_t, X_{t-1}) - G_w(X_t, X_{t-1})\} \right)^2 \right] = o(1).$$

En utilisant le lemme 1.2.2,

$$\begin{aligned} & n^{-1} \mathbb{E} \left[\left(\sum_t \{G_{h,w}(X_t, X_{t-1}) - G_w(X_t, X_{t-1})\} \right)^2 \right] \\ = & \mathbb{E} \left[\left(\int \{C_{h,w}(x, y) - C_w(x, y)\} dU_n(x, y) - 2 \int \{D_{h,w}(x) - D_w(x)\} dU_n(x) \right)^2 \right] \\ = & \int \{C_{h,w}(x, y) - C_w(x, y)\}^2 p(x, y) dx dy \\ & + 2 \frac{n-1}{n} \int \{C_{h,w}(x, y) - C_w(x, y)\} \{C_{h,w}(y, v) - C_w(y, v)\} p(x, y, v) dx dy dv \\ & - 3 \left(\int \{C_{h,w}(x, y) - C_w(x, y)\} p(x, y) dx dy \right)^2 \\ & + 4 \int \{D_{h,w}(x) - D_w(x)\}^2 p(x) dx - 4 \left(\int \{D_{h,w}(x) - D_w(x)\} p(x) dx \right)^2 \\ & - 4 \int \{C_{h,w}(x, y) - C_w(x, y)\} \{D_{h,w}(x) + D_{h,w}(y) - D_w(x) - D_w(y)\} p(x, y) dx dy \\ & + 8 \int \{C_{h,w}(x, y) - C_w(x, y)\} p(x, y) dx dy \int \{D_{h,w}(x) - D_w(x)\} p(x) dx. \quad (2.25) \end{aligned}$$

Comme $p_h(x, y) = \int k_h(x-u) k_h(y-v) p(u, v) dudv$ il s'ensuit de la continuité de p que $p_h(x, y) \rightarrow p(x, y)$ pour tout (x, y) si $h = h(n) \rightarrow 0$. Puisque $w(x)$ a un support compact S , nous avons la continuité uniforme de p par conséquent, la convergence uniforme de $p_h(x, y) \rightarrow p(x, y)$ sur $S \times S$ quand $n \rightarrow \infty$. Des définitions de $C_{h,w}$, C_w , $D_{h,w}$ et D_w il s'ensuit que $C_{h,w}(x, y) \rightarrow C_w(x, y)$ et $D_{h,w}(x, y) \rightarrow D_w(x, y)$ uniformément sur $S \times S$ et S dand $n \rightarrow \infty$. (il est vraiment nécessaire de redéfinire $C_{h,w}(x, y)$ et $D_w(x)$ aux bords de $S \times S$ et S , mais en supposant que ceux-ci ont une mesure de Lebesgue nulle cette contribution est négligeable). Comme les intégrales dans (2.25) sont sur S , $S \times S$ et $S \times S \times S$, et p et p_h sont bornées sur S grâce à la continuité il s'ensuit que pour un $\varepsilon > 0$ donné il existe

un N tel que pour $n > N$, $|C_{h,w}(x, y) - C_h(x, y)| < \varepsilon$ et $|D_{h,w}(x) - D_h(x)| < \varepsilon$ et il existe un M tel que $p(x) < M$ pour tout $x \in S$ et $(x, y) \in S \times S$. Donc chaque intégrale est majorée par $\varepsilon^p M^q |S|^q$ où $p = 1$ ou 2 , $q = 1, 2$ ou 3 et $|S|$ est la mesure de Lebesgue de S . Il résulte que (2.25) est de l'ordre $o(1)$ quand $n \rightarrow \infty$. ■

2.2 Test basé sur $I_{n,1}$

En se basant sur les résultats des trois sous sections précédentes, la statistique naturelle pour tester l'indépendance est

$$I_{n,1}^* = \{I_{n,1} - \mathbb{E}_n(I_{n,1})\} / SD_n(I_{n,1}), \quad (2.26)$$

où les expressions asymptotiques de $\mathbb{E}(I_{n,1})$ et $SD(I_{n,1})$ sont données par (2.18) et (2.21), et les estimateurs naturels dans le cas de “leave-one-out” sont donnés par

$$\mathbb{E}_n(I_{n,1}) = n^{-1} \left[2 \left\{ \frac{1}{n} \sum_t \widehat{p}(X_t) w(X_t) \right\}^2 - \frac{1}{n} \sum_t \widehat{p}^2(X_t) w(X_t) \right],$$

et

$$SD_n(I_{n,1}) = n^{-1/2} \left| \frac{1}{n} \sum_t \widehat{p}^2(X_t) w(X_t) - \left\{ \frac{1}{n} \sum_t \widehat{p}(X_t) w(X_t) \right\}^2 \right|. \quad (2.27)$$

Pour n grand on s'attend à ce que $I_{n,1}^*$ soit approximativement une loi normale réduite sous l'hypothèse nulle. Nous rejetons l'hypothèse nulle de l'indépendance si $I_{n,1}^* \geq u_\alpha$ où u_α est le fractile d'ordre $1 - \alpha$ de la loi normale standard.

3 Mesure basée sur l'information conditionnelle

Cette mesure est proche de celle introduite par Wu et al (1993) pour mesurer la dépendance conditionnelle. Elle est définie en termes de rapport des intégrales de corrélation. l'intégrale de corrélation trouve son origine dans l'étude des systèmes

chaotiques où elle est un moyen important pour caractériser les dynamiques des processus déterministes. Son estimation est relativement directe. La liaison entre les intégrales de corrélations et les quantités théoriques d'information établie par Prichard et Theiler (1995) montre que la mesure de Diks et Manzan (2002) correspond à l'information mutuelle conditionnelle de second ordre. Les quantités théoriques d'information utilisées par Granger et Lin (1994) sont aussi reliées à cette statistique. Cependant, leur test statistique est une généralisation de la fonction d'auto-corrélation, bien que celui de Diks et Manzan (2002), généralise la fonction d'auto-corrélation partielle. Ceci rend la mesure de Diks et Manzan (2002), plus adéquate pour étudier le retard de la dépendance, ce qui peut servir en tant que première étape à la sélection de l'ordre du modèle. Comme le nombre de paramètres dans les modèles paramétriques des séries chronologiques non linéaires (tel que *TAR*) augmente généralement rapidement avec le nombre de retards retenus, les critères de sélection de retrads sont importants pour la construction de modèles de séries chronologiques parcimonieux.

3.1 Définitions

La statistique de Diks et Manzan (2002) est basée sur l'information mutuelle conditionnelle. L'avantage de l'utiliser comme une statistique de test c'est qu'elle capture les dépendances tandis que le conditionnement sur les valeurs intermédiaires de la série temporelle donne également un aperçu sur l'ordre du processus sous-jacent.

Considérons l'information mutuelle conditionnelle dans le cadre des séries chronologiques. Pour une série temporelle stationnaire $\{X_t\}_{t=1, \overline{T}}$, définissons le vecteur de retard comme

$$X_t^m = (X_t, \dots, X_{t-m+1})', \quad (3.1)$$

où m est la dimension d'empilement. Le nombre total des vecteurs est N tel que $N = T - m + 1$.

L'information mutuelle conditionnelle entre X_t et X_{t-m} sachant une observation intermédiaire X_{t-1}^{m-1} est donnée par

$$\mathcal{I}(X_t, X_{t-m} | X_{t-1}^{m-1}) = -H(X_t^{m+1}) + 2H(X_t^m) - H(X_{t-1}^{m-1}), \quad (3.2)$$

où H est l'entropie de Shannon définie comme dans l'appendice.

L'information mutuelle conditionnelle a une interprétation particulière dans le contexte des séries temporelles, si X_t est un processus de Markov d'ordre k , la densité de probabilité conditionnelle dépend seulement des k dernières valeurs retardées de la série temporelle, les autres retards ne contiennent aucune information supplémentaire. L'information mutuelle conditionnelle entre X_t et X_{t-m} devient nulle pour $m > k$ et positive pour $m \leq k$. Dans ce sens l'information mutuelle conditionnelle peut être interprétée comme un identifiant de l'ordre.

Une autre interprétation utile est la suivante : La quantité moyenne de l'information sur X_t contenu dans X_{t-1}^m est donnée par $\mathcal{I}(X_t, X_{t-1}^m)$, tandis que la quantité moyenne de l'information de X_t dans X_{t-1}^{m-1} est seulement donnée par $\mathcal{I}(X_t, X_{t-1}^{m-1})$. Si nous soustrayons ces deux mesures d'information nous retrouvons

$$\mathcal{I}(X_t, X_{t-1}^m) - \mathcal{I}(X_t, X_{t-1}^{m-1}) = \mathcal{I}(X_t, X_{t-m} | X_{t-1}^{m-1}). \quad (3.3)$$

l'information mutuelle conditionnelle (voir l'appendice). Ceci montre clairement que l'information mutuelle conditionnelle quantifie la quantité moyenne de l'information supplémentaire de ce que X_{t-1}^m contient de X_t , en plus de l'information qui existe déjà dans X_{t-1}^{m-1} . Si X_{t-m} ne contient aucune information supplémentaire sur les valeurs de X_t en plus de celle de X_{t-1}^{m-1} , $\mathcal{I}(X_t, X_{t-m} | X_{t-1}^{m-1}) = 0$. Si d'autre part X_{t-m} contient de l'information supplémentaire sur X_t nous nous attendons à $\mathcal{I}(X_t, X_{t-m} | X_{t-1}^{m-1}) > 0$. Diks et Manzan (2002) proposent le test unilatérale basé

sur $\mathcal{I}(X_t, X_{t-m} | X_{t-1}^{m-1})$, estimée à partir des intégrales de corrélation.

Soit $C_m(\epsilon)$ et $\widehat{C}_m(\epsilon)$ les notations abrégées de $C_2(X_t^m; \epsilon)$ et de son estimateur $\widehat{C}_2(X_t^m; \epsilon)$ respectivement (voir la définition dans l'appendice), nous écrivons

$$\widehat{\mathcal{I}}(X_t, X_{t-m} | X_{t-1}^{m-1}) = -2 \ln \widehat{C}_m(\epsilon) + \ln \widehat{C}_{m+1}(\epsilon) + \ln \widehat{C}_{m-1}(\epsilon). \quad (3.4)$$

Pour la relation entre l'intégrale de corrélation et l'information voir l'appendice.

L'intégrale de corrélation d'ordre deux ($q = 2$) pour des vecteurs aléatoires de dimension m , X_t^m est

$$C_m(\epsilon) = \int \int \mathbf{1}_{\|x-x'\| \leq \epsilon} f_{X_m}(s) f_{X_m}(t) ds dt. \quad (3.5)$$

qui est l'espérance du noyau, $\mathbb{E}(\mathbf{1}_{\|X_m^1 - X_m^2\| \leq \epsilon})$, qui peut être estimée à partir des U -statistiques par

$$\widehat{C}_m(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathbf{1}_{(\|X_i^m - X_j^m\| \leq \epsilon)}.$$

où $\|\cdot\|$ désigne la norme L_∞ .

Notons que l'information mutuelle conditionnelle est une mesure non bornée de la dépendance conditionnelle. Ce qui conduit à l'utilisation de la version transformée donnée par

$$I_{n,2}(j, \epsilon) = 1 - \exp(-\widehat{\mathcal{I}}(X_t, X_{t-j} | X_{t-1}, \dots, X_{t-j+1})) = 1 - \frac{[\widehat{C}_j(\epsilon)]^2}{\widehat{C}_{j-1}(\epsilon) \widehat{C}_{j+1}(\epsilon)}, \quad (3.6)$$

qui prend des valeurs entre 0 et 1. L'utilisation de $I_{n,2}(j, \epsilon)$ a été proposé pour la première fois par Savit et Green (1991) pour déterminer la dimension de l'attracteur chaotique¹³. Notons dorénavant $I_{n,2}(j, \epsilon) \equiv I_{n,2}(j)$ puisque ϵ est fixé.

¹³Une des découvertes les plus spectaculaires des dernières années a été celle des attracteurs étranges (Chaotiques), ces objets géométriques issus de l'évolution de systèmes chaotiques. Dans le plan, ils sont formés d'une suite infinie de points $x_0, x_1, \dots, x_n, \dots$ qui dépendent de x_0 la valeur

Remarque 3.1.1 Dans la pratique la statistique $I_{n,2}(j, \epsilon)$ peut avoir des valeurs négatives (voir Savit and Green, 1991)

3.2 Distribution asymptotique

Wu et al (1993) ont montré que la distribution asymptotique de $I_{n,2}(j)$ sous H_0 est une normale de moyenne zéro et variance $V_{I_2(j)}$.

$$T^{1/2}I_{n,2}(j) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, V_{I_2(j)}) \quad (3.7)$$

où la variance asymptotique est donnée par

$$V_{I_2(j)} = 4 \left\{ \left(\frac{K_1(\epsilon)}{C_1(\epsilon)} \right)^{m-1} - \left[\left(\frac{K_1(\epsilon)}{C_1(\epsilon)} \right)^2 - 1 \right] \right\}^2$$

et $C_1(\epsilon)$ indique l'intégrale de corrélation lorsque la dimension d'empilement est égale à 1 tandis que $K_1(\cdot)$ est estimé par

$$\hat{K}_1(\epsilon) = \frac{2}{N(N-1)(N-2)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{k=j+1}^N \mathbf{1}_{(\|x_i - x_j\| \leq \epsilon)} \mathbf{1}_{(\|x_j - x_k\| \leq \epsilon)}$$

3.3 Test d'indépendance sérielle

Il est connu que l'approximation normale basée sur la distribution asymptotique ne donne pas toujours de bons résultats pour des tailles modérées. Diks et Manzan (2002) ont montré dans leurs études de simulation que ces problèmes se posent également pour les $I_{n,2}(j, \epsilon)$. C'est la principale motivation pour l'utilisation d'une approche bootstrap pour déterminer la distribution de la statistique de test sous

initiale. Au fur et à mesure que le nombre de points augmente, une image se forme dans le plan et devient de plus en plus nette. Cette image n'est pas une courbe ni une surface, c'est en fait un objet intermédiaire constitué de points avec entre eux des espaces inoccupés. L'objet est qualifié d'étrange en raison de sa structure pointilliste et de sa nature fractale. Une valeur différente de x_0 conduit à une toute autre suite qui après une courte phase, dessine la même image. D'où qu'on parte, on se retrouve toujours sur l'attracteur, c'est le côté prévisible de l'évolution. Où se retrouve-t-on exactement sur l'attracteur ? Il est impossible de répondre à la question, c'est le côté imprévisible de l'évolution.

l'hypothèse nulle.

Sous l'hypothèse nulle i.i.d. les statistiques d'ordre fournissent une statistique exhaustive minimale. Sous l'hypothèse nulle, et conditionnellement à la statistique d'ordre, chaque permutation des données observées est tout aussi probable, de sorte que le conditionnement par la statistique d'ordre conduit à un test de permutation¹⁴. Parce qu'un test de permutation peut être facilement mis en œuvre, il offre un moyen commode pour obtenir un test exact. $\widehat{C}_1(\epsilon)$ et $\widehat{K}_1(\epsilon)$ sont des fonctions de statistiques d'ordre donc sont invariantes quelque soit la permutation des séries temporelles sous H_0 . Dans ce cadre les p -valeurs ne sont pas affectées par les paramètres de position et d'échelle des statistiques. En plus les tests de permutation sont faciles à appliquer, et évitent les calculs embarrassants nécessaires pour obtenir la variance asymptotique. La procédure du test est telle que

1. Calculer $I_{n,2}(m, \epsilon)$ pour la série temporelle $\{X_t\}_{t=1, \overline{T}}$.
2. Permuter aléatoirement la série chronologique et obtenir $\{\widetilde{X}_t\}_{t=1, \overline{T}}$.
3. Calculer le test statistique basé sur $\{\widetilde{X}_t\}_{t=1, \overline{T}}$ noté par $\widetilde{I}_{n,2}(j, \epsilon)$.
4. Répéter les étapes 2 et 3 B fois.
5. Calculer la p -valeur unilatérale $\widehat{p} = \frac{1 + \#\{\widetilde{I}_{n,2}(m, \epsilon) \geq I_{n,2}(m, \epsilon)\}}{1 + B}$.
6. Rejeter l'hypothèse nulle de l'indépendance si $\widehat{p} \leq \alpha$, où α est le niveau de signification choisi.

Remarque 3.3.1 *Cette mesure est aussi utilisée pour tester la linéarité dans les modèles de séries chronologiques.*

4 Mesure basée sur l'entropie

Nous avons vu dans la section précédente pour des distributions continues, que l'écart entre la densité conjointe $f_j(\cdot)$ (on a changé de notation pour ne pas en-

¹⁴L'idée d'utiliser un test de permutation dans le contexte de l'indépendance sérielle remonte à Pitman (1937).

combrer la l'écriture) et le produit des densités marginales $p(\cdot)p(\cdot)$, est un signe de dépendance qui peut être mesurer de plusieurs façons. Dans cette section nous allons mesurer cette deviation en utilisant l'information de Kullback-Leibler (1959).

Soit $\{X_t\}$, un processus strictement stationnaire de densités marginale et conjointe $p(\cdot)$ et $f_j(\cdot)$ respectivement pour $Z_{jt} = (X_t, X_{t-j})'$. L'information de Kullback-Leibler est donnée par

$$I_3(j) = \int \ln \left(\frac{f_j(x, y)}{p(x)p(y)} \right) f_j(x, y) dx dy \quad j \in \mathbb{N}. \quad (4.1)$$

$I_3(j)$ n'est pas une métrique car elle n'est pas symétrique et ne satisfait pas l'inégalité triangulaire. $I_3(j)$ peut caractériser toutes les dépendances sérielles deux à deux car $I_3(j) \geq 0$ (facile de vérifier en utilisant l'inégalité de Jensen¹⁵) et $I_3(j) = 0$ si et seulement si X_t et X_{t-1} sont indépendantes. De plus elle possède une interprétation intéressante de la théorie de l'information ; elle mesure la prévisibilité par la différence d'entropie entre les distributions a priori et a posteriori, et elle est invariante sous n'importe quelle transformation continue et monotone de $\{X_t\}$. La propriété d'invariance est attractive parce que $\{X_t\}$ est i.i.d. si et seulement si toute transformation monotone et continue de $\{X_t\}$ est i.i.d.

Granger et Lin (1994) ont étudié les propriétés de $I_3(j)$ normalisée

$$\gamma^2(j) = 1 - e^{-2I_3(j)} \quad j \in \mathbb{N} \quad (4.2)$$

Elle est interprétée comme une auto-corrélation ombre "*shadow autocorrelation*" de $\{X_t\}$. Il ont proposé l'estimateur

$$\hat{\gamma}_n^2(j) = 1 - e^{-2I_{n,3}(j)} \quad j = 1, \dots, n-1 \quad (4.3)$$

où $I_{n,3}(j)$ est un estimateur non paramétrique lissé de l'entropie, basé sur l'échan-

¹⁵Si X is une variable aléatoire et φ une fonction convexe, alors $\varphi[\mathbb{E}(X)] \leq \mathbb{E}[\varphi(X)]$.

tillon $\mathcal{X} = \{X_t\}_{t=1,n}$

$$I_{n,3}(j) = \frac{1}{n-j} \sum_{t \in S_n(j)} \ln \frac{\widehat{f}_j(Z_{jt})}{\widehat{p}(X_t)\widehat{p}(X_{t-j})} \quad j = 1, \dots, n-1 \quad (4.4)$$

$\widehat{f}_j, \widehat{p}(\cdot)$ sont les estimateurs à noyaux de $f_j(\cdot)$ et $p(\cdot)$, $S_n(j) = \{t \in \mathbb{N} : j < t < n, f_{jt}(Z_{jt}) > 0, p_t(X_t) > 0, p_{t-j}(X_{t-j}) > 0\}$. Ils examinent la performance de $\widehat{\gamma}_n^2(j)$ pour des échantillons finis dans l'identification des retards importants pour de nombreux modèles de séries chronologiques non linéaires. Granger et Lin (1994), ont souligné, cependant, qu'aucune théorie pour la distribution limite n'existe pour $I_{n,3}(j)$ ou $\widehat{\gamma}_n^2(j)$. Ceci n'a pas permis l'application de leur mesure. En fait, Robinson (1991), a d'abord observé et également expliqué pourquoi aucune standardisation de $I_{n,3}(j)$ n'a une distribution limite connue sous H_0 . En particulier pas de standardisation de $I_{n,3}(j)$ n'a une distribution normale de moyenne zéro sous H_0 . Zheng (2000), dans le contexte i.i.d., observe également la difficulté d'obtenir la distribution limite sous H_0 de la statistique de test basée sur l'entropie de White (1982), pour la densité conditionnelle paramétrique. A fin d'éviter cette difficulté, Zheng utilise plutôt une mesure de divergence qui peut être considérée comme un terme de premier ordre modifié dans le développement de Taylor pour une statistique entropie. Cette mesure cependant, perd des propriétés importantes (exemple : l'invariance et l'interprétation théorique de l'information) de la mesure d'entropie.

Une difficulté énorme concernant la théorie de la distribution asymptotique pour $I_{n,3}(j)$ est que, sous H_0 , $I_{n,3}(j) \rightarrow 0$ à un taux égal à $n^{-\frac{1}{2}-\varepsilon_n}$, où $\varepsilon_n \geq c > 0$ dépend des paramètres lissés utilisés dans $\widehat{f}_j(\cdot)$ et $\widehat{p}(\cdot)$ respectivement, la normalisation par $n^{\frac{1}{2}}$ conduit à une statistique dégénérée car $I_{n,3}(j) \xrightarrow{p} 0$. Pour surmonter cette difficulté, Robinson (1991), propose un estimateur de l'entropie modifié de la forme suivante mais pour $j = 1$

$$I_{n,3,\gamma}(j) = \frac{1}{n-j} \sum_{t \in S_n(j)} C_t(\gamma) \ln \frac{\widehat{f}_{jt}(Z_{jt})}{\widehat{p}_t^2(X_t)} \quad j = 1, \dots, n-1 \quad (4.5)$$

où la fonction poids $C_t(\gamma) = 1 + j\gamma$ si $t = 1, \text{mod}(j+1)$ et $C_t(\gamma) = 1 - \gamma$ sinon, avec $\gamma \in]0, 1[$. Bien que l'estimateur de Robinson (1991), a une forme différente pour $j > 2$, (4.5) néanmoins définit un estimateur d'entropie utile. Robinson (1991), a souligné que l'utilisation de $C_t(\gamma)$ est une forme de fractionnement de l'échantillon; elle modifie le taux de convergence de $I_{n,3,\gamma}(j)$ sous H_0 , de sorte qu'une normalisation par $n^{1/2}$ donne une distribution limite normale non dégénérée. De plus, $I_{n,3,\gamma}(j)$ est encore consistante pour $I_3(j) \forall j > 0$ sous l'hypothèse d'ergodicité. Cette approche implique le choix du paramètre de tuning γ . De plus, cette approche ne fonctionne plus si X_t est uniformément distribuée sous H_0 , parce que dans ce cas $n^{1/2} I_{n,3,\gamma}(j) \xrightarrow{p} 0$ sous H_0 . Le test basé sur $I_{n,3,\gamma}(j)$ souffre d'une perte de puissance asymptotique locale (Hong and White, 2005).

4.1 Distribution asymptotique

Hong et White (2005) ont déterminé la distribution asymptotique de $I_{n,3}(j)$, en posant la condition suivante sur $\{X_t\}$.

Hypothèse 4.1.1 $\{X_t\}$ est strictement stationnaire de support sur $\mathbb{I} = [0, 1]$. Sa densité marginale $p : \mathbb{I} \rightarrow \mathbb{R}^+$ existe et vérifie $p(0) \geq p_0 > 0$ et elle est continûment différentiable jusqu'à l'ordre deux sur \mathbb{I} . De plus, $|p^{(2)}(x_1) - p^{(2)}(x_2)| \leq c|x_1 - x_2| \forall x_1, x_2 \in \mathbb{I}$ (i.e., $p^{(2)}$ est Lipchitzienne).

Tout au long, nous utilisons la convention suivante pour définir les dérivées aux bords de \mathbb{I}

$$\begin{aligned} p^{(d)}(0) &= \lim_{x \rightarrow 0^+} \frac{p^{(d-1)}(0+x) - p^{(d-1)}(0)}{x} & (d = 1, 2), \\ p^{(d)}(1) &= \lim_{x \rightarrow 0^-} \frac{p^{(d-1)}(1+x) - p^{(d-1)}(1)}{x} & (d = 1, 2). \end{aligned}$$

L'hypothèse 4.1.1, comme supposée dans Robinson (1991) et Hall (1988), évite l'épineux problème de traiter l'entropie dans les queues. Le support compact en premier

lieu apparaît restrictif, mais il peut toujours être assuré par une transformation continue strictement monotone telle que, la fonction logistique

$$X_t = \frac{1}{1 + e^{-Y_t}}, \quad (4.6)$$

où $\{Y_t\}$ est la série chronologique originale de support non borné. Aucune information n'est perdue dans (4.6) car $\{X_t\}$ est i.i.d. si et seulement si $\{Y_t\}$ est i.i.d., et $I_3(j)$ est invariante sous n'importe quelle transformation des données continue strictement monotone. De plus nous allons voir que la moyenne et la variance asymptotiques de $I_{n,3}(j)$ ont des distributions libres relativement à la loi sous-jacente du processus stochastique et donc sont invariantes sous n'importe quelle transformation continue strictement monotone. Ces caractéristiques rendent la mesure de l'entropie attrayante dans la pratique.

Pour voir que l'on peut souvent facilement faire en sorte que $p(x) \geq c < 0$ $\forall x \in \mathbb{I}$, soit Y_t de fonction de répartition $\tilde{\mathbb{P}}(\cdot)$ et de densité de probabilité $\tilde{p}(\cdot)$ et $F(\cdot)$ une fonction de répartition pré-spécifiée de densité $f(\cdot)$. Alors $X_t \equiv F(Y_t)$ est de support \mathbb{I} , et la fonction de répartition de X_t est donnée par

$$\mathbb{P}(x) = P(F(Y_t) \leq x) = \tilde{\mathbb{P}}(F^{-1}(x)) \quad x \in \mathbb{I}.$$

Donc

$$p(x) = \frac{d\mathbb{P}(x)}{dx} = \frac{\tilde{p}(F^{-1}(x))}{f(F^{-1}(x))} \quad x \in \mathbb{I}$$

Pour assurer que $\min_{x \in \mathbb{I}} p(x) \geq c$ il suffit que

$$f(y) \leq c^{-1} \tilde{p}(y) \quad -\infty < y < \infty.$$

Remarque 4.1.1 *Hong et White (2005), montrent qu'on pourrait permettre $p(x) \rightarrow 0$ aux bords avec une vitesse suffisamment lente et que leur théorie contiendrait à tenir mais sous des conditions fortes sur la fenêtre $h \equiv h_n$ utilisée dans*

l'estimation à noyau des densités.

Remarque 4.1.2 *Une alternative pour éviter l'hypothèse 4.1.1 est d'utiliser une fonction poids de support compact. Ceci permet à $\{X_t\}$ d'avoir un support non borné et/ou à p de s'annuler. Mais parce que l'information dans les queues est perdue, elle n'est pas une mesure consistante pour le test de la dépendance sérielle. Sinon nous pourrions considérer une suite de fonctions poids avec l'accroissement des supports compacts quand $n \rightarrow \infty$ comme discuté par Robinson (1991). Ceci donne une mesure consistante de la dépendance sérielle.*

4.2 Effets de bords et noyau de Jackknife

Nous estimerons les densités de probabilités via un noyau standard (i.e., noyau de second ordre).

Hypothèse 4.2.1 Soit $k : [-1, 1] \rightarrow \mathbb{R}^+$, une fonction de densité de probabilité symétrique et bornée.

L'hypothèse 4.2.1 implique que $\int_{-1}^1 k(u) du = 1$, $\int_{-1}^1 uk(u) du = 0$ et $\int_{-1}^1 u^2k(u) du < \infty$. Ceci et l'hypothèse 4.1.1 assurent que les biais des estimateurs à noyau des densités de probabilité sont de l'ordre $O(h^2)$ dans la région $]h, 1 - h[$. X_t est de support borné, les estimateurs à noyau des densités sont sujets à des effets de bords aux extrémités de \mathbb{I} . De tels effets de bords surviennent parce qu'il n'y a pas une couverture symétrique des données pour $k(\cdot)$ dans les régions limites $[0, h]$ et $[1 - h, 1]$ respectivement. Par conséquent, les estimateurs à noyau des densités dans les régions limites ne sont pas asymptotiquement sans biais quand $h \rightarrow 0$ (voir Härdle 1990, pp 130 – 133). Nous constatons que ces régions limites sont de longueur h et donc disparaissent quand $n \rightarrow \infty$. Pour $I_{n,3}(j)$, l'effet cumulatif des estimateurs à noyau des densités dans ces régions nulles peut influencer sur le comportement des estimateurs à noyau des densités dans la région $]h, 1 - h[$ en terme de l'erreur quadratique moyenne (EQM). En particulier, l'effet du biais cumulatif des estimateurs des

densités par le noyau dans les régions limites est $O_p(h)$ plutôt que l'usuel $O_p(h^2)$. Ceci ralentit la convergence de $I_{n,3}(j)$ vers $I_3(j)$ et complique l'analyse. Pour éviter cela, Hong et White (2005), proposent d'utiliser le noyau $k(\cdot)$ seulement à l'intérieur de la région $]h, 1-h[$; pour les régions limites $[0, h]$ et $[1-h, 1]$, ils utilisent le noyau de Jackknife

$$k_b(u) = (1+r) \frac{k(u)}{w_k(0,b)} - \frac{r}{\alpha} \frac{k(u/\alpha)}{w_k(0,b/\alpha)}, \quad (4.7)$$

où $w_k(l,b) = \int_{-b}^1 u^l k(u) du$ pour $l = 0, 1$, $r \equiv r(b)$ et $\alpha = \alpha(b)$ dépend de $b \in \mathbb{I}$, un indice dont la valeur dépend de la région où la densité est estimée, et qui ne va pas être intégré dans le calcul de l'estimateur de l'entropie et

$$r = \frac{w_k(1,b)/w_k(0,b)}{\alpha w_k(1,b/\alpha)/w_k(0,b/\alpha) - w_k(1,b)/w_k(0,b)}.$$

Comme cela a été proposé par Rice (1984), Hong et White (2005) ont choisi de prendre $\alpha = 2 - b$ tel que $b \in \mathbb{I}$, le support de k_b est $[-\alpha, \alpha]$ au lieu de $[-1, 1]$. Par conséquent, pour tout $b \in \mathbb{I}$,

$$\begin{aligned} \int_{-\alpha}^b k_b(u) du &= \int_{-b}^{\alpha} k_b(u) du = 1 & \int_{-\alpha}^b u k_b(u) du &= \int_{-b}^{\alpha} u k_b(u) du = 0 \\ \int_{-\alpha}^b u^2 k_b(u) du &= \int_{-b}^{\alpha} u^2 k_b(u) du > 0 & \int_{-\alpha}^b k_b^2(u) du &= \int_{-b}^{\alpha} k_b^2(u) du > 0 \end{aligned}$$

L'utilisation de k_b assure que le biais asymptotique des estimateurs à noyau des densités dans les régions limites est du même ordre que dans la région intérieure. Ainsi, l'effet cumulatif des estimateurs à noyau densités dans les régions limites $[0, h]$ et $[1-h, 1]$ est au plus du même ordre que les estimateurs à noyau des densités dans la région intérieure $]h, 1-h[$ en terme de l'*EQM*. Les effets de bords ne sont pas spécifiques aux estimateurs à noyau des densités. Ils surviennent quand le processus $\{X_t\}$ a un support borné et ils ne dépend pas du fait que $k(\cdot)$ a un support borné, (Hong and White, 2005).

Définissons les estimateurs à noyau leave-one out des densités marginales et bi-variée

$$\widehat{p}_t(X_t) = \frac{1}{n-1} \sum_{s=1}^n K_h(X_t, X_s) \mathbf{1}_{[s \neq t]} \quad (4.8)$$

$$\widehat{f}_{jt}(Z_{jt}) = \frac{1}{n_j-1} \sum_{s=j+1}^n K_h^{(2)}(Z_{jt}, Z_{js}) \mathbf{1}_{[s \neq t]} \quad (4.9)$$

où (et pour tout) $n_j = n - j$, $K_h^{(2)}(Z_{jt}, Z_{js}) = K_h(X_t, X_s) K_h(X_{t-j}, X_{s-j})$ et

$$K_h(x, y) = \begin{cases} h^{-1} k_{(x/h)}\left(\frac{x-y}{h}\right) & \text{si } x \in [0, h], \\ h^{-1} k\left(\frac{x-y}{h}\right) & \text{si } x \in]h, 1-h[, \\ h^{-1} k_{([1-x]/h)}\left(\frac{x-y}{h}\right) & \text{si } x \in [1-h, 1]. \end{cases} \quad (4.10)$$

$K_h(x, y) \neq K_h(y, x)$ malgré que $k(\cdot)$ est symétrique. Nous pouvons utiliser différentes fenêtres pour $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$, Hong et White (2005) utilisent la même fenêtre h , ce qui rend les biais de $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$ nuls sous H_0 , en menant à des conditions faibles sur h et une convergence rapide de $I_{n,3}(j)$. Si les estimateurs à noyau leave-one-out ne sont pas utilisés la moyenne asymptotique de $I_{n,3}(j)$ doit être modifiée, (Granger et Lin, 1994).

4.3 Développement asymptotique des estimateurs d'entropie

Pour obtenir la distribution limite de $I_{n,3}(j)$ Hong et white (2005) développent $I_{n,3}(j)$ en série de Taylor d'ordre deux et montrent que les deux premiers termes déterminent conjointement la distribution limite de $I_{n,3}(j)$. L'astuce consiste à retirer une moyenne non nulle de $I_{n,3}(j)$ et exploiter les conséquences de l'annulation des biais de $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$. $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$ influent sur la distribution limite de $I_{n,3}(j)$, même si elles ont des taux de convergence différents. Compte tenu de la difficulté bien connue de l'obtention de la distribution limite de $I_{n,3}(j)$, Hong et White (2005)

fournissent des heuristiques pour mieux comprendre leur approche.

Posons

$$\begin{aligned}
 I_{n,3}(j) &= \frac{1}{n_j} \sum_{t \in S_n(j)} \left\{ \ln \left(\frac{f_j(Z_{jt})}{p(X_t)p(X_{t-j})} \right) + \ln \left(\frac{\widehat{f}_{jt}(Z_{jt})}{f_j(Z_{jt})} \right) \right. \\
 &\quad \left. - \ln \left(\frac{\widehat{p}_t(X_t)}{p(X_t)} \right) - \ln \left(\frac{\widehat{p}_{t-j}(X_{t-j})}{p(X_{t-j})} \right) \right\} \\
 &\equiv \mathcal{I}_{jn}(f_j, p \circ p) + \mathcal{I}_{jn}(\widehat{f}_j, f_j) - \mathcal{I}_{1jn}(\widehat{p}, p) - \mathcal{I}_{3jn}(\widehat{p}, p)
 \end{aligned} \tag{4.11}$$

Pour le premier terme dans (4.11), nous avons $\mathcal{I}_{jn}(f_j, g \circ g) = 0$ presque sûrement sous H_0 . Pour le deuxième terme, en utilisant l'inégalité $|\ln(1+u) - u + \frac{1}{2}u^2| \leq u^3$ pour $|u| < 1$, nous obtenons

$$\begin{aligned}
 \mathcal{I}_{jn}(\widehat{f}_j, f_j) &= \frac{1}{n_j} \sum_{t \in S_n(j)} \ln \left(\frac{\widehat{f}_{jt}(Z_{jt}) + f_j(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right) \\
 &= \frac{1}{n_j} \sum_{t \in S_n(j)} \ln \left(1 + \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right) \\
 &= \frac{1}{n_j} \sum_{t \in S_n(j)} \left\{ \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} - \frac{1}{2} \left[\frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right]^2 \right. \\
 &\quad \left. + o \left(\left[\frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right]^2 \right) \right\} \\
 &= \widehat{W}_1(j) - \frac{1}{2} \widehat{W}_2(j) + \text{reste}
 \end{aligned} \tag{4.12}$$

sous H_0 , avec $\widehat{W}_1(j) = n_j^{-1} \sum_{t=j+1}^n \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})}$ et $\widehat{W}_2(j) = n_j^{-1} \sum_{t=j+1}^n \left\{ \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right\}^2$. $\widehat{W}_1(j)$ est le terme d'ordre un de notre développement de Taylor. Il peut être approximé comme suit

$$\widehat{W}_1(j) = \frac{1}{2} \widehat{H}_{1n}(j) + \text{reste} \tag{4.13}$$

où $\widehat{H}_{1n}(j)$ est la U -statistique d'ordre deux, provenant de l'interaction entre la variation d'échantillonnage de l'estimateur \widehat{f}_j et l'opération de calcul de la moyenne sur l'échantillon dans $\widehat{W}_1(j)$.

En effet, posons $\bar{f}_j(z_1) = \int_{\mathbb{I}^2} K_h^{(2)}(z_1, z_2) f_j(z_2) dz_2$, alors

$$\begin{aligned}
 \widehat{W}_1(j) &= n_j^{-1} \sum_{t=j+1}^n \frac{\widehat{f}_{jt}(Z_{jt}) - \bar{f}_j(Z_{jt})}{f_j(Z_{jt})} + \frac{\bar{f}_j(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \\
 &= n_j^{-1} (n_j - 1)^{-1} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \left[K_h^{(2)}(Z_{jt}, Z_{js}) - \sum_{s=j+1}^{t-1} K_h^{(2)}(Z_{jt}, Z_{js}) f_j(Z_{jt}) \right] \\
 &= 1/2 \binom{n_j}{2} \sum_{t=j+2}^n \sum_{s=j+1}^{t-1} \left[\left\{ \left[K_h^{(2)}(Z_{jt}, Z_{js}) - \sum_{s=j+1}^n K_h^{(2)}(Z_{jt}, Z_{js}) \right] \right. \right. \\
 &\quad \left. \left. + \sum_{s=j+1}^n K_h^{(2)}(Z_{jt}, Z_{js}) \right\} - \sum_{s=j+1}^{t-1} \left\{ K_h^{(2)}(Z_{jt}, Z_{js}) - \sum_{s=j+1}^n K_h^{(2)}(Z_{jt}, Z_{js}) \right. \right. \\
 &\quad \left. \left. + \sum_{s=j+1}^n K_h^{(2)}(Z_{jt}, Z_{js}) \right\} f_j(Z_{jt}) \right] / f_j(Z_{jt}) \\
 &\quad + 1/2 \binom{n_j}{2} \sum_{s=j+2}^n \sum_{t=j+1}^{s-1} \left[\left\{ \left[K_h^{(2)}(Z_{js}, Z_{jt}) - \sum_{t=j+1}^n K_h^{(2)}(Z_{js}, Z_{jt}) \right] \right. \right. \\
 &\quad \left. \left. + \sum_{t=j+1}^n K_h^{(2)}(Z_{js}, Z_{jt}) \right\} - \sum_{t=j+1}^{s-1} \left\{ K_h^{(2)}(Z_{js}, Z_{jt}) - \sum_{t=j+1}^n K_h^{(2)}(Z_{js}, Z_{jt}) \right. \right. \\
 &\quad \left. \left. + \sum_{t=j+1}^n K_h^{(2)}(Z_{js}, Z_{jt}) \right\} f_j(Z_{jt}) \right] / f_j(Z_{js})
 \end{aligned}$$

posons

$$\widetilde{K}_h^{(2)}(z_1, z_2) = K_h^{(2)}(z_1, z_2) - \int_{\mathbb{I}^2} K_h^{(2)}(z, z_2) dz$$

$$\widetilde{A}_{jn}(z_1, z_2) = \left[\widetilde{K}_h^{(2)}(z_1, z_2) - \int_{\mathbb{I}^2} \widetilde{K}_h^{(2)}(z_1, z_2) f_j(z) dz \right] / f_j(z_1)$$

$$B_{jn}(z_1) = \left[\int_{\mathbb{I}^2} K_h^{(2)}(z_1, z) f_j(z) dz - f_j(z_1) \right] / f_j(z_1)$$

$$H_{1jn}(z_1, z_2) = \widetilde{A}_{jn}(z_1, z_2) + \widetilde{A}_{jn}(z_2, z_1)$$

$$\gamma_{jn}(z_1, z_2) = \int_{\mathbb{I}^2} \left[K_h^{(2)}(z_1, z_2) - \int_{\mathbb{I}^2} K_h^{(2)}(z, z') f_j(z') dz' \right] dz / f_j(z_1)$$

en utilisant ces définitions, nous trouvons

$$\begin{aligned}
\widehat{W}_1(j) &= \frac{1}{2} \binom{n_j}{2} \sum_{t=j+2s=j+1}^n \sum_{t-1}^{t-1} \left[\widetilde{A}_{jn}(Z_{jt}, Z_{js}) + \widetilde{A}_{jn}(Z_{js}, Z_{jt}) \right] \\
&\quad + \frac{1}{2} \binom{n_j}{2} \sum_{t=j+2s=j+1}^n \sum_{t-1}^{t-1} \left[\gamma_{jn}(Z_{jt}, Z_{js}) + \gamma_{jn}(Z_{js}, Z_{jt}) \right] \\
&\quad + n_j^{-1} \sum_{t=j+2}^n B_{jn}(Z_{jt}) \\
&= \frac{1}{2} \widehat{H}_{1n}(j) + \frac{1}{2} \widehat{\Gamma}_n(j) + \widehat{B}_n(j)
\end{aligned}$$

Nous avons aussi

Lemme 4.3.1 (Hong et White, (2005)) *Supposons les hypothèses 4.1.1 et 4.2.1, $h \rightarrow 0$ et $j = o(n)$. Alors, $\widehat{\Gamma}_n(j) = 0$ pour tout n grand sous H_0 , avec une probabilité égale à 1.*

Le terme $\widehat{W}_2(j)$ est le second terme du développement de Taylor; peut être approximé par son analogue intégré

$$\begin{aligned}
\widehat{W}_2(j) &= n_j^{-1} \sum_{t=j+1}^n \int \left\{ \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right\}^2 f_j(z) dz + \text{reste} \\
&= n_j^{-1} \sum_{t=j+1}^n \int \mathbb{E} \left\{ \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right\}^2 f_j(z) dz \\
&\quad + n_j^{-1} \sum_{t=j+1}^n \int \left[\left\{ \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right\}^2 \right. \\
&\quad \left. - \mathbb{E} \left\{ \frac{\widehat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right\}^2 \right] f_j(z) dz + \text{reste} \\
&= L_n(j) + \widehat{H}_{2n}(j) + \text{reste.} \tag{4.14}
\end{aligned}$$

où $L_n(j)$ est *EQM* pondérée intégrée de \widehat{f}_{jt} , et $\widehat{H}_{2n}(j)$ est le second terme de la U -statistique.

Rassemblons les résultats de (4.12) à (4.14), ainsi posons $\widehat{H}_n(j) = \widehat{H}_{1n}(j) - \widehat{H}_{2n}(j)$, nous aurons

$$2\mathcal{I}_{jn}(\widehat{f}_j, f_j) = -L_n(j) + \widehat{H}_n(j) + \text{reste} \quad (4.15)$$

$\widehat{H}_{1n}(j)$ et $\widehat{H}_{2n}(j)$ sont de même ordre et conjointement déterminent la distribution limite de $\mathcal{I}_{jn}(\widehat{f}_j, f_j)$. Sous H_0 , $\widehat{H}_n(j) = O_p(n_j^{-1}h^{-1})$ est dominé par $L_n(j) = O(n_j^{-1}h^{-2} + h^4)$. Si $nh^4 \rightarrow \infty$, $nh^8 \rightarrow 0$ alors $n_j^{1/2}\mathcal{I}_{jn}(\widehat{f}_j, f_j) \xrightarrow{p} 0$. Quand $h \propto n_j^{-1/6}$, $\mathcal{I}_{jn}(\widehat{f}_j, f_j) = O_p(n_j^{-2/3})$ atteint son meilleur taux de convergence.

Par le même raisonnement nous pouvons obtenir

$$2\mathcal{I}_{ijn}(\widehat{p}, p) = -l_n(j) + \widehat{V}_{in}(j) + \text{reste} \quad i = 1, 2 \quad (4.16)$$

Sous H_0 , où $l_n(j)$ est *EQM* pondérée intégrée de $\widehat{p}(\cdot)$, et $\widehat{V}_{1n}(j)$ et $\widehat{V}_{2n}(j)$ sont les U -statistiques de second ordre. Celles-ci, sont de l'ordre $O_p(n_j^{-1}h^{-1/2})$ sous H_0 et sont dominées par $l_n(j)$. Si $n_j h^2 \rightarrow \infty$, $nh^8 \rightarrow 0$ alors $n_j^{1/2}\mathcal{I}_{ijn}(\widehat{p}, p) \xrightarrow{p} 0$. Quand $h \propto n_j^{-1/5}$, $\mathcal{I}_{ijn}(\widehat{p}, p) = O_p(n_j^{-4/5})$ atteint son taux de convergence optimal. Ceci est plus rapide que le taux optimal $O_p(n_j^{-2/3})$ de l'estimateur de l'entropie bivarié $\mathcal{I}_{jn}(\widehat{f}_j, f_j)$.

Maintenant, de (4.11), (4.15) et (4.16) nous obtenons

$$2I_{n,3}(j) = n_j^{-1}d_n^0 + \widehat{H}_n(j) + \text{reste} \quad (4.17)$$

Sous H_0 , où le facteur non stochastique $d_n^0 = (A_n^0 - 1)^2$ et

$$A_n^0 = (h^{-1} - 2) \int_{-1}^1 k^2(u) du + 2 \int_0^1 \int_{-1}^b k_b^2(u) dudb. \quad (4.18)$$

Le fait d'étudier la même fenêtre h , le carré du biais des termes de $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$ s'annulent à un ordre supérieur dans $I_{3,n}(j)$. Comme la suite d_n^0 dépend seulement

des composantes de la variance de $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$ et pas de leurs biais, le sous lissage n'est pas nécessaire pour enlever les biais et la fenêtre optimale pour $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$ est autorisée. Le plus important est que l'annulation des biais conduit à un taux de convergence rapide de $I_{n,3}(j)$ que pour $\mathcal{I}_{jn}(\widehat{f}_j, f_j)$ afin d'avoir $\mathcal{I}_n(j) = o_p(n_j^{-2/3})$. Ceci n'est vérifié que sous H_0 et une classe d'alternatives locales. D'autre part, la correction du terme apparent dans la composante variance A_n^0 dans (4.17) est dû à l'utilisation du noyau de Jackknife $k_b(\cdot)$ dans les régions limites $[0, h]$ et $[1 - h, 1]$. Cette correction n'est pas asymptotiquement négligeable et elle agit sur la moyenne asymptotique de $I_{n,3}(j)$.

4.4 Normalité asymptotique des estimateurs de l'entropie

Bien que le facteur non stochastique $n_j^{-1}d_n^0$ domine la U -statistique $\widehat{H}_n(j)$ dans l'ordre de grandeur, il agit sur la moyenne asymptotique de $I_{n,3}(j)$. La distribution limite de $I_{n,3}(j)$ centrée est déterminée par $\widehat{H}_n(j)$.

Théorème 4.4.1 (Hong et White, 2005) *Supposons les hypothèses 4.1.1, 4.2.1 et H_0 , $\frac{nh^4}{\ln n} \rightarrow \infty$ et $nh^7 \rightarrow 0$ quand $n \rightarrow \infty$, alors*

a) $2hn_j I_{n,3}(j) + hd_n^0 \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ pour tout retard $j = o(n)$ où

$$\sigma^2 = 2 \int_{-1}^1 \int_{-1}^1 \left[2k(u)k(u') - \int_{-1}^1 k(u+v)k(v)dv \int_{-1}^1 k(u'+v')k(v')dv' \right]^2 dudv'$$

b) Posons $I_{n,3} = [2hn_1 I_{3,n}(1) + hd_n^0 + \dots + 2hn_\nu I_{3,n}(\nu) + hd_n^0]$ où $\nu \in N^+$ fixé mais peut être arbitrairement grand, alors $I_{n,3} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 I_\nu)$, où I_ν est la matrice identité d'ordre ν .

Ainsi, par l'ajustement de la moyenne hd_n^0 et par la multiplication de $I_{n,3}(j)$ par $n_j h$ qui tend plus rapidement vers l'infini que $n_j^{1/2}$, nous obtenons la distribution limite de $I_{n,3}(j)$. Hong et White (2005) permettent et n'exigent pas que l'ordre $j \rightarrow \infty$ quand $n \rightarrow \infty$ pour $I_{n,3}(j)$ et la condition sur j est faible. De plus, la

condition sur h n'est pas restrictive. Les fenêtres optimales ($h \propto n_j^{-1/6}$ et $n_j^{-1/5}$ respectivement pour $\widehat{f}_{jt}(\cdot)$ et $\widehat{p}_t(\cdot)$ sont permises). Bien entendu, ces fenêtres optimales pour les estimateurs des densités ne sont pas les mêmes pour avoir la meilleure convergence de l'estimateur de l'entropie $I_{n,3}(j)$.

Le théorème 4.4.1(b) montre que sous H_0 la distribution finie dimensionnelle de $\{2hn_j I_{n,3}(j), j \in N^+\}$ est asymptotiquement une loi normale multivariée pour tout ensemble des retards distincts, et $Cov(hn_i I_{n,3}, hn_j I_{n,3}) \rightarrow 0$ pour $i \neq j$. Ceci entraîne un biais pour la construction d'un test portemanteau pour H_0 (voir section 4.5).

La particularité importante de $2hn_j I_{n,3}(j)$ est que sa moyenne asymptotique hd_n^0 et sa variance σ^2 sont des distributions libres et donc invariantes sous n'importe quelle transformation continue monotone. Aucune estimation n'est exigée, parce qu'elles sont connues en donnant $k(\cdot)$ et h . Ceci rend $I_{n,3}(j)$ plutôt attractive dans la pratique, il simplifie beaucoup le calcul des intervalles de confiance des estimateurs. De plus, pour calculer les p -values du bootstrap lissé proposé dans la section 4.5, nous n'avons besoin que de comparer la statistique non standardisée $n_j I_{n,3}$ avec son bootstrap équivalent à $n_j I_{n,3}^*(j)$. Numériquement les mêmes p -values de bootstrap sont obtenues si nous standardisons $n_j I_{n,3}(j)$ et $n_j I_{n,3}^*(j)$, et le caractère attrayant asymptotiquement du test est maintenu.

Le théorème 4.4.1 peut être utilisé pour obtenir la distribution limite de la mesure d'entropie normalisée $\widehat{\gamma}_n^2(j)$ de (4.3) de Granger et Lin (1994), qui est une autocorrelation ombre (shadow autocorrelation), et est utile pour identifier un retard important de la dépendance sérielle dans $\{X_t\}$.

Théorème 4.4.2 (Hong et White, 2005) *Supposons les conditions du théorème 4.4.1 et H_0 , alors*

$$a) \quad hn_j \widehat{\gamma}_n^2(j) + hd_n^0 \xrightarrow{\mathcal{L}} N(0, \sigma^2) \text{ pour tout retard } j = o(n)$$

$$b) \quad \text{Posons } \widehat{\gamma}_n^2 = [2hn_1 \widehat{\gamma}_n^2(1) + hd_n^0 + \dots + 2hn_\nu \widehat{\gamma}_n^2(\nu) + hd_n^0] \text{ où } \nu \in \mathbb{N}^+ \text{ fixé mais peut être grand, alors } \widehat{\gamma}_n^2 \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 I_\nu), \text{ où } I_\nu \text{ est la matrice identité d'ordre } \nu.$$

Diverses tentatives ont été faites dans la littérature pour développer une mesure de dépendance générale appropriée dans le contexte des séries chronologiques non linéaires, mais peu ont été aussi simples et informatives que la fonction d'auto-corrélation empirique. Le théorème 4.4.2 montre que la fonction d'auto-corrélation ombre estimée $\widehat{\gamma}_n^2(j)$; $j = 1, 2, \dots, \nu$ centrée et réduite est asymptotiquement $\mathcal{N}(0, 1)$ et les $\widehat{\gamma}_n^2(j)$ sont asymptotiquement indépendantes à travers les différents retards. Ces propriétés sont similaires à ceux de la fonction d'auto-corrélation empirique. Ainsi $\widehat{\gamma}_n^2(j)$ peut jouer un rôle important dans l'analyse des séries chronologiques non linéaires analogue à celui de l'auto-corrélation empirique dans l'analyse des séries chronologiques linéaires. Notons que la moyenne et la variance asymptotiques de $\widehat{\gamma}_n^2(j)$ sous H_0 sont $-n_j^{-1}d_n^0 \propto n_j^{-1}h^2$ et $(n_j h)^{-1} \sigma^2$. Ces taux diffèrent de ceux obtenus dans Granger et Lin (1994), à travers les simulations. Bien que la mesure $\gamma^2(j)$ dans (4.2) est positive, $\widehat{\gamma}_n^2(j)$ peut être négative sous H_0 .

Si comme dans Granger et Lin (1994), les estimateurs noyaux leave-one-out ne sont pas employés, les théorèmes 4.4.1 et 4.4.2 sont encore valables sauf que le facteur de décentrage d_n^0 doit être remplacé par

$$\widehat{d}_n^0 = d_n^0 - 2 \left[\{h^{-1}k(0) - 1\}^2 - 1 \right]$$

où le second terme est apporté par le terme du premier ordre $\widehat{W}_1(j)$ dans (4.12).

4.5 Tests d'hypothèses et le bootstrap lissé

4.5.1 Test d'hypothèse

Dans le contexte d'un i.i.d., l'entropie a été utilisée pour tester la normalité (Vasicek, 1976), la distribution uniforme (Dudewicz and van der Meulen, 1981) et la qualité de l'ajustement (Gokhale, 1983). Ces tests n'ont pas de distributions limites connues, mais les simulations montrent qu'ils ont une puissance importante. Ici, la capacité de $I_3(j)$ de capturer tous les écarts entre $f_j(\cdot)$ et $p(\cdot)p(\cdot)$, la rende

importante pour tester H_0 . Robinson (1991) a proposé un test asymptotique pour tester H_0 en utilisant $I_{n,3}(j)$ de (4.4).

Considérons un test pour H_0 qui est basé sur un retard d'ordre j

$$\bar{I}_{n,3}(j) = \sigma^{-1} [2hn_j I_{n,3}(j) + hd_n^0] \quad j \ll n \quad (4.19)$$

Du théorème 4.4.1, $\bar{I}_{n,3}(j) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ quand $n \rightarrow \infty$ sous H_0 . Nous obtenons ainsi un test asymptotique $\mathcal{N}(0, 1)$ pour H_0 en utilisant $I_{n,3}(j)$. Ni le partitionnement de l'échantillon ni le paramètre de tuning ne sont introduits, et $I_{n,3}(j)$ fonctionne même si X_t est i.i.d. uniforme. Hong et White (2005), ont montré que ce test est localement asymptotiquement plus puissant que le test de Robinson (1991) (voir la section 4.5.2).

Le test de Robinson, $\bar{I}_{n,3}(j)$ est un test qui se base sur un échantillon de grande taille et son niveau pour un échantillon fini est différent du niveau asymptotique. La qualité de l'approximation asymptotique dépend du terme d'ordre supérieur dans le développement de Taylor de $I_{n,3}(j)$ et le choix de la fenêtre h . Cette analyse suggère que la théorie asymptotique peut ne pas bien marcher pour des échantillon relativement grand, car les termes d'ordre élevé asymptotiquement négligeables dans $I_{n,3}(j)$ sont proches, selon l'ordre de grandeur, de la U -statistique dominante $\hat{H}_n(j)$ dans (4.16), qui détermine la distribution limite de $\bar{I}_{n,3}(j)$. Le même problème a été rencontré par Skaug et Tjøstheim (1993b, 1996), pour leurs tests. Heureusement parce que $\{X_t\}$ est i.i.d. sous H_0 , le bootstrap est bien approprié et fournit un moyen simple pour obtenir des valeurs critiques pour $\bar{I}_{n,3}(j)$. De façon intéressante, le bootstrap naïf (i.e., ré-échantillonner en remplaçant l'échantillon original $\mathcal{X} = \{X_t\}_{t=\overline{1,n}}$) ne donne pas une procédure consistante pour le test $\bar{I}_{n,3}(j)$, parce qu'il ne préserve pas les propriétés de la U -statistique dégénérée $\hat{H}_n(j)$ dans (4.16). Hong et White (2005), proposent plutôt la procédure de bootstrap lissé décrite comme suit :

1. Simuler l'échantillon de bootstrap $\mathcal{X}^* = \{X_t^*\}_{t=1, \dots, n}$ à partir de la densité estimée

$$\widehat{p}(x) = n^{-1} \sum K_h(x, X_t) \quad x \in \mathbb{I} \quad (4.20)$$

où $k(\cdot)$ et h sont les mêmes que ceux utilisés dans $I_{n,3}(j)$;

2. Calculer les statistiques d'entropie de bootstrap $I_{n,3}^*(j)$ de la même manière que $I_{n,3}(j)$ en remplaçant \mathcal{X} par \mathcal{X}^* . Les mêmes $k(\cdot)$ et h sont utilisés dans $I_{n,3}(j)$, $I_{n,3}^*(j)$ et $\widehat{p}(\cdot)$;
3. Répéter les étapes (1) et (2) B fois pour obtenir B statistiques de test de bootstrap $\{I_{n,3l}^*(j)\}_{l=1, \dots, B}$;
4. Calculer la p -value de bootstrap $p^* = \frac{1}{B} \sum_{l=1}^B \mathbf{1} [I_{n,3l}^*(j) > I_{n,3}(j)]$.

Pour obtenir des p -values de bootstrap bien précises, B doit être suffisamment grand. Notons que nous comparons l'estimateur de l'entropie $I_{n,3}(j)$ avec son équivalent de bootstrap $I_{n,3l}^*(j)$. Là, nous n'avons pas besoin de calculer la moyenne et la variance asymptotiques d_n^0 et σ^2 respectivement. Les p -values de bootstrap obtenues sont numériquement identiques à celles obtenues en comparant $\bar{I}_{n,3}(j)$ avec $\bar{I}_{n,3}^*(j)$. Tout ceci résulte de la propriété que la distribution asymptotique est libre de d_n^0 et σ^2 et de l'utilisation de la même fenêtre h dans le calcul de $\bar{I}_{n,3}(j)$ et $\bar{I}_{n,3}^*(j)$. Pour montrer que la procédure de bootstrap lissé est consistante, Hong et White (2005), ajoutent une condition sur le noyau $k(\cdot)$.

Hypothèse 4.5.1 Soit $k : [-1, 1] \rightarrow \mathbb{R}^+$ de classe C^2 sur \mathbb{I} , avec $k^{(d)}(-1) = k^{(d)}(1) = 0$ pour $d = 0, 1$, et $|k^{(2)}(u_1) - k^{(2)}(u_2)| \leq c|u_1 - u_2| \forall u_1, u_2 \in [-1, 1]$.

Théorème 4.5.2 (Hong et White, 2005) *Supposons les hypothèses 4.1.1–4.5.1 et H_0 . $nh^5 = O(1)$, $nh^7 \ln^3 n \rightarrow 0$ quand $n \rightarrow \infty$ et $j = o(n)$. Soit $\bar{I}_{n,3}^*(j)$ définie comme $\bar{I}_{n,3}(j)$ en remplaçant l'échantillon original \mathcal{X} par l'échantillon bootstrappé \mathcal{X}^* défini précédemment en utilisant les mêmes $k(\cdot)$ et h dans $\bar{I}_{n,3}(j)$, $\bar{I}_{n,3}^*(j)$ et*

$\widehat{p}(\cdot)$, alors $\bar{I}_{n,3}^*(j) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ quand $n \rightarrow \infty$ conditionnellement à \mathcal{X} .

Le théorème 4.5.2 montre que le bootstrap lissé fournit une approximation asymptotique de $\bar{I}_{n,3}(j)$ à la loi $\mathcal{N}(0,1)$, sous H_0 . Notons que le théorème 4.5.2 implique que $\bar{I}_{n,3}^*(j) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ inconditionnellement. Cependant, il n'indique pas le degré d'amélioration du bootstrap lissé sur l'approximation asymptotique de premier ordre. Comme $\bar{I}_{n,3}(j)$ est asymptotiquement une pivotale, il est plausible que le bootstrap lissé peut atteindre des niveaux de confiance raisonnablement précis pour $\bar{I}_{n,3}(j)$. Hong et White (2005), suggèrent l'utilisation du même noyau $k(\cdot)$ et la même fenêtre h dans le calcul de $\bar{I}_{n,3}(j)$, $\bar{I}_{n,3}^*(j)$ et $\widehat{p}(\cdot)$. Ceci n'est pas nécessaire mais devrait donner une meilleure approximation pour un échantillon fini. Nous avons examiné la performance du bootstrap lissé par des simulations et trouvé que le bootstrap lissé fournit un niveau raisonnable pour ce test pour des échantillons finis.

Ensuite nous considérons comportement asymptotique de $\bar{I}_{n,3}(j)$ en vertu d'une alternative globale à H_0 .

Hypothèse 4.5.3 Soit $\{X_t\}$, un processus strictement stationnaire de coefficient de mélange $\alpha(j) \leq Cj^{-\tau}$, $\tau > 2$, pour chaque $j \in \mathbb{N}^+$, la densité de probabilité jointe $f_j(\cdot)$ de Z_{jt} de support \mathbb{I}^2 est bornée loin de zéro et est continûment différentiable d'ordre deux sur \mathbb{I}^2 .

Théorème 4.5.4 (Hong et White, 2005) *Supposons les hypothèses 4.1.1–4.5.3, $nh^5 = O(1)$, $nh^7 \ln^3 n \rightarrow 0$ et $j = o(n)$; alors $P(\bar{I}_{n,3}(j) > C_n(j)) \rightarrow 1$ quand $n \rightarrow \infty$ pour toute suite de constantes $\{C_n(j) = o(n_j h)\}$ et $P(\bar{I}_{n,3}(j) > \bar{I}_{n,3}^*(j)) \rightarrow 1$, à condition que $f_j(\cdot) \neq p(\cdot)p(\cdot)$*

Ainsi, le test basé sur $\bar{I}_{n,3}(j)$ est consistant contre toutes les alternatives pour lesquelles $f_j(\cdot) \neq p(\cdot)p(\cdot)$. Peu importe, si une valeur critique asymptotique ou bootstrap est utilisée. Le théorème 4.5.4 implique que $\bar{I}_{n,3}(j) \rightarrow +\infty$ avec une

probabilité proche de 1, sous toutes alternatives tel que $f_j(\cdot) \neq p(\cdot)p(\cdot)$. Par conséquent les valeurs critiques supérieures sont appropriées. La valeur critique de la loi normale au niveau 5%, par exemple est de 1.645. Robinson (1991) assure la cohérence de son test pour un ordre de retard fixé, sous une condition de l'ergodicité faible qui tient compte des processus à mémoire longue. En revanche, Hong et White imposent une condition de α -mélange qui implique l'ergodicité et exclut les processus à longue mémoire. La condition de mélange est pratique dans l'analyse des séries chronologiques non linéaires. Elle est utilisée ici parce qu'elle permet à l'ordre j de tendre vers l'infini quand $n \rightarrow \infty$.

Le test de $\bar{I}_{n,3}(j)$ est informatif, il donne toute l'information concernant l'ordre pour lequel il existe une dépendance sérielle significative. Cependant, pour tester H_0 , il est possible que deux ordres de retard donnent des résultats contradictoires. Il est donc souhaitable d'avoir un test portemanteau qui utilise des retards multiples. Pour cela, soit

$$\mathcal{Q}_n(\nu) = \frac{1}{\sqrt{\nu}} \sum_{j=1}^{\nu} \bar{I}_{n,3}(j) \quad \nu \in \mathbb{N}^+ \quad (4.21)$$

Pour plus de simplicité, Hong et White (2005) considèrent que l'ordre de troncature du retard $\nu \in \mathbb{N}^+$ soit fixé. Il est possible de permettre à $\nu \rightarrow \infty$ quand $n \rightarrow \infty$ avec une pondération différente pour chaque retard j .

Théorème 4.5.5 (Hong et White, 2005) *i) Supposons les hypothèses 4.1.1–4.5.3*

et H_0 . Soit $\mathcal{Q}_n^(\nu)$ définie comme $\mathcal{Q}_n(\nu)$ en remplaçant l'échantillon original \mathcal{X} par l'échantillon bootstrap \mathcal{X}^* , les mêmes $k(\cdot)$ et h sont utilisés dans $\mathcal{Q}_n(\nu)$, $\mathcal{Q}_n^*(\nu)$ et $\hat{p}(\cdot)$, alors pour tout $\nu \in \mathbb{N}^+$ fixé $\mathcal{Q}_n(\nu) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ et $\mathcal{Q}_n^*(\nu) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ quand $n \rightarrow \infty$ conditionnellement à \mathcal{X} .*

ii) Supposons les conditions du théorème 4.5.4, alors $P(\mathcal{Q}_n(\nu) > C_n(\nu)) \rightarrow 1$ pour toute suite de constantes $\{C_n = o(nh)\}$ et $P(\mathcal{Q}_n(\nu) > \mathcal{Q}_n^(\nu)) \rightarrow 1$ pourvu que $f_j(\cdot) \neq p(\cdot)p(\cdot)$ pour $j \in \{1, 2, \dots, \nu\}$.*

Intuitivement, parce que $\bar{I}_{n,3}(j) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ et $Cov[\bar{I}_{n,3}(i), \bar{I}_{n,3}(j)] \rightarrow 0$ pour

$i \neq j$, sous H_0 , $\{\bar{I}_{n,3}(j)\}_{j=1,\bar{\nu}}$ est une suite de *v.a* asymptotiquement i.i.d. $\mathcal{N}(0, 1)$. Par conséquent $\mathcal{Q}_n(\nu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ quand $n \rightarrow \infty$ sous H_0 par le théorème central limite. La puissance de $\mathcal{Q}_n(\nu)$ dépend du choix de ν . Cependant, la dépendance de la puissance de $\mathcal{Q}_n(\nu)$ de ν est moins sensible que la dépendance de la puissance de $\bar{I}_{n,3}(j)$ de l'ordre j (Hong & White, 2005). Nous pouvons voir $\mathcal{Q}_n(\nu)$ comme un test portemanteau de la dépendance sérielle qui peut capturer la dépendance linéaire et non linéaire dans $\{X_t\}$. Notons que comme le test $\bar{I}_{2,n}(j)$, nous pouvons simplement comparer $\sum_{j=1}^{\nu} n_j \bar{I}_{n,3}(j)$ avec $\sum_{j=1}^{\nu} n_j \bar{I}_{n,3}^*(j)$ dans le calcul des p -values de bootstrap de $\mathcal{Q}_n(\nu)$.

Autrement, il est possible de construire un test statistique asymptotique χ_{ν}^2 , $\mathcal{K}_n(\nu) = \sum_{j=1}^{\nu} \bar{I}_{n,3}^2(j)$. Ce test est asymptotiquement moins efficace que $\mathcal{Q}_n(\nu)$ car $\mathcal{K}_n(\nu)$ n'exploite pas la nature unilatérale (i.e., $\bar{I}_{n,3}(j) \rightarrow +\infty$) du test statistique $\bar{I}_{3,n}(j)$ sous l'alternative de H_0 .

4.5.2 Test pour l'hypothèse i.i.d. $\mathcal{U}[0, 1]$

Diebold et *al* (1998), ont proposé une méthode graphique pour évaluer les prévisions de densité qui sont utiles par exemple dans la gestion des risques financiers, choix du prix et le contrôle de la politique macroéconomique. Ils montrent que si des prévisions de densité de prévision pour l'horizon un spécifient correctement la structure dynamique du processus générateur de données, alors les séries de transformations intégrales de probabilité en ce qui concerne les densités prévision sont i.i.d. Ainsi, $\bar{I}_{n,3}(j)$ peut être utilisée pour évaluer si une suite de prévisions de densité spécifie correctement la structure dynamique. Diebold et *al* (1998), montrent de plus que, si une série de densité de prévisions pour l'horizon un coïncide avec la densité conditionnelle générant les données, alors la série de transformation intégrale de probabilité sachant les densités prévision est i.i.d. $\mathcal{U}[0, 1]$. Dans ce cas, les densités prévision sont optimales en ce qui concerne le critère de la minimisation de la perte moyenne. Ainsi, pour évaluer si une série de densité prévision coïncide avec

la véritable série de densité conditionnelles, il suffit d'évaluer si la série de transformation intégrale de probabilité est i.i.d. $\mathcal{U}[0, 1]$. Le test $\bar{I}_{n,3}(j)$ n'est pas adapté à cette hypothèse car il n'intègre pas les informations sur la distribution $\mathcal{U}[0, 1]$. Cependant, Hong et White (2005) proposent un test approprié pour cette hypothèse conjointe. Lorsque un retard d'ordre j est utilisé, soit la statistique de test

$$\bar{I}_{n,3}^{\mathcal{U}}(j) = \sigma^{-1} \left\{ 2h \sum_{t \in S_n(j)} \ln \hat{f}_j(Z_{jt}) + h \left[(A_n^0)^2 - 1 \right] \right\} \quad (4.22)$$

où A_n^0 est comme dans (4.18) et $S_n(j) = \{t \in \mathbb{N} : j < t \leq n, \hat{f}_j(Z_{jt}) > 0\}$. Notons que nous n'avons pas besoin d'estimer la densité marginale $p(\cdot)$ qui est inconnue sous les hypothèses i.i.d. et $\mathcal{U}[0, 1]$. A cause de ceci, la constante de centrage dans $\bar{I}_{n,3}^{\mathcal{U}}(j)$ diffère un petit peu de celle de $\bar{I}_{n,3}(j)$. Pour un retard multiple la statistique de test est donnée par

$$\mathcal{Q}_n^{\mathcal{U}}(\nu) = \frac{1}{\sqrt{\nu}} \sum_{j=1}^{\nu} \bar{I}_{n,3}^{\mathcal{U}}(j) \quad \nu \in \mathbb{N}^+ \quad (4.23)$$

Nous pouvons aussi envisager une procédure de bootstrap pour $\bar{I}_{n,3}^{\mathcal{U}}(j)$ et $\mathcal{Q}_n^{\mathcal{U}}(\nu)$. Ici, nous générons l'échantillon de bootstrap \mathcal{X}^* de la distribution $\mathcal{U}[0, 1]$. Ni le bootstrap naïf ni le bootstrap lissé dans (4.20) n'offrira une procédure cohérente pour $\bar{I}_{n,3}^{\mathcal{U}}(j)$ et $\mathcal{Q}_n^{\mathcal{U}}(\nu)$. Une procédure suffisante, beaucoup plus simple est spécifiée ci-dessous. Comme les tests $\bar{I}_{n,3}^{\mathcal{U}}(j)$ et $\mathcal{Q}_n^{\mathcal{U}}(\nu)$, nous pouvons simplement comparer $\sum_{t \in S_n(j)} \ln \hat{f}_j(Z_{jt})$ et $\sum_{j=1}^{\nu} \sum_{t \in S_n(j)} \ln \hat{f}_j(Z_{jt})$ avec leurs homologues respectivement dans le calcul des p -values de bootstrap de $\bar{I}_{n,3}^{\mathcal{U}}(j)$ et $\mathcal{Q}_n^{\mathcal{U}}(\nu)$.

Théorème 4.5.6 (Hong et White, 2005) *Supposons les hypothèses 4.2.1, 4.5.3 et l'hypothèse $H_0^{\mathcal{U}}$ que $\{X_t\}$ est i.i.d. $\mathcal{U}[0, 1]$ et $nh^5 = O(1)$, $h \rightarrow 0$. Soit $\bar{I}_{n,3}^{\mathcal{U}}(j)^*$ et $\mathcal{Q}_n^{\mathcal{U}}(\nu)^*$ définies comme $\bar{I}_{n,3}^{\mathcal{U}}(j)$ et $\mathcal{Q}_n^{\mathcal{U}}(\nu)$ respectivement avec l'échantillon de bootstrap \mathcal{X}^* , remplaçant l'échantillon \mathcal{X} , où \mathcal{X}^* est un échantillon i.i.d. $\mathcal{U}[0, 1]$. Les mêmes $k(\cdot)$ et h sont utilisés dans les tests statistiques. Alors*

- i) pour tout retard $j = o(n)$, $\bar{I}_{n,3}^{\mathcal{U}}(j) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ et $\bar{I}_{n,3}^{\mathcal{U}}(j)^* \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ conditionnellement à \mathcal{X} ,
- ii) pour tout $\nu \in \mathbb{N}^+$ fixé $\mathcal{Q}_n^U(\nu) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ et $\mathcal{Q}_n^U(\nu)^* \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ quand $n \rightarrow \infty$ conditionnellement à \mathcal{X} .

Le test basé sur $\bar{I}_{n,3}^{\mathcal{U}}(j)$ peut détecter tous les écarts de i.i.d. et de la distribution uniforme $\mathcal{U}[0,1]$ deux à deux. Comme l'ont souligné Diebold et al (1998), il n'y a pas de test statistique approprié pour $H_0^{\mathcal{U}}$ dans la littérature qui soit important pour évaluer si ces densités prévision sont optimales. $\bar{I}_{n,3}^{\mathcal{U}}(j)$ et $\mathcal{Q}_n^U(\nu)$ sont des tests adéquats pour telle hypothèse jointe. Notons qu'il n'y a pas un test de type Robinson (1991) correspondant à $H_0^{\mathcal{U}}$ car la division de l'échantillon ne marche pas dans ce cas.

4.6 Puissance locale asymptotique

Le test statistique de Robinson (1991), et le test statistique de Hong et White (2005) sont asymptotiquement $\mathcal{N}(0,1)$ sous H_0 , et les deux sont consistants contre tous les processus j -dépendants qui satisfont l'hypothèse 4.5.3. Pour examiner leurs qualités, Hong et White (2005), ont étudié leurs puissances asymptotiques locales. Comme Tjøstheim (1996), l'a souligné, l'analyse de la puissance asymptotique locale est plutôt difficile dans les tests non-paramétriques de l'indépendance sérielle. Pour simplifier, ils considèrent une classe de processus localement j -dépendants pour lesquels il existe une dépendance au retard j seulement, mais j peut tendre vers l'infini quand $n \rightarrow \infty$. Ainsi, ils supposent que la densité de probabilité jointe de Z_{jt} est donnée par

$$H_{jn}(a_n) : f_j(z) = p(x)p(y)[1 + a_n q_j(z) + r_{jn}(z)], \quad z \equiv (x, y)' \in \mathbb{I}^2 \quad (5.1)$$

où $q_j : \mathbb{I}^2 \rightarrow \mathbb{R}$ est une fonction caractérisant l'écart à H_0 , r_{jn} est un reste découlant du développement de Taylor de $f_j(\cdot) \equiv f_{jn}(\cdot)$ et la constante $a_n \rightarrow 0$, régit la vitesse

à laquelle l'alternative locale $H_{jn}(a_n)$ converge vers l'hypothèse nulle H_0 .

- Hypothèse 4.6.1**
- a) $1 + a_n q_j(z) + r_{jn}(z) \geq 0$ pour tout $z \in \mathbb{I}^2$, tout $n, j \in \mathbb{N}$,
 - b) $\int_{\mathbb{I}^2} q_j(z) p(x)p(y) dz = 0$ et $\int_{\mathbb{I}^2} r_{jn}(z) p(x)p(y) dz = 0$ pour tout $n, j \in \mathbb{N}$,
 - c) $p(\cdot)$ et $q_j(\cdot)$ sont continuellement différentiables d'ordre deux sur \mathbb{I} et \mathbb{I}^2 respectivement, $|p^{(2)}(x_1) - p^{(2)}(x_2)| \leq c|x_1 - x_2|$ pour tout $x_1, x_2 \in \mathbb{I}$ et certains $c \in]0, \infty[$,
 - d) $\int_{\mathbb{I}^2} |q_j(z)|^3 p(x)p(y) dz \leq C$ et $\int_{\mathbb{I}^2} |r_{jn}(z)|^3 p(x)p(y) dz = o(a_n^2)$ uniformément en $j \in \mathbb{N}$.

L'hypothèse 4.6.1 (a, b) assure que $f_j(\cdot)$ est une densité de probabilité bi-variée pour tout $n, j \in \mathbb{N}^+$. L'hypothèse 4.6.1 (d) assure que le rest r_{jn} n'a pas d'impact sur la distribution limite de ce test. Notons que la densité marginale $p_n(\cdot)$ de X_t peut dépendre de n sous $H_{jn}(a_n)$ et peut ne pas coïncider avec $p(\cdot)$, la densité marginale de X_t sous H_0 .

Deux exemples de $H_{1n}(a_n)$ sont : un processus $MA(1)$

$$X_t = a_n \varepsilon_{t-1} + \varepsilon_t \quad (4.24)$$

et un processus de type $ARCH(1)$

$$X_t = \varepsilon_t \sqrt{1 + a_n X_{t-1}^2} \quad (4.25)$$

où ε_t est un processus i.i.d. $(0, \sigma_{\varepsilon_t}^2)$ de densité marginale bornée inférieurement. Nous avons $q_1(z) = xy$ pour (4.24) et $q_1(z) = (x^2 - \sigma_{\varepsilon_t}^2)(y^2 - \sigma_{\varepsilon_t}^2)$ pour (4.25)

Hong et White (2005) ne considèrent que $\bar{I}_{n,3}(j)$. Les conclusions pour la puissance locale de $\mathcal{Q}_n(\nu)$ sont similaires.

Théorème 4.6.2 (Hong et White, 2005) *Supposons les hypothèses 4.2.1 et 4.6.1, $nh^4/\ln n \rightarrow \infty$, $nh^7 \rightarrow 0$ et $j = o(n)$ alors $\bar{I}_{n,3}(j) - \mu_j \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ sous $H_{jn}(n^{-\frac{1}{2}}h^{-\frac{1}{2}})$,*

où $\mu_j = \sigma^{-1} \int_{\mathbb{I}^2} q_j^2(z) p(x)p(y) dz$.

Ainsi, $\bar{I}_{n,3}(j)$ a une puissance non triviale contre $H_{jn} \left(n^{-\frac{1}{2}} h^{-\frac{1}{2}} \right)$ chaque fois que $q_j \neq 0$. Le taux $a_n = n^{-\frac{1}{2}} h^{-\frac{1}{2}}$ est plus petit que le taux paramétrique $n^{-\frac{1}{2}}$ quand $h \rightarrow 0$, mais il est plus rapide que $n^{-\frac{1}{4}}$ car $nh^3 \rightarrow \infty$. Par exemple quand $h \propto n^{-\frac{1}{6}}$ (le taux optimal pour $\hat{f}_{jt}(\cdot)$) est utilisé, nous avons $n^{-\frac{1}{2}} h^{-\frac{1}{2}} \propto n^{-\frac{5}{12}}$ qui est légèrement plus lent que $n^{-\frac{1}{2}}$ mais plus rapide que $n^{-\frac{1}{4}}$. En fait les taux admissibles pour a_n pourraient encore être améliorés vers le taux paramétrique $n^{-\frac{1}{2}}$ en assouplissant la condition sur la fenêtre h , qui pourrait être atteint par l'utilisation des noyaux d'ordre supérieur. Le taux a_n pourrait ainsi être arbitrairement proche de $n^{-\frac{1}{2}}$ mais il est toujours plus petit que $n^{-\frac{1}{2}}$ en raison du lissage. Dans la pratique, nous devons choisir h pour équilibrer le niveau et la puissance dans les échantillons finis. Nous allons utiliser la méthode de driven-data pour choisir h dans notre simulation et l'application empirique.

Chapitre

3

Simulations et application

Dans ce chapitre nous allons présenter l'évaluation numérique des mesures de dépendance sérielles décrites dans le chapitre 2. Nous avons élaboré plusieurs programmes informatiques en langage Matlab R2011a. Des simulations intensives pour des modèles de séries chronologiques linéaires et non linéaires tels que *AR*, *Bilinéaire*, *TGARCH*, *TAR*, *ARCH*. ont été menées. Nous avons comparé, par la suite, les puissances de ces statistiques de test pour différentes tailles d'échantillons. Nous avons, également, fait une application sur la série de données réelles, *SP500* à la fermeture, du 1er Janvier 1992 au 31 Décembre 2003, excluant les week-ends et les jours fériés, ce qui donne un total de 3026 observations.

1 Mesure de Hoeffding

Rappelons la statistique de test I_n de Skaug et Tjøstheim (1993b), basée sur la mesure de Hoeffding (1948b),

$$I_n = \frac{1}{n} \sum_{t=1}^n \{F_n(X_{t-1}, X_t) - F_n(X_{t-1}, \infty)F_n(\infty, X_t)\}^2.$$

Nous avons vu que cette mesure converge en loi vers une somme de khi deux pondérée $\sum_{j,k} \lambda_{jk} W_{jk}^2$, théorème **1.2.1** du chapitre **2**.

Sous l'hypothèse H_0 que $\{X_t, t \geq 0\}$ est i.i.d., le test, pour un niveau approximatif α , rejette H_0 si $nI_n > u_\alpha$ où u_α est le quantile d'ordre $1 - \alpha$ de la loi limite de nI_n sous l'hypothèse de l'indépendance.

Contrairement au cas d'une distribution limite gaussienne, le premier et le deuxième moment de $\sum_{j,k} \lambda_{jk} W_{jk}^2$, qui peuvent être calculés exactement, en fonction des $\{\lambda_{jk}\}$, sont insuffisants pour calculer les valeurs critiques u_α .

Les poids $\{\lambda_{jk}\}$ peuvent être trouvés, dans le cas où X_0 est continuellement distribuée, comme dans la preuve de la proposition 1 p. 213 dans Shorack & Wellner (1986). Nous aurons $\eta_m = (m\pi)^{-2}$, et les coefficients dans le théorème **1.2.1** sont donnés par $\lambda_{jk} = (jk\pi^2)^{-2}$, ce qui montre que I_n a la même distribution limite que la statistique de test de Blum et al (1961). Dans le cas discret, on les trouve numériquement à partir des valeurs propres de la matrice estimée de Σ . Après avoir trouvé les poids, nous pouvons, soit, utiliser les simulations de Monté Carlo ou l'inversion numérique de la fonction caractéristique¹⁶ de $\sum_{j,k} \lambda_{jk} W_{jk}^2$

$$\varphi_X(t) = \mathbb{E} \left\{ \exp \left(it \sum_{j,k \geq 1} \lambda_{jk} W_{jk}^2 \right) \right\} = \prod_{j,k \geq 1} (1 - 2i\lambda_{jk}t)^{-1/2}.$$

pour obtenir les valeurs critiques u_α de la distribution limite.

Nous utilisons ces valeurs critiques u_α de la distribution limite au lieu de la distribution exacte de nI_n . Ceci conduit à un niveau de signification différent du niveau α souhaité. Cet effet peut être étudié au travers de simulations de la probabilité de rejet sous l'hypothèse H_0 . La table 1 donne des valeurs approchées de la fonction de répartition lorsque la variable numérique varie de 0.30 à 7.35 avec un pas de 0.05. Signalons que le temps d'exécution des programmes est long. La table 3 donne la

¹⁶**Théorème (Gil-Pelaez) :** Si f et $\varphi \in L^2$, $\mathbb{E}(X)$ existe et est finie, alors

$$F_X(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{+\infty} \frac{e^{itx} \varphi_X(-t) - e^{-itx} \varphi_X(t)}{it} dt.$$

taille du test $P_{H_0}(nI_n \geq u_\alpha)$, où les u_α sont tirées de la table 2 pour $X_0 \cong \mathcal{N}(0, 1)$ et pour différentes valeurs de n et α .

| y | $F(y)$ | y | $F(y)$ | y | $F(y)$ |
|------|---------|------|---------|------|---------|
| .30 | .000125 | 2.35 | .901125 | 4.30 | 0,9905 |
| .35 | .000375 | 2.40 | .90925 | 6.5 | .99925 |
| .40 | .001 | 2.45 | .916 | 6.55 | .998875 |
| .45 | .007 | 2.50 | .91775 | 6.6 | .999125 |
| .50 | .0135 | 2.55 | .9215 | 6.65 | .999125 |
| .55 | .03075 | 2.60 | .923625 | 6.7 | .999375 |
| .60 | .053375 | 2.65 | .930125 | 6.75 | .99925 |
| .65 | .080375 | 2.70 | .937 | 6.8 | .99925 |
| .70 | .118625 | 2.75 | .941125 | 6.85 | .9995 |
| .75 | .1555 | 2.80 | .944625 | 6.9 | .99925 |
| .80 | .199875 | 2.85 | .950625 | 6.95 | .999 |
| .85 | .239625 | 2.90 | .951 | 7 | .9995 |
| .90 | .28725 | 3 | .9555 | 7.05 | .999875 |
| .95 | .326 | 4 | .981375 | 7.10 | .999875 |
| 1 | .384 | 4.05 | .984625 | 7.15 | .999625 |
| 1.05 | .41425 | 4.1 | .983875 | 7.2 | .999875 |
| 1.10 | .453625 | 4.15 | .9885 | 7.25 | .99975 |
| 1.15 | .48525 | 4.2 | .988875 | 7.3 | .999375 |
| 1.20 | .541875 | 4.25 | 0,98725 | 7.35 | 1 |

Table 1 : $F(y)$ en appliquant le théorème de Gil-Pelaez

Nous remarquons une légère diminution de l'ordre de 10^{-4} pour 6.75, de 2.10^{-4} pour 6.95 et 7.15, 4.10^{-4} en 7.3. Ceci est probablement dû, d'une part à la troncature adoptée dans $\varphi_x(x)$ et d'autre part au fait que le matériel utilisé, un laptop pavilion dv5 "core 2 duo 2 GHz" peut devenir moins performant après 24 heures d'exécution

continues. Signalons, que ces valeurs sont très proches de celles données dans Blum et al (1961).

| $1 - \alpha$ | $F^{-1}(1 - \alpha)$ |
|--------------|----------------------|
| .90 | 2.29774 |
| .95 | 2.84479 |
| .99 | 4.29230 |

Table 2 : Valeurs critiques de I_n pour $X_0 \cong \mathcal{N}(0, 1)$

Les probabilités dans le tableau suivant ont été trouvées en faisant 8000 itérations du test sur des données simulées. Remarquons que le seuil réel du test n'est pas très éloigné du seuil nominal.

| n | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.1$ |
|-----|-----------------|-----------------|----------------|
| 50 | .015 | .063 | .123 |
| 100 | .0114 | .057 | .114 |
| 250 | .0095 | .049 | .100 |

Table 3 : Niveau exact de I_n pour $X_0 \cong \mathcal{N}(0, 1)$

1.1 Puissance du test basé sur I_n

L'étude analytique de la puissance est, en général, difficile. Une étude de simulation a été entreprise. Pour ce faire, nous avons choisi quelques modèles de séries chronologiques (étudiés dans le chapitre 1), comme des alternatives à l'hypothèse H_0 .

c) Modèle Autorégressif à seuil d'ordre un $TAR(1)$

$$X_t = \begin{cases} \varphi X_{t-1} + \varepsilon_t & \text{si } X_{t-1} > 1 \\ 0.4X_{t-1} + \varepsilon_t & \text{si } X_{t-1} \leq 1 \end{cases} \quad |\varphi| < 1$$

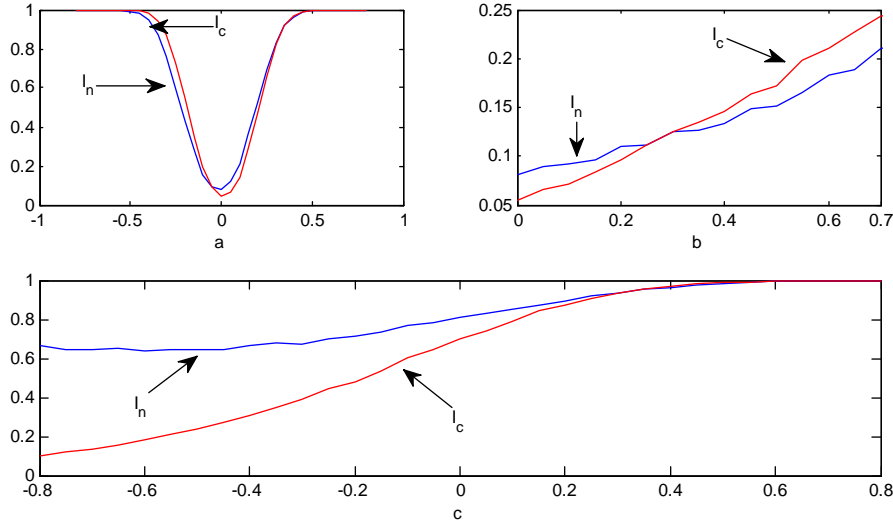


FIG. 1.1: Puissance du test I_n .

Dans la FIG 1.1, les fonctions puissance pour les tests I_n et I_c (coefficient de corrélation)¹⁷ sont tracées en fonction de φ pour $n = 100$. Pour chaque φ , nous avons utilisé 8000 simulations pour l'estimation de la puissance.

On remarque que la puissance du test I_n est proche de celle du coefficient de corrélation I_c pour les modèles linéaires. Pour les modèles non linéaires le test I_n est plus puissant que I_c .

Pour un retard $j > 1$, la statistique est donnée par

$$G_n(m) = \sum_{j=1}^m I_n(j). \tag{1.1}$$

¹⁷Le coefficient de corrélation est donné par

$$I_c = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_{i-1} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

où $I_n(j) = \int \{F_{j,n}(z^{(1)}, z^{(2)}) - F_{j,n}(z^{(1)}, \infty) F_{j,n}(\infty, z^{(2)})\}^2 dF_{j,n}(z^{(1)}, z^{(2)})$. (voir chapitre 2)

Cette statistique converge en loi vers une somme de loi de probabilité du khi deux pondérées à m degrés de liberté (voir le théorème 1.5.1 du chapitre 2)

Un test qui rejette H_0 si $G_n(m) > \mu_{m,\alpha}$, où $\mu_{m,\alpha}$ est le quantile d'ordre $1-\alpha$ de la distribution limite de $nG_n(m)$, a un niveau proche de α . Lorsque G^{-1} existe, comme pour $m=1$, $nG_n(m)$ est une distribution libre relativement à la loi sous-jacente du processus stochastique, et nous avons $\lambda_{jk} = (jk\pi^2)^{-2}$. La table des valeurs critiques est obtenue en prenant la moyenne de 20 valeurs critiques, chacune étant obtenue à partir de 8000 réalisations.

Dans notre cas, nous ne nous intéressons qu'au premier retard $j=1$.

2 Mesure basée sur la différence pondérée des densités

Passons maintenant à la deuxième mesure de dépendance, celle de la différence des densités pondérées donnée par

$$I_{n,1} = \frac{1}{n} \sum_t \{\hat{p}_2(X_t, X_{t-1}) - \hat{p}(X_t)\hat{p}(X_{t-1})\} w_2(X_t, X_{t-1})$$

Les densités sont estimées par la méthode du noyau. Nous utilisons l'estimation à noyau leave one out donné par (2.16) du chapitre 2. Pour nos simulations nous choisissons le noyau gaussien $k(x) = (2\pi)^{-1/2} \exp\{-\frac{1}{2}x^2\}$ de fenêtre optimale $h = n^{-1/6}$ (voir Silverman, 1986).

Nous présentons la puissance du test $P_{H_1}(I_{n,1} > u_\varepsilon)$ en faisant 8000 réalisations pour différentes tailles d'échantillons, au niveau de signification 5% pour des innovations i.i.d $\mathcal{N}(0, 1)$.

Les valeurs critiques, u_ε sont calculées en faisant 8000 itérations, sous H_0 .

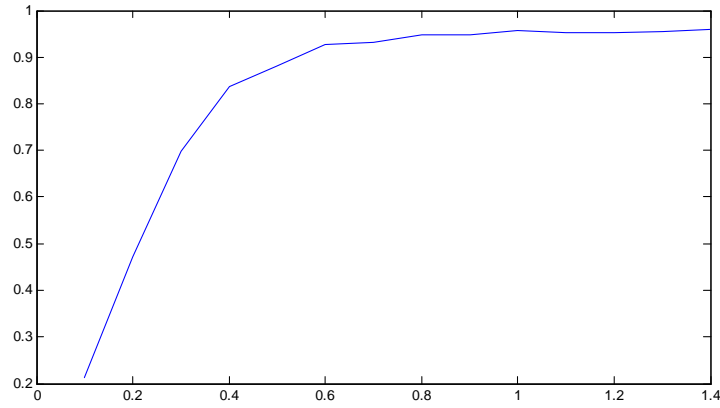


FIG. 2.1: Puissance du test en fonction de la fenêtre au niveau de signification 5% pour le modèle $AR(1)$ pour $n = 100$

Dans le chapitre précédent, nous avons étudié la théorie asymptotique de $I_{n,1}$ (biais, variance et distribution asymptotique), peut-on l'utiliser en pratique ?

En se basant sur les résultats de la deuxième section du chapitre **2**, la statistique naturelle pour tester la dépendance est

$$I_{n,1}^* = \left\{ I_{n,1} - \widehat{\mathbb{E}}(I_{n,1}) \right\} / \widehat{SD}(I_{n,1}) \quad (2.1)$$

où les expressions de $\mathbb{E}(I_{n,1})$ et $SD(I_{n,1})$ sont données dans le chapitre **2**, et où leurs estimations naturelles, dans le cas leave-one-out, est donnée par

$$\widehat{\mathbb{E}}(I_{n,1}) = n^{-1} \left[2 \left\{ \frac{1}{n} \sum_t \widehat{p}(X_t) w(X_t) \right\}^2 - \frac{1}{n} \sum_t \widehat{p}^2(X_t) w(X_t) \right]$$

et

$$\widehat{SD}(I_{n,1}) = n^{-1/2} \left| \frac{1}{n} \sum_t \widehat{p}^2(X_t) w(X_t) - \left\{ \frac{1}{n} \sum_t \widehat{p}(X_t) w(X_t) \right\}^2 \right|. \quad (2.2)$$

Pour n grand $I_{n,1}^*$ suit une loi normale centrée et réduite, sous H_0 . On rejette l'hypothèse nulle de l'indépendance pour $I_{n,1}^* \geq u_\alpha$, pour un niveau approximatif α .

Pour comparer la proximité entre le niveau réel et le niveau approximatif, nous

examinons d'abord les approximations asymptotiques de la moyenne et de l'écart type (cas de leave-one-out). Les résultats sont représentés dans la FIG 2.2 en fonction de n pour un processus $\{e_t\}$, i.i.d. tel que $\mathbb{E}(e_t) = 0$ et $SD(e_t) = 1$; la fonction poids, choisie, est $w(x) \equiv 1$. La moyenne et l'écart type simulés de $I_{n,1}$ sont représentés sur la même figure. Ils ont été calculés à partir de 8000 réalisations de longueur n . Nous avons utilisé à nouveau la fenêtre $h = n^{-1/6}$.

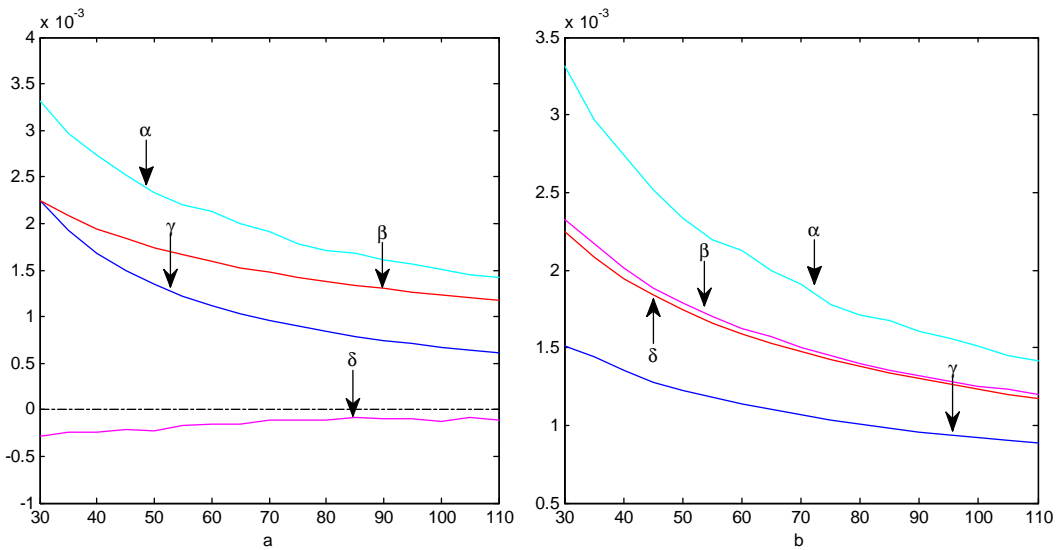


FIG. 2.2: La performance des formules asymptotiques quand n varie pour $X_t \cong i.i.d N(0, 1)$. (a) α : $SD(I_{n,1})$ simulé, β : $SD(I_{n,1})$ donné par la formule (2.21), γ : $\mathbb{E}(I_{n,1})$ simulée, δ : $\mathbb{E}(I_{n,1})$ donnée par la formule (2.18); (b) α : $SD(I_{n,1})$ simulé, β : $SD(I_{n,1})$ donné par la formule (2.21), γ : $\widehat{SD}(I_{n,1})_{corr}$, δ : $\widehat{SD}(I_{n,1})$.

De la FIG 2.2.a, de la courbe de la moyenne nous pouvons remplacer $\mathbb{E}_n(I_{n,1})$ par zéro dans (2.18) du chapitre 2. L'écart type est beaucoup plus gênant, cependant. Pour $n = 100$ on voit que l'utilisation de (2.21) du chapitre 2, conduit à une sous-estimation claire. Des améliorations peuvent être obtenues en incluant des termes d'ordre trois dans le développement asymptotique qui, sont pour $h = n^{-1/6}$, les termes dûs au biais dans $\widehat{p}(x)$ (2.23) d'ordre $O(n^{-1/2}h)$ et le premier terme en n^{-1} de l'écart type de (2.18) qui est de l'ordre $n^{-1}h^{-1}$. Pour $n = 100$, $n^{-1/2}h = n^{-2/3} \approx 0.046$ et $n^{-1}h^{-1} = n^{-5/6} \approx 0.022$ de sorte qu'il y a relativement de petites différences dans l'ordre entre ces termes et le premier terme d'ordre $n^{-1/2} = 0.1$.

Dans la pratique les intégrales entrant dans (2.21) sont estimées par (2.2). La mise en œuvre de cette expression diminue la qualité de l'approximation comme le montre FIG 2.3. La raison est que, dans le cas normal, les valeurs de $\int p^3(x) dx$ et $\{\int p^2(x) dx\}^2$ sont très proches les unes des autres, de sorte que l'estimation de leur différence est dominée par le biais des estimations des intégrales. Nous pouvons ajuster les biais à l'aide de (2.20) de sorte que \widehat{p} est remplacé par $\widehat{p} - \frac{1}{2}h^2\widehat{p}''$ et \widehat{p}^2 par $\widehat{p}^2 - h^2\widehat{p}\widehat{p}''$, où $\widehat{p}''(x)$ est l'estimateur par la méthode du noyau de $p''(x)$. Cela donne une version corrigée du biais de $\widehat{SD}(I_{n,1})$

$$\begin{aligned} \widehat{SD}(I_{n,1})_{corr} &= n^{-1/2} \left| \frac{1}{n} \sum_t \left[\widehat{p}^2(X_t) - h^2\widehat{p}(X_t)\widehat{p}''(X_t) \right] w(X_t) \right. \\ &\quad \left. - \left\{ \frac{1}{n} \sum_t \left[\widehat{p}(X_t) - \frac{1}{2}h^2\widehat{p}''(X_t) \right] w(X_t) \right\}^2 \right|. \end{aligned} \quad (2.3)$$

Nous sommes alors très proches de l'expression du premier ordre de (2.21) comme le montre la figure FIG 2.2.b.

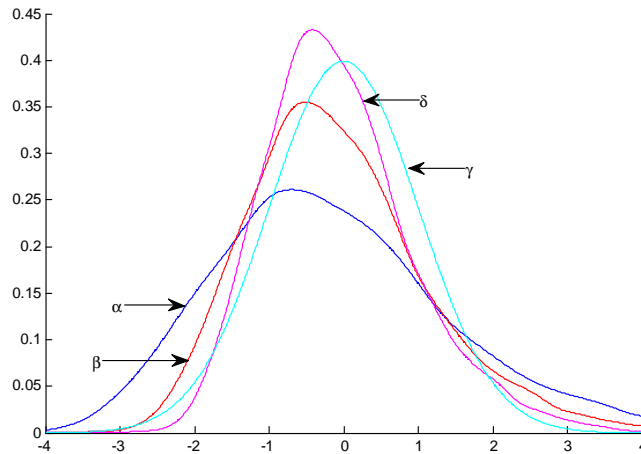


FIG. 2.3: La distribution asymptotique de $I_{n,1}$ sous H_0 pour $n + 1 = 100$ et X_t de loi $N(0, 1)$. La densité de $I_{n,1}/SD$ avec SD est simulé par $\alpha = \widehat{SD}(I_{n,1})$ donné par (2.2), $\beta = \widehat{SD}(I_{n,1})_{corr}$ donné par (2.3), $\gamma =$ l'écart type simulé de $I_{n,1}$, δ : la densité normale standard.

L'approximation de la densité de $I_{n,1}$ par la loi normale est étudiée dans la

figure FIG 2.3, où la représentation de la densité simulée de la statistique de test standardisé $I_{n,1}^*$ est montrée à la fois lorsque $SD(I_{n,1})$ est approché par simulations et lorsque $SD(I_{n,1})$ est estimé par (2.2) et par (2.3), en utilisant 8000 réalisations. La distribution normale standard est tracée à titre de référence. L'approximation à la normalité est raisonnablement bonne, mais il est clair que la normalisation fondée sur les (2.2) et (2.3) mènera à un bon niveau du test si les fractiles de la distribution normale standard sont utilisés. Les niveaux de signification approximatifs pour 0.01, 0.05 et 0.1 pour la statistique de test $I_{n,1}^*$ en utilisant (2.3) sont donnés par 0.048, 0.093 et 0.135. Et l'utilisation de l'écart type simulé d'autre part a donné 0.021, 0.059 et 0.096. Ceci et le fait que, sous H_0 , $\{X_t\}$ est i.i.d. suggèrent fortement l'utilisation de la méthode de bootstrap. Puisque l'approximation normale semble adéquate, seul l'écart-type dans (2.3) est bootstrapé.

La distribution de $I_{n,1}^*$ simulée pour le processus gaussien $\{e_t\}$ avec $\widehat{\mathbb{E}}(I_{n,1}) = 0$ et $\widehat{SD}_{corr}(I_{n,1})$ calculé à partir de 50 répliques est illustré sur la figure FIG 2.4.

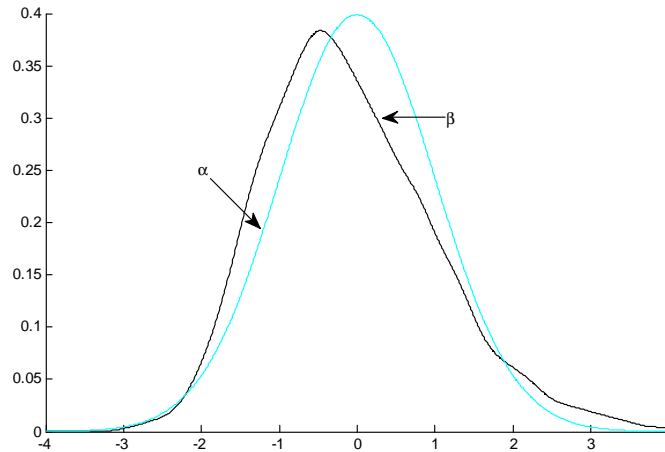


FIG. 2.4: La distribution asymptotique de $I_{n,1}$ sous H_0 pour $n + 1 = 100$ et X_t de loi $N(0, 1)$. La densité de $I_{n,1}/SD$ avec SD est simulé par $\alpha = \widehat{SD}_{corr}(I_{n,1})$ bootstrapé, $\beta =$ la densité normale standard.

Comparons la puissance de test basé sur $I_{n,1}$ est le test de corrélation, I_c Pour la statistique $I_{n,1}$ nous considérons deux versions, la première, $I_{n,1}$ standardisée

en utilisant l'écart type donné par (2.3) notée $I_{n,1}^*$ et la deuxième, en utilisant la méthode du bootstrap décrite auparavant, notée $I_{n,1}^\clubsuit$. Le bootstrap est aussi utilisé pour le calcul de la valeur critique du test I_c . Pour le modèle $AR(1)$ au niveau de 5% les résultats sont présentés dans le tableau suivant

| Test | $\varphi = .2$ | $\varphi = .5$ |
|---------------------|----------------|----------------|
| $I_{n,1}^*$ | .216 | .908 |
| $I_{n,1}^\clubsuit$ | .159 | .875 |
| I_c | .644 | .996 |

Table 4 La puissance des tests au niveau 5% pour $n + 1 = 100$ pour $AR(1)$

Il est clair que la puissance du coefficient de corrélation est la meilleure.

3 Mesure basée sur l'information conditionnelle

Avant d'examiner la puissance du test basé sur la statistique $I_{n,2}(j, \epsilon)$ pour différents modèles, nous examinons d'abord le niveau exact du test. Rappelons que la vérification du niveau exact du test de permutation pour l'indépendance n'est pas nécessaire, puisque le test de permutation, par construction, a un niveau exact.

Le tableau 5 montre le niveau exact du test asymptotique pour des tailles d'échantillon de 100 et 200 pour ϵ égal à 0.5, 1.0, 1.5 et 2 sur la base de 1000 simulations. Dans tous les cas le test asymptotique a tendance à accepter l'hypothèse nulle. L'augmentation de la taille de la série n'améliore pas le niveau du test pour $\epsilon = 2$. Le niveau du test augmente en augmentant j .

| $I_{n,2}(j)$ | | | | | | | | |
|--------------|-----------------|----------------|------------------|----------------|-----------------|----------------|------------------|----------------|
| j | $n = 100$ | | | | $n = 200$ | | | |
| | $\epsilon = .5$ | $\epsilon = 1$ | $\epsilon = 1.5$ | $\epsilon = 2$ | $\epsilon = .5$ | $\epsilon = 1$ | $\epsilon = 1.5$ | $\epsilon = 2$ |
| 1 | .0470 | .0340 | .0480 | .0550 | .0440 | .0470 | .0460 | .0470 |
| 2 | .0870 | .0760 | .0930 | .0980 | .1030 | .1070 | .1050 | .0870 |
| 3 | .1400 | .1280 | .1380 | .1430 | .1520 | .1620 | .1600 | .1400 |
| 4 | .1910 | .1740 | .1850 | .1950 | .2190 | .2120 | .2090 | .1900 |

Table 5 : Niveau exact du test de $I_{n,2}(j, \epsilon)$ en faisant varier la taille de l'échantillon, la fenêtre ϵ et la dimension d'empilement j , au seuil .05 pour des v.a i.i.d. gaussiennes

Nous étudions la performance de ce test pour les alternatives données auparavant. Tout au long, nous utilisons 1000 simulations pour chaque cas, en gardant le nombre de réplifications du bootstrap fixé à $B = 199$. Les résultats sont reportés dans la table 6.

Pour le modèle $AR(1)$ le test de permutation a une puissance proche de l'unité pour un retard d'ordre un pour les différentes tailles de l'échantillon. Le taux de rejet est proche du niveau nominal ce qui confirme la capacité du test dans la détection de la dépendance conditionnelle. Notons que la puissance du test pour des retards supérieurs à 1 est encore plus petite que le niveau du test. C'est peut-être dû au fait qu'il y a une dépendance conditionnelle dans ce processus pour un retard plus élevé.

Le modèle *Bilinéaire* présente une dépendance conditionnelle à travers les deux premiers retards. Pour ce modèle, l'augmentation dans la taille de l'échantillon améliore la puissance du test. Comme prévu, le test a une puissance contre cette alternative seulement pour le premier et le deuxième retard.

Le test a aussi une puissance contre le modèle TAR d'ordre un pour $\epsilon = 1$ le taux de rejet est de .682 pour une taille de 100 et .899 pour $n = 200$. Le cas $\epsilon = 1.5$ montre moins de puissance que la petite fenêtre pour le retard d'ordre un. Ici, il peut également être observé qu'il existe une certaine puissance dans le

| modèles | | $I_{n,2}(j)$ | | | | | | | | | | | |
|-------------------------------|-----|--------------|-------|-------|-------|-------|-------|-----------|-------|---|--|--|--|
| | | $n = 100$ | | | | | | $n = 200$ | | | | | |
| | | ϵ | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | | | |
| <i>AR</i> (1) | .5 | .8870 | .0250 | .0160 | .0110 | .9990 | .0250 | .0100 | .0050 | | | | |
| | 1 | .9650 | .0200 | .0210 | .0090 | 1.000 | .0240 | .0180 | .0120 | | | | |
| | 1.5 | .9790 | .0250 | .0230 | .0150 | 1.000 | .0350 | .0160 | .0130 | | | | |
| <i>ARCH</i> (1) | .5 | .6330 | .0360 | .0190 | .0280 | .9560 | .0590 | .0190 | .0040 | | | | |
| | 1 | .8500 | .0470 | .0150 | .0150 | .9900 | .0830 | .0280 | .0170 | | | | |
| | 1.5 | .8800 | .0520 | .0240 | .0160 | .9920 | .0880 | .0230 | .0180 | | | | |
| <i>Bilinéaire AR</i> (2) | .5 | .3060 | .2670 | .0170 | .0310 | .5850 | .7070 | .0450 | .0110 | | | | |
| | 1 | .4530 | .7290 | .0340 | .0110 | .7440 | .9780 | .0610 | .0110 | | | | |
| | 1.5 | .4850 | .8500 | .0370 | .0120 | .7650 | .9910 | .0440 | .0170 | | | | |
| <i>Threshold AR</i> (1) | .5 | .5850 | .0380 | .0170 | .0260 | .8850 | .0510 | .0250 | .0190 | | | | |
| | 1 | .6820 | .0480 | .0290 | .0340 | .8980 | .0740 | .0320 | .0280 | | | | |
| | 1.5 | .5660 | .0600 | .0300 | .0330 | . | . | . | . | | | | |
| <i>Threshold GARCH</i> (1, 1) | .5 | .4870 | .1190 | .0260 | .0700 | .7940 | .1200 | .0470 | .0280 | | | | |
| | 1 | .6660 | .2960 | .0800 | .0250 | .9300 | .6900 | .2070 | .0370 | | | | |
| | 1.5 | .7270 | .4270 | .1490 | .0420 | .9490 | .7800 | .3760 | .1010 | | | | |

Table 6 : Puissance du test basé sur de $I_{n,2}(j, \epsilon)$ en variant la taille de l'échantillon, la fenêtre ϵ et la dimension d'empilement j , au seuil 5% pour des v.a i.i.d. gaussiennes

retard d'ordre deux. Ceci est dû à la considération d'une fenêtre trop importante par rapport à l'écart type de la distribution conditionnelle de X sachant le passé.

Pour le modèle $ARCH(1)$ le test à une puissance remarquablement élevé pour une taille d'échantillon $n = 100$ pour $\epsilon = 1$ il va de .85 à .99 pour la taille $n = 200$. Certaine puissance marginale est également détectée dans le second retard.

Le test a aussi une puissance contre l'alternative *Threshold GARCH*. Pour le modèle *Threshold GARCH(1, 1)* le test a une puissance pour les 4 retards étudiés. Dans ce cas, l'interprétation en terme d'ordre, n'est pas possible puisque le modèle se présente comme une somme infinie de Markov.

4 Mesure basée sur l'entropie

La statistique de test est donnée par

$$I_{n,3}(j) = \frac{1}{n-j} \sum_{t \in S_n(j)} \ln \frac{\hat{f}_j(Z_{jt})}{\hat{p}(X_t)\hat{p}(X_{t-j})} \quad j = 1, \dots, n-1$$

Pour tout les PGD la condition que la densité marginale $p(\cdot) > c$ dans l'hypothèse *A.1*, cède après la transformation logistique dans (4.6). Une expérience préliminaire montre que la théorie asymptotique fournit une approximation assez faible pour les niveaux du test pour des échantillons finis considérés. Pour assurer une comparaison équitable pour tout les tests, nous utilisons le bootstrap modifié (lissé) proposé dans la section 4 du chapitre 2.

Pour cela, nous transformerons d'abord les données via la fonction logistique donnée dans (4.6), afin que l'échantillon $\mathcal{X} = \{X_t\}_{t=1, \dots, n}$ ait un support sur \mathbb{I} . Ceci induit une dépendance de chaque X_t sur l'ensemble de l'échantillon \mathcal{X} , mais cette effet devrait être asymptotiquement négligeable. L'entropie $I_3(j)$ est invariante en vertu de cette transformation. Pour ce test nous utilisons le noyau quartique $k(u) = \frac{15}{16} (1 - u^2)^2 \mathbf{1}_{(|u| \leq 1)}$ et la même fenêtre h . Nous utilisons la méthode driven-data dans le choix de h qui est plus objective que le choix arbitraire. De (4.17), une fenêtre

asymptotiquement optimale qui donne le taux de convergence optimal pour le critère d'information de Kullback-Leibler pour $\hat{p}_t(\cdot)$ est

$$h^o = \left\{ \left[\int_{-1}^1 k^2(u) du \right]^{-1} \left[\int_{-1}^1 u^2 k^2(u) du \right]^2 \int_0^1 \left[\frac{p^{(2)}(x)}{p(x)} \right]^2 p(x) dx \right\}^{-\frac{1}{5}} n^{-\frac{1}{5}} \quad (6.1)$$

où $k(\cdot)$ est le noyau utilisé dans l'estimation de $p(\cdot)$. Pour le noyau quartique, nous avons

$$h_Q^o = 2,0236 \left\{ \int_0^1 \left[\frac{p^{(2)}(x)}{p(x)} \right]^2 p(x) dx \right\}^{-\frac{1}{5}} n^{-\frac{1}{5}} \quad (6.2)$$

Cette fenêtre optimale ne peut pas être utilisée puisqu'elle contient $p(\cdot)$ et $p^{(2)}(\cdot)$ inconnues. Nous utilisons donc la méthode de substitution pour obtenir une fenêtre data-driven

$$\hat{h}_Q = 2,0236 \left\{ n^{-1} \sum_{t \in \tilde{S}_n} \left[\frac{\tilde{p}_{nt}^{(2)}(X_t)}{\tilde{p}_{nt}(X_t)} \right]^2 \right\}^{-\frac{1}{5}} n^{-\frac{1}{5}} \quad (6.3)$$

où $\tilde{p}_{nt}(X_t)$ et $\tilde{p}_{nt}^{(2)}(X_t)$ sont les estimateurs de $p(\cdot)$ et $p^{(2)}(\cdot)$, $\tilde{S}_n = \{t \in \mathbb{N} : 1 \leq t \leq n : X_t \in [h_0, 1 - h_0]\}$ est l'ensemble des indices où X_t est dans la région $[h_0, 1 - h_0]$ et $h_0 = h_0(n)$ est la fenêtre préliminaire utilisée dans $\tilde{p}_{nt}(\cdot)$ et $\tilde{p}_{nt}^{(2)}(\cdot)$. L'utilisation de \tilde{S}_n nous permet d'éviter les effets de bord pour $\tilde{p}_{nt}^{(2)}(\cdot)$, qui existe encore malgré la prise en charge de l'effet de bords de $\tilde{p}_{nt}(\cdot)$ par l'utilisation du noyau de Jacknaife $k_b(\cdot)$. Si $\tilde{p}_{nt}(\cdot)$ et $\tilde{p}_{nt}^{(2)}(\cdot)$ sont consistants pour $p(\cdot)$ et $p^{(2)}(\cdot)$, \hat{h}_Q va être consistante pour h_Q^o . Pour plus de comodité, nous utilisons aussi le noyau quartique pour $\tilde{p}_{nt}(X_t)$ et $\tilde{p}_{nt}^{(2)}(X_t)$. Pour examiner la sensibilité du test pour le choix préliminaire de la fenêtre h_0 , posons $h_0 = \hat{S}_X n^{-\frac{1}{1+\lambda}}$ pour $\lambda = 1, 2, 3, 4, 5$, où \hat{S}_X est l'écart-type de l'échantillon \mathcal{X} . Ceci couvre un intervalle suffisamment large des taux pour h_0 . Ces taux assurent la consistance de $\tilde{p}_{nt}(X_t)$ et $\tilde{p}_{nt}^{(2)}(X_t)$ pour $p(\cdot)$ et $p^{(2)}(\cdot)$ si $p(\cdot)$ est continuellement différentiable d'ordre quatre, sur \mathbb{I} . Nous reportons seulement les résultats pour $\lambda = 5$.

Pour examiner les niveaux bootstrapés du test, nous simulons 1000 réalisations de

l'échantillon \mathcal{X} sous PGD0. Posons $B = 100$, le nombre des itérations de bootstrap pour chaque itération de simulation.

Nous avons élaboré un programme informatique en langage Matlab afin de mettre en pratique $I_{n,3}$. Ce programme tourne mais nécessite deux journées entières pour effectuer 1%.

La table 7 rapporte la puissance des trois mesures I_n , $I_{n,1}$ et $I_{n,2}$ pour les différents modèles définis auparavant. Pour $n = 100$, le test basé sur I_n est le plus puissant pour les modèles $AR(1)$ et Bilinéaire $AR(2)$. le test basé sur $I_{n,1}$ est plus puissant pour $ARCH(1)$ et Threshold $GARCH(1,1)$. Pour le test basé sur $I_{n,2}$ est plus puissant pour Threshold $AR(2)$.

| modèles | $n = 100$ | | | $n = 200$ | | |
|------------------------------|-----------|--------------|--------------|-----------|--------------|--------------|
| | $I_n(1)$ | $I_{n,1}(1)$ | $I_{n,2}(1)$ | $I_n(1)$ | $I_{n,1}(1)$ | $I_{n,2}(1)$ |
| <i>AR</i> (1) | .9995 | .9760 | .8870 | 1.0000 | .9992 | .9990 |
| <i>ARCH</i> (1) | .1853 | .8056 | .6330 | .2360 | .9396 | .9560 |
| Bilinéaire <i>AR</i> (2) | .8521 | .4773 | .3060 | .9985 | .5916 | .5850 |
| Threshold <i>AR</i> (2) | .1736 | .4284 | .5850 | .1797 | .6319 | .8850 |
| Threshold <i>GARCH</i> (1,1) | .1782 | .6780 | .4870 | .2502 | .8408 | .7940 |

Table 7 : Puissance des tests en variant la taille de l'échantillon, au seuil 5% pour un retard égal à 1

5 Application sur une série de données réelles

Nous appliquons les tests étudiés auparavant sur la série de données réelles, *SP500* à la fermeture (P_t), du 1^{er} Janvier 1992 au 31 Décembre 2003, excluant les week-ends et les jours fériés, ce qui donnerait un total de 3026 observations. Nous posons $X_t = 100 \ln(P_t/P_{t-1})$, (on la transforme par la fonction logistique afin d'avoir une support $[0, 1]$ pour le test basé sur l'entropie). Il est connu que l'indice boursier présente un regroupement de volatilité fort, Nous l'ajustons par le modèle *ARMA* (3, 3) à erreur *GARCH* (1, 1) :

$$X_t = .000551 + .7616X_{t-1} + .6790X_{t-2} - .7974X_{t-3} - .7657\varepsilon_{t-1} - .6790\varepsilon_{t-2} + .7612\varepsilon_{t-3} + \varepsilon_t h_t^{1/2}$$

$$h_t = 5.75 \cdot 10^{-7} + .06859\varepsilon_{t-1}^2 + .9284h_{t-1}.$$

Nous appliquons les tests étudiés sur la série (les rendements logarithmiques) et sur les résidus standardisés du modèle ajusté. Les résultats sont donnés dans le tableau suivant :

| j | <i>SP500</i> | | | Les résidus de <i>ARMA</i> (3, 3) – <i>GARCH</i> (1, 1) | | |
|-----|--------------|--------------|--------------|---|--------------|--------------|
| | $I_n(j)$ | $I_{n,1}(j)$ | $I_{n,2}(j)$ | $I_n(j)$ | $I_{n,1}(j)$ | $I_{n,2}(j)$ |
| 1 | .00 | .00 | .05 | 1.00 | 1.00 | .9800 |
| 2 | .00 | .00 | .05 | 1.00 | 1.00 | .0300 |
| 3 | .00 | .00 | .05 | 1.00 | 1.00 | .5250 |
| 4 | .00 | .00 | .05 | 1.00 | 1.00 | .1700 |

Table 8 : Les p -values bootstrapées des tests pour la série *SP500* et les résidus standardisés du modèle ajusté

Pour les rendements logarithmiques de la série *SP500*, les trois tests rejettent l'hypothèse d'indépendance aux différents retards (1 – 4). L'application sur les résidus standardisés du modèle ajusté, montre que, les tests basés sur I_n et $I_{n,1}$ acceptent l'hypothèse d'indépendance pour les quatre retards or le test basé sur $I_{n,2}$ la rejette

pour l'ordre deux. Il semble que ce dernier détecte la présence d'une dépendance conditionnelle pour un retard $j = 2$.

Appendice

Equation intégrale linéaire

De nombreux problèmes physiques sont généralement résolus par des méthodes d'équations différentielles peuvent être résolus plus efficacement par des méthodes d'équations intégrales. En effet, celles-ci ont fait leur apparition dans la littérature actuelle jusqu'ici pas résoluble par des méthodes classiques d'équations différentielles. Ces problèmes ne manquent pas dans de nombreux appliquée, et les types de solutions explorées ici sera utile en particulier dans les mathématiques appliquées, mécanique théoriques et la physique mathématique.

Definition

Une équation intégral est une équation dans laquelle une fonction inconnue apparaît sous un ou plusieurs signes d'intégrale. Par exemple, pour $a \leq s \leq b$, $a \leq t \leq b$, les équations

$$f(s) = \int_a^b K(s, t) g(t) dt \quad (1)$$

$$g(s) = f(s) + \int_a^b K(s, t) g(t) dt \quad (2)$$

$$g(s) = \int_a^b K(s, t) [g(t)]^2 dt \quad (3)$$

où la fonction $g(s)$ est une fonction inconnue tandis que toutes les autres fonctions sont connues, sont des équations intégrales. Ces fonction peuvent être à valeurs

complexes pour les variables réels s et t .

Les équations intégrales apparaissent dans de nombreux domaines de la mécanique et la physique mathématique. Elles surgissent aussi comme une représentation des solutions des équations différentielles. En effet, une équation différentielle peut être remplacé par une équation intégrale qui comprend ses conditions limites. En tant que tel, chaque solution de l'équation intégrale satisfait automatiquement ces conditions limites. Les équations intégrales font également l'une des outils utiles les plus utilisés dans de nombreuses branches de l'analyse pure, telles que les théories de l'Analyse fonctionnelle et les processus stochastiques.

On peut aussi considérer les équations intégrales dans laquelle la fonction inconnue dépend non seulement d'une seule variable, mais de plusieurs variables. Par exemple,

$$g(s) = f(s) + \int_{\Omega} K(s, t) g(t) dt \quad (4)$$

où s et t sont des vecteurs de dimension n . De même, on peut aussi considérer les systèmes d'équations intégrales avec plusieurs fonctions inconnues.

Une équation intégrale est dite linéaire si seulement les opérations linéaires sont effectuées sur la fonction inconnue. Les équations (1) et (2) sont linéaires, tandis (3) est non linéaire. En effet les équations (1) et (2) peuvent être écrite comme

$$L[g(s)] = f(s) \quad (5)$$

où L est l'opérateur intégral approprié. Alors, pour toutes constantes c_1 et c_2 , nous avons

$$L[c_1 g_1(s) + c_2 g_2(s)] = c_1 L[g_1(s)] + c_2 L[g_2(s)]. \quad (6)$$

C'est le critère général pour un opérateur linéaire.

Le type le plus général de l'équation intégrale linéaire est de la forme

$$h(s)g(s) = f(s) + \lambda \int_a^b K(s,t)g(t)dt$$

b peut être variable ou fixe. Les fonctions f, h et K sont des fonctions connues, tandis g est à déterminer ; λ un paramètre réel ou complexe non nul. La fonction $K(s, t)$ est dite noyau.

Conditions de régularité

Pour une fonction de carré intégrable $g(t)$

$$\int_a^b |g(t)|^2 dt < \infty. \quad (7)$$

C'est ce qu'on appelle une fonction de \mathcal{L}_2 . Les conditions de régularité sur le noyau $K(s, t)$ comme une fonction de deux variables sont similaires. C'est une fonction de \mathcal{L}_2 si

– $\forall s \in [a, b], \forall t \in [a, b]$

$$\int_a^b \int_a^b |K(s, t)|^2 ds dt < \infty.$$

– $\forall s \in [a, b]$

$$\int_a^b |K(s, t)|^2 dt < \infty.$$

– $\forall t \in [a, b]$

$$\int_a^b |K(s, t)|^2 ds < \infty.$$

Types particulier de noyaux

Noyau dégénéré

Un noyau $K(s, t)$ est dégénéré ou séparable si il peut être exprimé comme une somme finie du produit d'une fonction de s seulement et d'une fonction de t seuleu-

ment i.e.,

$$K(s, t) = \sum_{i=1}^n a_i(s) b_i(t) \quad (8)$$

Les fonctions $a_i(s)$ peut être supposées linéairement indépendantes ($c_1 a_1 + c_2 a_2 + \dots + c_n a_n = 0$, où c_i sont des constantes arbitraires, alors $c_1 = c_2 = \dots = c_n = 0$), autrement le nombre du terme dans la relation (8) peut être réduit.

Noyau symétrique

Une fonction à valeurs complexes $K(s, t)$ est symétrique ou Hermitienne si $K(s, t) = K^*(t, s)$, où $*$ désigne le conjugué complexe. Pour un noyau réel, il est symétrique si $K(s, t) = K(t, s)$.

$O(\cdot)$ et $o(\cdot)$ stochastiques

Une suite de v.a $\{X_n\}$ de fonction de répartition $\{F_n\}$, est dite bornée en probabilité si pour tout $\varepsilon > 0$ il existe M_ε et N_ε tel que

$$F_n(M_\varepsilon) - F_n(-M_\varepsilon) > 1 - \varepsilon, \quad \forall n > N_\varepsilon.$$

pour deux suites de v.a $\{U_n\}$ et $\{V_n\}$, la notation $U_n = O_p(V_n)$ signifie que $\{U_n/V_n\}$ est $O_p(1)$. De plus, $U_n = o_p(V_n)$ signifie que $U_n/V_n \xrightarrow{\mathcal{L}} 0$. $U_n = o_p(V_n) \Rightarrow U_n = O_p(V_n)$.

Théorème de Mercer

Le terme noyau, dans ce contexte, est une fonction continue telle que $h(x, s) = h(s, x)$, i.e., symétrique.

h est dit de type positif si

$$\sum_{i=1}^n \sum_{j=1}^n h(x_i, x_j) c_i c_j \geq 0.$$

pour toute suites finies de points x_1, \dots, x_n de $[a, b]$ et tout choix des réels c_1, \dots, c_n .

À h on associe l'opérateur intégral B défini par :

$$[Bf](x) = \int h(x, s) f(s) ds.$$

Pour des raisons techniques, supposons que f peut parcourir l'espace $L^2[a, b]$ des fonctions réelles de carré intégrable.

Théorème 0.0.1 *Soit K une fonction noyau continue symétrique de type positif.*

Alors

- les valeurs propres de B sont des réels positifs
- il existe une base hilbertienne $(e_i)_i$ de $L^2[a, b]$ constituée de fonctions propres
- les fonctions propres correspondantes à des valeurs propres strictement positives sont continues sur $[a, b]$
- h admet la représentation

$$h(s, t) = \sum_i \lambda_i e_i(s) e_i(t).$$

où la convergence est absolue et uniforme.

En mathématiques et en analyse fonctionnelle le théorème de Mercer est une représentation d'une fonction symétrique définie positive par le carré d'une série convergente de produits de fonctions. Ce théorème est l'un des résultats phares de James Mercer. C'est un outil théorique important dans la théorie des équations intégrales. Il est aussi utilisé dans la théorie hilbertienne des processus stochastiques.

Généralisation

Le théorème se généralise en remplaçant l'intervalle $[a, b]$ par un espace compact, la mesure de Lebesgue sur $[a, b]$ étant remplacée par une mesure finie dénombrablement additive μ sur les boréliens de X . Supposons que le support de μ est X , c'est-à-dire que $\mu(U) > 0$ pour tout ouvert U non vide de X . Alors on a essentiellement le même résultat :

Théorème 0.0.2 *Soit K un noyau continu symétrique de type positif sur X . Alors les valeurs propres de T_K sont des réels positifs ou nuls et il existe une base hilbertienne $(e_i)_i$ de $L^2\mu(X)$ constituée de fonctions propres pour T_K . Les fonctions propres correspondant aux valeurs propres non nulles sont continues sur X et K s'écrit*

$$K(s, t) = \sum_{j=1}^{\infty} \lambda_j e_j(s) e_j(t)$$

où la convergence est absolue et uniforme sur X .

La généralisation suivante porte sur la représentation des noyaux mesurables.

Soit (X, M, μ) un espace mesuré σ -fini. Un noyau L^2 sur X est une fonction

$$K \in L^2_{\mu \otimes \mu}(X, X)$$

Tout noyau L^2 définit un opérateur borné T_K , par la formule :

$$T_K \varphi \psi = \int_{X \times X} K(y, x) \varphi(y) \psi(x) d[\mu \otimes \mu](y, x)$$

T_K est un opérateur compact (c'est même un opérateur de Hilbert-Schmidt). Si le noyau K est symétrique, par le théorème spectral pour les opérateurs compacts normaux sur un espace de Hilbert, il existe une base hilbertienne constituée de vecteurs propres pour T_K . Les vecteurs de cette base correspondant aux valeurs propres non nulles forment une famille dénombrable $(e_i)_i$.

Théorème 0.0.3 *Si K est un noyau symétrique de type positif sur (X, M, μ) , alors on a*

$$K(y, x) = \sum_{i \in \mathbb{N}} \lambda_i e_i(y) e_i(x)$$

qui converge en norme L^2 . Noter que si le noyau n'est plus supposé continu, le développement ne converge plus nécessairement uniformément.

Théorème de Karhunen–Loève

Théorème 0.0.4 *soit $\{X_t, t \in [a, b]\}$ un processus aléatoire de moyenne nulle et de carré intégrable défini sur l'espace probabilisé (Ω, F, P) , de fonction covariance continue $K_X(s, t)$.*

Alors $K_X(s, t)$ est le noyau de Mercer et soit e_k une base orthonormale de $L^2([a, b])$ formée par les fonction propres de T_{K_X} de valeur propres respectives λ_k , X_t admet la représentation suivante

$$\sum_{k=1}^{\infty} Z_k e_k(t)$$

d'où la convergence dans L^2 , uniformément en t et

$$Z_k = \int_{[a, b]} X_t e_k(t) dt$$

De plus les v.a Z_k ont une moyenne nulle, sont non corrélées et de variance λ_k .

Notons que par la généralisation du théorème de Mercer on peut remplacer l'intervalle $[a, b]$ par des espaces compacts C et la mesure de sur $[a, b]$ par une mesure de Borel de support C .

Théorie de l'information

La théorie de l'information a été introduite par Shannon et Wiener et ses applications statistiques dues à Kullback (1959). (suite de l'introduction)

soit X une variable (vecteur) aléatoire continue de fonction de densité $f_X(x)$. L'entropie de Shannon est définie par

$$H(X) = - \int \ln(f_X(x)) f_X(x) dx \quad (9)$$

qui n'est rien d'autre que l'espérance de $-\ln f_X$, $-\mathbb{E}(\ln f_X)$. Notons que l'entropie de Shannon est indépendante de l'échelle. Par exemple la variable transformée aX , où a est une constante réelle, son entropie devient $H(aX) = H(X) + m \ln a$; où m est la dimension de x .

Pour le couple de *v.a* X, Y de fonction de densité jointe $f_{X,Y}(x, y)$, l'entropie jointe

$$H(X, Y) = - \int \int \ln(f_{X,Y}(x, y)) f_{X,Y}(x, y) dx dy \quad (10)$$

L'entropie conditionnelle de X sachant Y est l'entropie moyenne de X conditionnellement à Y :

$$H(X|Y) = - \int \int \ln(f_{X|Y}(x|y)) f_{X,Y}(x, y) dx dy \quad (11)$$

où $f_{X|Y}(x|y)$ est la fonction de densité conditionnelle de X sachant $Y = y$. Il est facile de vérifier que $H(X|Y) = H(X, Y) - H(Y)$. Notons que $H(X|Y)$ est invariante en changeant ses arguments. Cependant, l'information mutuelle

$$\mathcal{I}(X, Y) = \int \int \ln \left(\frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)} \right) f_{X,Y}(x, y) dx dy \quad (12)$$

est une mesure de dépendance symétrique entre X et Y . L'information mutuelle mesure l'information moyenne contenue dans une *v.a* à propos de l'autre. La symé-

trie résulte directement de la définition et devient évidente après l'avoir exprimé en terme d'entropie : $\mathcal{I}(X, Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$. L'information mutuelle est invariante non seulement sous des transformations d'échelle de X et de Y , mais plus généralement sous toute transformation continue strictement monotone de X et de Y . L'information mutuelle n'est jamais négative, $\mathcal{I}(X, Y) \geq 0$ (on peut la montrer facilement en utilisant l'inégalité de Jensen), nous aurons l'égalité si et seulement si $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Cette propriété la rend une quantité utile pour tester les hypothèses d'indépendance. L'information mutuelle est un cas particulier de l'information de Kullback-Leibler définie pour deux fonctions de densité $f(x)$ et $g(x)$, respectivement, comme

$$\mathcal{I}_{KL} = \int \ln \left(\frac{f(x)}{g(x)} \right) f(x) dx \quad (13)$$

Robinson (1991) a proposé un test pour l'indépendance qui se base sur l'information de Kullback-Leibler, qui se réduit dans ce cas à l'information mutuelle.

Pour tester l'indépendance conditionnelle entre X et Y , sachant une troisième variable Z , il est utile de considérer l'information mutuelle conditionnelle définie par

$$\mathcal{I}(X, Y|Z) = \int \int \int \ln \left(\frac{f_{X|Y,Z}(x|y, z)}{f_{X|Z}(x|z)} \right) f_{X,Y,Z}(x, y, z) dx dy dz \quad (14)$$

L'information mutuelle conditionnelle quantifie la moyenne de l'information conditionnelle dans Y concernant X , sachant l'information sur X déjà contenue dans Z . Ceci peut être vu en l'exprimant comme $\mathcal{I}(X, Y|Z) = H(X|Z) - H(X|Y, Z) = -H(X, Y, Z) + H(X, Z) + H(Y, Z) - H(Z)$. Par ce que l'entropie n'est plus supérieure à celle conditionnée, $\mathcal{I}(X, Y|Z) \geq 0$, avec une égalité si et seulement si X et Y sont conditionnellement indépendants sachant Z .

Intégrales de corrélation

L'intégrale de corrélation généralisée d'ordre q de X est définie comme

$$C_q(X; \epsilon) = \left[\int \left(\int \mathbf{1}_{\|x-x'\| \leq \epsilon} f_X(x') dx' \right)^{q-1} f_X(x) dx \right]^{\frac{1}{q-1}} \quad (15)$$

où $\mathbf{1}_{(\cdot)}$ est la fonction indicatrice, et $\|\cdot\|$ dénote la norme L_∞

$$\|X\| = \sup_{i=1, \dots, \dim X} |x_i| \quad (16)$$

Le paramètre ϵ joue le rôle d'une fenêtre. L'estimation de l'intégrale de corrélation est souvent utilisée dans la théorie du chaos pour étudier les structures fractals et pour caractériser les séries temporelles déterministes. Les intégrales de corrélation sont aussi utiles pour tester l'indépendance sérielle, par ce que les intégrales de corrélation généralisées factorisent quand les éléments de X sont i.i.d. La factorisation pour $q = 2$ est utilisée dans les tests *BDS* pour l'indépendance basée sur $C_2(X; \epsilon)$.

Pour décrire la liaison entre les quantités théoriques d'information et les intégrales de corrélation, il est commode de noter que l'entropie de Shannon est un cas particulier de l'entropie généralisé, l'entropie de Renyi définie par

$$H_q(X) = -\frac{1}{q-1} \ln \int (f_X(x))^{q-1} f_X(x) dx \quad (17)$$

où q indique l'ordre de l'entropie de Renyi. En effet, En effet, en prenant la limite pour $q \rightarrow 1$, on obtient en utilisant règle de l'Hospital

$$\lim_{q \rightarrow 1} H_q(X) = -\frac{d}{dq} \Big|_{q=1} \int f_X(x)^{q-1} f_X(x) dx = -\int \ln f_X(x) f_X(x) dx \quad (18)$$

qui est l'entropie de Shannon. Prenons le logarithme de l'équation (15), obtenons

$$-\ln C_q(X; \epsilon) = -\frac{1}{q-1} \ln \left[\int \left(\int \mathbf{1}_{\|x-x'\| \leq \epsilon} f_X(x') dx' \right)^{q-1} f_X(x) dx \right] \quad (19)$$

qui est similaire à l'entropie généralisé de Renyi donnée dans l'équation (17), la seule différence est de remplacer $f_X(x)$ à l'intérieur des parenthèses par l'intégrale de $f_X(x')$ sur la boule de centre x et de rayon ϵ . L'intégrale intérieur de l'équation (15), se comporte comme $\epsilon^m f_X(x)$ pour un ϵ petit, où m est la dimension de X ¹⁸. Ainsi, à un facteur d'échelle dépend de ϵ , l'intégrale de corrélation va correspondre à l'intégrale dans l'équation (17). La relation entre $H_q(X)$ et $C_q(X; \epsilon)$ pour un ϵ petit est

$$H_q(X) \simeq -\ln C_q(X; \epsilon) + m \ln \epsilon \quad (20)$$

Ceci montre que l'intégrale de corrélation estimée fourni un estimateur non paramétrique de $H_q(x)$ et vice versa. Pour donner un exemple comment ceci mène à l'estimation des quantités théoriques de l'information, considérons $\mathcal{I}_q(X, Y)$ l'information mutuelle d'ordre q entre X et Y , donner par

$$\mathcal{I}_q(X, Y) = H_q(X) + H_q(Y) - H_q(X, Y)$$

Soit $\widehat{C}_q(X; \epsilon)$, $\widehat{C}_q(Y; \epsilon)$, $\widehat{C}_q(X, Y; \epsilon)$ les estimateurs des intégrales de corrélation, un estimateur de $\mathcal{I}_q(X, Y)$ est donné par

$$\widehat{\mathcal{I}}_q(X, Y) = \ln \widehat{C}_q(X, Y; \epsilon) - \ln \widehat{C}_q(X; \epsilon) - \ln \widehat{C}_q(Y; \epsilon) \quad (21)$$

Notons que les termes proportionnels à $m \ln \epsilon$ s'annulent car la dimension de (X, Y) est la somme de celle de X et de Y . Une opposition similaire se produit dans l'information mutuelle conditionnelle pour laquelle nous obtenons

¹⁸En appliquant le théorème de la moyenne.

$$\widehat{\mathcal{I}}_q(X, Y|Z) = \ln \widehat{C}_q(X, Y, Z; \epsilon) - \ln \widehat{C}_q(X, Z; \epsilon) - \ln \widehat{C}_q(Y, Z; \epsilon) + \ln \widehat{C}_q(Z; \epsilon) \quad (22)$$

Pour plus de détail sur la liaison entre les intégrales de corrélation et la théorie d'information voir Pichard et Theiler (1995).

Le choix de $q = 2$ est de beaucoup le plus utilisé dans l'analyse du chao, comme il est accordé aux algorithmes d'estimation efficace. L'information mutuelle conditionnelle $\mathcal{I}_q(X, Y|Z)$ n'est pas définie positive pour $q \neq 1$ (voir Dicks et Manzan, 2002). Ceci veut dire qu'il est possible de construire des exemples de *v.a* X et Y qui peuvent être dépendantes conditionnellement donnant une variables Z et pour laquelle $I_2(X, Y|Z)$ est nulle ou négative. Si $I_2(X, Y|Z)$ est nulle, le test basé sur \mathcal{I}_2 asymptotiquement, n'a pas de puissance contre les alternatives. Cette situation semble être très exceptionnelle, et généralement $\mathcal{I}_2(X, Y; \epsilon)$ est soit positive soit négative. Ceci suggère qu'un test unilatéral, rejetant pour $\mathcal{I}_2(X, Y; \epsilon)$ grande, n'est pas toujours optimal. Dans la pratique \mathcal{I}_2 se comporte comme \mathcal{I}_1 puisque, généralement nous observons une grande puissance pour les testes unilatéraux que pour les tests bilatéraux. Ceci nous conduit de prendre $q = 2$, ainsi que l'application des tests unilatéraux.

Pichard et Theiler (1995) utilisent le noyau de fonction indicatrice standard qu'il se pourrait bien ne pas être optimal au point de vu statistique. L'estimation en utilisant les U -statistiques donne dans le cas naturel le noyau de fonction indicatrice. Dans le contexte présent, nous n'avons aucune raison de supposer que la modification de l'intégrale de corrélation en utilisant un autre noyau conduira aux tests plus puissants. Ceci pourrait paraître non intuitif dans le cadre non paramétrique, comme il est connu que certains noyaux sont mieux que d'autre pour l'estimation des fonctions non paramétriques. Cependant il y a des différences importantes entre les problèmes de l'estimation de la fonction non paramétrique et les tests non paramétriques. Par

exemple, dans le problème d'estimation d'une fonction non paramétrique la fenêtre doit tendre vers zéro à un taux asymptotiquement adéquat pour obtenir la consistance. Pour l'analyse ceci n'a pas besoin d'être le cas comme il est illustré clairement par le fait que la théorie asymptotique pour les tests de *BDS* maintienne pour toute valeur fixée, il faut garder dans la tête que nous voulons estimer la fonctionnelle de la fonction densité et pas la densité elle-même.

Références

- [1] Ahmad, I. A and Lin, P. E. (1976). A nonparametric estimation of the entropy for absolutely continuous distribution. *IEEE Trans. Inform. Theory*, vol. IT-**22**, pp. 372-375.
- [2] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain ‘goodness of fit’ criteria based on stochastic processes. *Ann. Math. Statist.*, **23**, 193-212.
- [3] Aparicio, F. M. and Escribano, A. (1998). Information-theoretic analysis of serial dependence and cointegration. *Studies in Nonlinear Dynamics and Econometrics*, **3**, number 3, 119-140.
- [4] Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series : First order characterization and order determination. *Biometrika*, **77**, number 4, 669-687.
- [5] Avouyi- Dovi, S. , Jondeau, E. (1999). La modélisation de la volatilité des bourses asiatiques *Note d’étude de recherche. Banque de France*.
- [6] Barnard, G. A (1963). Discussion of a paper by M. S. Bartlett. *Journal of the Royal Statistical Society B*, **25**, 294.
- [7] Basharin, G. P. (1956). On statistical estimate for the entropy of a sequence of independent random variables, *Theory prob. Appl*, vol **4**, 333-336.
- [8] Berk, A. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist* , **37**, 51-58.
- [9] Besag, J., and P. J. Diggle (1977). Simple Monte Carlo tests for spatial pattern. *Applied Statistics*, **26**, 327–33.

-
- [10] Blom, G.(1976). Some properties of incomplete U -statistics, *Biometrika*, **63**, 573-580.
- [11] Blum, J.R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on sample distribution function. *Annals of Mathematical Statistics* **32**, 485-498.
- [12] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.
- [13] Bougerol, P. and Picard, N. (1992). Stationarity of *GARCH* processes and of some nonnegative time series. *Journal of Econometrics* **52**,115-27.
- [14] Brigé, L (1985). Estimation de l'entropie d'une densité, in Exposé Sém. d'Osay, Paris, France.
- [15] Brockwell, P. J and Davis, R. A. (1987). *Time Series : Theory and Methods*. New York : Springer-Verlag.
- [16] Brown, B. M. and Kildea, D. G. (1978). Reduced U -statistics and the Hodges-Lehmann estimator, *Ann. Statist.*, **6**, 828-835.
- [17] Cantelli, F.P. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giornale dell'Istituto Italiano degli Attuari* **4**, 421-424.
- [18] Carlstien, E. (1988). Degenerate U-statistics based on non-independent observations. *Calcutta Statist. Assoc. Bull.* **37**, 55-65.
- [19] Chan, N. H. and Tran, L.T. (1992). Nonparametric tests for serial dependence. *Journal of Time Series Analysis*, **13**, number 1, 19-28.
- [20] Chen, R. and Tsay, R.S. (1991) On the ergodicity of *TAR*(1) processes. *Ann. Appl. Probab.* **1**. 613-634.
- [21] Cheng, B. and Tong, H. (1991). On consistent nonparametric order determination and chaos. *Journal of the Royal Statistical Society B*, **45**, number 2, 427-449.

-
- [22] Cheng, B., and Robinson, P.M. (1991). Density estimation in strongly dependent non linear time series, *Statistica Sinica* **1**, 335-359.
- [23] Chung, K. L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans. Am. Math. Soc.*, **67**, 36-50.
- [24] DasGupta, A (2008). *Asymptotic theory of statistics and probability*. Springer.
- [25] Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Mult. Anal.* **11**, 102-113.
- [26] Deheuvels, P. (1992). Functional laws of the iterated logarithm for large increments of empirical and quantile processes. *Stochastic Processes Appl.*, **43**(1), 133-163.
- [27] Deheuvels, P. (2000). Chung type functional laws of the iterated logarithm for tail empirical processes. *Ann.I.H.P. B*, **36**, 583-616.
- [28] Deheuvels, P. et Mason, D. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Prob.*, **20**, 1248-1287.
- [29] Deheuvels, P. et Mason, D. (2004). General confidence bounds for nonparametric functional estimators. *Stat. Inf. for Stoch. Proc.*, **7**, 225-277.
- [30] Delgado, M. A (1996). Testing serial independence using the sample distribution function, *Journal of Time Series Analysis*, **17**, number 3, 271-285.
- [31] Demitriev, Y. G., and Tarasenko, F.P. (1973). On the estimation of functionals of the probability density and its derivatives. *Theory Prob. Appl.*, vol **18**, 628-633.
- [32] Devroye, L., and Györfi, L. (1985). *A test for normality based on sample entropy*; The L1 view. New York.
- [33] Diebold, F. X., T. Gunther, and A. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review*, **39**, 863-883.

-
- [34] Diks, C. (1999). consistent testing for serial independence. *Technical Report. CENDEF working paper 99-02, University of Amsterdam.*
- [35] Diks, C. and H. Tong (1999). A test for symmetries of multivariate probability distributions. *Biometrika*, **86**, 605–614.
- [36] Diks, C., and Manzan, S. (2002). Tests for serial independence and linearity based on correlation integrals, *Studies in Nonlinear Dynamics and Econometrics*, vol **6**, number 2, article 2.
- [37] Diks, C., and Panchenko, V. (2007). Nonparametric tests for serial independence based on quadratic forms. *Statistica Sinica* **17**, 81–97.
- [38] Dionísio, A., Menezes, R., and Mendes, D. A. (2006). Entropy based independence test, *Nonlinear Dynamics* **44**, 351–357.
- [39] Donsker, M.D. (1951). Justification and extension of Doob’s heuristic approach to the Kolmogorov-Smirnov theorems. *annals of Mathematical Statistics* **23**, 22-28.
- [40] Dudewicz, E. J., and van der Meulen, E.C. (1981). Entropy based tests of uniformity, *Journal of the American Statistical Association*, **76**, 967–974.
- [41] Dunford N., and Schwartz J. (1958). *Linear operators*. Part 1 : general theory. New York.
- [42] Einmahl, U. et Mason, D. (2000). An empirical process approach to the uniform consistency of kernel type estimators. *Journ. Theoretic. Probab.*, **13**, 1-13.
- [43] Einmahl, U. et Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.*, **33**(3), 1380-1403.
- [44] Engen, S. and M. Lillegård (1997). Stochastic simulations conditioned on sufficient statistics, *Biometrika*, **84**, 235–40.
- [45] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica* **50**, 987–1007.

-
- [46] Fan, J. and Q, Yao (2003) *Nonlinear Time Series*. Springer, New York.
- [47] Fernandes, M.(2001). Nonparametric entropy based tests of independence between stochastic processes, Working paper.
- [48] Fernandes, M., and Néri, B. (2010). Nonparametric entropy based tests of independence between stochastic processes, *Econometric Reviews*, **29**, 276-306.
- [49] Finkelstein, H. (1971). The law of iterated logarithm for empirical distributions. *Ann. Math. Stat.*, **42**, 607–625.
- [50] Gebhard Kirchgässner and Jürgen Wolters (2007). Introduction to the moderne time series.
- [51] Glivenko, V. (1933). Sulla determinazione empirica della leggi di probabilità. *Giornale dell'Istituto Italiano degli Attuari* **4**, 92-99.
- [52] Gokhale, D. V. (1983). On Entropy-Based Goodness-of-Fit Tests, *Computational Statistics and Data Analysis*, **1**, 157–165.
- [53] Granger, C. W. J., and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Gottingen.
- [54] Granger, C. W. and Lin, J. L. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, **15**, number 4, 371-384.
- [55] Hall, P. (1988). Estimating the Direction in Which a Data Set Is Most Interesting, *Probability Theory and Related Fields*, **80**, 51–77.
- [56] Halmos, P.R. (1946). The entropy of unbiased estimation, *Ann. Maths. Statist.*, **17**, 34-43.
- [57] Härdle, W. (1990). *Applied Nonparametric Regression*. New York : Cambridge University Press.
- [58] Hardy G.H, Littlwood J.E., and Polia G. (1967), *Inequalities*. New York, Cambridge Univ.

-
- [59] Hoeffding, W. (1948a). A class of statistics with asymptotically normal distribution, *Ann. Maths. Statist.*, **19**, 293-325.
- [60] Hoeffding, W. (1948b). A non-parametric test of independence. *Ann. Math. Statist.* **19**, 546-57.
- [61] Hoeffding, W. (1961). The strong law of large numbers for U -statistics. Univ. of North Carolina Institute of Statistics Mimeo series, No. 302.
- [62] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, **58**, 13-30.
- [63] Hong, Y. (1998). Testing for pairwise independence via the empirical distribution function, *Journal of the Royal Statistical Society. B*, **60**, 429-453.
- [64] Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function : A generalized spectral density approach, *Journal of the American Statistical Association*, vol. **94**, 1201-1220.
- [65] Hong, Y., and White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence, *Econometrica*, **73**, 837-901.
- [66] Hutcheson, K., and Shenton, L. R. (1974). Some moments of an estimate of Shannon's measure of information. *Comm. Stat.*, vol. **3**, 89-94.
- [67] Imhof, J.P. (1961). Computing The distribution of quadratic forms in normal variables. *Biometrika*, **48**, 417-426.
- [68] Joag-Dev, K. (1991). Measures of dependence. In *Handbook of Statistics 4. Nonparametric methods*, ed. P. R. Krishnaiah and P. K. Sen, 79-88. Amsterdam : North-Holland.
- [69] Joe, H. (1989a). Relative entropy measures of multivariate dependence, *Journal of the American Statistical Association*, **84**, 157-164.
- [70] Joe, H. (1989b). Estimation of entropy and other functionals of a multivariate density, *Annals of the Institute of Statistical Mathematics*, **41**, 683-697.

-
- [71] Kac, M. and Siegert, A. J. F. (1947). An explicit representation of stationary Gaussian process, *Ann. Math. Statist.* **18**, 438-442.
- [72] Kanwal. R. (1971). *Linear Integral Equations Theory and Technique*, Academic Press, New York.
- [73] Karlin. S. and Taylor. H. M. (1975). *A first course in stochastic processes*. Academic press, USA.
- [74] Kosorok, M R. (2006). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science+Business Media, INC.
- [75] Kullback, S. (1959). *Information Theory and statistics*. New York : Wiley & Sons.
- [76] Landesberg, P. T. (1985). Usage et limite du concept d'entropie, in : *Communications*, **41**, pp. 63 80.
- [77] Lay, D. C. (2003). *Linear algebra and its applications*. Pearson Education, 3rd edition.
- [78] Ling, S. (1999) On probability properties of a double threshold *ARMA* conditional heteroskedasticity model. *J. Appl. Probab.* **36**. 688-705.
- [79] Maassoumi, E. and J. Racine (2002). Entropy and predictability of stock returns, *Journal of Econometrics*, **107**, 175 189.
- [80] Matilla García, M. and M.R. Marín, (2008). A nonparametric independence test using permutation entropy, *Journal of Econometrics*, **144**(1), 139–155.
- [81] Mercer, J (1909). Functions of positive and negative type and their connection to the theory of integral equations, *Phil. Trans.Royal Soc. London A* **209**, 415, 446.
- [82] Mokkaem. A.(1989a). Estimation of the entropy and information of absolutely continuous random variables, *IEEE Transactions on Information Theory*, **35**, 193 196.

-
- [83] Mokkadem. A. (1989*b*). Critères de mélange pour des processus stationnaires. Estimation sous des hypothèses de mélange. Entropie des processus linéaires. Thèse doctorat d'état, Univ. Paris XI, Orsay, France.
- [84] Petrucci, J. D. et Woolford, S. W. (1984). A threshold $AR(1)$ model. *Journal of Applied Probability* **21**, 270–286.
- [85] Pham, D.T. (1985) Bilinear Markovian Representation of Bilinear Models. *Stochastic Processes and Their Applications* **20**, 295-306.
- [86] Pham, D. T. (1986). The mixing property of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and Their Applications* **23**, 291–300.
- [87] Prichard, D. and Theiler, J. (1995). Generalized redundancies for time series analysis. *Physica D*, **84**, 476-493.
- [88] Rao. B. L. S. P. (1983). *Nonparametric functional estimation*. Orlando, FL : Academic.
- [89] Rice, J. (1984). Boundary Modification for Kernel Regression, *Communications in Statistics*, Ser. A,), 893–900.
- [90] Robinson, P.M. (1991). Consistent nonparametric entropy-based testing. *Review of Economic Studies*, **58**, 437-453.
- [91] Rosenblatt. M. (1956). Remark on some nonparametric estimates of density function. *Ann. Math. Statist.*, vol **27**, pp. 832-835.
- [92] Rosenblatt. M. (1956). A central limit theorem and a strong mixing condition *Proc; Nat. Acad. Sci., U.S.A.*, vol **42**, pp. 43-47.
- [93] Rosenblatt. M. (1975). Quadratic Measure of Deviation of Two-Dimensional Density Estimates and a Test of Independence. *Annals of Statistics* **3**, 1-14.
- [94] Sanvi Avouyi-Dovi et Eric Jondeau (1999). La modélisation de la volatilité des bourses asiatiques. Notes d'étude et de recherche.

-
- [95] Savit, R. and Green, M. (1991). Time series and dependent variables. *Physica D*, **50**, number 50, i 95-116.
- [96] Scargle. J. D. (1981). Studies in astronomical time series analysis. I. Modeling random processes in the time domain. *Astrophys. J. Suppl. Ser.*, **45**, 1-75.
- [97] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.
- [98] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- [99] Sievers, G. L. (1978). Weighted rank statistics for simple linear regression, *J. Ann. Statist. Assoc.*, **73**, 628-631.
- [100] Shapiro, C. P. and Hupert, L. (1979). Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions, *Ann. Statist.*, **6**, 788-794.
- [101] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Application to Statistics*. Wiley New York.
- [102] Skaug, H. J. and Tjøstheim. D (1993a). A nonparametric test of serial dependence based on the empirical distribution function, *Biometrika*, **80**, 591-602.
- [103] Skaug, H. J. and Tjøstheim. D (1993b). Nonparametric tests of serial independence In *Developments in Time Series Analysis : In Honour of Maurice B. Priestley* (ed. T. Subba Rao), pp. 207-230. Chapman & Hall, London.
- [104] Skaug, H. J (1993c). The limit distribution of the Hoeffding statistic for test of serial independence. Unpublished report.
- [105] Skaug, H. J. and D. Tjøstheim, (1996). Measures of distance between densities with application to testing for serial independence, in *Time Series Analysis in memory of E. J. Hannan, P. Robinson and M. Rosenblatt* (eds.), Springer : New York, pp. 363-377.

-
- [106] Slepian, D. (1968). A numerical method for determining the eigenvalues and eigenfunctions of analytic kernels. *SIAMS J. NUMER. ANAL.* v **5**, N° 3, 568-600.
- [107] Stephens, M. A. (1976). Asymptotic results for goodness of fit statistics with unknown parameters, *Ann. Statist.*, **4**, 357-369.
- [108] Stute, W. and Schuman, G. (1980). The general Glivenko-Cantelli theorem for stationary sequences of random observations. *Scand. J. Statist.* **7**, 102-4.
- [109] Stute, W. (1982a). The law of the iterated logarithm for kernel density estimators. *Ann. Prob.*, **10**, 414-422.
- [110] Stute, W. (1982b). The oscillation behavior of empirical processes. *Ann. Prob.*, **10**, 86-107.
- [111] Stute, W. (1986a). Conditional empirical processes. *Ann. Prob.*, **14**(2), 638-647.
- [112] Stute, W. (1986b). On almost sure convergence of conditional empirical distribution functions. *Ann. Prob.*, **14**(3), 891-901.
- [113] Su, L. and White. H. (2007). A Consistent Characteristic Function-Based Test for Conditional Independence. *Journal of Econometrics* **141**, 807-834.
- [114] Su, L. and White. H. (2008). A Nonparametric Hellinger Metric Test for Conditional Independence *Econometric Theory* **24**, 829-864.
- [115] Tjøstheim, D. (1996). Measures and Tests of Independence : A Survey, *Statistics*, **28**, 249-284.
- [116] Tjøstheim, D and Hufthammer, H.O (2013). Local Gaussian correlation : A new measure of dependence. *Journal of Econometrics* **172**, 33-48
- [117] Tong, H. (1978). On a threshold model. In C. H. Chen (ed.), *Pattern Recognition and Signal Processing. Sijhoff & Noordhoff, Amsterdam*.
- [118] Tong, H. (1990). *Non-Linear Time Series : A Dynamical System Approach*. Oxford University Press, Oxford, UK.

-
- [119] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics*, **52**, 479-487.
- [120] Tschernig, R. and Yang, L. (2000). Nonparametric lags selection for time series. *Journal of Time Series Analysis*, **21**, number 4, 457-487.
- [121] Ullah, A. (1993). Entropy divergence and distance measures with econometric applications, *Journal of Statistical Planning and Inference*, **49**, 137-162.
- [122] Vapnik, V. (1982). *Estimation of dependences*. Springer-Verlag, N.Y.
- [123] Vapnik, V. and Červonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probab. Appl.*, **16**, 264-280.
- [124] Vapnik, V. and Červonenkis, A. Y. (1981). Necessary and sufficient conditions for uniform convergence of means to mathematical expectations. *Theor. Probab. Appl.*, **26**, 147-163.
- [125] van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes*. Springer, New York.
- [126] van der Vaart, A. (1998). *Asymptotic statistic*. Cambridge University Press.
- [127] Vasicek, O. (1976). A Test for Normality Based on Sample Entropy, *Journal of the Royal Statistical Society, Ser. B*, **38**, 54-59.
- [128] Veysseyre, R. (2006). *Statistique et probabilité ppxe*, PhD thesis, Department of Mathematics, University of California, San Diego.
- [129] Watson, G. S. (1961). Goodness-of-fit tests on a circle, *Biometrika*, **48**, 109-114.
- [130] White, H. (1982). Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, **50**, 1-25.
- [131] Wu, K., Savit, R. and Brock, W. (1993). Statistical tests for deterministic effects in broad band time series. *Physica D*, **69**, 172-188.

- [132] Zakoian, J. M. (1994). Threshold heteroscedastic models. *Journal of Economic Dynamics and Control* **18**, 931–955.
- [133] Zheng, X. (2000). A Consistent Test of Conditional Parametric Distributions, *Econometric Theory*, **16**, 667–691.