

Université de Montréal

**Estimation des autocovariances et des
autocorrélations d'un modèle saisonnier**

par

Alain Latour

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en informatique

(Février 1986)

© Alain Latour, 1986

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée

Estimation des autocovariances et des autocorrélations d'un modèle saisonnier

présentée par

Alain Latour

a été évaluée par un jury composé des personnes suivantes :

Robert Cléroux

(président-rapporteur)

Roch Roy

(directeur de recherche)

Urs Mag

(membre du jury)

Guy Mélard

(examineur externe)

Marc Moore

(représentant du doyen de la FESP)

Table des matières

Sommaire	ix
Summary	xiii
Liste des tableaux	xvii
Liste des figures	xix
Préface	xxi
Préface	xxi
Chapitre 1. Introduction	1
Résumé	1
1.1. Quelques notations	1
1.2. Résultats relatifs aux deux premiers moments	3
1.3. Résultats relatifs aux distributions asymptotiques	5
Chapitre 2. Marche aléatoire saisonnière	9
Résumé	9
2.1. Introduction	9
2.2. Deux premiers moments de la marche aléatoire saisonnière	10
2.3. Autocovariances échantillonnales avec données centrées	18
2.3.1. Statistiques usuelles	18
2.3.2. Espérance et variance des autocovariances	20
2.4. Généralisation à une transformation linéaire de la marche aléatoire saisonnière	24
2.4.1. Espérance de l'autocovariance	27

2.4.2.	Covariance entre deux autocovariances	31
2.4.3.	Variance de la covariance échantillonnale	33
2.5.	Covariance entre deux autocovariances échantillonnales de délais multiples de la période	35
2.5.1.	Contribution des blocs diagonaux	36
2.5.2.	Contribution des blocs de la partie triangulaire supérieure	37
2.5.3.	Contribution des blocs de la partie triangulaire inférieure	39
2.6.	Autocovariances avec des données non centrées	43
2.7.	Approximation des deux premiers moments de $r(ks)$	45
Chapitre 3.	Théorème de représentation	47
	Résumé	47
3.1.	Introduction	47
3.2.	Représentation d'un processus saisonnier	52
3.3.	Autocovariances échantillonnales du processus saisonnier général	54
3.3.1.	Espérance asymptotique des autocovariances d'échantillonnales	54
3.3.2.	Covariance asymptotique entre deux autocovariances échantillonnales	59
3.3.3.	Autocovariances avec données non centrées	63
Chapitre 4.	Les résultats de Dickey	65
	Résumé	65
4.1.	Introduction	65
4.2.	Variables de base	66
4.2.1.	Diagonalisation de Γ_n	66
4.2.2.	Structure de T_n	71
4.2.3.	Structure de W_n	72
4.2.4.	Distribution asymptotique de Γ_n , T_n et W_n	73
Chapitre 5.	Distributions asymptotiques	75
	Résumé	75
5.1.	Introduction	75
5.2.	Distribution asymptotique de $ns(1 - r'(ds))$ dans le contexte de la marche aléatoire saisonnière	76

5.2.1.	Dénominateur de $1 - r'(ds)$	77
5.2.2.	Numérateur de $ns(1 - r'(ds))$	78
5.3.	Distribution asymptotique de $ns(1 - r(ds))$ dans le contexte du modèle plus général	85
5.3.1.	Numérateur de $ns(1 - r(ds))$	85
5.3.2.	Dénominateur de $1 - r(ds)$	101
5.3.3.	Distribution asymptotique	104
5.4.	Distribution asymptotique de $ns(1 - r(ds))$ dans le contexte de la marche aléatoire saisonnière	112
5.5.	Distribution asymptotique de $ns(1 - r'(ds))$ dans le contexte du modèle plus général	115
5.6.	Distribution asymptotique des autocovariances de la marche aléatoire saisonnière	116
5.6.1.	Avec observations centrées	116
5.6.2.	Avec observations non centrées	118
5.7.	Quelques comparaisons	119
5.7.1.	Processus autorégressif d'ordre 1	119
5.7.2.	Modèle saisonnier simple	120
Chapitre 6. Résultats numériques		123
	Résumé	123
6.1.	Introduction	123
6.2.	Détails de la mise en œuvre	125
6.2.1.	Générateur aléatoire de variables normales	125
6.2.2.	Simulation des distributions asymptotiques	126
6.3.	Programmes de simulation	130
6.4.	Résultats des simulations	130
Annexe A. Utilisation de MacSyma		151
A.1.	Description du logiciel	151
A.2.	Exemple de séance MacSyma	152
A.3.	Autres programmes MacSyma	158

A.3.1. ProgA	158
A.3.2. ProgB.....	161
Annexe B. Programmes reliés à la simulation	171
B.1. Description du logiciel.....	171
B.2. Programmes préliminaires.....	171
B.2.1. Programme ConsDef.....	171
B.2.2. Le program ConsDick.....	182
B.3. Programmes de simulation	188
B.3.1. Loi asymptotique avec données centrées	188
B.3.2. Loi asymptotique avec données non centrées.....	198
B.3.3. Loi exacte avec données centrées	199
B.3.4. Loi exacte avec données non centrées.....	200
Annexe C. Autre approche pour la simulation des lois asymptotiques	203
C.1. Description du logiciel.....	203
C.2. Moments et distributions.....	203
C.2.1. La variable T_r	204
C.2.2. La variable W_r	204
C.2.3. La variable Γ_r	205
C.3. Simulation du triplet $(T_r, W_r, \Gamma_r)'$	206
Remerciements.....	211
Bibliographie	213
Liste des ouvrages cités	213

Sommaire

L'ouvrage de BOX et JENKINS (1976) a eu un impact considérable dans l'application et le développement de techniques de séries chronologiques. L'approche temporelle qu'ils suggèrent pour l'étude d'une série est un processus itératif comportant trois étapes. La première consiste à identifier un modèle permettant de décrire la série, et la seconde à estimer les paramètres du modèle identifié. À la troisième étape, on vérifie si les propriétés du modèle ajusté sont en accord avec les propriétés connues des observations. Si le modèle semble adéquat, on peut l'utiliser à des fins de prévision ; sinon, il faut reprendre le processus à la première étape.

C'est à l'étape d'identification que le statisticien analyse les autocovariances et les autocorrélations échantillonnales. Dans le cas des séries stationnaires, les propriétés de ces statistiques sont relativement bien connues, comme le prouvent ANDERSON (1971), FULLER (1976) et PRIESTLEY (1981). Notons cependant que la majorité de ces résultats sont asymptotiques. En ce qui concerne les séries non stationnaires, peu de résultats étaient connus au début de la seconde moitié des années soixante-dix. Notre travail se situe dans le cadre de l'étude des propriétés des statistiques usuelles calculées à partir de données non stationnaires.

Les résultats relatifs aux autocovariances et aux autocorrélations échantillonnales sont présentés en deux volets. Dans l'un d'eux, nous considérons les statistiques calculées avec des données centrées, tandis que dans l'autre, les données ne sont pas centrées. Notre démarche se fait en cinq étapes correspondant aux différents chapitres de cette thèse.

Au chapitre deux, nous introduisons la notion de promenade ou marche aléatoire saisonnière. Il s'agit essentiellement d'un processus stochastique pour lequel une différence d'ordre s produit un bruit blanc. Nous déterminons dans un premier temps les deux premiers moments de ce processus. Nous nous intéressons par la suite aux autocovariances échantillonnales avec les données centrées. Nous obtenons une formule exacte pour l'espérance de l'autocovariance de délai k (dénotée par $c(k)$),

avec $k \geq 0$, ainsi qu'une formule exacte pour la variance d'une autocovariance dont le délai est un multiple de la période (s).

Nous généralisons ces résultats à une transformation linéaire spécifique de la marche aléatoire saisonnière. L'intérêt de cette transformation devient évident au chapitre trois lorsque nous obtenons les propriétés des statistiques descriptives usuelles dans un contexte plus général où une différence d'ordre s produit un processus linéaire stationnaire. L'avant-dernière section du chapitre deux traite du cas où nous calculons les autocovariances avec les données non centrées (dénotées par $c'(k)$).

Pour terminer le chapitre deux, nous obtenons une approximation de l'espérance et de la variance d'un coefficient d'autocorrélation dont le délai est un multiple de la période via les formules de Kendall pour le calcul approché des moments d'un rapport de formes quadratiques. L'erreur de ces approximations est de l'ordre de $1/N$, N étant la longueur de la série.

Signalons qu'un bon nombre de calculs du chapitre deux ont nécessité l'utilisation du programme de calcul symbolique MacSyma.

Au chapitre trois, nous étudions un processus saisonnier plus général pour lequel une différence d'ordre s engendre un processus linéaire stationnaire. Dans un premier temps, nous montrons que le processus saisonnier est à toutes fins pratiques la somme d'une transformation linéaire spécifique de la marche aléatoire saisonnière et d'un processus stationnaire. Il s'agit d'une généralisation d'un résultat utilisé par ROY et LEFRANÇOIS (1981) ainsi que par HASZA (1980). Nous explicitons alors les deux premiers moments asymptotiques des autocovariances échantillonnales préalablement divisées par la longueur de la série. Nous terminons le chapitre trois en indiquant ce qu'il advient de ces résultats lorsque nous utilisons des autocovariances échantillonnales avec des données non centrées.

Certains résultats de DICKEY (1976), utilisés au chapitre cinq pour obtenir la distribution asymptotique d'une fonction simple de l'autocorrélation de délai ds , sont résumés au chapitre quatre. Deux coefficients d'autocorrélation sont envisagés : $r(k)$, calculé à partir des données centrées, et $r'(k)$ calculé avec des données non centrées.

Nous savons que dans le cas d'un processus stationnaire de moyenne nulle, $r(k)$ et $r'(k)$ ont la même distribution asymptotique. En mettant en évidence la distribution asymptotique de $N(1 - r(ds))$ et celle de $N(1 - r'(ds))$ pour $N = ns$ et $d \geq 0$, nous constatons que dans le cas de données non stationnaires, il existe des différences structurelles entre celles-ci. Nous obtenons ces résultats autant pour la promenade aléatoire saisonnière que pour le processus saisonnier plus général. Toujours dans le

contexte de la marche aléatoire saisonnière, les distributions asymptotiques de $c(k)$ et de $c'(k)$ sont explicitées lorsque k est un multiple de la période.

Nous terminons avec le chapitre six où les différentes distributions rencontrées dans les premiers chapitres sont évaluées par la méthode de Monte Carlo. Ces simulations permettent au lecteur de juger de l'impact de l'utilisation de données centrées ou non centrées dans le calcul des autocorrélations, tant au niveau des moments exacts et asymptotiques qu'au niveau des distributions. Elles nous donnent aussi une idée de la vitesse de convergence des lois exactes vers les lois asymptotiques.

Summary

The work by BOX et JENKINS (1976) had a considerable impact on the application and development of time series techniques. The temporal approach they suggest for studying a series is an iterative process consisting of three steps. The first step is to identify a model that describes the series, and the second step is to estimate the parameters of the identified model. In the third step, one checks if the properties of the fitted model align with the known properties of the observations. If the model appears adequate, it can be used for forecasting; otherwise, the process must be restarted from the first step.

It is at the identification stage that the statistician analyzes sample autocovariances and autocorrelations. In the case of stationary series, the properties of these statistics are relatively well-known, as evidenced by ANDERSON (1971), FULLER (1976), and PRIESTLEY (1981). However, most of these results are asymptotic. Regarding non-stationary series, few results were known at the beginning of the second half of the 1970s. Our work is situated within the study of the properties of standard statistics calculated from non-stationary data.

The results related to sample autocovariances and autocorrelations are presented in two parts. In one part, we consider statistics calculated with centered data, while in the other part, the data are not centered. Our approach is divided into five steps corresponding to the different chapters of this thesis.

In Chapter Two, we introduce the notion of seasonal random walk. This is essentially a stochastic process for which a difference of order s produces white noise. We first determine the first two moments of this process. We then focus on sample autocovariances with centered data. We obtain an exact formula for the expectation of the autocovariance of lag k (denoted by $c(k)$), with $k \geq 0$, as well as an exact formula for the variance of an autocovariance with lag being a multiple of the period (s).

We generalize these results to a specific linear transformation of the seasonal random walk. The interest of this transformation becomes evident in Chapter Three when we obtain the properties of usual descriptive statistics in a more general context where a difference of order s produces a stationary linear process. The penultimate section of Chapter Two deals with the case where we calculate autocovariances with non-centered data (denoted by $c'(k)$).

To conclude Chapter Two, we obtain an approximation of the expectation and variance of an autocorrelation coefficient with a lag that is a multiple of the period using Kendall's formulas for the approximate calculation of moments of a ratio of quadratic forms. The error of these approximations is of the order of $1/N$, where N is the length of the series.

It should be noted that many of the calculations in Chapter Two required the use of the symbolic computation program MacSyma.

In Chapter Three, we study a more general seasonal process for which a difference of order s results in a stationary linear process. Initially, we show that the seasonal process is, for all practical purposes, the sum of a specific linear transformation of the seasonal random walk and a stationary process. This generalizes a result used by ROY et LEFRANÇOIS (1981) and HASZA (1980). We then explicitly describe the asymptotic first two moments of the sample autocovariances divided by the length of the series. We finish Chapter Three by explaining what happens to these results when using sample autocovariances with non-centered data.

Some results from DICKEY (1976), used in Chapter Five to obtain the asymptotic distribution of a simple function of the autocorrelation with delay ds , are summarized in Chapter Four. Two autocorrelation coefficients are considered: $r(k)$, calculated from centered data, and $r'(k)$ calculated from non-centered data.

We know that in the case of a stationary process with zero mean, $r(k)$ and $r'(k)$ have the same asymptotic distribution. By highlighting the asymptotic distribution of $N(1 - r(ds))$ and that of $N(1 - r'(ds))$ for $N = ns$ and $d \geq 0$, we observe that, in the case of non-stationary data, there are structural differences between them. We obtain these results for both the seasonal random walk and the more general seasonal process. Still within the context of the seasonal random walk, the asymptotic distributions of $c(k)$ and $c'(k)$ are detailed when k is a multiple of the period.

We conclude with Chapter Six, where the various distributions encountered in the previous chapters are evaluated using Monte Carlo methods. These simulations allow the reader to assess the impact of using centered versus non-centered data in

calculating autocorrelations, both in terms of exact and asymptotic moments, as well as distributions. They also provide insight into the speed of convergence of exact laws to asymptotic laws.

Liste des tableaux

6.1	Nombre de délais considérés pour les différentes valeurs de N et s	125
6.2	Table de constantes nécessaires à l'exécution de l'algorithme M de Knuth	126
6.3	Série $\sum_{i=1}^{\infty} \gamma_i^k$ pour différentes valeurs de k où $\gamma_i = \frac{2(-1)^{i+1}}{\pi(2i-1)}$	126
6.4	Les deux premiers moments des distributions asymptotiques	134
6.5	Percentiles des distributions asymptotiques	134
6.6	Les deux premiers moments des distributions exactes et asymptotiques de $N(1 - r(ds))$ et de $N(1 - r'(ds))$	135
6.7	Percentiles des distributions exactes et asymptotiques de $N(1 - r(ds))$ pour $s = 1$	136
6.8	Percentiles des distributions exactes et asymptotiques de $N(1 - r'(ds))$ pour $s = 1$	137
6.9	Percentiles des distributions exactes et asymptotiques de $N(1 - r(ds))$ et $N(1 - r'(ds))$ pour $s = 4$	138
6.10	Percentiles des distributions exactes et asymptotiques de $N(1 - r(ds))$ et $N(1 - r'(ds))$ pour $s = 12$	139

Liste des figures

1	Échantillon du texte produit en 1986	xxii
2	Exemple de graphique produit en 1986	xxiii
5.1	Matrice \mathbf{P} du lemme 5.3.....	83
6.1	Distribution de $R(d, s)$ lorsque $s = 1$	131
6.2	Distribution de $R'(d, s)$ lorsque $s = 1$	131
6.3	Distribution de $R(d, s)$ lorsque $s = 4$	132
6.4	Distribution de $R'(d, s)$ lorsque $s = 4$	132
6.5	Distribution de $R(d, s)$ lorsque $s = 12$	133
6.6	Distribution de $R'(d, s)$ lorsque $s = 12$	133
6.7	Distribution de $N(1 - r(ds))$ lorsque $s = 1$ et $N = 60$	141
6.8	Distribution de $N(1 - r'(ds))$ lorsque $s = 1$ et $N = 60$	142
6.9	Distribution de $N(1 - r(ds))$ lorsque $s = 1$ et $N = 108$	142
6.10	Distribution de $N(1 - r'(ds))$ lorsque $s = 1$ et $N = 108$	143
6.11	Distribution de $N(1 - r(ds))$ lorsque $s = 1$ et $N = 204$	143
6.12	Distribution de $N(1 - r'(ds))$ lorsque $s = 1$ et $N = 204$	144
6.13	Distribution de $N(1 - r(ds))$ lorsque $s = 4$ et $N = 60$	144
6.14	Distribution de $N(1 - r'(ds))$ lorsque $s = 4$ et $N = 60$	145
6.15	Distribution de $N(1 - r(ds))$ lorsque $s = 4$ et $N = 108$	145
6.16	Distribution de $N(1 - r'(ds))$ lorsque $s = 4$ et $N = 108$	146
6.17	Distribution de $N(1 - r(ds))$ lorsque $s = 4$ et $N = 204$	146
6.18	Distribution de $N(1 - r'(ds))$ lorsque $s = 4$ et $N = 204$	147
6.19	Distribution de $N(1 - r(ds))$ lorsque $s = 12$ et $N = 60$	147

6.20	Distribution de $N(1 - r'(ds))$ lorsque $s = 12$ et $N = 60$	148
6.21	Distribution de $N(1 - r(ds))$ lorsque $s = 12$ et $N = 108$	148
6.22	Distribution de $N(1 - r'(ds))$ lorsque $s = 12$ et $N = 108$	149
6.23	Distribution de $N(1 - r(ds))$ lorsque $s = 12$ et $N = 204$	149
6.24	Distribution de $N(1 - r'(ds))$ lorsque $s = 12$ et $N = 204$	150

Préface

Політика — це не точна наука. Ось
чому в школі я обожнював математику.
У математиці для мене все було
зрозуміло.

Володимир Зеленський^a

*a. La politique n'est pas une science
exacte. C'est pourquoi, à l'école, j'adorais les
mathématiques.*

Tout en mathématiques était clair pour moi.

Volodymyr Ze-
lensky

En triant mes archives, j'ai retrouvé une copie de ma thèse de doctorat, rédigée en 1986, il y a près de 40 ans. La redécouverte de ce document, aux pages légèrement jaunies, m'a replongé dans une époque révolue, pleine de souvenirs. Tout comme mon mémoire de maîtrise (LATOURE (2025)), cette thèse représente une étape clé dans mes premiers pas en mathématiques et en informatique. Je me suis remémoré la richesse de cette période, marquée autant par l'acquisition de nouvelles connaissances que par la consolidation de celles déjà acquises.

En 1978, à l'UQAM, le mémoire numéro 383 (LATOURE (1978)) a marqué un tournant décisif dans l'utilisation de la mise en page assistée par ordinateur ; sa préface décrit l'état du traitement de texte à cette époque. L'arrivée des Macintosh, en janvier 1984, a transformé l'accès aux outils de traitement de texte. Ces ordinateurs compacts, dont la base mesurait à peine plus qu'une feuille A4¹, pouvaient exécuter des programmes avancés, inaugurant ainsi une nouvelle ère technologique. Qui aurait pu prédire, à l'époque, l'immense impact de ces innovations sur la production de documents académiques et scientifiques ?

1. En réalité, la base de cet ordinateur mesurait 24,4 cm × 27,5 cm, pour une hauteur de 34,5 cm.

Corollaire 2.1: *Dans le contexte de la proposition 2.1*

$$\begin{aligned}
 1) \lim_{t \rightarrow \infty} \text{Var}[X_t] &= \infty, \\
 2) \lim_{t \rightarrow \infty} \text{Cov}[X_t, X_{t'}] &= \begin{cases} 0 & r \neq r', \\ \infty & r = r' \text{ et } \nu \leq \nu' \end{cases} \\
 3) \rho(X_t, X_{t'}) &= \begin{cases} 0 & r \neq r', \\ \{(\nu + 1)/(\nu' + 1)\}^{1/2} & r = r' \text{ et } \nu \leq \nu'. \end{cases}
 \end{aligned}$$

Supposons que $\nu' \geq \nu$, et posons $\nu' - \nu = m$. On note alors que pour ν fixé,

$$\lim_{m \rightarrow \infty} \rho(X_t, X_{t'}) = 0,$$

tandis que pour m fixé,

$$\lim_{\nu \rightarrow \infty} \rho(X_t, X_{t'}) = \begin{cases} 0 & r \neq r', \\ 1 & r = r'. \end{cases}$$

Figure 1 – Échantillon du texte produit en 1986

L'utilisation de ces machines, bien que facilitant la mise en page, nécessitait encore des efforts considérables, notamment pour la gestion des formules mathématiques, qu'il fallait définir manuellement. Malgré ces défis, le résultat final était un document de qualité tout à fait acceptable pour l'époque.

Un extrait du document de 1986 est présenté dans la Figure 1. Vous pouvez le comparer à la page 12 de cette nouvelle édition, et constater l'énorme différence avec la mise en page actuelle réalisée avec L^AT_EX. Les outils modernes ont considérablement évolué, offrant des possibilités inimaginables à l'époque.

Aujourd'hui, une panoplie d'outils à la pointe de la technologie est à notre disposition. Des environnements comme R-Studio permettent de créer en quelques minutes des graphiques d'une grande précision esthétique.

Au chapitre 6, les graphiques ont été mis à jour grâce à de nouvelles simulations. Pour assurer une comparaison avec la version originale, la figure 6.7 de la page 141 remplace celle présentée à la figure 2.

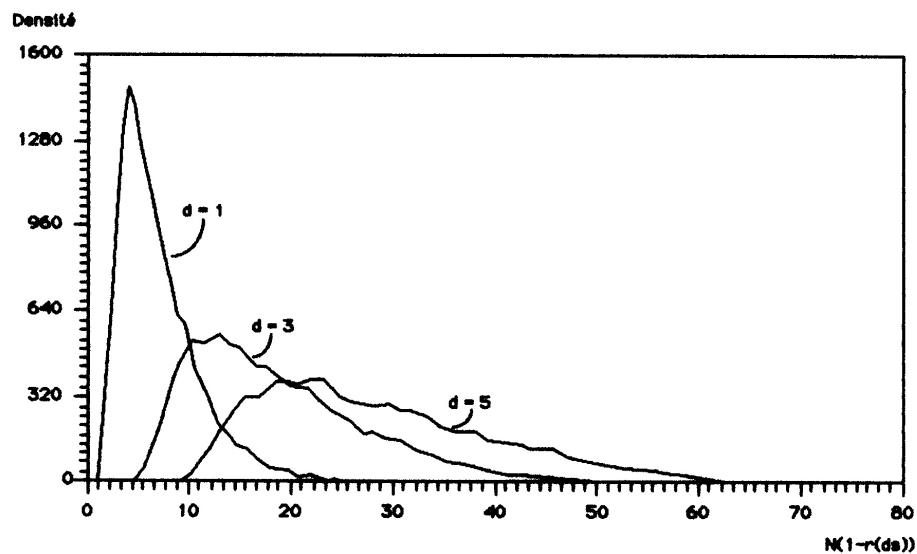


Figure 2 – Exemple de graphique produit en 1986

Ces mises à jour ont été réalisées à l’aide de la fonction `density` de R, qui permet d’estimer la densité de probabilité d’une distribution à partir d’un échantillon de données. Cette fonction applique un lissage par noyau afin de produire une courbe continue et lissée représentant la distribution des valeurs.

L’intelligence artificielle joue un rôle de plus en plus central dans la recherche, et FERRY (2025) propose une réflexion approfondie sur cette évolution. Dans cette perspective, le code utilisé pour générer ces graphiques a été développé au cours d’un échange avec ChatGPT². Lors de cette interaction, nous avons décrit le contexte et précisé nos attentes pour orienter la conception du script.

L’importation de données depuis des fichiers Excel se fait sans difficulté et dans un format impeccable. Même les programmes sources, qu’il s’agisse de Fortran ou d’autres langages, peuvent être formatés de manière attrayante. Ces innovations, qui auraient été des trésors dans les années 70 ou 80, ont considérablement transformé et simplifié le travail des chercheurs. Nous en sommes aujourd’hui à la pratique de la *recherche reproductible* définie comme suit :

La recherche reproductible désigne une pratique scientifique essentielle qui consiste à s’assurer que les résultats d’une étude peuvent être reproduits en suivant la même méthodologie, les mêmes données et les mêmes analyses que celles décrites par les auteurs d’origine. Cela

2. ChatGPT est un modèle d’intelligence artificielle développé par OPENAI (2024), conçu pour interpréter et produire du texte de manière fluide, répondre à des questions et assister dans diverses tâches rédactionnelles et analytiques.

*permet de vérifier la validité des conclusions et de renforcer la confiance dans les découvertes scientifiques.*³

Ma passion pour les mathématiques et l'informatique, qui a marqué cette période particulièrement enrichissante, n'a fait que croître avec le temps. J'ai eu la chance d'étudier à une époque où l'éducation représentait encore un puissant moteur de mobilité sociale. Cet amour de l'apprentissage a jalonné toute ma carrière, et c'est avec une immense satisfaction que je continue à espérer le transmettre aujourd'hui. Et pour reprendre, avec une touche d'humour, l'esprit de Platon :

- À quoi bon, Alain, apprendre à faire des maths, puisque tu vas mourir ?
- À connaître le plaisir de faire des maths avant de mourir.

Selon ChatGPT OPENAI (2024), Aristote aurait exprimé l'idée que

« c'est dans le cheminement, la quête de compréhension et la résolution des problèmes que l'esprit humain trouve son épanouissement, bien plus que dans la simple possession de savoirs acquis. »

C'est avec cet état d'esprit que j'ai toujours envisagé mon rôle d'enseignant. À ce titre, je me permets de citer Aristophane⁴ :

« Enseigner, ce n'est pas remplir un vase, c'est allumer un feu... »

Comme indiqué en introduction de cette préface, j'ai retrouvé une photocopie du document de 1986. Malheureusement, les fichiers originaux, enregistrés sur des disquettes de 3,5 pouces, étaient devenus inaccessibles aux logiciels modernes. Le document a donc été numérisé à l'aide d'outils courants, produisant des fichiers PDF. Cependant, la reconnaissance optique des caractères n'était pas parfaite. L'objectif ultime était de créer une version en \LaTeX du texte. À ma grande surprise, des applications récentes, telles que MathPix 3.4.11, permettent désormais de générer directement un document \LaTeX à partir d'un PDF.

MathPix utilise l'intelligence artificielle pour accomplir cette tâche. Cette application analyse les images contenant des équations manuscrites ou imprimées et en extrait le contenu mathématique avec une grande précision. Sa technologie, connue sous le nom de *reconnaissance optique des mathématiques*, est spécialement conçue pour interpréter des symboles mathématiques complexes, tels que les intégrales, fractions, racines carrées, etc.

3. Définition donnée par ChatGPT.

4. Aristophane : 445 à 386 av. J.-C.

Comme complément, il est également possible de recourir à ChatGPT pour reformater certaines formules, réviser le texte, ou corriger les erreurs d'orthographe et de grammaire. Le résultat final répond pleinement à nos attentes. Ce processus illustre de manière exemplaire une application réussie de l'intelligence artificielle.

J'espère que reproduire cette thèse après tant d'années, en utilisant des outils contemporains, saura inspirer les esprits curieux et illustrer le remarquable progrès des technologies de production textuelle.

A. Latour – 18 avril 2025

Chapitre 1

Introduction

Résumé

Dans ce chapitre, nous introduisons les concepts fondamentaux et les notations nécessaires à l'étude des propriétés distributionnelles des autocovariances et des autocorrélations échantillonnales. Nous débutons par une revue succincte des résultats connus dans la littérature, en particulier ceux relatifs aux processus autorégressifs à moyenne mobile intégrés (ARIMA), et nous précisons les définitions des autocovariances et des coefficients d'autocorrélation. Cette mise en contexte nous permet de situer notre travail dans le cadre des développements récents sur les modèles de séries chronologiques, notamment les modèles saisonniers.

Nous présentons ensuite une synthèse des principaux résultats relatifs aux deux premiers moments des autocovariances et des autocorrélations échantillonnales pour certains modèles spécifiques, notamment les processus ARIMA et leurs généralisations aux processus saisonniers. Enfin, nous introduisons les résultats relatifs aux distributions asymptotiques de ces quantités, en nous appuyant sur les travaux de référence dans le domaine. Ce chapitre jette ainsi les bases des développements ultérieurs en mettant en évidence les résultats théoriques qui serviront d'appui à nos propres contributions.

1.1. Quelques notations

Afin de présenter le problème que nous traiterons, nous situons notre recherche dans son contexte historique en faisant une brève revue des résultats connus. Cette revue n'est pas exhaustive; elle se veut plutôt indicatrice du développement des

dernières années relativement aux propriétés distributionnelles des autocovariances et des autocorrélations échantillonnales. Auparavant, introduisons quelques notations.

Soit Y_1, \dots, Y_N , une série chronologique. Nous désignerons l'autocovariance de délai k calculée à partir de données centrées par $c(k)$, c'est-à-dire,

$$c(k) = \frac{1}{N} \sum_{i=1}^{N-k} (Y_i - \bar{Y}) (Y_{i+k} - \bar{Y}), \quad 0 \leq k \leq N-1$$

où $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$ est la moyenne arithmétique de la série. L'autocovariance calculée à partir de données non centrées est

$$c'(k) = \frac{1}{N} \sum_{i=1}^{N-k} Y_i Y_{i+k}, \quad 0 \leq k \leq N-1$$

Ces deux autocovariances donnent lieu à deux coefficients d'autocorrélation, à savoir :

$$r(k) = \frac{\sum_{i=1}^{N-k} (Y_i - \bar{Y}) (Y_{i+k} - \bar{Y})}{\sum_{i=1}^N (Y_i - \bar{Y})^2}, \quad 1 \leq k \leq N-1 \quad (1.1)$$

et

$$r'(k) = \frac{\sum_{i=1}^{N-k} Y_i Y_{i+k}}{\sum_{i=1}^N Y_i^2}, \quad 1 \leq k \leq N-1$$

BOX et JENKINS (1976) ont introduit la notation $\text{ARIMA}(p, d, q)$ pour désigner un processus autorégressif à moyenne mobile intégré, c'est-à-dire, un processus de la forme $\{Y_t : t \in \mathbb{N}\}$ admettant la représentation

$$\Phi(B)\nabla^d Y_t = \theta_0 + \Theta(B)e_t$$

où B est l'opérateur retard, c'est-à-dire, $B[Y_t] = Y_{t-1}$; $\nabla = \mathbf{1} - B$, $\mathbf{1}$ étant l'opérateur identité, et où

$$\Phi(B) = \mathbf{1} - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

et

$$\Theta(B) = \mathbf{1} - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

sont respectivement l'opérateur autorégressif et l'opérateur moyenne mobile. Le paramètre θ_0 est un paramètre de tendance. La suite $\{e_t\}_{t \geq 0}$ est une suite de variables aléatoires indépendantes identiquement distribuées de moyenne nulle et de variance σ^2 . On retrouve dans MÉLARD (1985) une synthèse intéressante traitant de ces processus.

1.2. Résultats relatifs aux deux premiers moments

WICHERN (1973), ainsi que (BOX et JENKINS, 1976, p. 200), ont étudié le comportement des autocorrélations d'une série générée par un processus ARIMA(0, 1, 1), décrit par le modèle suivant :

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1}, \quad t \geq 1$$

Ils constatent que $r(k)$ décroît de façon linéaire avec k . Le même modèle est aussi considéré par ROY (1977). Il obtient les deux premiers moments asymptotiques de la fonction d'autocovariance et montre que :

$$\lim_{N \rightarrow \infty} E[c(k)/N] = \frac{\sigma^2(1-\theta)^2}{6}, \quad k \geq 0 \quad (1.2)$$

et

$$\lim_{N \rightarrow \infty} \text{Cov}[c(k)/N, c(h)/N] = \frac{\sigma^4(1-\theta)^4}{45}, \quad k, h \geq 0 \quad (1.3)$$

où N désigne la longueur de la série. Dans le cas où $\theta = 0$, le modèle décrit une marche aléatoire. Nous généraliserons (1.2) et (1.3) à la marche aléatoire saisonnière.

ANDERSON (1979) déduit des formules donnant la moyenne, la variance et les autocovariances de séries générées par un processus ARIMA(p, d, q), pour $d = 0$ et 1, en termes des autocovariances du processus ARMA(p, q) sous-jacent.

DE GOOIJER (1980) s'intéresse aux autocorrélations échantillonnales. Il exprime, sous une forme impliquant une intégrale, les quatre premiers moments de $r(k)$, dans le contexte où les observations proviennent d'un processus ARIMA(p, d, q), $d = 0$ et 1. Néanmoins, les formules fournies permettent l'évaluation numérique de ces moments. Il compare alors les moments exacts avec des approximations dues à KENDALL (1954) et à MARRIOTT et POPE (1954), dans le contexte d'un modèle ARIMA(1, 0, 1) et d'un ARIMA(0, 1, 1). Nous reprendrons ces approximations à la section 2.6.

DE GOOIJER (1981) étudie les moments des autocovariances et des autocorrélations d'un processus ARIMA($p, 0, q$). Il y annonce des résultats développés dans ANDERSON et DE GOOIJER (1983) où figurent des formules explicites pour $\text{Var}[c(k)]$ et $\text{Cov}[c(k), c(0)]$ dans le contexte d'un ARIMA($p, 0, q$) et indique comment on peut obtenir ces quantités dans le contexte d'un ARIMA($p, 1, q$). Il précise cependant qu'il s'agit de calculs d'une complexité énorme.

ROY et LEFRANÇOIS (1981) publient un article où les deux premiers moments asymptotiques des autocovariances échantillonnales sont mis en évidence dans le contexte où la série observée, $Y_t, t = 1, \dots, N$, provient d'un ARIMA($p, 1, q$). Leur démarche repose en partie sur l'utilisation d'une approximation de Y_t en termes

d'une promenade aléatoire. Ils montrent que :

$$\lim_{N \rightarrow \infty} E[c(k)/N] = \frac{S}{6}, \quad (1.4)$$

et

$$\lim_{N \rightarrow \infty} \text{Cov}[c(k)/N, c(h)/N] = \frac{S^2}{45} \quad (1.5)$$

où $S = \sum_{j=-\infty}^{\infty} \gamma_j$, γ_j étant la j -ème autocovariance du processus stationnaire ∇Y_t .

L'étude de cet article nous indique que la connaissance des propriétés des autocovariances échantillonnelles d'une marche aléatoire permet de mieux cerner celles d'un processus non stationnaire. Aussi, nous mettrons en évidence une approximation exprimant un processus saisonnier en termes d'une marche aléatoire saisonnière, nous conduisant ainsi à une généralisation de (1.4) et (1.5).

Comme nous le démontrons au chapitre deux, les deux premiers moments exacts d'une marche aléatoire saisonnière s'obtiennent facilement. Nous obtenons aussi des formules dans le cas d'un processus $\{Y_t : t \in \mathbb{N}\}$ pour lequel $\omega_t = Y_t - Y_{t-s}$, $t \geq 1$, est un ARIMA($p, 0, q$).

Évidemment, elles sont d'une plus grande complexité que les premières. PEÑA (1984) et GODOLPHIN (1977) obtiennent des formules intéressantes pour un modèle multiplicatif.

Le premier volet du chapitre deux présente des résultats relatifs aux deux premiers moments des autocovariances échantillonnelles de la marche aléatoire saisonnière. Nous obtenons une formule explicite pour $E[c(k)]$, quelle que soit la valeur de k , et une formule tout aussi explicite pour $\text{Var}[c(k)]$, lorsque k est un multiple de la période (s). Nous reprenons, au second volet, le travail pour une transformation linéaire particulière de la marche aléatoire saisonnière. En plus de fournir la formule exacte de $E[c(k)]$, $k \geq 0$, nous donnons une formule explicite de $\text{Cov}[c(ks), c(hs)]$. Disposant de ces quantités, nous pouvons utiliser les formules de Kendall pour approximer $E[r(k)]$ et $\text{Var}[r(k)]$ lorsque k est un multiple de la période. Ces quelques résultats élémentaires nous montrent que $r(ds)$ converge en probabilité vers 1, pour $d \geq 0$. Des résultats relatifs à $c'(k)$ sont aussi disponibles.

Au chapitre trois, nous démontrons qu'un processus saisonnier ayant des propriétés assez générales, peut s'écrire, à toutes fins pratiques, comme la somme d'une promenade aléatoire saisonnière et d'un processus stationnaire. Nous exploitons alors les propriétés de la promenade aléatoire saisonnière pour déduire les deux premiers moments asymptotiques du processus saisonnier plus général.

1.3. Résultats relatifs aux distributions asymptotiques

Une bonne partie des résultats de la thèse de doctorat de DICKEY (1976) exposés dans DICKEY et FULLER (1979) forment la pierre angulaire d'une série d'articles axés sur la caractérisation de la distribution asymptotique de certains estimateurs des moindres carrés. DICKEY et FULLER (1979) étudient la distribution asymptotique de l'estimateur du paramètre autorégressif d'un processus autorégressif d'ordre 1 sous l'hypothèse que ce dernier est égal à 1 en valeur absolue.

De façon plus précise, ils envisagent différents modèles dont :

$$\text{a) } Y_t = \rho Y_{t-1} + e_t, \quad t \geq 1 \quad (1.6)$$

et

$$\text{b) } Y_t = \mu + \rho Y_{t-1} + e_t, \quad t \geq 1 \quad (1.7)$$

où $\{e_t\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$, et où $Y_0 = 0$. Dénotons par r , l'estimateur des moindres carrés de ρ . On constate que pour le modèle (1.6), r a une forme très proche de celle de $r'(1)$. En effet, nous avons :

$$r'(1) = \frac{\sum_{i=1}^{N-1} Y_i Y_{i+1}}{\sum_{i=1}^N Y_i^2}$$

tandis que

$$r = \frac{\sum_{i=1}^{N-1} Y_i Y_{i+1}}{\sum_{i=1}^{N-1} Y_i^2}$$

Cependant, bien que les formules soient similaires, nous verrons au chapitre cinq que les distributions asymptotiques sont différentes.

Notons que dans sa thèse de doctorat, Dickey construit des tables des principaux percentiles des lois exactes et asymptotiques des différentes statistiques étudiées. Ces tables sont obtenues par la méthode de Monte Carlo. De leur côté, EVANS et SAVIN (1981) ont créé une table de la fonction de répartition de $N(1-r)/\sqrt{2}$ par une méthode d'intégration numérique, dans le contexte du modèle (1.6), sous l'hypothèse que $\rho = 1$. Ils utilisent la distribution asymptotique fournie par RAO (1978). DICKEY et FULLER (1981) étudient le test du rapport de vraisemblance pour éprouver l'hypothèse $H : (\mu, \rho) = (0, 1)$ dans le contexte du modèle (1.7).

HASZA et FULLER (1979) présentent une généralisation des résultats de DICKEY et FULLER (1979). On y travaille avec un processus ARIMA($p, 0, 0$) dont l'opérateur

autorégressif a deux racines sur le cercle unitaire. Ils déduisent les distributions asymptotiques des estimateurs des moindres carrés sous cette hypothèse.

SAID et DICKEY (1981) obtiennent la distribution asymptotique des estimateurs des moindres carrés des paramètres d'un modèle ARIMA($p, 1, q$). Par la suite, SAID et DICKEY (1984) développent un test basé sur une approximation d'un processus ARIMA($p, 1, q$) par un processus ARIMA($p', 0, 0$) (p' étant fonction du nombre d'observations), permettant d'éprouver l'hypothèse qu'une des racines de l'opérateur autorégressif est sur le cercle unitaire.

L'article de HASZA (1980) marque un intérêt croissant pour les résultats de Dickey dans l'obtention de résultats relatifs à la distribution asymptotique des autocorrélations. Il ouvre la voie en résolvant le problème dans le contexte d'un processus ARIMA($p, 1, q$) en utilisant (1.1) comme définition du coefficient $r(k)$. Il trouve des approximations simples pour le numérateur et le dénominateur de $N(1 - r(k))$, ce qui lui permet de caractériser la loi asymptotique de cette statistique. Ses résultats seront généralisés au chapitre cinq dans le cadre d'un processus saisonnier.

Dans cette lignée d'articles, celui de HASZA et FULLER (1982) est le premier à traiter des modèles saisonniers. On considère le modèle :

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-s} + \alpha_3 Y_{t-s-1} + e_t, \quad t \geq 1$$

$\{e_t\}$ étant une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance σ^2 . Une méthode pour tester l'hypothèse $H : (\alpha_1, \alpha_2, \alpha_3) = (1, 1, -1)$ y est suggérée.

Entre autres modèles, DICKEY ET COLLAB. (1984) étudient :

$$\text{a) } Y_t = \rho Y_{t-s} + e_t, \quad t \geq 1$$

et

$$\text{b) } Y_t - \mu = \rho(Y_{t-s} - \mu) + e_t \quad t \geq 1 \quad (1.8)$$

où $\{e_t\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance σ^2 , et où Y_{-s+1}, \dots, Y_0 sont les valeurs initiales. Un ensemble d'estimateurs des paramètres des différents modèles est suggéré et les auteurs caractérisent la distribution asymptotique de ceux-ci sous l'hypothèse que $\rho = 1$, en se basant sur les résultats précédents de DICKEY (1976), DICKEY et FULLER (1979), de même que HASZA et FULLER (1979).

Il est intéressant de remarquer que pour le modèle (1.8), l'estimateur des moindres carrés de ρ a une forme très voisine du coefficient d'autocorrélation de délai s calculé

sans centrer les données :

$$r = \frac{\sum_{i=1}^N Y_{i-s} Y_i}{\sum_{i=1}^N Y_{i-s}^2}$$

Nous verrons cependant que les distributions asymptotiques sont différentes.

Au chapitre cinq, nous considérons un processus saisonnier $\{Y_t : t \in \mathbb{N}\}$, pour lequel $\omega_t = Y_t - Y_{t-s}$ est un processus stationnaire. Exploitant certains résultats de DICKEY (1976), résumés au chapitre quatre, nous obtenons la distribution asymptotique de $N(1 - r(ds))$ et celle de $N(1 - r'(ds))$, pour $d \geq 0$, et nous constatons que ces deux distributions sont différentes, même si la moyenne du processus original est nulle. Nous mettons aussi en évidence la distribution asymptotique de $c(ds)$ et celle de $c'(ds)$ lorsque les données proviennent d'une marche aléatoire saisonnière.

Au chapitre six, nous évaluerons par la méthode de Monte Carlo les différentes distributions rencontrées au chapitre cinq. Nous pourrons alors juger des écarts existant entre les distributions exactes et les distributions asymptotiques, de même qu'entre les distributions des statistiques basées sur des données centrées et celles des statistiques calculées avec des données non centrées.

Nous avons pris connaissance de l'article de YAJIMA (1985) au moment où nous faisons la rédaction finale de cette thèse. Dans cet article, il s'attaque à la détermination des propriétés asymptotiques des autocorrélations et des autocorrélations partielles d'un processus saisonnier. Il fait appel aux travaux de HASZA et FULLER (1979) et caractérise la distribution asymptotique de $r(k)$, lorsque k n'est pas un multiple de la période, et celle de $N(1 - r(k))$ lorsque k est un multiple de la période. Cependant, il suppose que $\nabla^d \nabla_s^D Y_t$ est un processus ARIMA($p, 0, q$), ∇_s désignant l'opérateur différence saisonnière d'ordre s , c'est-à-dire $\nabla_s Y_t = Y_t - Y_{t-s}$. Dans le cas que nous traitons, nous supposons simplement que $\nabla_s Y_t$ est un processus linéaire, ce qui est une hypothèse moins forte que celle de Yajima. Par conséquent, la caractérisation des distributions asymptotiques qu'il obtient et celle que nous fournissons ne sont pas identiques. Néanmoins, il obtient des résultats fort intéressants relativement aux autocorrélations partielles. Cependant, il ne donne aucun résultat analytique concernant les moments des autocovariances ou des autocorrélations et il travaille uniquement avec des autocorrélations calculées à partir de données centrées.

L'article de TIAO et BOX (1981) a stimulé la recherche sur les propriétés des modèles multivariés. Ainsi, SAMARANAYAKE et HASZA (1982) étudient les propriétés asymptotiques des corrélations de séries multivariées non stationnaires générées par un processus autorégressif d'ordre p . DAVIES ET COLLAB. (1984) obtiennent des

expressions analytiques pour les moments exacts des corrélations croisées échantillonales d'un processus autorégressif à moyenne mobile multivarié. Ils développent aussi une méthode permettant l'évaluation numérique de celles-ci. Cependant, dans cette thèse, nous nous limiterons aux modèles univariés.

Chapitre 2

Marche aléatoire saisonnière

Résumé

Ce chapitre traite de la marche aléatoire saisonnière, un processus stochastique défini par la relation $\nabla_s X_t = e_t$, où e_t est un bruit blanc et ∇_s l'opérateur de différenciation saisonnière. Nous explorons les propriétés fondamentales de ce modèle, notamment ses moments d'ordre un et deux, ainsi que son comportement asymptotique. L'analyse met en évidence la non-stationnarité du processus et ses implications pour la modélisation des séries temporelles présentant une structure saisonnière marquée.

2.1. Introduction

Dans un premier temps, notre étude portera sur le modèle

$$\nabla_s X_t = e_t$$

où $\nabla_s = (\mathbb{1} - B^s)$, B étant l'opérateur de retard, c'est-à-dire : $B\{X\}_t = \{X_{t-1}\}$, tandis que $\mathbb{1}$ est l'opérateur identité. La suite $\{e_t\}$ est un bruit blanc, c'est-à-dire, une suite de variables aléatoires indépendantes et identiquement distribuées (que nous supposerons gaussiennes lorsque cela sera nécessaire) de moyenne nulle et de variance σ^2 . Nous désignerons le processus $\{X_t\}$ par « *marche (ou promenade) aléatoire saisonnière*. » Nous obtiendrons les deux premiers moments de X_t , et nous reprendrons le travail dans le cas où $\nabla_s X_t$ est un processus ARMA(p, q) stationnaire. Nous aurons alors

$$\Phi(B)\nabla_s X_t = \Theta(B)e_t$$

où

$$\Phi(B) = \mathbf{1} - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

est l'opérateur autorégressif et

$$\Theta(B) = \mathbf{1} - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

est l'opérateur moyenne mobile.

Finalement, nous étudierons les autocovariances échantillonnelles du processus $\nabla_s X_t = e_t$.

2.2. Deux premiers moments de la marche aléatoire saisonnière

Le lecteur pourra se référer au livre de GRAHAM (1981) pour la définition et les principales propriétés du produit de Kronecker, noté \otimes . L'emploi de ce produit met en évidence les structures sous-jacentes au processus $\{X_t\}$.

Soit $\mathbf{X} = (X_0, \dots, X_N)'$, une série chronologique et $\mathbf{X}_0 = (X_{-s}, \dots, X_{-1})'$, l'ensemble des valeurs initiales de la série. Pour simplifier l'écriture et le développement, nous supposerons que le vecteur \mathbf{X} est de longueur ns . Cela signifie que $N = ns - 1$, et que la longueur de la série est un multiple de la période. Dénotons par \mathbf{e} le vecteur $(e_0, \dots, e_{ns-1})'$, pour lequel $\mu_e = \mathbf{0}$ et $\Sigma_e = \sigma^2 \mathbf{I}_{ns} = \sigma^2 (\mathbf{I}_n \otimes \mathbf{I}_s)$.

Dans ce cas, il est facile de voir que

$$\begin{aligned} X_0 &= X_{-s} + e_0 \\ X_1 &= X_{-s+1} + e_1 \\ &\vdots \\ X_{s-1} &= X_{-1} + e_{s-1} \\ \hline X_s &= X_{-s} + e_0 + e_s \\ X_{s+1} &= X_{-s+1} + e_1 + e_{s+1} \\ &\vdots \\ X_{2s-1} &= X_{-1} + e_{s-1} + e_{2s-1} \\ \hline X_{2s} &= X_{-s} + e_0 + e_s + e_{2s} \\ X_{2s+1} &= X_{-s+1} + e_1 + e_{s+1} + e_{2s+1} \end{aligned}$$

$$\vdots$$

$$X_{3s-1} = X_{-1} + e_{s-1} + e_{2s-1} + e_{3s-1}$$

etc.

Si nous posons $j = \nu s + r$, avec $0 \leq r \leq s - 1$, nous obtenons

$$X_j = X_{\nu s + r} = X_{-s+r} + \sum_{i=0}^{\nu} e_{is+r}, \quad (2.1)$$

et en définissant

$$\begin{aligned} \mathbf{X}_{k+1} &= (X_{ks}, X_{ks+1}, \dots, X_{(k+1)s-1}), \quad 0 \leq k \leq n-1, \\ \mathbf{e}_{k+1} &= (e_{ks}, e_{ks+1}, \dots, e_{(k+1)s-1}), \quad 0 \leq k \leq n-1, \end{aligned}$$

nous pouvons écrire

$$\mathbf{X}_k = \mathbf{X}_0 + \sum_{i=1}^k \mathbf{e}_i, \quad k = 1, \dots, n.$$

On obtient finalement

$$\mathbf{X} = \mathbf{1}_n \otimes \mathbf{X}_0 + \mathbf{T}_n \otimes \mathbf{I}_s \mathbf{e} \quad (2.2)$$

où $\mathbf{1}_n$ est un vecteur de n éléments tous égaux à 1, \mathbf{T}_n est une matrice $n \times n$ triangulaire inférieure dont tous les éléments situés sur et sous la diagonale principale sont égaux à 1. Les espérances que nous calculerons seront conditionnelles au vecteur \mathbf{X}_0 . On a :

$$\mathbb{E}[\mathbf{X}] = \mathbf{1}_n \otimes \mathbf{X}_0 = \boldsymbol{\mu}_X$$

et

$$\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)'] = \sigma^2 (\mathbf{T}_n \otimes \mathbf{I}_s) (\mathbf{T}_n \otimes \mathbf{I}_s)' = \sigma^2 (\mathbf{T}_n \mathbf{T}_n' \otimes \mathbf{I}_s) \quad (2.3)$$

On remarque que :

$$\mathbf{T}_n \mathbf{T}_n' \otimes \mathbf{I}_s = \begin{bmatrix} \mathbf{1I}_s & \mathbf{1I}_s & \cdots & \mathbf{1I}_s \\ \mathbf{1I}_s & \mathbf{2I}_s & \cdots & \mathbf{2I}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1I}_s & \mathbf{2I}_s & \cdots & \mathbf{nI}_s \end{bmatrix}$$

et la proposition suivante est démontrée.

Proposition 2.1. *Soit $\{X_t\}$ un processus stochastique tel que*

$$\nabla_s X_t = e_t, \quad t \geq 0$$

avec $\{e_t\}$ un bruit blanc de variance σ^2 et dénotons par X_{-s}, \dots, X_{-1} les s valeurs initiales. Si nous écrivons t et t' sous la forme

$$t = \nu s + r \quad \text{et} \quad t' = \nu' s + r'$$

alors :

$$\begin{aligned} 1) \quad \mathbb{E}[X_t] &= X_{r-s} & t \geq 0 \\ 2) \quad \text{Var}[X_t] &= \sigma^2(\nu + 1) & t \geq 0 \\ 3) \quad \text{Cov}[X_t, X_{t'}] &= \begin{cases} 0 & \text{si } r \neq r', \\ \sigma^2(\nu + 1) & \text{si } r = r' \text{ et } \nu \leq \nu'. \end{cases} \end{aligned}$$

La proposition 2.1 peut aussi se démontrer directement à partir de (2.1). Le corollaire suivant en est une conséquence immédiate.

Corollaire 2.1. *Dans le contexte de la proposition 2.1 :*

$$\begin{aligned} 1) \quad \lim_{t \rightarrow \infty} \text{Var}[X_t] &= \infty \\ 2) \quad \lim_{t \rightarrow \infty} \text{Cov}[X_t, X_{t'}] &= \begin{cases} 0 & r \neq r', \\ \infty & r = r' \text{ et } \nu \leq \nu'. \end{cases} \\ 3) \quad \rho(X_t, X_{t'}) &= \begin{cases} 0 & r \neq r', \\ \left(\frac{\nu + 1}{\nu' + 1}\right)^{1/2} & r = r' \text{ et } \nu \leq \nu'. \end{cases} \end{aligned}$$

Supposons que $\nu' \geq \nu$, et posons $\nu' - \nu = m$. On note alors que pour ν fixé,

$$\lim_{m \rightarrow \infty} \rho(X_t, X_{t'}) = 0$$

tandis que pour m fixé :

$$\lim_{\nu \rightarrow \infty} \rho(X_t, X_{t'}) = \begin{cases} 0 & r \neq r', \\ 1 & r = r'. \end{cases}$$

Notons que X_t n'est pas asymptotiquement stationnaire (voir ROY (1977)).

Supposons maintenant qu'une différence saisonnière d'ordre s appliquée à la série $\{X_t\}$ génère ω_t , un processus ARMA(p, q) stationnaire de moyenne nulle dont les autocovariances sont $\gamma_0, \gamma_1, \dots$. Découpons la suite des observations par tranches de longueur s , comme dans le cas de la marche aléatoire saisonnière. Soit $\mathbf{X}_0 = (X_{-s}, \dots, X_{-1})'$ le vecteur des valeurs initiales, $\mathbf{X}_i = (X_{(i-1)s}, \dots, X_{is-1})'$ et $\boldsymbol{\omega}_i = (\omega_{(i-1)s}, \dots, \omega_{is-1})'$, $i \geq 0$. Si nous concaténons tous les vecteurs \mathbf{X}_i pour

obtenir $\mathbf{X}' = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)$ et que nous faisons de même avec les $\boldsymbol{\omega}_i$, afin d'obtenir $\boldsymbol{\omega}$, nous aurons :

$$\mathbf{X} = \mathbf{1}_n \otimes \mathbf{X}_0 + (\mathbf{T}_n \otimes \mathbf{I}_s) \boldsymbol{\omega}$$

Par conséquent,

$$\mathbb{E}[\mathbf{X}] = \mathbf{1}_n \otimes \mathbf{X}_0 + (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbb{E}[\boldsymbol{\omega}] = \mathbf{1}_s \otimes \mathbf{X}_0$$

et

$$\text{Var}[\mathbf{X}] = \boldsymbol{\Sigma}_X = (\mathbf{T}_n \otimes \mathbf{I}_s) \boldsymbol{\Sigma}_\omega (\mathbf{T}_n \otimes \mathbf{I}_s)'$$

où

$$\boldsymbol{\Sigma}_\omega = \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_{n-1} \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{n-2} \\ \boldsymbol{\Gamma}_2 & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{n-1} & \boldsymbol{\Gamma}_{n-2} & \boldsymbol{\Gamma}_{n-3} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix}$$

et

$$\boldsymbol{\Gamma}_i = \begin{bmatrix} \gamma_{is} & \gamma_{is+1} & \cdots & \gamma_{(i+1)s-1} \\ \gamma_{is-1} & \gamma_{is} & \cdots & \gamma_{(i+1)s-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{(i-1)s+1} & \gamma_{(i-1)s+2} & \cdots & \gamma_{is} \end{bmatrix}$$

Par conséquent,

$$\boldsymbol{\Sigma}_X = \begin{bmatrix} \mathbf{I}_s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_s & \mathbf{I}_s & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_s & \mathbf{I}_s & \cdots & \mathbf{I}_s \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{n-1} \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{n-1} & \boldsymbol{\Gamma}_{n-2} & \cdots & \boldsymbol{\Gamma}_0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_s & \mathbf{I}_s & \cdots & \mathbf{I}_s \\ \mathbf{0} & \mathbf{I}_s & \cdots & \mathbf{I}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_s \end{bmatrix}$$

Pour obtenir la covariance entre \mathbf{X}_t et $\mathbf{X}_{t'}$, nous pourrions examiner la structure de

$$(\mathbf{T}_n \otimes \mathbf{I}_s) \boldsymbol{\Sigma}_\omega (\mathbf{T}_n \otimes \mathbf{I}_s)'$$

Cependant, nous préférons utiliser le fait que

$$X_j = X_{\nu s+r} = X_{-s+r} + \sum_{i=0}^{\nu} \omega_{is+r} \quad (2.4)$$

L'équation 2.4) est obtenue par le même raisonnement qui nous a conduit à (2.1).

On a

$$\text{Var}[X_t] = \text{Var}[X_{\nu s+r}]$$

$$\begin{aligned}
&= \text{Var} \left[\sum_{i=0}^{\nu} \omega_{is+r} \right] \\
&= \text{E} \left[\left(\sum_{i=0}^{\nu} \omega_{is+r} \right)^2 \right] \\
&= \sum_{i=0}^{\nu} \sum_{j=0}^{\nu} \text{E}[\omega_{is+r} \omega_{js+r}] \\
&= \sum_{i=0}^{\nu} \sum_{j=0}^{\nu} \gamma_{(j-i)s} \\
&= (\nu + 1) \sum_{k=-\nu}^{\nu} \gamma_{ks} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks} \tag{2.5}
\end{aligned}$$

Puisque ω_t est un processus ARMA(p, q) stationnaire, les $\gamma_j, j \geq q + 1$, satisfont l'équation aux différences

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \cdots + \varphi_p \gamma_{k-p}$$

On peut en conclure que $\sum_{k=0}^{\infty} \gamma_k$ et $\sum_{k=0}^{\infty} k \gamma_k$ convergent absolument (cf. ROY et LEFRANÇOIS (1981)). Par conséquent, chacune des sommes dans (2.5) converge absolument lorsque $\nu \rightarrow \infty$. Néanmoins, (2.5) nous indique que $\lim_{t \rightarrow \infty} \text{Var}[X_t]$ n'existe pas.

Au niveau des covariances, nous avons

$$\begin{aligned}
\text{Cov}[X_t, X_{t'}] &= \text{Cov}[X_{\nu s+r}, X_{\nu' s+r'}] \\
&= \text{Cov} \left[\sum_{i=0}^{\nu} \omega_{is+r}, \sum_{j=0}^{\nu'} \omega_{js+r'} \right] \\
&= \text{E} \left[\sum_{i=0}^{\nu} \omega_{is+r} \sum_{j=0}^{\nu'} \omega_{js+r'} \right] \\
&= \sum_{i=0}^{\nu} \sum_{j=0}^{\nu'} \text{E}[\omega_{is+r} \omega_{js+r'}] \\
&= \sum_{i=0}^{\nu} \sum_{j=0}^{\nu'} \gamma_{(j-i)s+(r'-r)}
\end{aligned}$$

Sans perdre de généralité, nous pouvons supposer que $\nu' > \nu$. Posons temporairement $\zeta = r' - r$, et développons la double sommation :

$$\begin{array}{cccccc}
\gamma_\zeta & +\gamma_{s+\zeta} & +\cdots +\gamma_{\nu s+\zeta} & +\gamma_{(\nu+1)s+\zeta} +\gamma_{(\nu+2)s+\zeta} +\cdots +\gamma_{\nu' s+\zeta} \\
+\gamma_{-s+\zeta} +\gamma_\zeta & & +\cdots +\gamma_{(\nu-1)s+\zeta} +\gamma_{\nu s+\zeta} & +\gamma_{(\nu+1)s+\zeta} +\cdots +\gamma_{(\nu'-1)s+\zeta} \\
+\gamma_{-2s+\zeta} +\gamma_{-s+\zeta} & & +\cdots +\gamma_{(\nu-2)s+\zeta} +\gamma_{(\nu-1)s+\zeta} +\gamma_{\nu s+\zeta} & +\cdots +\gamma_{(\nu'-2)s+\zeta} \\
\vdots & \vdots & \ddots \vdots & \vdots & \vdots & \ddots \vdots \\
+\gamma_{-\nu s+\zeta} +\gamma_{-(\nu+1)s+\zeta} +\cdots +\gamma_\zeta & & +\gamma_{s+\zeta} & +\gamma_{2s+\zeta} & +\cdots +\gamma_{(\nu'-\nu)s+\zeta}
\end{array}$$

Elle devient donc :

$$\text{Cov}[X_{\nu s+r}, X_{\nu' s+r'}] = \sum_{k=-\nu}^{\nu} (\nu+1 - |k|) \gamma_{ks+\zeta} + \sum_{k=1}^{\nu'-\nu} \sum_{j=0}^{\nu} \gamma_{(k+j)s+\zeta}. \quad (2.6)$$

L'équation (2.6) met en évidence le fait que la covariance entre $X_{\nu s+r}$ et $X_{\nu' s+r'}$ est beaucoup plus complexe que dans le cas de la marche aléatoire saisonnière. Des équations (2.5) et (2.6), nous pouvons obtenir les coefficients d'autocorrélation

$$\begin{aligned}
\rho(X_{\nu s+r}, X_{\nu' s+r'}) = & \\
& \frac{\sum_{k=-\nu}^{\nu} (\nu+1 - |k|) \gamma_{ks+r'-r} + \sum_{k=1}^{\nu'-\nu} \sum_{j=0}^{\nu} \gamma_{(k+j)s+r'-r}}{\left((\nu+1) \sum_{k=-\nu}^{\nu} \gamma_{ks} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks} \right)^{1/2} \left((\nu'+1) \sum_{k=-\nu'}^{\nu'} \gamma_{ks} - \sum_{k=-\nu'}^{\nu'} |k| \gamma_{ks} \right)^{1/2}} \quad (2.7)
\end{aligned}$$

L'équation (2.7) ne nous permet pas de voir ce qui se passe lorsque t ou t' deviennent grands. Nous allons donc l'examiner plus en détail. Supposons que ν' soit suffisamment plus grand que ν (il suffit que $\nu' > 2\nu$) et que $\nu' = \nu + m$. Nous pouvons alors écrire (2.6) sous la forme

$$\begin{aligned}
\text{Cov}[X_t, X_{t'}] &= \sum_{k=-\nu}^{\nu} (\nu+1) \gamma_{ks+\zeta} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks+\zeta} + \sum_{k=1}^m \sum_{j=0}^{\nu} \gamma_{(k+j)s+\zeta} \\
&= \sum_{k=-\nu}^{\nu} (\nu+1) \gamma_{ks+\zeta} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks+\zeta} \\
&\quad + \sum_{k=1}^{\nu} k \gamma_{ks+\zeta} + (\nu+1) \sum_{k=1}^{m-\nu} \gamma_{(k+\nu)s+\zeta} + \sum_{k=1}^{\nu} k \gamma_{(\nu+m-k+1)s+\zeta} \quad (2.8)
\end{aligned}$$

Si nous maintenons fixés ν , r et r' et que nous faisons tendre m vers l'infini, toutes les sommes apparaissant en (2.8) seront bien définies. Cependant, le dénominateur

de (2.7) tend vers l'infini. Par conséquent, pour ν , r et r' fixés,

$$\lim_{m \rightarrow \infty} \left| \text{Cov} \left[X_{\nu s+r}, X_{(\nu+m)s+r'} \right] \right| < \infty$$

et

$$\lim_{m \rightarrow \infty} \rho \left(X_{\nu s+r}, X_{(\nu+m)s+r'} \right) = 0.$$

Nous allons maintenant faire tendre ν vers l'infini en fixant m , r , et r' . Puisque ν tend vers l'infini et que m est fixé, nous pouvons supposer que $m < \nu$ et écrire (2.6) sous la forme

$$\begin{aligned} \text{Cov} [X_{\nu s+r}, X_{\nu' s+r'}] &= \sum_{k=-\nu}^{\nu} (\nu + 1 - |k|) \gamma_{ks+\zeta} + \sum_{k=1}^m k \gamma_{ks+\zeta} \\ &\quad + m \sum_{k=m+1}^{\nu} \gamma_{ks+\zeta} + \sum_{k=1}^m (m + 1 - k) \gamma_{(k+\nu)s+\zeta} \\ &= (\nu + 1) \sum_{k=-\nu}^{\nu} \gamma_{ks+\zeta} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks+\zeta} + \sum_{k=1}^m k \gamma_{ks+\zeta} \\ &\quad + m \sum_{k=m+1}^{\nu} \gamma_{ks+\zeta} + \sum_{k=1}^m (m + 1 - k) \gamma_{(k+\nu)s+\zeta} \quad (2.9) \end{aligned}$$

Les troisième et cinquième sommations n'impliquent qu'un nombre fini de termes. Les autres sommes convergent absolument lorsque ν tend vers l'infini. Par conséquent, (2.9) nous indique que

$$\lim_{\nu \rightarrow \infty} \text{Cov} [X_{\nu s+r}, X_{(\nu+m)s+r'}]$$

n'existe pas nécessairement.

L'existence de cette limite dépend de $\sum_{k \geq 0} \gamma_{ks+\zeta}$. Dans le cas de la marche aléatoire saisonnière, seul $\gamma_0 = \sigma^2 \neq 0$, de sorte que $\text{Cov} [X_{\nu s+r}, X_{(\nu+m)s+r'}]$ est égal à 0 ou $\sigma^2(\nu + 1)$, selon que $\zeta = 0$ ou non. En établissant (2.9), on suppose implicitement que $m \geq 1$, mais si $m = 0$, les trois dernières sommes n'apparaissent pas. Ainsi, si $m = 0$,

$$\text{Cov} [X_{\nu s+r}, X_{\nu s+r'}] = (\nu + 1) \sum_{k=-\nu}^{\nu} \gamma_{ks+r'-r} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks+r'-r}.$$

D'où,

$$\lim_{\nu \rightarrow \infty} \text{Cov} [X_{\nu s+r}, X_{\nu s+r'}]$$

n'existe pas nécessairement.

En ce qui concerne la corrélation, si $m = 0$, on a

$$\rho(X_{\nu s+r}, X_{\nu s+r'}) = \frac{\sum_{k=-\nu}^{\nu} \gamma_{ks+r-r'} - \frac{1}{\nu+1} \sum_{k=-\nu}^{\nu} |k| \gamma_{ks+r-r'}}{\sum_{k=-\nu}^{\nu} \gamma_{ks} - \frac{1}{\nu+1} \sum_{k=-\nu}^{\nu} |k| \gamma_{ks}} \quad (2.10)$$

De (2.10), on déduit que

$$\lim_{\nu \rightarrow \infty} \rho(X_{\nu s+r}, X_{\nu s+r'}) = \frac{\sum_{k=-\infty}^{\infty} \gamma_{ks+r'-r}}{\sum_{k=-\infty}^{\infty} \gamma_{ks}}.$$

Dans le cas où $r' - r = 0$, cette limite vaut 1. Si $m \geq 1$, on montre de manière analogue que

$$\lim_{\nu \rightarrow \infty} \rho(X_{\nu s+r}, X_{(\nu+m)s+r'}) = \frac{\sum_{k=-\infty}^{\infty} \gamma_{ks+r'-r}}{\sum_{k=-\infty}^{\infty} \gamma_{ks}}.$$

qui est aussi égale à 1 lorsque $r' = r$. La proposition 2.2 résume ces derniers résultats.

Proposition 2.2. *Soit $\{X_t : t \geq 1\}$ un processus stochastique régi par le modèle*

$$\nabla_s X_t = \omega_t$$

où $\{\omega_t\}$ est un processus ARMA(p, q) stationnaire de moyenne nulle, dont les autocovariances sont $\gamma_0, \gamma_1, \dots$, et désignons par X_{-s}, \dots, X_{-1} les s valeurs initiales. Si nous écrivons t et t' sous la forme

$$t = \nu s + r \quad \text{et} \quad t' = \nu' s + r',$$

en posant $\zeta = r' - r$, alors :

$$(1) \quad \mathbb{E}[X_t] = X_{r-s}, \quad t \geq 0,$$

$$(2) \quad \text{Var}[X_t] = (\nu + 1) \sum_{k=-\nu}^{\nu} \gamma_{ks} - \sum_{k=-\nu}^{\nu} |k| \gamma_{ks}, \quad t \geq 0,$$

$$(3) \quad \text{Cov}[X_t, X_{t'}] = \sum_{k=-\nu}^{\nu} (\nu + 1 - |k|) \gamma_{ks+\zeta} + \sum_{k=1}^{\nu'-\nu} \sum_{j=0}^{\nu} \gamma_{(k+j)s+\zeta}, \quad \nu \leq \nu',$$

$$(4) \quad \lim_{m \rightarrow \infty} \rho(X_{\nu s+r}, X_{(\nu+m)s+r'}) = 0, \quad \nu, r \text{ et } r' \text{ fixés},$$

$$(5) \lim_{\nu \rightarrow \infty} \rho(X_{\nu s+r}, X_{(\nu+m)s+r'}) = \frac{\sum_{k=-\infty}^{\infty} \gamma_{ks+r'-r}}{\sum_{k=-\infty}^{\infty} \gamma_{ks}}, \quad m, r \text{ et } r' \text{ fixés}$$

et cette limite vaut 1 quand $r' = r$.

PEÑA (1984) obtient des formules pour un modèle saisonnier multiplicatif. On peut aussi se référer à GODOLPHIN (1977).

2.3. Autocovariances échantillonnables avec données centrées

2.3.1. Statistiques usuelles

Tout comme pour les séries stationnaires, nous utiliserons les statistiques descriptives suivantes :

$$\bar{X} = \frac{1}{ns} \sum_{t=0}^{ns-1} X_t$$

et

$$c(k) = \frac{1}{ns} \sum_{t=0}^{ns-k-1} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad k \geq 0$$

Ainsi, \bar{X} est la moyenne échantillonnable, tandis que $c(k)$ est l'autocovariance échantillonnable de délai k , $k \geq 0$. On peut cependant donner une forme plus commode à $c(k)$. En effet, définissons $\mathbf{P}_{ns}(k)$, la matrice $ns \times ns$, dont les éléments sont 1 sur la k -ième diagonale supérieure et 0 ailleurs. On peut alors se convaincre que

$$c(k) = \frac{1}{ns} (\mathbf{X} - \mathbf{1}_{ns}\bar{X})' \mathbf{P}_{ns}(k) (\mathbf{X} - \mathbf{1}_{ns}\bar{X})$$

On préfère toujours utiliser une matrice symétrique dans la définition d'une forme quadratique. Dans notre cas, cela donne

$$c(k) = \frac{1}{2ns} (\mathbf{X} - \mathbf{1}_{ns}\bar{X})' (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) (\mathbf{X} - \mathbf{1}_{ns}\bar{X})$$

Si nous posons

$$\mathbf{A} = \mathbf{I}_n \otimes \mathbf{I}_s - (\mathbf{J}_n/n) \otimes (\mathbf{J}_s/s) \quad (2.11)$$

où \mathbf{J}_k est une matrice $k \times k$ dont tous les éléments sont égaux à 1, nous voyons aisément que

$$(\mathbf{X} - \mathbf{1}_{ns}\bar{X}) = \mathbf{A}\mathbf{X}$$

De plus, $\mathbf{X} = \mathbf{1}_n \otimes \mathbf{X}_0 + (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}$. On obtient

$$c(k) = \frac{1}{2ns} \mathbf{X}' \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} \mathbf{X}$$

Pour simplifier les calculs, posons temporairement

$$\mathbf{Q} = \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A}$$

Alors,

$$\begin{aligned} c(k) &= \frac{1}{2ns} (\mathbf{1}_n \otimes \mathbf{X}_0 + (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e})' \mathbf{Q} (\mathbf{1}_n \otimes \mathbf{X}_0 + (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}) \\ &= \frac{1}{2ns} \left[(\mathbf{1}_n \otimes \mathbf{X}_0)' \mathbf{Q} (\mathbf{1}_n \otimes \mathbf{X}_0) + ((\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e})' \mathbf{Q} ((\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}) \right. \\ &\quad \left. + 2 (\mathbf{1}_n \otimes \mathbf{X}_0)' \mathbf{Q} ((\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}) \right] \end{aligned} \quad (2.12)$$

Pour calculer $E[c(k)]$ sachant que $\boldsymbol{\mu}_e = \mathbf{0}$, et que $\boldsymbol{\Sigma}_e = \sigma^2 \mathbf{I}_n \otimes \mathbf{I}_s$, on peut utiliser un résultat sur les moments de formes quadratiques ((ANDERSON, 1971, lemmes 3.4.3, 3.4.4 et 3.4.5)).

Lemme 2.1. *Soit $Q = \mathbf{u}' \mathbf{B} \mathbf{u}$, où $\mathbf{B} = \{b_{i,j}\}$ est une matrice symétrique $n \times n$, et \mathbf{u} est un vecteur aléatoire de longueur n . Soit $R = \mathbf{u}' \mathbf{C} \mathbf{u}$, où $\mathbf{C} = \{c_{i,j}\}$ est une matrice symétrique $n \times n$. Supposons que les composantes u_i , $i = 1, \dots, n$, du vecteur \mathbf{u} sont mutuellement indépendantes et que*

$$E[u_i] = 0, \quad E[u_i^2] = \sigma^2 \quad \text{et} \quad E[u_i^4] = \kappa_4 + 3\sigma^4$$

Alors,

- (1) $E[Q] = \sigma^2 \text{tr}[\mathbf{B}]$,
- (2) $\text{Var}[Q] = \kappa_4 \left(\sum_{i=1}^n b_{i,i}^2 \right) + 2\sigma^4 \text{tr}[\mathbf{B}^2]$,
- (3) $\text{Cov}[Q, R] = \kappa_4 \left(\sum_{i=1}^n b_{i,i} c_{i,i} \right) + 2\sigma^4 \text{tr}[\mathbf{B}\mathbf{C}]$.

L'espérance de $c(k)$ sera obtenue en évaluant l'espérance de chacun des termes de (2.12). On trouve :

$$\begin{aligned} E[c(k)] &= \left\{ (\mathbf{1}_n \otimes \mathbf{X}_0)' \mathbf{Q} (\mathbf{1}_n \otimes \mathbf{X}_0) + E[\mathbf{e}' (\mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{Q} (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}] \right. \\ &\quad \left. + 2 (\mathbf{1}_n \otimes \mathbf{X}_0)' \mathbf{Q} (\mathbf{T}_n \otimes \mathbf{I}_s) E[\mathbf{e}] \right\} / (2ns) \\ &= \alpha + \sigma^2 \text{tr} \left\{ (\mathbf{T}_n \otimes \mathbf{I}_s)' \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} (\mathbf{T}_n \otimes \mathbf{I}_s) \right\} / (2ns) \\ &= \alpha + \sigma^2 \text{tr} \left\{ \mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A} \right\} / (ns), \end{aligned}$$

où

$$\alpha = (\mathbf{1}_n \otimes \mathbf{X}_0)' \mathbf{Q} (\mathbf{1}_n \otimes \mathbf{X}_0) / (2ns) \quad (2.13)$$

est une quantité dépendant de \mathbf{X}_0 , k , n et s . Notons qu'il s'agit du seul terme faisant intervenir les valeurs initiales. Posons

$$\bar{X}_0 = \frac{1}{s} \sum_{t=1}^s X_{-t}$$

et $x_i = X_i - \bar{X}_0$, la $i^{\text{ième}}$ composante de $\mathbf{A} (\mathbf{I}_n \otimes \mathbf{X}_0) = \mathbf{1}_n \otimes \mathbf{X}_0 - \mathbf{1}_{ns} \otimes \bar{X}_0$.

Sachant que

$$\begin{aligned} \alpha &= (\mathbf{1} \otimes \mathbf{X}_0)' \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} (\mathbf{1} \otimes \mathbf{X}_0) / (2ns) \\ &= (\mathbf{1} \otimes \mathbf{X}_0)' \mathbf{A} \mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{1} \otimes \mathbf{X}_0) / (ns) \\ &= \frac{1}{ns} \sum_{i=1}^{ns-k} (x_i x_{i+k}) \end{aligned}$$

en appliquant l'inégalité de Cauchy, on trouve

$$\begin{aligned} |\alpha| &\leq \frac{1}{ns} \left(\sum_{i=1}^{ns-k} x_i^2 \right)^{1/2} \left(\sum_{i=1}^{ns-k} x_{i+k}^2 \right)^{1/2} \\ |\alpha| &\leq \frac{1}{ns} \sum_{i=1}^{ns} x_i^2 \\ |\alpha| &= \frac{1}{s} \sum_{i=1}^s (X_i - \bar{X}_0)^2 \\ &= \lambda^*, \end{aligned}$$

où λ^* est indépendant de n . On peut finalement écrire

$$\mathbb{E}[c(k)] \leq \lambda^* + \sigma^2 \text{tr}\{\mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}\} / (ns).$$

Nous nous intéresserons éventuellement à $\mathbb{E}[c(k)/n]$. Nous savons évidemment que

$$\mathbb{E}[c(k)/n] \leq \lambda^*/n + \sigma^2 \text{tr}\{\mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}\} / (n^2 s).$$

Lorsque $n \rightarrow \infty$, il est évident que seul le second terme domine, car $\lambda^*/n \rightarrow 0$. Par conséquent, il suffit d'analyser le second terme pour déterminer le comportement de $\mathbb{E}[c(k)]$.

2.3.2. Espérance et variance des autocovariances

De fait, nous voulons évaluer la trace de $\mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}$. Lorsque $k = 0$, il s'agit de trouver la trace de $\mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}$, et si $k \geq 1$, il s'agit d'additionner les

éléments de la k -ième sous-diagonale de $\mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}$. Nous dénotons maintenant $\mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}$ par \mathbf{V} . En utilisant la définition de \mathbf{A} , nous allons expliciter la matrice \mathbf{V} . Nous avons

$$\begin{aligned} \mathbf{V} &= (\mathbf{I}_n \otimes \mathbf{I}_s - \mathbf{J}_n \otimes \mathbf{J}_s / (ns))' (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) (\mathbf{I}_n \otimes \mathbf{I}_s - \mathbf{J}_n \otimes \mathbf{J}_s / (ns)) \\ &= \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s - \mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{J}_s / (ns) - \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s / (ns) \\ &\quad + \mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s \mathbf{J}_s / (ns)^2 \\ &= \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s + \left\{ \mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n / (n^2 s) - (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n + \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n) / (ns) \right\} \otimes \mathbf{J}_s, \end{aligned}$$

car $\mathbf{J}_s \mathbf{J}_s = s \mathbf{J}_s$. Notons que la matrice $\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n$ est une matrice ayant toutes ses lignes identiques. Le j -ième élément de l'une d'elles est

$$\frac{j}{2}(2n - j + 1).$$

En effet, la j -ième colonne de $\mathbf{T}_n \mathbf{T}'_n$ est un vecteur de la forme : $(1, 2, \dots, j - 1, j, j, \dots, j)'$. La somme de ses composantes est

$$\frac{j}{2}(j + 1) + (n - j)j = j \left(\frac{j + 1}{2} + (n - j) \right) = \frac{j}{2}(2n - j + 1).$$

Or, la somme des composantes de la j -ième colonne de $\mathbf{T}_n \mathbf{T}'_n$ est précisément le j -ième élément d'une ligne de $\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n$.

Lorsque nous nous intéressons à $\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n$, nous constatons qu'il s'agit d'une matrice dont tous les éléments sont égaux. Pour obtenir cette valeur commune, il suffit de faire la somme des éléments d'une ligne de $\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n$. Un simple calcul montre que

$$\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n = \frac{n(2n + 1)(n + 1)}{6} \mathbf{J}_n \quad (2.14)$$

Puisque la matrice $\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n$ est une matrice ayant toutes ses lignes identiques, le j -ième élément est

$$\frac{j}{2}(2n - j + 1).$$

L'élément (i, j) de la matrice $\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n + \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n$ est donné par

$$\frac{j}{2}(2n - j + 1) + \frac{i}{2}(2n - i + 1) = \frac{1}{2} \left\{ (i + j)(2n + 1) - (i^2 + j^2) \right\}.$$

Cela nous permet de voir que l'élément r_{ij} de la matrice

$$\mathbf{R}_n = \left\{ \mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n / (n^2 s) - (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n + \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n) / (ns) \right\} \quad (2.15)$$

est

$$r_{ij} = n(2n + 1)(n + 1) / (6n^2 s) - \left\{ (i + j)(2n + 1) - (i^2 + j^2) \right\} / (2ns)$$

$$\begin{aligned}
&= \left\{ (2n+1)(n+1) - 3(i+j)(2n+1) + 3(i^2+j^2) \right\} / (6ns) \\
&= \left\{ (2n+1)(n-3(i+j)+1) + 3(i^2+j^2) \right\} / (6ns) \\
&= \left\{ 2n^2 + (-6j-6i+3)n + 3j^2 - 3j + 3i^2 - 3i + 1 \right\} / (6ns) \tag{2.16}
\end{aligned}$$

Maintenant, nous avons

$$\mathbf{A}(\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A} = \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s + \mathbf{R}_n \otimes \mathbf{J}_s.$$

Dénotons par t_i le total des éléments de la i -ième sous-diagonale de la matrice \mathbf{R}_n .

On a

$$\begin{aligned}
t_i &= \sum_{j=1}^{n-i} r_{i+j,j} \\
&= \frac{1}{6ns} \sum_{j=1}^{n-i} \left(2n^2 + 3(-4j-2i+1)n + 6j^2 + 6(i-1)j + 3(i^2-i) + 1 \right) \\
&= -\frac{1}{6ns} \left\{ 2n^3 - (i-3)n^2 - (3i^2+3i-1)n + 2i^3 - i \right\}. \tag{2.17}
\end{aligned}$$

Maintenant, $E[c(ms)]$, pour $0 \leq m \leq n-1$, peut être calculé facilement. En effet,

$$\begin{aligned}
E[c(ms)] &= \alpha + \sigma^2 \operatorname{tr}[\mathbf{P}_{ns}(ms) \mathbf{A}(\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) \mathbf{A}] / (ns) \\
&= \alpha + \sigma^2 \operatorname{tr}[\mathbf{P}_{ns}(ms) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) + \mathbf{P}_{ns}(ms) (\mathbf{R}_n \otimes \mathbf{J}_s)] / (ns).
\end{aligned}$$

En examinant la forme de $\mathbf{P}_{ns}(ms) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s)$, on constate que la contribution de ce terme est $s(n-m)(n-m+1)/2$. De plus, faire la somme des éléments de la ms -ième diagonale de $\mathbf{R}_n \otimes \mathbf{J}_s$ revient à additionner s fois les éléments de la m -ième diagonale de \mathbf{R}_n . On obtient alors

$$E[c(ms)] = \alpha + \sigma^2 \{s(n-m)(n-m+1)/2 + st_m\} / (ns). \tag{2.18}$$

De manière plus explicite,

$$\begin{aligned}
E[c(ms)] &= \alpha + \frac{\sigma^2}{ns} \left\{ s(n-m)(n-m+1)/2 - s \left\{ 2n^3 + (-m+3)n^2 \right. \right. \\
&\quad \left. \left. + (-3m^2 - 3m + 1)n + 2m^3 - m \right\} / (6n) \right\} \\
&= \alpha + \frac{\sigma^2}{6n^2s} \left\{ (3nm^2 - (6n^2 + 3n)m + 3n^3 + 3n^2)s - 2m^3 \right. \\
&\quad \left. + 3nm^2 + (n^2 + 3n + 1)m - 2n^3 - 3n^2 - n \right\} \\
&= \alpha + \frac{\sigma^2}{6n^2s} \left\{ (3s-2)n^3 + ((-6s+1)m + 3s-3)n^2 \right. \\
&\quad \left. + ((3s+3)m^2 + 3(1-s)m - 1)n - 2m^3 + m \right\} \tag{2.19}
\end{aligned}$$

Aussi, si les valeurs initiales sont nulles, $\alpha = 0$ et on a un résultat exact et explicite pour $E[c(ms)]$. De plus, on voit que

$$\lim_{n \rightarrow \infty} E[c(ms)/(ns)] = \lim_{n \rightarrow \infty} \sigma^2 \{n^3(3s-2)\} / (6n^3s^2) = \frac{\sigma^2}{6s^2}(3s-2) \quad (2.20)$$

Dans le cas particulier où $s = 1$, on retrouve

$$\lim_{n \rightarrow \infty} E[c(m)/n] = \sigma^2/6,$$

ce qui est conforme à ROY (1977), voir (1.2).

Il reste à expliciter $E[c^{(ns)}(ms+r)]$, pour $0 \leq m \leq n-2$ et $1 \leq r < s$. Par inspection de la structure de

$$\mathbf{P}_{ns}(k) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s + \mathbf{R}_n \otimes \mathbf{J}_s),$$

on se rend compte que la matrice $\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s$ n'apporte aucune contribution, puisque les éléments de la $(ms+r)$ -ième diagonale de $\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s$ sont nuls pour $r \neq 0$. Ainsi, pour $k = ms+r$, ($s > 1$), en examinant la structure de $\mathbf{R}_n \otimes \mathbf{J}_s$, nous avons

$$\begin{aligned} E[c(ms+r)] &= \alpha + \frac{\sigma^2}{ns} \{(s-r)t_m + rt_{m+1}\} \\ &= \alpha - \frac{\sigma^2}{6n^2s^2} \left[\left\{ (2m^3 + n(-3m^2 - 3m + 1) - m + n^2(3-m) \right. \right. \\ &\quad \left. \left. + 2n^3)(s-r) \right\} - \left\{ r \left(n(-3(m+1)^2 - 3(m+1) + 1) \right. \right. \right. \\ &\quad \left. \left. \left. + 2(m+1)^3 - m + n^2(2-m) + 2n^3 - 1 \right) \right\} \right] \\ &= \alpha - \frac{\sigma^2}{6n^2s^2} \left[\left\{ 2sn^3 - (sm - 3s + r)n^2 - (3sm^2 + (3s + 6r)m \right. \right. \right. \\ &\quad \left. \left. \left. + s - 6r)n + 2sm^3 + 6rm^2 + (-s + 6r)m + r \right\} \right] \quad (2.21) \end{aligned}$$

Le coefficient de n^3 au numérateur de (2.21) est $-2s$, de sorte que

$$\lim_{n \rightarrow \infty} E[c(ms+r)/(ns)] = -2\sigma^2 / (6s^2) = -\sigma^2 / (3s^2).$$

On voit que le rapport des espérances tend vers 1 lorsque le délai est un multiple de la période, tandis qu'il tend vers

$$\frac{-\sigma^2}{3s^2} \div \frac{\sigma^2(3s-2)}{6s^2} = \frac{-2}{3s-2}$$

lorsqu'il ne s'agit pas d'un multiple de la période.

La prochaine section contiendra une généralisation de ces résultats dans le cas où nous avons une transformation linéaire de la marche aléatoire saisonnière. Entre autres, nous obtiendrons la covariance entre les autocovariances échantillonales dont

les délais sont des multiples de la période. En spécialisant ces prochains résultats, nous obtenons

$$\begin{aligned}
\text{Var}(c(ms)) &= \frac{\sigma^4}{180n^4s^2} \left\{ [60s - 56]n^6 \right. \\
&+ [104m - 120 - s(120m - 120)]n^5 \\
&+ [55m^2 + 180m - 110 - s(90m^2 + 180m - 120)]n^4 \\
&- [120m^3 - 90m^2 - 170m + 60 + s(-360m^3 + 90m^2 + 240m - 60)]n^3 \\
&- [100m^4 + 180m^3 - 35m^2 - 90m + 14 \\
&\quad + (225m^4 - 180m^3 - 135m^2 + 90m)s]n^2 \\
&+ [102m^5 - 90m^3 + 14m]n \\
&\left. + m^2 [26m^4 - 30m^2 + 14] \right\} \tag{2.22}
\end{aligned}$$

et si nous posons $m = 0$, l'équation (2.22) devient

$$\begin{aligned}
\text{Var}(c(0)) &= \frac{\sigma^4}{90n^2s^2} \left(2(15s - 14)n^4 + 60(s - 1)n^3 \right. \\
&\quad \left. + 5(12s - 11)n^2 + 30(s - 1)n - 7 \right). \tag{2.23}
\end{aligned}$$

Notons aussi que

$$\lim_{n \rightarrow \infty} \text{Var}(c(ms)/(ns)) = \frac{\sigma^4(15s - 14)}{45s^4}.$$

Pour $s = 1$, on retrouve (1.3) évaluée en $\theta = 0$.

2.4. Généralisation à une transformation linéaire de la marche aléatoire saisonnière

Nous allons maintenant compléter l'étude des autocovariances en considérant une transformation linéaire d'une marche aléatoire saisonnière. Cette étude est motivée par le fait que nous obtiendrons au chapitre trois un théorème de représentation d'un processus saisonnier où apparaîtra une transformation linéaire particulière de la marche aléatoire saisonnière. Nous expliciterons donc l'espérance de $c(k)$ et nous déterminerons la covariance entre $c(h)$ et $c(k)$, lorsque h et k sont des multiples de s , dans ce contexte plus général.

Par l'équation (2.1),

$$\mathbf{X}_t = \mathbf{X}_0 + \sum_{i=1}^t \mathbf{e}_i.$$

Nous allons maintenant supposer que $\mathbf{X}_0 = \mathbf{0}$, et considérer $\mathbf{Y}_t = \mathbf{H}_0 \mathbf{X}_t$ où \mathbf{H}_0 est une matrice circulante, c'est-à-dire une matrice de la forme

$$\mathbf{H}_0 = \begin{bmatrix} a_1 & a_2 & \cdots & a_{s-1} & a_0 \\ a_2 & a_3 & \cdots & a_0 & a_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{s-1} & a_0 & \cdots & a_{s-3} & a_{s-2} \\ a_0 & a_1 & \cdots & a_{s-2} & a_{s-1} \end{bmatrix}.$$

Nous exploiterons alors les propriétés intéressantes du carré de cette matrice explicitées par le lemme suivant.

Lemme 2.2. *Soit \mathbf{H}_0 la matrice circulante introduite précédemment. Alors*

$$\mathbf{J} \mathbf{H}_0^2 = \left(\sum_{i=0}^{s-1} a_i \right)^2 \mathbf{J},$$

où \mathbf{J} est une matrice ayant tous ses éléments égaux à 1. En d'autres termes, la somme des éléments d'une ligne ou d'une colonne quelconque de \mathbf{H}_0^2 est une constante g_0^2 égale à $\left(\sum_{i=0}^{s-1} a_i \right)^2$.

Preuve. Puisque \mathbf{H}_0 est une matrice symétrique, l'élément en position (i, j) de la matrice \mathbf{H}_0^2 est obtenu en faisant le produit scalaire de la colonne i par la colonne j de \mathbf{H}_0 . Il est clair que la somme des éléments de la ligne i de \mathbf{H}_0^2 sera

$$\sum_{j=1}^s \mathbf{c}'_i \mathbf{c}_j,$$

où \mathbf{c}_k est la k -ème colonne de la matrice \mathbf{H}_0 . Définissons \mathbf{P} , une matrice $s \times s$, de la façon suivante :

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

La matrice \mathbf{P} est orthogonale. Lorsque nous faisons le produit $\mathbf{P} \mathbf{A}$, où \mathbf{A} est une matrice $s \times q$, la nouvelle matrice est obtenue en faisant monter d'une position les lignes de \mathbf{A} et en remplaçant la dernière ligne par la première. Aussi, $\mathbf{P}^k \mathbf{A}$ est obtenue de la matrice \mathbf{A} en faisant monter de k positions les lignes de \mathbf{A} et en remplaçant les k dernières lignes par les k premières.

et il devient évident que

$$\begin{array}{cccccccccccc}
\delta_0^2 & + & \delta_1^2 & + & \delta_2^2 & + & \dots & + & \delta_{s-1}^2 & = & h_0 \\
\delta_1^2 & + & \delta_0^2 & + & \delta_1^2 & + & \dots & + & \delta_{s-2}^2 & = & h_0 \\
\delta_2^2 & + & \delta_1^2 & + & \delta_0^2 & + & \dots & + & \delta_{s-3}^2 & = & h_0 \\
\vdots & & \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
\delta_{s-2}^2 & + & \delta_{s-3}^2 & + & \delta_{s-4}^2 & + & \dots & + & \delta_1^2 & = & h_0 \\
\delta_{s-1}^2 & + & \delta_{s-2}^2 & + & \delta_{s-3}^2 & + & \dots & + & \delta_0^2 & = & h_0
\end{array}$$

Ceci termine la démonstration du lemme. \square

2.4.1. Espérance de l'autocovariance

Nous nous intéressons maintenant à l'autocovariance échantillonnale obtenue à partir des Y_t . La concaténation des n vecteurs $\mathbf{Y}_t = \mathbf{H}_0 \mathbf{X}_t$ de dimension s nous donne un nouveau vecteur \mathbf{Y} de dimension $n \times s$. On remarque que $\mathbf{Y} = \mathbf{I}_n \otimes \mathbf{H}_0 \mathbf{X}$, et nous obtenons comme expression

$$c(k) = \frac{1}{2ns} \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{H}_0)' \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X}.$$

En utilisant le fait que $\mathbf{X} = (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}$, que $\boldsymbol{\mu}_e = \mathbf{0}$, et que $\boldsymbol{\Sigma}_e = \sigma^2 \mathbf{I}_{ns}$, on obtient

$$\begin{aligned}
c(k) &= \frac{1}{2ns} \mathbf{e}' (\mathbf{T}_n \otimes \mathbf{I}_s)' (\mathbf{I}_n \otimes \mathbf{H}_0)' \mathbf{A} (\mathbf{P}_{ns}(k) \\
&\quad + \mathbf{P}'_{ns}(k)) \mathbf{A} (\mathbf{I}_n \otimes \mathbf{H}_0) (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}, \quad (2.24)
\end{aligned}$$

et en appliquant le lemme 2.1, on trouve

$$\begin{aligned}
\mathbb{E}[c(k)] &= \frac{\sigma^2}{2ns} \text{tr}\{(\mathbf{T}'_n \otimes \mathbf{H}_0) \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} (\mathbf{T}_n \otimes \mathbf{H}_0)\} \\
&= \frac{\sigma^2}{ns} \text{tr}\{\mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A}\} \quad (2.25)
\end{aligned}$$

Posons

$$\mathbf{W} = \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A}.$$

Alors,

$$\begin{aligned}
\mathbf{W} &= (\mathbf{I}_n \otimes \mathbf{I}_s - \mathbf{J}_n \otimes \mathbf{J}_s / (ns)) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) (\mathbf{I}_n \otimes \mathbf{I}_s - \mathbf{J}_n \otimes \mathbf{J}_s / (ns)) \\
&= \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 - \frac{1}{ns} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) (\mathbf{J}_n \otimes \mathbf{J}_s) - \frac{1}{ns} (\mathbf{J}_n \otimes \mathbf{J}_s) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \\
&\quad + \frac{1}{n^2 s^2} (\mathbf{J}_n \otimes \mathbf{J}_s) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) (\mathbf{J}_n \otimes \mathbf{J}_s) \\
&= \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 - \frac{1}{ns} (\mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{H}_0^2 \mathbf{J}_s) - \frac{1}{ns} (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{J}_s \mathbf{H}_0^2) \\
&\quad + \frac{1}{n^2 s^2} (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s \mathbf{H}_0^2 \mathbf{J}_s)
\end{aligned}$$

Sachant que g_0^2 est la somme des éléments d'une ligne (ou d'une colonne) de \mathbf{H}_0^2 , on a

$$\begin{aligned}
\mathbf{W} &= \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 - g_0^2 (\mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s) / (ns) - g_0^2 (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{J}_s) / (ns) \\
&\quad + s g_0^2 (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s) / (ns)^2 \\
&= \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2 [(\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s) / (n^2 s) \\
&\quad - ((\mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n \otimes \mathbf{J}_s) + (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{J}_s)) / (ns)].
\end{aligned}$$

Puisque $\mathbf{R}_n = \mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n / (n^2 s) - (\mathbf{J}_n \mathbf{T}_n \mathbf{T}'_n + \mathbf{T}_n \mathbf{T}'_n \mathbf{J}_n) / (ns)$,

$$\mathbf{W} = \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2 \mathbf{R}_n \otimes \mathbf{J}_s. \quad (2.26)$$

2.4.1.1. Cas où le délai k est un multiple de la période. Les calculs sont simplifiés de façon importante puisque dans ce cas,

$$\mathbf{P}_{ns}(k) = \mathbf{P}_{ns}(ms) = \mathbf{P}_n(m) \otimes \mathbf{I}_s.$$

Nous devons évaluer

$$\begin{aligned}
&\text{tr}\{\mathbf{P}_{ns}(ms)\mathbf{W}\} \\
&= \text{tr}\{\mathbf{P}_n(m)\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2 \mathbf{P}_n(m)\mathbf{R}_n \otimes \mathbf{J}_s\}, \\
&\quad \text{puisque } \text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A}) \text{tr}(\mathbf{B}) \\
&= \text{tr}\{\mathbf{P}_n(m)\mathbf{T}_n \mathbf{T}'_n\} \text{tr}\{\mathbf{H}_0^2\} + g_0^2 \text{tr}\{\mathbf{P}_n(m)\mathbf{R}_n\} \text{tr}\{\mathbf{J}_s\} \\
&= s\delta_0(n-m)(n-m+1)/2 + g_0^2 st_m.
\end{aligned} \quad (2.27)$$

où δ_0 est l'élément diagonal de la matrice \mathbf{H}_0^2 , tandis que t_m a été défini par l'équation (2.17). Ainsi,

$$\begin{aligned} & \text{tr}\left\{\mathbf{P}_n(m)\mathbf{T}_n\mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2\mathbf{P}_n(m)\mathbf{R}_n \otimes \mathbf{J}_s\right\} \\ &= s\delta_0(n-m)(n-m+1)/2 - g_0^2s\left\{2n^3 - (m-3)n^2\right. \\ &\quad \left. - (3m^2 + 3m - 1)n + 2m^3 - m\right\}/(6ns) \\ &= -\left\{(2g_0^2 - 3s\delta_0)n^3 - ((m-3)g_0^2 - 6\delta_0sm + 3s\delta_0)n^2\right. \\ &\quad \left. - ((3m^2 + 3m - 1)g_0^2 + 3sm\delta_0(m-1))n + m(2m^2 - 1)g_0^2\right\}/(6n). \end{aligned}$$

On conclut que

$$\begin{aligned} \text{E}[c(ms)] &= -\frac{\sigma^2}{6n^2s}\left\{(2g_0^2 - 3s\delta_0)n^3 - ((m-3)g_0^2 - 6\delta_0sm + 3s\delta_0)n^2\right. \\ &\quad \left. - ((3m^2 + 3m - 1)g_0^2 + 3sm\delta_0(m-1))n + m(2m^2 - 1)g_0^2\right\}. \end{aligned}$$

En divisant par ns et en prenant la limite, on trouve

$$\lim_{n \rightarrow \infty} \text{E}[c(ms)/(ns)] = \frac{\sigma^2}{s} \left(\frac{\delta_0}{2} - \frac{g_0^2}{3s} \right). \quad (2.28)$$

Remarquons que si $s = 1$, $g_0^2 = \delta_0$ puisque $\mathbf{H}_0^2 = \delta_0$. Ainsi, la limite devient $\sigma^2 g_0^2/6$, tel que l'indiquent ROY et LEFRANÇOIS (1981), voir (1.4).

2.4.1.2. Cas où le délai n'est pas un multiple de la période. Nous allons introduire une notation qui nous facilitera l'évaluation des différentes traces.

Définition 2.1. Soit \mathbf{A} une matrice $q \times q$. La somme des éléments de la ν -ième diagonale inférieure sera notée par $\delta^-(\mathbf{A}, \nu)$, et la somme des éléments de la ν -ième diagonale supérieure sera notée par $\delta^+(\mathbf{A}, \nu)$ pour $1 \leq \nu \leq q - 1$. De plus, nous convenons que $\delta^-(\mathbf{A}, 0) = \delta^+(\mathbf{A}, 0) = \text{tr}(\mathbf{A})$, tandis que $\delta^+(\mathbf{A}, q) = \delta^-(\mathbf{A}, q) = 0$. \square

La matrice $\mathbf{T}_n\mathbf{T}'_n \otimes \mathbf{H}_0^2$ a la structure suivante :

$$\begin{bmatrix} 1\mathbf{H}_0^2 & 1\mathbf{H}_0^2 & 1\mathbf{H}_0^2 & \cdots & 1\mathbf{H}_0^2 \\ 1\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & \cdots & 2\mathbf{H}_0^2 \\ 1\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & 3\mathbf{H}_0^2 & \cdots & 3\mathbf{H}_0^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & 3\mathbf{H}_0^2 & \cdots & n\mathbf{H}_0^2 \end{bmatrix}$$

Par les équations (2.25) et (2.26), nous avons

$$\begin{aligned} \mathbb{E}[c(m+s)] &= \frac{\sigma^2}{ns} \left\{ \text{tr}(\mathbf{P}_{ns}(ms+r)(\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)) \right. \\ &\quad \left. + g_0^2 \text{tr}(\mathbf{P}_{ns}(ms+r)(\mathbf{R}_n \otimes \mathbf{J}_s)) \right\} \quad (2.29) \end{aligned}$$

où $r \neq 0$, et $1 \leq m \leq n-1$. On se rend compte que de chercher la somme des éléments de la $(ms+r)$ -ième diagonale inférieure de $\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2$ revient à chercher la somme des éléments sur des diagonales inférieures ou supérieures de \mathbf{H}_0^2 . Puisque \mathbf{H}_0^2 est symétrique, $\delta^-(\mathbf{H}_0^2, \nu) = \delta^+(\mathbf{H}_0^2, \nu)$, pour $1 \leq \nu \leq n-1$. Pour $\nu = ms+r$, on voit que $\text{tr}\{\mathbf{P}_{ns}(\nu)(\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)\}$ s'obtient en évaluant

$$\begin{aligned} &1\delta^-(\mathbf{H}_0^2, r) + 1\delta^+(\mathbf{H}_0^2, s-r) + 2\delta^-(\mathbf{H}_0^2, r) + 2\delta^+(\mathbf{H}_0^2, s-r) + \dots \\ &+ (n-m-1)\delta^-(\mathbf{H}_0^2, r) + (n-m-1)\delta^+(\mathbf{H}_0^2, s-r) + (n-m)\delta^-(\mathbf{H}_0^2, r) \\ &= \left\{ \delta^-(\mathbf{H}_0^2, r) + \delta^+(\mathbf{H}_0^2, s-r) \right\} \\ &\quad \times (n-m-1)(n-m)/2 + (n-m)\delta^-(\mathbf{H}_0^2, r). \end{aligned}$$

En exploitant la forme de la matrice \mathbf{H}_0^2 , on peut se convaincre que $\delta^-(\mathbf{H}_0^2, r) = (s-r)\delta_r$ et que $\delta^+(\mathbf{H}_0^2, s-r) = r\delta_r$. Par conséquent,

$$\delta^-(\mathbf{H}_0^2, r) + \delta^+(\mathbf{H}_0^2, s-r) = s\delta_r.$$

Ainsi,

$$\text{tr}(\mathbf{P}_{ns}(\nu)(\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)) = s\delta_r(n-m-1)(n-m)/2 + (s-r)\delta_r(n-m).$$

Dans le cas où $r = 0$, nous aurions

$$\begin{aligned} &\text{tr}(\mathbf{P}_{ns}(ms)(\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)) \\ &= \left(\delta^-(\mathbf{H}_0^2, 0) + \delta^+(\mathbf{H}_0^2, s) \right) (n-m-1)(n-m)/2 + (n-m)\delta^-(\mathbf{H}_0^2, 0) \\ &= \delta^-(\mathbf{H}_0^2, 0) \left((n-m-1)(n-m)/2 + (n-m) \right) \\ &= \delta^-(\mathbf{H}_0^2, 0) (n-m) \left((n-m-1)/2 + 1 \right) \\ &= \delta^-(\mathbf{H}_0^2, 0) (n-m)(n-m+1)/2 \\ &= s\delta_0(n-m)(n-m+1)/2 \end{aligned}$$

ce qui est effectivement la contribution de $\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2$ dans le cas où le délai est un multiple de la période. Quant au terme $\text{tr}(\mathbf{P}_{ns}(ms+r)\mathbf{R}_n \otimes \mathbf{J}_s)$, nous avons montré, voir équation (2.21), qu'il valait

$$-\frac{1}{6ns} \left\{ 2sn^3 - ((m-3)s+r)n^2 - (3sm^2 + 3(s+2r)m - s + 6r)n \right.$$

$$+2sm^3 + 6rm^2 - (s - 6r)m + r\}$$

Ce qui nous permet d'écrire, selon (2.29),

$$\begin{aligned} \mathbb{E}[c(ms + r)] &= \frac{\sigma^2}{ns} \left\{ s\delta_r \frac{(n-m-1)(n-m)}{2} + (n-m)(s-r)\delta_r \right. \\ &\quad \left. - \frac{g_0^2}{6ns} \left[2sn^3 - ((m-3)s+r)n^2 - (3sm^2 + 3(s+2r)m - s + 6r) \right. \right. \\ &\quad \left. \left. + 2sm^3 + 6rm^2 - (s-6r)m + r \right] \right\} \\ &= \frac{\sigma^2}{6s^2n^2} \left[(3s^2\delta_r - 2g_0^2s) n^3 \right. \\ &\quad + \left(3(s^2(1-2m) - 2rs) \delta_r + g_0^2(s(m-3) + r) \right) n^2 \\ &\quad + \left((3s^2m^2 - 3s(s-2r)m) \delta_r + g_0^2(3sm^2 + (3s+6r)m - s + 6r) \right) n \\ &\quad \left. - g_0^2(2sm^3 + 6rm^2 - (s-6r)m + r) \right]. \end{aligned} \quad (2.30)$$

En divisant par ns et en prenant la limite, on trouve

$$\lim_{n \rightarrow \infty} \mathbb{E}[c(ms + r)/(ns)] = \frac{\sigma^2}{s} \left(\frac{\delta_r}{2} - \frac{g_0^2}{3s} \right). \quad (2.31)$$

Il est intéressant de noter que lorsque nous remplaçons r par 0, nous obtenons l'expression (2.28).

2.4.2. Covariance entre deux autocovariances

L'autocovariance échantillonnale est donnée par l'équation (2.24). Pour obtenir la covariance entre deux autocovariances, nous devons calculer la covariance entre deux formes quadratiques. Posons

$$\begin{aligned} \mathbf{P} &= \frac{1}{2ns} (\mathbf{T}'_n \otimes \mathbf{H}_0) \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} (\mathbf{T}_n \otimes \mathbf{H}_0), \\ \mathbf{Q} &= \frac{1}{2ns} (\mathbf{T}'_n \otimes \mathbf{H}_0) \mathbf{A} (\mathbf{P}_{ns}(h) + \mathbf{P}'_{ns}(h)) \mathbf{A} (\mathbf{T}_n \otimes \mathbf{H}_0). \end{aligned}$$

Par conséquent,

$$c(k) = \mathbf{e}' \mathbf{P} \mathbf{e}, \quad \text{et} \quad c(h) = \mathbf{e}' \mathbf{Q} \mathbf{e}.$$

Pour simplifier les calculs, nous supposons que $\kappa_4 = 0$ (ce qui est vrai lorsque le vecteur \mathbf{e} est distribué normalement). Dans ce cas, en utilisant le lemme 2.1, on trouve

$$\text{Cov}[c(k), c(h)] = 2\sigma^4 \text{tr}(\mathbf{P}\mathbf{Q}).$$

Posons temporairement

$$\mathbf{Q}_k = \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A}.$$

On obtient alors pour la covariance

$$\begin{aligned} \text{Cov}[c(k), c(h)] &= \frac{\sigma^4}{2n^2s^2} \text{tr}\{(\mathbf{T}'_n \otimes \mathbf{H}_0) \mathbf{Q}_k (\mathbf{T}_n \otimes \mathbf{H}_0) (\mathbf{T}'_n \otimes \mathbf{H}_0) \mathbf{Q}_h (\mathbf{T}_n \otimes \mathbf{H}_0)\} \\ &= \frac{\sigma^4}{2n^2s^2} \text{tr}\{\mathbf{Q}_k (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{Q}_h (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)\}. \end{aligned}$$

Posons $\mathbf{M} = (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)$, qui est une matrice symétrique. Alors,

$$\begin{aligned} \text{Cov}[c(k), c(h)] &= \sigma^4 \text{tr}\{\mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} \mathbf{M}\} / (2n^2s^2) \\ &= \sigma^4 \text{tr}\{\mathbf{A} \mathbf{P}_{ns}(k) \mathbf{A} \mathbf{M}\} / (n^2s^2) \\ &= \sigma^4 \text{tr}\{\mathbf{A} \mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} (\mathbf{P}_{ns}(h) + \mathbf{P}'_{ns}(h)) \\ &\quad \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)\} / (n^2s^2) \\ &= \sigma^4 \left\{ \text{tr}(\mathbf{A} \mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} \mathbf{P}_{ns}(h) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)) \right. \\ &\quad \left. + \text{tr}(\mathbf{A} \mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} \mathbf{P}'_{ns}(h) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2)) \right\} / (n^2s^2) \\ &= \sigma^4 \left\{ \text{tr}(\mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} \mathbf{P}_{ns}(h) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A}) \right. \\ &\quad \left. + \text{tr}(\mathbf{P}_{ns}(k) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} \mathbf{P}'_{ns}(h) \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A}) \right\} / (n^2s^2) \end{aligned}$$

Soit $\mathbf{W} = \mathbf{A} (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2) \mathbf{A} = \mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2 \mathbf{R}_n \otimes \mathbf{J}_s$, et soit $w_{i,j}$ l'élément en position (i, j) de la matrice \mathbf{W} . Nous avons montré (cf. équation (2.16)) que

$$r_{i,j} = \frac{(2n+1)(n-3(i+j)+1) + 3(i^2+j^2)}{6ns},$$

et la constante g_0^2 est liée à \mathbf{H}_0^2 . Ainsi,

$$\begin{aligned} \text{Cov}[c(k), c(h)] &= \frac{\sigma^4}{n^2s^2} \sum_{i=1}^{ns-h} \sum_{j=1}^{ns-k} (w_{i+h, j} w_{j+k, i} + w_{i, j} w_{j+k, i+h}) \\ &= \frac{\sigma^4}{n^2s^2} \sum_{i=1}^{ns-h} \sum_{j=1}^{ns-k} (w_{i, j} w_{i+h, j+k} + w_{i+h, j} w_{i, j+k}), \end{aligned} \quad (2.32)$$

car \mathbf{W} est symétrique.

Bien que nous ayons une formule relativement simple pour la covariance entre deux autocovariances, nous n'avons pas réussi à trouver une formule plus explicite. Nous allons cependant l'utiliser dans certains cas bien précis.

2.4.3. Variance de la covariance échantillonnale

Afin d'obtenir la variance de $c(0)$, il suffit d'appliquer la formule (2.32) en posant h et k égaux à zéro. On obtient

$$\text{Var}[c(0)] = 2 \frac{\sigma^4}{n^2 s^2} \sum_{i=1}^{ns} \sum_{j=1}^{ns} w_{i,j}^2. \quad (2.33)$$

De plus, $w_{i,j}$ est l'élément en position (i, j) de la matrice $\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2 \mathbf{R}_n \otimes \mathbf{J}_s$, à savoir :

$$\begin{bmatrix} 1\mathbf{H}_0^2 & 1\mathbf{H}_0^2 & 1\mathbf{H}_0^2 & \cdots & 1\mathbf{H}_0^2 \\ 1\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & \cdots & 2\mathbf{H}_0^2 \\ 1\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & 3\mathbf{H}_0^2 & \cdots & 3\mathbf{H}_0^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1\mathbf{H}_0^2 & 2\mathbf{H}_0^2 & 3\mathbf{H}_0^2 & \cdots & n\mathbf{H}_0^2 \end{bmatrix} + g_0^2 \begin{bmatrix} r_{11}\mathbf{J}_s & r_{12}\mathbf{J}_s & r_{13}\mathbf{J}_s & \cdots & r_{1n}\mathbf{J}_s \\ r_{21}\mathbf{J}_s & r_{22}\mathbf{J}_s & r_{23}\mathbf{J}_s & \cdots & r_{2n}\mathbf{J}_s \\ r_{31}\mathbf{J}_s & r_{32}\mathbf{J}_s & r_{33}\mathbf{J}_s & \cdots & r_{3n}\mathbf{J}_s \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1}\mathbf{J}_s & r_{n2}\mathbf{J}_s & r_{n3}\mathbf{J}_s & \cdots & r_{nn}\mathbf{J}_s \end{bmatrix}$$

Pour évaluer la sommation apparaissant en (2.33), nous la diviserons en morceaux. À cause de la symétrie de la matrice, nous ne considérerons que les blocs $s \times s$ de la diagonale principale et de la partie triangulaire supérieure.

2.4.3.1. Contribution des blocs de la diagonale. La contribution des blocs de la diagonale sera donnée par

$$\begin{aligned} & \sum_{i=1}^n \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) (i\delta_{|\lambda|} + g_0^2 r_{ii})^2 \\ &= \sum_{i=1}^n \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) (i^2 \delta_{|\lambda|}^2 + 2\delta_{|\lambda|} g_0^2 i r_{ii} + g_0^4 r_{ii}^2) \\ &= \sum_{i=1}^n i^2 \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \delta_{|\lambda|}^2 + 2g_0^2 \sum_{i=1}^n i r_{ii} \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \delta_{|\lambda|} \\ & \quad + g_0^4 \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \sum_{i=1}^n r_{ii}^2 \\ &= sh_0 n(n+1)(2n+1)/6 - g_0^4 (n+1) (3n^2 + 2n + 1) / 6 \\ & \quad + g_0^4 s^2 (n+1)(2n+1) (6n^2 + n + 3) / (60ns^2), \end{aligned}$$

où h_0 est la somme des carrés des éléments d'une ligne (ou d'une colonne) de \mathbf{H}_0^2 .
On obtient finalement

$$\begin{aligned} & \sum_{i=1}^n \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left(i\delta_{|\lambda|} + g_0^2 r_{ii} \right)^2 \\ &= \frac{(20h_0s - 18g_0^4)n^4 - (30h_0s - 30g_0^4)n^3 + (10h_0s - 15g_0^4)n^2 + 3g_0^4}{60n} \end{aligned}$$

2.4.3.2. Contribution des blocs hors diagonale. Les éléments en question sont ceux des blocs (i, j) , $2 \leq j \leq n$ et $1 \leq i \leq j-1$, dont la contribution est double. De manière plus précise, la contribution du bloc (i, j) est

$$\begin{aligned} & \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left(i\delta_{|\lambda|} + g_0^2 r_{ij} \right)^2 \\ &= \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left(i^2 \delta_{|\lambda|}^2 + 2\delta_{|\lambda|} g_0^2 i r_{ij} + g_0^4 r_{ij}^2 \right) \\ &= i^2 s h_0 + 2s g_0^4 i r_{ij} + g_0^4 s^2 r_{ij}^2. \end{aligned}$$

Ainsi, la contribution des éléments hors diagonale est

$$2 \sum_{j=2}^n \sum_{i=1}^{j-1} \left(i^2 s h_0 + 2s g_0^4 i r_{ij} + g_0^4 s^2 r_{ij}^2 \right). \quad (2.34)$$

On peut vérifier que

$$\begin{aligned} (1) \quad & \sum_{j=2}^n \sum_{i=1}^{j-1} i^2 = \frac{n^4 - n^2}{12}, \\ (2) \quad & \sum_{j=2}^n \sum_{i=1}^{j-1} i r_{ij} = -\frac{28n^4 + 15n^3 - 20n^2 - 15n - 8}{360s}, \\ (3) \quad & \sum_{j=2}^n \sum_{i=1}^{j-1} r_{ij}^2 = \frac{28n^5 + 24n^4 - 5n^3 - 15n^2 - 23n - 9}{360ns^2}. \end{aligned}$$

En remplaçant ces quantités dans l'équation (2.34), on trouve

$$\frac{(30h_0s - 28g_0^4)n^5 - 6g_0^4n^4 + (35g_0^4 - 30h_0s)n^3 + 15g_0^4n^2 - 7g_0^4n - 9g_0^4}{180n}.$$

Après avoir multiplié par $2\sigma^4 / (n^2s^2)$ la somme des contributions et après simplification, la variance est

$$\begin{aligned} \text{Var}[c(0)] &= \frac{\sigma^4}{90n^2s^2} \left((30h_0s - 28g_0^4)n^4 + (60h_0s - 60g_0^4)n^3 \right. \\ &\quad \left. + (60h_0s - 55g_0^4)n^2 + (30h_0s - 30g_0^4)n - 7g_0^4 \right). \end{aligned}$$

Si nous considérons une marche aléatoire saisonnière, alors $g_0^4 = h_0 = 1$. En remplaçant ces valeurs dans la dernière équation, on obtient l'équation (2.23).

Selon (FULLER, 1976, page 378),

$$\begin{aligned} \mathbb{E}\left[\sum_{t=1}^n (X_t - \bar{X})^2\right] &= \sigma^2 \frac{n^2 - 1}{6}, \\ \text{Var}\left[\sum_{t=1}^n (X_t - \bar{X})^2\right] &= \sigma^4 \frac{(n^2 - 1)(2n^2 + 7)}{90}, \end{aligned}$$

lorsque les observations proviennent du modèle

$$X_t = X_{t-1} + \varepsilon_t, \quad t = 1, \dots, n,$$

où les variables $\{\varepsilon_t\}$ sont indépendantes et identiquement distribuées selon $\mathcal{N}(0, \sigma^2)$ et $X_{-1} = 0$.

Cela nous permet de conclure que :

$$\begin{aligned} (1) \quad \mathbb{E}[c(0)] &= \sigma^2 \frac{n^2 - 1}{6n}, \\ (2) \quad \text{Var}[c(0)] &= \sigma^4 \frac{(n^2 - 1)(2n^2 + 7)}{90n^2}. \end{aligned}$$

Il est intéressant de noter que si nous nous plaçons dans le même contexte, la constante α de (2.19) vaut 0 (car $X_{-1} = 0$) et nous devons poser $s = 1$ et $m = 0$. L'équation (2.19) devient :

$$\mathbb{E}[c(0)] = \sigma^2 \frac{n^3 - n}{6n^2} = \sigma^2 \frac{n^2 - 1}{6n}.$$

En faisant le même travail au niveau de l'équation (2.23), nous obtenons

$$\text{Var}[c(0)] = \sigma^4 \frac{2n^4 + 5n^2 - 7}{90n^2} = \sigma^4 \frac{(n^2 - 1)(2n^2 + 7)}{90n^2}.$$

Ainsi, en utilisant nos résultats, nous retrouvons ceux mentionnés par Fuller.

2.5. Covariance entre deux autocovariances échantillonnables de délais multiples de la période

Selon l'équation (2.32),

$$\text{Cov}(c(hs), c(ks)) = \frac{\sigma^4}{n^2 s^2} \sum_{i=1}^{(n-h)s} \sum_{j=1}^{(n-k)s} (w_{i+hs,j} w_{i,j+ks} + w_{i,j} w_{i+hs,j+ks}). \quad (2.35)$$

où, $w_{i,j}$ est l'élément (i, j) de la matrice $\mathbf{W} = \mathbf{T}_n \mathbf{T}_n' \mathbf{H}_0^2 + g_0^2 \mathbf{R}_n \otimes \mathbf{J}_s$.

La matrice \mathbf{W} étant de taille $ns \times ns$, elle peut être partitionnée en blocs $s \times s$. Ainsi partitionnée, nous avons une matrice $n \times n$ dont les éléments sont des matrices $s \times s$, et nous pouvons parler des « *diagonales-blocs* » de cette matrice $n \times n$.

Afin d'évaluer la double sommation apparaissant en (2.35), nous procéderons en trois étapes :

- (1) Évaluation de la contribution des blocs diagonaux ($s \times s$).
- (2) Évaluation de la contribution des blocs ($s \times s$) de la partie triangulaire supérieure.
- (3) Évaluation de la contribution des blocs ($s \times s$) de la partie triangulaire inférieure.

Le lecteur trouvera à l'annexe A les programmes ProgA et ProgB qui ont permis l'évaluation des formules décrites aux sections 2.5.1 à 2.5.3.

2.5.1. Contribution des blocs diagonaux

Sans perte de généralité, nous pouvons supposer que $h \leq k$. Le terme général de la sommation dépend de quatre éléments de la matrice $\mathbf{T}_n \mathbf{T}'_n \mathbf{H}_0^2 + g_0^2 \mathbf{R}_n \otimes \mathbf{J}_s$, à savoir :

$$\begin{array}{cc} w_{i,j} & w_{i,j+ks} \\ w_{i+hs,j} & w_{i+hs,j+ks}. \end{array}$$

Il est naturel de se référer à ces quatre éléments via leur position : nord-ouest, nord-est, sud-ouest et sud-est.

Nous savons de plus que

$$\mathbf{H}_0^2 = \begin{bmatrix} \delta_0 & \delta_1 & \dots & \delta_{s-2} & \delta_{s-1} \\ \delta_1 & \delta_0 & \dots & \delta_{s-3} & \delta_{s-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \delta_{s-2} & \delta_{s-3} & \dots & \delta_0 & \delta_1 \\ \delta_{s-1} & \delta_{s-2} & \dots & \delta_1 & \delta_0 \end{bmatrix}$$

et que la somme des éléments d'une ligne de \mathbf{H}_0^2 est une constante notée g_0^2 . De plus, la somme des carrés des éléments d'une ligne de \mathbf{H}_0^2 est une constante notée h_0 . Si nous considérons les éléments du i -ème bloc de la diagonale-bloc principale de

$$\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{H}_0^2 + g_0^2 \mathbf{R}_n \otimes \mathbf{J}_s,$$

nous aurons :

$$\begin{array}{cc} i\delta_\lambda + g_0^2 r_{i,i} & i\delta_\lambda + g_0^2 r_{i,i+k} \\ i\delta_\lambda + g_0^2 r_{i+h,i} & (i+h)\delta_\lambda + g_0^2 r_{i+h,i+k} \end{array}$$

puisque l'élément en position (i, j) de $\mathbf{T}_n \mathbf{T}'_n = \min\{i, j\}$. La contribution du i -ème bloc est

$$\begin{aligned} & \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left[(i\delta_{|\lambda|} + g_0^2 r_{i,i}) \left((i+h)\delta_{|\lambda|} + g_0^2 r_{i+h,i+k} \right) \right. \\ & \quad \left. + (i\delta_{|\lambda|} + g_0^2 r_{i+h,i}) \left(i\delta_{|\lambda|} + g_0^2 r_{i,i+k} \right) \right] \\ &= \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left[(2i^2 + ih) \delta_{|\lambda|}^2 \right. \\ & \quad \left. + g_0^2 ((i+h)r_{i,i} + i(r_{i+h,i+k} + r_{i,i+k} + r_{i+h,i})) \delta_{|\lambda|} \right. \\ & \quad \left. + g_0^4 (r_{i,i} r_{i+h,i+k} + r_{i,i+k} r_{i+h,i}) \right] \\ &= h_0 s (2i^2 + ih) + s g_0^4 ((i+h)r_{i,i} + i(r_{i+h,i+k} + r_{i,i+k} + r_{i+h,i})) \\ & \quad + s^2 g_0^4 (r_{i,i} r_{i+h,i+k} + r_{i,i+k} r_{i+h,i}). \end{aligned} \tag{2.36}$$

Il reste donc à sommer ces termes pour $i \in \{1, 2, \dots, n-k\}$. Ce travail sera aisément effectué en utilisant le programme MacSyma.

2.5.2. Contribution des blocs de la partie triangulaire supérieure

Pour évaluer cette contribution, nous considérons les $n - k - 1$ diagonales-blocs de la partie triangulaire supérieure. Deux cas sont examinés.

2.5.2.1. L'élément en position sud-ouest n'est pas au-dessus de la diagonale des blocs. Soit ν le numéro de la diagonale. On voit aisément que les indices des blocs à considérer sont de la forme $(i, i + \nu)$, l'indice ν étant un élément de $\{1, 2, \dots, n - k - 1\}$, tandis que pour ν fixé, $i \in \{1, 2, \dots, n - k - \nu\}$. En effet, le fait que $h \geq k$ implique que les quatre éléments du terme de la sommation apparaissant dans (2.35) déterminent un rectangle dont la base est au moins aussi grande que la hauteur. Quand i augmente, ce rectangle glisse dans la direction sud-est et son déplacement est limité par la droite. L'élément nord-est réside dans le bloc $(i, i + \nu + k)$. Par conséquent, il faut que $i + \nu + k \leq n$, ce qui signifie que $i \leq n - k - \nu$. Si l'élément en position sud-ouest n'est pas au-dessus de la diagonale

des blocs, cela signifie que son indice de ligne est plus grand ou égal à son indice de colonne, c'est-à-dire que $i + h \geq i + \nu$, donc $\nu \leq h$.

Les quatre éléments intervenants sont de la forme :

$$\begin{array}{cc} i\delta_\lambda + g_0^2 r_{i,i+\nu} & i\delta_\lambda + g_0^2 r_{i,i+\nu+k} \\ (i+\nu)\delta_\lambda + g_0^2 r_{i+h,i+\nu} & (i+h)\delta_\lambda + g_0^2 r_{i+h,i+\nu+k} \end{array}$$

de sorte que la contribution du bloc $(i, i + \nu)$ est :

$$\begin{aligned} & \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left[(i\delta_{|\lambda|} + g_0^2 r_{i,i+\nu}) \left((i+h)\delta_{|\lambda|} + g_0^2 r_{i+h,i+\nu+k} \right) \right. \\ & \quad \left. + \left((i+\nu)\delta_{|\lambda|} + g_0^2 r_{i+h,i+\nu} \right) \left(i\delta_{|\lambda|} + g_0^2 r_{i,i+\nu+k} \right) \right] \\ &= \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left[(2i^2 + i(\nu + h)) \delta_{|\lambda|}^2 \right. \\ & \quad \left. + g_0^2 \left((i+h)r_{i,i+\nu} + (i+\nu)r_{i,i+\nu+k} + i(r_{i+h,i+\nu+k} + r_{i+h,i+\nu}) \right) \delta_{|\lambda|} \right. \\ & \quad \left. + g_0^4 (r_{i,i+\nu} r_{i+h,i+\nu+k} + r_{i,i+\nu+k} r_{i+h,i+\nu}) \right] \\ &= h_0 s (2i^2 + i(\nu + h)) + s g_0^4 \left((i+h)r_{i,i+\nu} + (i+\nu)r_{i,i+\nu+k} \right. \\ & \quad \left. + i(r_{i+h,i+\nu+k} + r_{i+h,i+\nu}) \right) + s^2 g_0^4 (r_{i,i+\nu} r_{i+h,i+\nu+k} + r_{i,i+\nu+k} r_{i+h,i+\nu}). \quad (2.37) \end{aligned}$$

Il restera à faire une double sommation de ces derniers termes en faisant varier l'indice ν dans $\{1, \dots, h\}$, et pour chacune de ces valeurs de ν , l'indice i dans $\{1, \dots, n - k - \nu\}$.

2.5.2.2. L'élément en position sud-ouest est au-dessus de la diagonale des blocs. Si l'élément en position sud-ouest est au-dessus de la diagonale des blocs, cela signifie que son indice de ligne est plus petit que son indice de colonne, c'est-à-dire que $i + h < i + \nu$, et donc $\nu > h$. Les quatre éléments intervenants sont de la forme :

$$\begin{array}{cc} i\delta_\lambda + g_0^2 r_{i,i+\nu} & i\delta_\lambda + g_0^2 r_{i,i+\nu+k} \\ (i+h)\delta_\lambda + g_0^2 r_{i+h,i+\nu} & (i+h)\delta_\lambda + g_0^2 r_{i+h,i+\nu+k} \end{array}$$

De sorte que la contribution du bloc $(i, i + \nu)$ est :

$$\begin{aligned} & \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left[(i\delta_{|\lambda|} + g_0^2 r_{i,i+\nu}) \left((i+h)\delta_{|\lambda|} + g_0^2 r_{i+h,i+\nu+k} \right) \right. \\ & \quad \left. + \left((i+h)\delta_{|\lambda|} + g_0^2 r_{i+h,i+\nu} \right) \left(i\delta_{|\lambda|} + g_0^2 r_{i,i+\nu+k} \right) \right] \\ &= \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left[(2i^2 + 2ih) \delta_{|\lambda|}^2 \right. \\ & \quad \left. + g_0^2 \left[(i+h)(r_{i,i+\nu} + r_{i,i+\nu+k}) + i(r_{i+h,i+\nu+k} + r_{i+h,i+\nu}) \right] \delta_{|\lambda|} \right] \end{aligned}$$

$$\begin{aligned}
& + g_0^4 [r_{i,i+\nu} r_{i+h,i+\nu+k} + r_{i,i+\nu+k} r_{i+h,i+\nu}] \\
= & h_0 s (2i^2 + 2ih) + s g_0^2 [(i+h)(r_{i,i+\nu} + r_{i,i+\nu+k}) + i(r_{i+h,i+\nu+k} + r_{i+h,i+\nu})] \\
& + s^2 g_0^4 [r_{i,i+\nu} r_{i+h,i+\nu+k} + r_{i,i+\nu+k} r_{i+h,i+\nu}]. \tag{2.38}
\end{aligned}$$

Il restera à faire une double sommation de ces derniers termes en faisant varier l'indice ν dans $\{h+1, \dots, n-k-1\}$ et pour chacune de ces valeurs de ν , l'indice i dans $\{1, \dots, n-k-\nu\}$.

Finalement, par les équations (2.37) et (2.38), la contribution totale de la partie triangulaire supérieure est :

$$\begin{aligned}
s h_0 & \left[\sum_{\nu=1}^h \sum_{i=1}^{n-k-\nu} (2i^2 + (\nu+h)i) + 2 \sum_{\nu=h+1}^{n-k-1} \sum_{i=1}^{n-k-\nu} (i^2 + ih) \right] \\
& + s g_0^4 \left[\sum_{\nu=1}^{n-k-1} \sum_{i=1}^{n-k-\nu} i (r_{i,i+\nu} + r_{i,i+\nu+k} + r_{i+h,i+\nu+k} + r_{i+h,i+\nu}) \right. \\
& \quad \left. + \sum_{\nu=1}^h \sum_{i=1}^{n-k-\nu} (\nu r_{i,i+\nu+k} + h r_{i,i+\nu}) + h \sum_{\nu=h+1}^{n-k-1} \sum_{i=1}^{n-k-\nu} (r_{i,i+\nu+k} + r_{i,i+\nu}) \right] \\
& + s^2 g_0^4 \left[\sum_{\nu=1}^{n-k-1} \sum_{i=1}^{n-k-\nu} (r_{i,i+\nu} r_{i+h,i+\nu+k} + r_{i,i+\nu+k} r_{i+h,i+\nu}) \right]. \tag{2.39}
\end{aligned}$$

L'évaluation de (2.39) sera faite à l'aide de MacSyma.

2.5.3. Contribution des blocs de la partie triangulaire inférieure

Ce cas-ci est légèrement plus complexe, car il faut considérer trois classes de diagonales.

2.5.3.1. L'élément apparaissant en position nord-est n'est pas au-dessus de la diagonale principale des blocs ($s \times s$). Supposons que l'élément nord-est est dans le bloc d'indice $(\nu+j, j)$. Les quatre blocs intervenants sont

$$\begin{array}{cc}
(\nu+j, j) & (\nu+j, j+k) \\
(\nu+h+j, j) & (\nu+h+j, j+k)
\end{array}$$

Dans ce cas, $\nu+h+j$ doit être inférieur ou égal à n , de sorte que $j \leq n-h-\nu$. Donc, $j \in \{1, \dots, n-h-\nu\}$. Or, l'élément nord-est n'est pas au-dessus de la diagonale si,

et seulement si, $\nu + j \geq j + k \Leftrightarrow \nu \geq k$. À l'intérieur du bloc, nous avons

$$\begin{array}{cc} j\delta_\lambda + g_0^2 r_{\nu+j,j} & (j+k)\delta_\lambda + g_0^2 r_{\nu+j,j+k} \\ j\delta_\lambda + g_0^2 r_{\nu+h+j,j} & (j+k)\delta_\lambda + g_0^2 r_{\nu+h+j,j+k} \end{array}$$

Le terme à considérer est :

$$\begin{aligned} & (j\delta_\lambda + g_0^2 r_{\nu+j,j}) \left((j+k)\delta_\lambda + g_0^2 r_{\nu+h+j,j+k} \right) \\ & + (j\delta_\lambda + g_0^2 r_{\nu+h+j,j}) \left((j+k)\delta_\lambda + g_0^2 r_{\nu+j,j+k} \right) \\ & = 2[jk + j^2]\delta_\lambda^2 + g_0^2 [jr_{\nu+h+j,j+k} + (k+j)r_{\nu+h+j,j} \\ & \quad + jr_{\nu+j,j+k} + (k+j)r_{\nu+j,j}] \delta_\lambda + g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}] \end{aligned}$$

D'où

$$\begin{aligned} & \sum_{\lambda=-1}^{s-1} (s - |\lambda|) \left\{ 2(jk + j^2)\delta_\lambda^2 + g_0^2 [jr_{\nu+h+j,j+k} + (k+j)r_{\nu+h+j,j} \right. \\ & \quad \left. + jr_{\nu+j,j+k} + (k+j)r_{\nu+j,j}] \delta_\lambda + g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}] \right\} \\ & = 2sh_0 [jk + j^2] + sg_0^4 [jr_{\nu+h+j,j+k} + (k+j)r_{\nu+h+j,j} + jr_{\nu+j,j+k} \\ & \quad + (k+j)r_{\nu+j,j}] + s^2 g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}] \end{aligned} \tag{2.40}$$

Il restera à faire une double sommation de ces derniers termes en faisant varier l'indice ν dans $\{k, \dots, n-h-1\}$ et, pour chacune de ces valeurs de ν , l'indice j dans $\{1, \dots, n-h-\nu\}$.

2.5.3.2. L'élément apparaissant en position nord-est est au-dessus de la diagonale principale des blocs ($s \times s$), mais celui du coin sud-est est en dessous de celle-ci. Les indices des blocs intervenants sont

$$\begin{array}{cc} (\nu + j, j) & (\nu + j, j + k) \\ (\nu + h + j, j) & (\nu + h + j, j + k) \end{array}$$

Dans ce cas-ci,

$$j + k > \nu + j \Rightarrow k > \nu \quad \text{et} \quad \nu + h + j > j + k \Rightarrow \nu > k - h$$

et donc $k - h + 1 \leq \nu \leq k - 1$. Les quatre éléments à l'intérieur du bloc sont de la forme :

$$\begin{array}{cc} j\delta_\lambda + g_0^2 r_{\nu+j,j} & (j+\nu)\delta_\lambda + g_0^2 r_{\nu+j,j+k} \\ j\delta_\lambda + g_0^2 r_{\nu+h+j,j} & (j+k)\delta_\lambda + g_0^2 r_{\nu+h+j,j+k} \end{array}$$

L'indice j doit être un élément de $\{1, \dots, n - h - \nu\}$. Le terme à considérer est :

$$\begin{aligned} & (j\delta_\lambda + g_0^2 r_{\nu+j,j}) \left((j+k)\delta_\lambda + g_0^2 r_{\nu+h+j,j+k} \right) \\ & + (j\delta_\lambda + g_0^2 r_{\nu+h+j,j}) \left((j+\nu)\delta_\lambda + g_0^2 r_{\nu+j,j+k} \right) \\ & = [j(\nu+k) + 2j^2] \delta_\lambda^2 + g_0^2 [jr_{\nu+h+j,j+k} + (\nu+j)r_{\nu+h+j,j} \\ & + jr_{\nu+j,j+k} + (k+j)r_{\nu+j,j}] \delta_\lambda + g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}]. \end{aligned}$$

D'où

$$\begin{aligned} & \sum_{\lambda=-s+1}^{s-1} (s-|\lambda|) \left\{ [j(\nu+k) + 2j^2] \delta_\lambda^2 \right. \\ & + g_0^2 [jr_{\nu+h+j,j+k} + (\nu+j)r_{\nu+h+j,j} + jr_{\nu+j,j+k} + (k+j)r_{\nu+j,j}] \delta_\lambda \\ & \left. + g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}] \right\} \\ & = sh_0 [j(\nu+k) + 2j^2] + sg_0^4 [jr_{\nu+h+j,j+k} + (\nu+j)r_{\nu+h+j,j} \\ & + jr_{\nu+j,j+k} + (k+j)r_{\nu+j,j}] + s^2 g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}]. \quad (2.41) \end{aligned}$$

Il restera à faire une double sommation de ces derniers termes en faisant varier l'indice ν dans $\{k-h+1, \dots, k-1\}$ et, pour chacune de ces valeurs de ν , l'indice j dans $\{1, \dots, n-h-\nu\}$.

2.5.3.3. L'élément apparaissant en position sud-est n'est pas au-dessous de la diagonale principale des blocs ($s \times s$). Dans ce cas-ci

$$\nu + h + j \leq j + k \Rightarrow \nu + h \leq k \Rightarrow \nu \leq k - h$$

et

$$j + k \leq n \Rightarrow j \leq n - k$$

Les quatre éléments à l'intérieur du bloc sont de la forme :

$$\begin{array}{cc} j\delta_\lambda + g_0^2 r_{\nu+j,j} & (j+\nu)\delta_\lambda + g_0^2 r_{\nu+j,j+k} \\ j\delta_\lambda + g_0^2 r_{\nu+h+j,j} & (j+\nu+h)\delta_\lambda + g_0^2 r_{\nu+h+j,j+k} \end{array}$$

L'indice j doit être un élément de $\{1, \dots, n-k\}$. Le terme à considérer est :

$$\begin{aligned} & (j\delta_\lambda + g_0^2 r_{\nu+j,j}) \left((j+\nu+h)\delta_\lambda + g_0^2 r_{\nu+h+j,j+k} \right) \\ & + (j\delta_\lambda + g_0^2 r_{\nu+h+j,j}) \left((j+\nu)\delta_\lambda + g_0^2 r_{\nu+j,j+k} \right) \\ & = [j(2\nu+h) + 2j^2] \delta_\lambda^2 + g_0^2 [jr_{\nu+h+j,j+k} + (\nu+j)r_{\nu+h+j,j} \\ & + jr_{\nu+j,j+k} + (j+\nu+h)r_{\nu+j,j}] \delta_\lambda \\ & + g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}]. \end{aligned}$$

D'où

$$\begin{aligned}
& \sum_{\lambda=-s+1}^{s-1} (s - |\lambda|) \left\{ \left[j(2v + h) + 2j^2 \right] \delta_\lambda^2 \right. \\
& \quad + g_0^2 [jr_{\nu+h+j,j+k} + (\nu + j)r_{\nu+h+j,j} + jr_{\nu+j,j+k} + (j + \nu + h)r_{\nu+j,j}] \delta_\lambda \\
& \quad \left. + g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}] \right\} \\
& = sh_0 \left[j(2v + h) + 2j^2 \right] + sg_0^4 [jr_{\nu+h+j,j+k} + (\nu + j)r_{\nu+h+j,j} \\
& \quad + jr_{\nu+j,j+k} + (j + \nu + h)r_{\nu+j,j}] \\
& \quad + s^2 g_0^4 [r_{\nu+j,j}r_{\nu+h+j,j+k} + r_{\nu+j,j+k}r_{\nu+h+j,j}]. \tag{2.42}
\end{aligned}$$

Il restera à faire une double sommation de ces derniers termes en faisant varier l'indice ν dans $\{1, \dots, k - h\}$ et, pour chacune de ces valeurs de ν , l'indice j dans $\{1, \dots, n - k\}$.

La somme des contributions explicites par les équations (2.40), (2.41) et (2.42) nous donne la contribution de la partie triangulaire supérieure.

Toutes les contributions dont il est question dans les sections 2.5.1 à 2.5.3 ont été évaluées par le programme MacSyma. L'expression finale de $\text{Cov}[c(hs), c(ks)]$ est obtenue en faisant leur somme et en multipliant celle-ci par $\sigma^4 / (n^2 s^2)$. Après simplification, on obtient

$$\begin{aligned}
& \text{Cov}[c(hs), c(ks)] \\
& = \frac{\sigma^4}{n^2 s^2} \left\{ \left[56g_0^4 - 60sh_0 \right] n^6 \right. \\
& \quad + \left[(-52k - 52h + 120)g_0^4 + (60k + 60h - 120)sh_0 \right] n^5 \\
& \quad + \left[(-60k^2 + (65h - 90)k - 60h^2 - 90h + 110) g_0^4 \right. \\
& \quad \left. + (90k^2 + (-90h + 90)k + 90h^2 + 90h - 120) sh_0 \right] n^4 \\
& \quad + \left[(50k^3 - 90k^2 + (30h^2 + 90h - 90)k + 40h^3 - 90h^2 - 80h + 60) g_0^4 \right. \\
& \quad + (-120k^3 + 90k^2 + (-180h^2 - 90h + 150)k \\
& \quad \left. - 60h^3 + 90h^2 + 90h - 60) sh_0 \right] n^3 \\
& \quad + \left[(15k^4 + (-10h + 60)k^3 + (90h^2 - 45)k^2 \right. \\
& \quad + (-10h^3 + 90h^2 + 55h - 60)k + 15h^4 + 30h^3 - 45h^2 - 30h + 14) g_0^4 \\
& \quad \left. + (30k^4 + (30h - 60)k^3 + (90h^2 - 30)k^2 + (60h^3 - 90h^2 - 90h + 60)k \right.
\end{aligned}$$

$$\begin{aligned}
& +15h^4 - 30h^3 - 15h^2 + 30h) sh_0] n^2 \\
& + [-9k^5 + (-60h^2 + 25)k^3 + (-30h^4 + 60h^2 - 14)k - 3h^5 + 5h^3] g_0^4 n \\
& + [-3hk^5 + (-20h^3 + 15h)k^3 + (-3h^5 + 15h^3 - 14h)k] g_0^4 \left. \vphantom{sh_0} \right\} \frac{1}{180n^2} \quad (2.43)
\end{aligned}$$

De plus,

$$\lim_{n \rightarrow \infty} \text{Cov} \left(\frac{c(hs)}{n}, \frac{c(ks)}{n} \right) = \frac{\sigma^4 (15sh_0 - 14g_0^4)}{45s^2}. \quad (2.44)$$

Dans le cas où $s = 1$, pour une marche aléatoire univariée où $h_0 = g_0^4$, l'équation (2.44) devient

$$\lim_{n \rightarrow \infty} \text{Cov} \left(\frac{c(h)}{n}, \frac{c(k)}{n} \right) = \frac{\sigma^4 g_0^4}{45}.$$

tel que démontré dans ROY et LEFRANÇOIS (1981), voir (1.5).

Si nous posons $h = k$ dans (2.43), nous obtenons la formule de la variance de $c(ks)$:

$$\begin{aligned}
& \text{Var}[c(ks)] \\
& = \frac{\sigma^4}{180n^4 s^2} \left\{ 4 [15sh_0 - 14g_0^4] n^6 \right. \\
& \quad - [120(k-1)sh_0 - 4(26k-30)g_0^4] n^5 \\
& \quad - [30(3k^2+6k-4)sh_0 - 5(11k^2+36k-22)g_0^4] n^4 \\
& \quad + [30(12k^3-3k^2-8k+2)sh_0 - 10(12k^3-9k^2-17k+6)g_0^4] n^3 \\
& \quad - [15(15k^4-12k^3-9k^2+6k)sh_0 \\
& \quad \quad + (100k^4+180k^3-35k^2-90k+14)g_0^4] n^2 \\
& \quad + [2(51k^5-45k^3+7k)g_0^4] n \\
& \quad \left. + [2(13k^6-15k^4+7k^2)g_0^4] \right\}.
\end{aligned}$$

2.6. Autocovariances avec des données non centrées

Dans certaines situations, on peut être amené à considérer les autocovariances échantillonnables calculées à partir des données non centrées, c'est-à-dire :

$$c'(k) = \frac{1}{N} \sum_{t=1}^{N-k} X_t X_{t+k},$$

où N est le nombre d'observations.

Il est possible de retrouver les deux premiers moments de ces autocovariances échantillonnables. Dans un tel cas, la matrice \mathbf{A} dont il est question en (2.11) devient

tout simplement :

$$\mathbf{A} = \mathbf{I}_n \otimes \mathbf{I}_s.$$

En reprenant le cheminement de la section 2.3, on constate que la matrice \mathbf{R}_n (cf. (2.15)) devient une matrice dont les éléments sont nuls. Par conséquent, $t_i = 0$, $i = 0, \dots, n-1$ (cf. (2.17)). Les équations (2.18) et (2.21) deviennent alors :

$$\mathbb{E}[c'(ms+r)] = \begin{cases} \alpha + \sigma^2 \frac{(n-m)(n-m+1)}{2n} & r = 0, \\ \alpha & 1 \leq r \leq s-1. \end{cases} \quad (2.45)$$

Notons que si les valeurs initiales sont nulles, alors $\alpha = 0$. Dans le contexte plus général de la section 2.4, où nous étudions une transformation linéaire de la marche aléatoire saisonnière, nous avons (cf. (2.27) et (2.30))

$$\mathbb{E}[c'(ms+r)] = \frac{\sigma^2 \delta_r (n-m)(s(n-m+1) - 2r)}{2ns}, \quad 0 \leq r \leq s-1.$$

Cependant, pour retrouver l'expression fournissant la covariance entre deux autocovariances calculées à partir de données non centrées, nous devons reprendre les calculs de la section 2.5 en posant $r_{i,j} = 0$, pour $1 \leq i, j \leq n$.

La contribution des blocs diagonaux devient, selon (2.36) :

$$h_0 s \sum_{i=1}^{n-k} (2i^2 + ih).$$

La contribution des blocs de la partie triangulaire supérieure, donnée par (2.39), devient :

$$h_0 s \left[\sum_{\nu=1}^h \sum_{i=1}^{n-k-\nu} (2i^2 + (\nu+h)i) + 2 \sum_{\nu=h+1}^{n-k-1} \sum_{i=1}^{n-k-\nu} (i^2 + hi) \right]$$

tandis que (2.40), (2.41) et (2.42) nous permettent d'établir que la contribution des blocs de la partie triangulaire inférieure est donnée par :

$$h_0 s \left[2 \sum_{\nu=k}^{n-h-1} \sum_{j=1}^{n-k-\nu} (j^2 + kj) + \sum_{\nu=k-h+1}^{k-1} \sum_{j=1}^{n-h-\nu} (2j^2 + (\nu+k)j) + \sum_{\nu=1}^{k-h} \sum_{j=1}^{n-k} (2j^2 + (2\nu+h)j) \right].$$

Il vaut mieux utiliser MacSyma pour compléter le travail. À l'annexe A, nous donnons la trace documentée de la séance de travail liée à l'évaluation des expressions apparaissant dans cette section. En multipliant la somme des trois contributions par

$\sigma^4 / (n^2 s^2)$, on trouve :

$$\begin{aligned} \text{Cov}[c'(hs), c'(ks)] = & \frac{\sigma^4 h_0}{12n^2 s} \left[4n^4 + (-12k + 4h + 8)n^3 \right. \\ & + (18k^2 + (-6h - 30)k - 18h^2 + 18h + 8)n^2 \\ & + (-12k^3 + 42k^2 + (24h^2 - 18h - 30)k \\ & \quad \left. + 12h^3 - 30h^2 + 14h + 4)n \right. \\ & + 2k^4 + (-2h - 20)k^3 + (22 - 6h^2)k^2 \\ & \left. + (-4h^3 + 30h^2 + 6h - 4)k - 5h^4 + 2h^3 - 19h^2 - 2h \right]. \end{aligned}$$

On obtient la variance de l'autocovariance en substituant h à k dans la dernière expression :

$$\begin{aligned} \text{Var}[c'(hs)] &= \frac{\sigma^4 h_0}{12n^2 s} \left(4n^4 + 8(1 - h)n^3 - 2(3h^2 + 6h - 4)n^2 \right. \\ & \quad \left. + 2(12h^3 - 3h^2 - 8h + 2)n - 3(5h^4 - 4h^3 - 3h^2 + 2h) \right). \end{aligned} \quad (2.46)$$

En remplaçant h par 0 dans la dernière expression, on obtient :

$$\text{Var}[c(0)] = \frac{\sigma^4 h_0 (n + 1) (n^2 + n + 1)}{3ns}.$$

2.7. Approximation des deux premiers moments de $r(ks)$

Puisque nous disposons des deux premiers moments de $c(ks)$, nous pouvons obtenir une approximation pour la moyenne et la variance de $r(ks)$. La méthode est due à KENDALL (1954) et a été reprise à plusieurs occasions, e.g. DE GOOIJER (1980) et ANDERSON et DE GOOIJER (1983). On a :

$$\text{E}[r(ks)] \approx \frac{\text{E}[c(ks)]}{\text{E}[c(0)]} \left[1 - \frac{\text{E}[c(ks)c(0)]}{\text{E}[c(ks)]\text{E}[c(0)]} + \frac{\text{E}[c^2(0)]}{\text{E}^2[c(0)]} \right] \quad (2.47)$$

$$\begin{aligned} \text{Var}[r(ks)] \approx & \frac{\text{E}^2[c(ks)]}{\text{E}^2[c(0)]} \left[\frac{\text{E}[c^2(ks)]}{\text{E}^2[c(ks)]} - \frac{2\text{E}[c(ks)c(0)]}{\text{E}[c(ks)]\text{E}[c(0)]} + \frac{\text{E}[c^2(0)]}{\text{E}^2[c(0)]} \right. \\ & \left. - \left[\frac{\text{E}[c(ks)c(0)]}{\text{E}[c(ks)]\text{E}[c(0)]} - \frac{\text{E}[c^2(0)]}{\text{E}^2[c(0)]} \right]^2 \right] \end{aligned} \quad (2.48)$$

Évidemment, l'évaluation de ces expressions peut s'avérer longue et ardue bien que l'algèbre impliquée soit élémentaire. Le programme ProgF de la section I.3

permet d'obtenir les quantités de base. En suivant les directives des deux séances de travail suivant la liste de ce programme, nous obtenons les formes explicites cherchées. Cependant, nous ne les reproduirons pas ici puisqu'il faut une page complète pour écrire la formule obtenue pour (2.47), et un peu plus de sept pages pour (2.48).

Néanmoins, l'approximation de $E[r(ks)]$ se présente comme le quotient de polynômes en n de degré 6. Le coefficient de n^6 est au numérateur et au dénominateur :

$$-135\delta_0^3 s^3 + 270\delta_0^2 g_0^2 s^2 - 180\delta_0 g_0^4 s + 40g_0^6$$

Cela signifie donc que

$$\lim_{n \rightarrow \infty} E[r(ks)] = 1.$$

De plus, en évaluant l'approximation de $\text{Var}[r(ks)]$, nous obtenons un quotient de polynômes en n . Au numérateur, il est de degré 10, tandis qu'au dénominateur, il est de degré 12. Ainsi, la variance de $r(ks)$ tend vers 0 quand n tend vers l'infini. On a donc la proposition suivante :

Proposition 2.3 : Dans ce contexte précis, le coefficient de corrélation échantillonnal d'un délai multiple de la période converge en probabilité vers 1.

Chapitre 3

Théorème de représentation

Résumé

Ce chapitre traite du théorème de représentation des processus saisonniers, un résultat fondamental permettant d'exprimer ces processus comme la somme d'une marche aléatoire saisonnière et d'un processus strictement stationnaire. Nous introduisons tout d'abord le cadre théorique en définissant le modèle sous-jacent et en précisant les conditions garantissant l'existence d'une représentation sous forme de série infinie. À partir de ces hypothèses, nous établissons une décomposition qui met en évidence le rôle des innovations et des coefficients de la structure autorégressive.

Cette représentation joue un rôle clé dans l'étude des propriétés statistiques des processus saisonniers, en particulier pour l'analyse des moments asymptotiques des autocovariances échantillonales et la caractérisation des distributions asymptotiques des coefficients d'autocorrélation. Les résultats présentés ici généraliseront les travaux précédents sur les processus autorégressifs intégrés et serviront de base aux développements ultérieurs de cette thèse.

3.1. Introduction

Considérons le modèle suivant

$$\nabla_s Z_t = \omega_t, \quad t \in \mathbb{N} \quad (3.1)$$

où ω_t peut s'exprimer sous la forme

$$\omega_t = \sum_{j=0}^{\infty} \Psi_j e_{t-j}, \quad t \in \mathbb{Z} \quad (3.2)$$

et où :

- a) e_t est une suite de variables aléatoires indépendantes et identiquement distribuées, de moyenne nulle et de variance σ^2 ,
- b) $\sum_{j=0}^{\infty} j |\Psi_j| < \infty$, avec $\Psi_0 = 1$.

On remarque qu'un processus satisfaisant (3.2), ainsi que les conditions a) et b), est stationnaire au sens large puisque

$$\sum_{j=0}^{\infty} j |\Psi_j| < \infty \implies \sum_{j=0}^{\infty} |\Psi_j| < \infty \implies \sum_{j=0}^{\infty} |\Psi_j|^2 < \infty$$

Si ω_t est un processus ARMA(p, q) stationnaire, alors il admet une représentation unique sous la forme (3.2), où $\{e_t\}$ est le processus d'innovation.

En effet, dans ce cas, ω_t satisfait

$$\varphi_p(B)\omega_t = \theta_q(B)e_t, \quad t \in \mathbb{Z}.$$

Puisque le processus est stationnaire, $\varphi_p(z)$ est un polynôme ayant toutes ses racines à l'extérieur du cercle unitaire, et il existe une unique représentation sous la forme (3.2). Nous savons que les Ψ_j satisfont l'équation aux différences

$$\varphi_p(B)\Psi_j = 0, \quad j > \max\{p-1, q\} \quad (3.3)$$

La fonction $\varphi_p(B)$ peut être exprimée comme un produit de monômes de la forme suivante :

$$\varphi_p(B) = \prod_{j=1}^{\tau} (1 - G_j B)^{e(j)},$$

où $e(j)$ représente la multiplicité associée à G_j . Évidemment, $\sum_{i=1}^{\tau} e(i) = p$. Mais puisque Ψ_j satisfait (3.3), nous pouvons le représenter comme une combinaison linéaire de termes de la forme $j^m G_h^j$, avec $0 \leq m \leq e(h) - 1$ (voir JORDÁN (1965)). Ainsi,

$$\begin{aligned} j^m |G_i^j| &\leq j^{e(i)-1} |G_i^j|, \\ \Rightarrow j^m |G_i^j| &\leq j^{p-1} |G_i^j| \\ \Rightarrow j^m |G_i^j| &\leq j^{p-1} M^j, \quad \text{où } M = \max\{|G_i| : 1 \leq i \leq \tau\} \end{aligned}$$

Utilisant la notation de Landau, on peut écrire

$$\Psi_j = O(j^{p-1} M^j)$$

et comme l'indiquent ROY et LEFRANÇOIS (1981),

$$\sum_{j=1}^{\infty} j |\Psi_j| \leq \text{constante} \sum_{j=1}^{\infty} j^{p-1} M^j < \infty$$

ce qui montre que la condition b) est satisfaite pour un processus ARMA(p, q) stationnaire. Supposons maintenant que $\{Z_{-s+1}, Z_{-s+2}, \dots, Z_0\}$ sont des valeurs initiales. Nous avons

$$\begin{aligned} Z_1 &= Z_{-s+1} + \omega_1 \\ Z_2 &= Z_{-s+2} + \omega_2 \\ &\vdots \\ Z_s &= Z_0 + \omega_s \end{aligned}$$

$$\begin{aligned} Z_{s+1} &= Z_{-s+1} + \omega_1 + \omega_{s+1} \\ Z_{s+2} &= Z_{-s+2} + \omega_2 + \omega_{s+2} \\ &\vdots \\ Z_{2s} &= Z_0 + \omega_s + \omega_{2s} \end{aligned}$$

$$\begin{aligned} Z_{2s+1} &= Z_{-s+1} + \omega_1 + \omega_{s+1} + \omega_{2s+1} \\ Z_{2s+2} &= Z_{-s+2} + \omega_2 + \omega_{s+2} + \omega_{2s+2} \\ &\vdots \\ Z_{3s} &= Z_0 + \omega_s + \omega_{2s} + \omega_{3s} \end{aligned}$$

etc.

De manière plus générale, en écrivant $t = ks + r$ avec $r \in \{1, \dots, s\}$, nous avons

$$Z_{ks+r} = Z_{-s+r} + \sum_{i=0}^k \omega_{is+r}.$$

Ainsi,

$$\mathbf{Z}_t = \mathbf{Z}_0 + \sum_{j=1}^t \boldsymbol{\omega}_j, \quad (3.4)$$

où

$$\begin{aligned} \mathbf{Z}_0 &= (Z_{-s+1}, Z_{-s+2}, \dots, Z_0)', \\ \boldsymbol{\omega}_j &= (\omega_{(j-1)s+1}, \omega_{(j-1)s+2}, \dots, \omega_{js})', \\ \mathbf{Z}_t &= (Z_{(t-1)s+1}, Z_{(t-1)s+2}, \dots, Z_{ts})'. \end{aligned}$$

Sachant que

$$\omega_m = \sum_{i=0}^{\infty} \Psi_i e_{m-i},$$

on obtient

$$\omega_{is+r} = \sum_{j=0}^{\infty} \Psi_j e_{is+r-j}.$$

Pour $r = 1, 2, \dots, s$, cette relation donne :

$$\begin{bmatrix} \omega_{is+1} \\ \omega_{is+2} \\ \vdots \\ \omega_{(i+1)s-1} \\ \omega_{(i+1)s} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \Psi_0 & \Psi_1 & \cdots & \Psi_s & \cdots \\ 0 & 0 & \cdots & \Psi_0 & \Psi_1 & \Psi_2 & \cdots & \Psi_{s+1} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \Psi_0 & \cdots & \Psi_{s-3} & \Psi_{s-2} & \Psi_{s-1} & \cdots & \Psi_{2s-2} & \cdots \\ \Psi_0 & \Psi_1 & \cdots & \Psi_{s-2} & \Psi_{s-1} & \Psi_s & \cdots & \Psi_{2s-1} & \cdots \end{bmatrix} \begin{bmatrix} e_{(i+1)s} \\ e_{(i+1)s-1} \\ \vdots \\ e_{is+1} \\ \vdots \end{bmatrix}$$

Posons

$$\begin{aligned} \mathbf{e}_j &= (e_{js}, e_{js-1}, \dots, e_{(j-1)s+1})', \\ \Psi_0 &= \begin{bmatrix} 0 & 0 & \cdots & 0 & \Psi_0 \\ 0 & 0 & \cdots & \Psi_0 & \Psi_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \Psi_0 & \cdots & \Psi_{s-3} & \Psi_{s-2} \\ \Psi_0 & \Psi_1 & \cdots & \Psi_{s-2} & \Psi_{s-1} \end{bmatrix}, \\ \Psi_k &= \begin{bmatrix} \Psi_{(k-1)s+1} & \Psi_{(k-1)s+2} & \cdots & \Psi_{ks} \\ \Psi_{(k-1)s+2} & \Psi_{(k-1)s+3} & \cdots & \Psi_{ks+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{ks} & \Psi_{ks+1} & \cdots & \Psi_{(k+1)s-1} \end{bmatrix}. \end{aligned}$$

On peut alors écrire :

$$\omega_j = \sum_{i=0}^{\infty} \Psi_i \mathbf{e}_{j-i} \quad (3.5)$$

Nous verrons plus tard que ω_j est défini avec probabilité 1, puisque nous mettrons en évidence le fait que la suite des matrices $\{\Psi_i\}_{i=0}^{\infty}$ définit un filtre linéaire sommable.

De plus,

$$\mathbf{Z}_t = \mathbf{Z}_0 + \sum_{j=1}^t \omega_j = \mathbf{Z}_0 + \sum_{j=1}^t \sum_{i=0}^{\infty} \Psi_i \mathbf{e}_{j-i}$$

Développons la double sommation :

$$\begin{aligned}
& \sum_{j=1}^t \sum_{i=0}^{\infty} \Psi_i \mathbf{e}_{j-i} \\
&= \Psi_0 \sum_{j=1}^t \mathbf{e}_j + \sum_{j=1}^t \sum_{i=1}^{\infty} \Psi_i \mathbf{e}_{j-i} \\
&= \Psi_0 \sum_{j=1}^t \mathbf{e}_j + \sum_{i=1}^{\infty} \Psi_i \left(\sum_{j=1}^t \mathbf{e}_{j-i} + \sum_{k=t-i+1}^t \mathbf{e}_k - \sum_{k=t-i+1}^t \mathbf{e}_k \right) \\
&= \Psi_0 \sum_{j=1}^t \mathbf{e}_j + \sum_{i=1}^{\infty} \Psi_i \left(\sum_{j=1}^{t+i} \mathbf{e}_{j-i} - \sum_{j=0}^{i-1} \mathbf{e}_{t-j} \right) \\
&= \Psi_0 \sum_{j=1}^t \mathbf{e}_j + \sum_{i=1}^{\infty} \Psi_i \left(\sum_{j=1-i}^t \mathbf{e}_j - \sum_{j=0}^{i-1} \mathbf{e}_{t-j} \right) \\
&= \Psi_0 \sum_{j=1}^t \mathbf{e}_j + \sum_{i=1}^{\infty} \Psi_i \left(\sum_{j=1-i}^0 \mathbf{e}_j + \sum_{j=1}^t \mathbf{e}_j - \sum_{j=0}^{i-1} \mathbf{e}_{t-j} \right) \\
&= \sum_{i=0}^{\infty} \Psi_i \sum_{j=1}^t \mathbf{e}_j + \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \Psi_i \mathbf{e}_{-j} - \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \Psi_i \mathbf{e}_{t-j} \\
&= \sum_{i=0}^{\infty} \Psi_i \sum_{j=1}^t \mathbf{e}_j + \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} \Psi_i \mathbf{e}_{-j} - \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} \Psi_i \mathbf{e}_{t-j}.
\end{aligned}$$

Ainsi,

$$\mathbf{Z}_t = \mathbf{Z}_0 + \mathbf{H}_0 \mathbf{X}_t + \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{e}_{-j} - \mathbf{D}_t, \quad t = 1, 2, \dots \quad (3.6)$$

où

$$\mathbf{H}_j = \sum_{i=j}^{\infty} \Psi_i, \quad j \geq 0, \quad (3.7)$$

$$\mathbf{D}_t = \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{e}_{t-j}, \quad t \geq 1. \quad (3.8)$$

Dans la représentation (3.6), \mathbf{X}_t est une marche aléatoire saisonnière. Nous allons maintenant montrer que \mathbf{D}_t est défini avec probabilité 1, et qu'il s'agit d'un processus stationnaire au sens strict. Cependant, pour cette démonstration, nous devons faire appel à la théorie des filtres linéaires. Nous référons le lecteur à (BRILLINGER, 1981, section 2.7).

Soit $\mathbf{A}_0, \mathbf{A}_1, \dots$, une suite de matrices réelles $s \times r$. Considérons \mathbf{Y}_t , un vecteur aléatoire de longueur s , et \mathbf{X}_t , un vecteur aléatoire de longueur r . Définissons la norme d'une matrice de la façon suivante :

Définition 3.1 (Norme). Soit \mathbf{A} une matrice $s \times r$. La norme de la matrice \mathbf{A} , notée par $\|\mathbf{A}\|$, est donnée par

$$\|\mathbf{A}\| = \sum_{i=1}^s \sum_{j=1}^r |a_{i,j}|$$

où $a_{i,j}$ est l'élément (i,j) de la matrice \mathbf{A} . \square

Définition 3.2 (Filtre linéaire sommable). Soit $\mathbf{A}_0, \mathbf{A}_1, \dots$, une suite de matrices réelles $s \times r$. Si

$$\sum_{i=0}^{\infty} \|\mathbf{A}_i\| < \infty,$$

alors la suite $\mathbf{A}_0, \mathbf{A}_1, \dots$, constitue un filtre linéaire sommable. \square

Supposons que

$$\mathbf{Y}_t = \sum_{i=0}^{\infty} \mathbf{A}_i \mathbf{X}_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots$$

Le processus \mathbf{Y}_t est alors défini comme l'image du processus \mathbf{X}_t par le filtre linéaire déterminé par les matrices \mathbf{A}_i . Selon (BRILLINGER, 1981, lemme 3.1, p. 31), ce filtre linéaire préserve certaines propriétés du processus \mathbf{X}_t et garantit l'existence de \mathbf{Y}_t .

Lemme 3.1. : Soit $\mathbf{X}_t, t = 0, \pm 1, \pm 2, \dots$, un processus stationnaire au sens strict, de dimension r . Supposons que $E[\|\mathbf{X}_t\|] < \infty$ et que $\{\mathbf{A}_i\}$ détermine un filtre linéaire sommable. Alors,

$$\mathbf{Y}_t = \sum_{i=0}^{\infty} \mathbf{A}_i \mathbf{X}_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots$$

est un processus de dimension s , défini avec probabilité 1, et stationnaire au sens strict. De plus, si pour $k > 0$, $E[\|\mathbf{X}_t\|^k] < \infty$, alors $E[\|\mathbf{Y}_t\|^k] < \infty$.

On constate ainsi que la suite $\{\Psi_j\}_{j=0}^{\infty}$ définissant la série intervenant en (3.2) est un filtre linéaire sommable. Puisque la suite $\{e_t\}_{t \geq 0}$ est un processus stationnaire au sens strict, ω_t est défini avec probabilité 1.

3.2. Représentation d'un processus saisonnier

Nous allons maintenant établir un théorème qui permet de représenter un processus saisonnier (segmenté en tranches de longueur correspondant à la période) comme la somme d'une marche aléatoire saisonnière et d'un processus strictement stationnaire. Ce résultat sera utilisé pour déterminer les deux premiers moments

asymptotiques des autocovariances échantillonnelles, ainsi que la distribution asymptotique de $ns(1 - r(ds))$, où $r(ds)$ désigne le coefficient d'autocorrélation au délai ds , calculé à partir d'une série de longueur ns . Ce théorème constitue une généralisation des équations (2.5) de ROY et LEFRANÇOIS (1981) et HASZA (1980).

Théorème 3.1. : Soit $\{Z_t\}$ un processus satisfaisant les hypothèses de la section 3.1. Soit \mathbf{Z}_t le vecteur de longueur s défini par (3.4). Alors, \mathbf{Z}_t peut s'écrire sous la forme

$$\mathbf{Z}_t = \mathbf{Z}_0 + \mathbf{H}_0 \mathbf{X}_t + \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{e}_{t-j} - \mathbf{D}_t, \quad t = 1, 2, \dots$$

où \mathbf{Z}_0 est un vecteur de valeurs initiales, \mathbf{H}_j , $j \geq 0$ sont définies par (3.7), \mathbf{X}_t est une marche aléatoire saisonnière, et \mathbf{D}_t , $t = 0, 1, 2, \dots$, est défini avec probabilité 1 et est stationnaire au sens strict. De plus, si $E[\|\mathbf{e}_t\|^k] < \infty$, alors $E[\|\mathbf{D}_t\|^k] < \infty$.

Preuve. Par définition (voir (3.8)), nous avons :

$$\mathbf{D}_t = \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{e}_{t-j}.$$

De plus, par construction, la suite $\{\mathbf{e}_t\}$ constitue une suite de vecteurs aléatoires stationnaires au sens strict. Par conséquent, $\{\mathbf{D}_t\}$ est également un processus stationnaire au sens strict.

Pour démontrer la proposition, il suffit de montrer que la suite des matrices \mathbf{H}_j constitue un filtre linéaire sommable, puis d'appliquer le lemme 3.1. Il faut donc montrer que

$$\sum_{u=1}^{\infty} \|\mathbf{H}_u\| < \infty.$$

Nous avons

$$\begin{aligned} \mathbf{H}_u &= \sum_{i=u}^{\infty} \Psi_i, \\ \Rightarrow \|\mathbf{H}_u\| &\leq \sum_{i=u}^{\infty} \|\Psi_i\|, \\ \Rightarrow \sum_{u=1}^{\infty} \|\mathbf{H}_u\| &\leq \sum_{u=1}^{\infty} \sum_{i=u}^{\infty} \|\Psi_i\| = \sum_{j=1}^{\infty} j \|\Psi_j\| = \sum_{j=1}^{\infty} j |\Psi_j| \end{aligned} \quad (3.9)$$

La dernière égalité peut sembler peu intuitive. Examinons la matrice Ψ_j .

Sur la diagonale qui s'étend du coin inférieur gauche au coin supérieur droit, on trouve Ψ_{j_s} . Cela montre que leur contribution est donnée par $j_s |\Psi_{j_s}|$.

Les éléments de la partie triangulaire supérieure gauche ont la forme $\Psi_{(j-1)s+m}$, avec $1 \leq m \leq s-1$. Ces mêmes éléments apparaissent également dans la partie triangulaire inférieure droite de la matrice Ψ_{j-1} . Par conséquent, leur contribution totale est :

$$jm \left| \Psi_{(j-1)s+m} \right| + (j-1)(s-m) \left| \Psi_{(j-1)s+m} \right|,$$

où le premier terme provient de Ψ_j et le second de Ψ_{j-1} . Donc,

$$\begin{aligned} & jm \left| \Psi_{(j-1)s+m} \right| + (j-1)(s-m) \left| \Psi_{(j-1)s+m} \right| \\ &= (jm + (j-1)(s-m)) \left| \Psi_{(j-1)s+m} \right| \\ &= ((j-1)s + m) \left| \Psi_{(j-1)s+m} \right|. \end{aligned}$$

Ce qui explique la dernière égalité apparaissant en (3.9). Par hypothèse, $\sum_{j=1}^{\infty} j \left| \Psi_j \right| < \infty$, et ainsi la suite des matrices $\{\mathbf{H}_u\}_{u \geq 1}$ détermine un filtre linéaire sommable. Cela complète la démonstration. \square

3.3. Autocovariances échantillonnables du processus saisonnier général

Nous allons maintenant utiliser les résultats relatifs à la marche aléatoire saisonnière pour déduire les moments asymptotiques des autocovariances du processus saisonnier plus général.

3.3.1. Espérance asymptotique des autocovariances d'échantillonnables

Dans ce qui suit, si \mathbf{c} est un vecteur de dimension m , on note par \bar{c} la moyenne arithmétique de ses m éléments. Posons

$$\sigma^2(\mathbf{c}) = \frac{1}{m} \sum_{i=1}^m (c_i - \bar{c})^2 \quad (3.10)$$

(Il s'agit, bien sûr, d'un abus de notation). De plus, nous désignons le terme $\mathbf{Z}_0 + \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{e}_{-j}$ par $\mathbf{\Lambda}$, qui est en fait un vecteur constant. Nous avons donc

$$\mathbf{Z}_t = \mathbf{\Lambda} + \mathbf{H}_0 \mathbf{X}_t - \mathbf{D}_t.$$

Si nous posons $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$, $\mathbf{Z} = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_n)'$ et $\mathbf{D} = (\mathbf{D}'_1, \dots, \mathbf{D}'_n)'$, nous pouvons écrire

$$\mathbf{Z} = \mathbf{1}_n \otimes \mathbf{\Lambda} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} - \mathbf{D},$$

et l'autocovariance échantillonnale de délai k de \mathbf{Z} est

$$\begin{aligned}
c_{\mathbf{Z}}(k) &= \mathbf{Z}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \mathbf{Z} / (2ns) \\
&= \{ \mathbf{1}_n \otimes \mathbf{\Lambda} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} - \mathbf{D} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \\
&\quad \times \{ \mathbf{1}_n \otimes \mathbf{\Lambda} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} - \mathbf{D} \} / (2ns) \\
&= \{ \mathbf{1}_n \otimes \mathbf{\Lambda} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \\
&\quad \times \{ \mathbf{1}_n \otimes \mathbf{\Lambda} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \} / (2ns) \\
&\quad + \mathbf{D}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \mathbf{D} / (2ns) \\
&\quad - \{ \mathbf{1}_n \otimes \mathbf{\Lambda} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \}' \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} \mathbf{D} / (ns) \\
&= \{ \mathbf{1}_n \otimes \mathbf{\Lambda} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \{ \mathbf{1}_n \otimes \mathbf{\Lambda} \} / (2ns) \\
&\quad + \{ (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \{ (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \} / (2ns) \\
&\quad + \mathbf{D}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \mathbf{D} / (2ns) \\
&\quad + \{ \mathbf{I}_n \otimes \mathbf{\Lambda} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \{ (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \} / (ns) \\
&\quad - \{ (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \}' \mathbf{A} (\mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k)) \mathbf{A} \mathbf{D} / (ns) \\
&\quad - \{ \mathbf{1}_n \otimes \mathbf{\Lambda} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \mathbf{D} / (ns) \tag{3.11} \\
&= T_1 + T_2 + T_3 + T_4 - T_5 - T_6.
\end{aligned}$$

Le premier terme, T_1 , est une quantité dépendant de \mathbf{A} , k , n et s . Tout comme nous l'avons fait pour la constante définie en (2.13), on peut montrer que T_1 est majoré par une quantité ne dépendant pas de n . Ainsi,

$$\lim_{n \rightarrow \infty} T_1/n = 0.$$

Le second terme, T_2 , a été étudié au chapitre deux, tandis que les termes T_4 et T_6 ont une espérance nulle, puisque \mathbf{X} et \mathbf{D} ont des moyennes nulles.

Le lemme suivant sera utile pour analyser le comportement du terme T_3 .

Lemme 3.2. *Soit X_1, X_2, \dots, X_n une série chronologique stationnaire dont les autocovariances $\gamma_j, j \geq 0$, satisfont*

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty.$$

Alors,

$$\text{Var}(\bar{X}) \rightarrow 0.$$

La démonstration du lemme 3.2 se trouve dans (ROY, 1980, p. 14).

Examinons maintenant le terme T_3 . Si nous convenons de désigner par d_i le i -ième élément du vecteur \mathbf{D} , la forme quadratique de ce troisième terme est donnée par

$$T_3 = \frac{1}{ns} \sum_{t=1}^{ns-k} (d_t - \bar{d}) (d_{t+k} - \bar{d}), \quad (3.12)$$

où \bar{d} représente la moyenne des d_t . La principale difficulté vient du fait que le vecteur \mathbf{D} ne peut être vu comme une série chronologique de longueur ns générée par un processus stationnaire univarié. Pour s'en convaincre, il suffit d'examiner $E[\mathbf{D}\mathbf{D}']$ et de remarquer que les autocovariances ne dépendent pas uniquement du délai.

En utilisant l'inégalité de Cauchy-Schwarz, on peut montrer que

$$\begin{aligned} |T_3| &= \frac{1}{ns} \left| \sum_{t=1}^{ns-k} (d_t - \bar{d}) (d_{t+k} - \bar{d}) \right| \\ &\leq \frac{1}{ns} \left\{ \sum_{t=1}^{ns-k} (d_t - \bar{d})^2 \right\}^{1/2} \left\{ \sum_{t=1}^{ns-k} (d_{t+k} - \bar{d})^2 \right\}^{1/2} \\ &\leq \frac{1}{ns} \left\{ \sum_{t=1}^{ns} (d_t - \bar{d})^2 \right\} = \sigma^2(\mathbf{D}) \\ &= \frac{1}{ns} \left\{ \sum_{j=1}^n \sum_{i=1}^s (d_{(j-1)s+i} - \bar{d})^2 \right\} \\ &= \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n (d_{(j-1)s+i} - \bar{d})^2 \right\} \end{aligned}$$

D'où

$$|T_3| \leq \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n (d_{(j-1)s+i} - \bar{d})^2 \right\} = \sigma^2(\mathbf{D}).$$

Cependant,

$$\bar{d} = \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n d_{(j-1)s+i} \right\} = \frac{1}{s} \sum_{i=1}^s \bar{d}_i, \quad \text{où} \quad \bar{d}_i = \frac{1}{n} \sum_{j=1}^n d_{(j-1)s+i}.$$

Ainsi,

$$\begin{aligned}
|T_3| &\leq \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n (d_{(j-1)s+i} - \bar{d})^2 \right\} \\
&= \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n \left((d_{(j-1)s+i} - \bar{d}_i) + (\bar{d}_i - \bar{d}) \right)^2 \right\} \\
&= \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n \left((d_{(j-1)s+i} - \bar{d}_i)^2 + 2(d_{(j-1)s+i} - \bar{d}_i)(\bar{d}_i - \bar{d}) \right. \right. \\
&\quad \left. \left. + (\bar{d}_i - \bar{d})^2 \right) \right\},
\end{aligned}$$

et puisque $\sum_{i=1}^s \sum_{j=1}^n (d_{(j-1)s+i} - \bar{d}_i)(\bar{d}_i - \bar{d}) = 0$, nous obtenons :

$$\begin{aligned}
|T_3| &\leq \frac{1}{s} \sum_{i=1}^s \left\{ \frac{1}{n} \sum_{j=1}^n (d_{(j-1)s+i} - \bar{d}_i)^2 \right\} + \frac{1}{s} \sum_{i=1}^s (\bar{d}_i - \bar{d})^2 \\
&= \frac{1}{s} \sum_{i=1}^s c_i(0) + \frac{1}{s} \sum_{i=1}^s (\bar{d}_i - \bar{d})^2.
\end{aligned} \tag{3.13}$$

En (3.13), $c_i(0)$ est l'autocovariance échantillonnale de délai 0 de la i -ème composante du processus vectoriel \mathbf{D}_t . Nous savons que $c_i(0)$ est un estimateur convergent (en moyenne quadratique) de $\gamma_i(0)$, voir ANDERSON (1971).

D'autre part, si nous posons =

$$\alpha_j = \begin{cases} 1 - 1/s, & j = 1 \\ -1/s, & j \neq 1 \end{cases},$$

nous avons

$$\bar{d}_i - \bar{d} = \bar{d}_i - \frac{1}{s} \sum_{j=1}^s \bar{d}_j = \frac{1}{s} \sum_{j=1}^s \alpha_j \bar{d}_j.$$

Donc,

$$(\bar{d}_i - \bar{d})^2 = \frac{1}{s^2} \left(\sum_{j=1}^s \alpha_j \bar{d}_j \right)^2 \leq \frac{1}{s^2} \left(\sum_{j=1}^s \alpha_j^2 \right) \left(\sum_{j=1}^s \bar{d}_j^2 \right) \leq \frac{1}{s} \sum_{j=1}^s \bar{d}_j^2,$$

de sorte qu'au niveau des espérances, nous avons

$$\mathbb{E} \left[(\bar{d}_i - \bar{d})^2 \right] \leq \frac{1}{s} \sum_{j=1}^s \mathbb{E} [\bar{d}_j^2],$$

et

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[(\bar{d}_i - \bar{d})^2 \right] \leq \frac{1}{s} \sum_{j=1}^s \lim_{n \rightarrow \infty} \mathbb{E} [\bar{d}_j^2].$$

Du fait que $\sum_{j=0}^{\infty} |\Psi_j| < \infty$, on conclut que $\sum_{j=0}^{\infty} |\gamma_j| < \infty$. En effet,

$$\begin{aligned} \gamma_k &= \sigma^2 \sum_{j=0}^{\infty} \Psi_j \Psi_{j+k}, \\ \text{et, } \sum_{k=0}^{\infty} |\gamma_k| &\leq \sigma^2 \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} |\Psi_j| |\Psi_{j+k}| \\ &\leq \sigma^2 \sum_{k=0}^{\infty} |\Psi_k| \sum_{j=0}^{\infty} |\Psi_j| \\ &= \sigma^2 \left(\sum_{k=0}^{\infty} |\Psi_k| \right)^2 < \infty. \end{aligned}$$

Ainsi, par le lemme 3.2, $\text{Var}[\bar{d}_j] \rightarrow 0$. Or, $\mathbb{E}[\bar{d}_j^2]$ est la variance de \bar{d}_j , \bar{d}_j étant la moyenne arithmétique d'une série chronologique stationnaire de moyenne nulle.

Reprenant (3.13), nous avons

$$|T_3| \leq \frac{1}{s} \sum_{i=1}^s c_i(0) + \frac{1}{s} \sum_{i=1}^s (\bar{d}_i - \bar{d})^2 \leq \frac{1}{s} \sum_{i=1}^s c_i(0) + \frac{1}{s^2} \sum_{i=1}^s \sum_{j=1}^s \bar{d}_j^2.$$

De sorte qu'au niveau des espérances,

$$\mathbb{E}[|T_3|] \leq \frac{1}{s} \sum_{i=1}^s \mathbb{E}[c_i(0)] + \frac{1}{s^2} \sum_{i=1}^s \sum_{j=1}^s \mathbb{E}[\bar{d}_j^2]. \quad (3.14)$$

En divisant par n et en prenant la limite, nous obtenons

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[|T_3|] = 0.$$

Notons que nous avons montré que $\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{\sigma^2(\mathbf{D})}{n} \right] = 0$.

Pour le cinquième terme, nous avons

$$\begin{aligned} T_5 &= \frac{1}{ns} \{ (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \mathbf{D} \\ &= \frac{1}{ns} \{ \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{A} \mathbf{P}_{ns}(k) \mathbf{A} \mathbf{D} + \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{A} \mathbf{P}'_{ns}(k) \mathbf{A} \mathbf{D} \}. \end{aligned}$$

Afin de simplifier les calculs, posons $\mathbf{Y} = (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X}$. Nous avons

$$\begin{aligned}
|T_5| &= \left| \frac{1}{ns} \left[\sum_{t=1}^{ns-k} (y_t - \bar{y}) (d_{t+k} - \bar{d}) + \sum_{t=1}^{ns-k} (d_t - \bar{d}) (y_{t+k} - \bar{y}) \right] \right| \\
&\leq \left| \frac{1}{ns} \sum_{t=1}^{ns-k} (y_t - \bar{y}) (d_{t+k} - \bar{d}) \right| + \left| \frac{1}{ns} \sum_{t=1}^{ns-k} (d_t - \bar{d}) (y_{t+k} - \bar{y}) \right| \\
&\leq \frac{1}{ns} \left(\left| \sum_{t=1}^{ns-k} (y_t - \bar{y})^2 \right|^{1/2} \left| \sum_{t=1}^{ns-k} (d_{t+k} - \bar{d})^2 \right|^{1/2} \right. \\
&\quad \left. + \left| \sum_{t=1}^{ns-k} (d_t - \bar{d})^2 \right|^{1/2} \left| \sum_{t=1}^{ns-k} (y_{t+k} - \bar{y})^2 \right|^{1/2} \right) \\
&\leq \frac{2}{ns} \left(\left| \sum_{t=1}^{ns} (y_t - \bar{y})^2 \right|^{1/2} \left| \sum_{t=1}^{ns} (d_t - \bar{d})^2 \right|^{1/2} \right).
\end{aligned}$$

Au niveau des espérances, en utilisant encore une fois l'inégalité de Cauchy-Schwarz, nous avons

$$\mathbb{E}[|T_5|] \leq 2 \left(\mathbb{E} \left[\frac{1}{ns} \sum_{t=1}^{ns} (y_t - \bar{y})^2 \right] \right)^{1/2} \left(\mathbb{E} \left[\frac{1}{ns} \sum_{t=1}^{ns} (d_t - \bar{d})^2 \right] \right)^{1/2},$$

d'où

$$\begin{aligned}
\frac{1}{n} \mathbb{E}[|T_5|] &\leq \frac{2}{\sqrt{n}} \left(\mathbb{E} \left[\frac{1}{n^2 s} \sum_{t=1}^{ns} (y_t - \bar{y})^2 \right] \right)^{1/2} \left(\mathbb{E} \left[\frac{1}{ns} \sum_{t=1}^{ns} (d_t - \bar{d})^2 \right] \right)^{1/2}, \\
\frac{1}{n} \mathbb{E}[|T_5|] &\leq \frac{2}{\sqrt{ns}} (\mathbb{E}[c_{\mathbf{Y}}(0)/n])^{1/2} (\mathbb{E}[c_{\mathbf{D}}(0)])^{1/2}
\end{aligned}$$

et on voit que $\mathbb{E}[|T_5|]/n$ tend vers zéro lorsque n tend vers l'infini, car $\mathbb{E}[c_{\mathbf{Y}}(0)/n]$ et $\mathbb{E}[c_{\mathbf{D}}(0)]$ convergent.

Ainsi, seul T_2 contribue à la limite de l'espérance. Comme nous l'avons montré à la section 2.4.1, voir équation (2.31),

$$\lim_{n \rightarrow \infty} \mathbb{E}[c_{\mathbf{Z}}(ms+r)/ns] = \frac{\sigma^2}{s} \left[\frac{\delta_r}{2} - \frac{g_0^2}{3s} \right].$$

3.3.2. Covariance asymptotique entre deux autocovariances échantillonnables

Posons

$$Q(\mathbf{u}, \mathbf{v}, k) = \frac{1}{2ns} \mathbf{u}' \mathbf{A} \{ \mathbf{P}_{ns}(k) + \mathbf{P}'_{ns}(k) \} \mathbf{A} \mathbf{v}, \quad (3.15)$$

où $\mathbf{A} = \mathbf{I}_n \otimes \mathbf{I}_s - (\mathbf{J}_n/n) \otimes (\mathbf{J}_s/s)$, $0 \leq k \leq ns - 1$, \mathbf{u} et \mathbf{v} étant deux vecteurs de longueur ns .

Lemme 3.3. : Soient $\mathbf{u}, \mathbf{v}, \boldsymbol{\zeta}$ et $\boldsymbol{\xi}$ des vecteurs de longueur ns . Alors,

$$|Q(\mathbf{u}, \mathbf{v}, k)Q(\boldsymbol{\zeta}, \boldsymbol{\xi}, h)| \leq \sigma(\mathbf{u})\sigma(\mathbf{v})\sigma(\boldsymbol{\zeta})\sigma(\boldsymbol{\xi}),$$

où $\sigma(\mathbf{x})$ est défini en (3.10).

Preuve. Il suffit d'utiliser l'inégalité de Cauchy-Schwarz. En effet,

$$\begin{aligned} & |Q(\mathbf{u}, \mathbf{v}, k)Q(\boldsymbol{\zeta}, \boldsymbol{\xi}, h)| \\ &= \frac{1}{(ns)^2} \left| \sum_{i=1}^{ns-k} (u_i - \bar{u})(v_{i+k} - \bar{v}) \sum_{i=1}^{ns-h} (\zeta_i - \bar{\zeta})(\xi_{i+h} - \bar{\xi}) \right| \\ &\leq \frac{1}{(ns)^2} \left[\sum_{i=1}^{ns-k} (u_i - \bar{u})^2 \sum_{i=1}^{ns-k} (v_{i+k} - \bar{v})^2 \sum_{i=1}^{ns-h} (\zeta_i - \bar{\zeta})^2 \sum_{i=1}^{ns-h} (\xi_{i+h} - \bar{\xi})^2 \right]^{1/2} \\ &\leq \frac{1}{(ns)^2} \left[\sum_{i=1}^{ns} (u_i - \bar{u})^2 \sum_{i=1}^{ns} (v_i - \bar{v})^2 \sum_{i=1}^{ns} (\zeta_i - \bar{\zeta})^2 \sum_{i=1}^{ns} (\xi_i - \bar{\xi})^2 \right]^{1/2} \\ &= \sigma(\mathbf{u})\sigma(\mathbf{v})\sigma(\boldsymbol{\zeta})\sigma(\boldsymbol{\xi}). \end{aligned}$$

□

Corollaire 3.1. Si $\boldsymbol{\eta}$ est un vecteur de longueur ns et s'il existe un vecteur $\boldsymbol{\tau}$ de longueur s tel que $\boldsymbol{\eta} = \mathbf{1}_n \otimes \boldsymbol{\tau}$, alors $\sigma(\boldsymbol{\eta}) = \sigma(\boldsymbol{\tau})$.

Preuve. Un calcul direct montre que

$$\begin{aligned} \sigma^2(\boldsymbol{\eta}) &= \frac{1}{ns} \sum_{i=1}^{ns} (\eta_i - \bar{\eta})^2 \\ &= \frac{1}{ns} \sum_{j=1}^n \sum_{i=1}^s (\eta_{(j-1)s+i} - \bar{\eta})^2 \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{s} \sum_{i=1}^s (\tau_i - \bar{\tau})^2 \\ &= \sigma^2(\boldsymbol{\tau}). \end{aligned}$$

Ce qui termine la démonstration du corollaire 3.1. □

Maintenant, puisque par la relation (3.11),

$$\begin{aligned} c_{\mathbf{Z}}(k) &= Q(\mathbf{I}_n \otimes \mathbf{A}, \mathbf{I}_n \otimes \mathbf{A}, k) + Q(\mathbf{Y}, \mathbf{Y}, k) + Q(\mathbf{D}, \mathbf{D}, k) \\ &\quad + 2Q(\mathbf{I}_n \otimes \mathbf{A}, \mathbf{Y}, k) - 2Q(\mathbf{Y}, \mathbf{D}, k) - 2Q(\mathbf{I}_n \otimes \mathbf{A}, \mathbf{D}, k), \end{aligned}$$

le produit $c_{\mathbf{Z}}(k)c_{\mathbf{Z}}(h)$ est une somme de 36 termes de la forme $Q(\cdot, \cdot, k) \times Q(\cdot, \cdot, h)$.

Pour l'instant, nous écrivons

$$c_{\mathbf{Z}}(k)c_{\mathbf{Z}}(h) = Q(\mathbf{Y}, \mathbf{Y}, k)Q(\mathbf{Y}, \mathbf{Y}, h) + S(h, k, n).$$

Il est possible de montrer que

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[S(n, k, n)]}{n^2} = 0$$

Cela se fait en considérant chacun des 35 termes impliqués. Il est évident que le premier terme, c'est-à-dire, $Q(\mathbf{1}_n \otimes \mathbf{A}, \mathbf{1}_n \otimes \mathbf{A}, k) Q(\mathbf{1}_n \otimes \mathbf{A}, \mathbf{1}_n \otimes \mathbf{A}, h) / n^2 \leq \sigma^4(\mathbf{A}) / n^2$, converge vers 0 lorsque n tend vers l'infini. Pour le terme $Q(\mathbf{D}, \mathbf{D}, k) Q(\mathbf{D}, \mathbf{D}, h)$, nous avons que

$$\begin{aligned} \frac{|Q(\mathbf{D}, \mathbf{D}, k) Q(\mathbf{D}, \mathbf{D}, h)|}{n^2} &\leq \sigma^4(\mathbf{D}), \\ \Rightarrow \frac{|Q(\mathbf{D}, \mathbf{D}, k) Q(\mathbf{D}, \mathbf{D}, h)|}{n^2} &\leq \left(\frac{\sigma^2(\mathbf{D})}{n} \right)^2 \end{aligned}$$

et en étudiant T_3 , nous avons montré que $\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{\sigma^2(\mathbf{D})}{n} \right] = 0$. Par conséquent,

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}[Q(\mathbf{D}, \mathbf{D}, k) Q(\mathbf{D}, \mathbf{D}, h)]}{n} = 0.$$

Pour le terme $Q(\mathbf{1}_n \otimes \mathbf{A}, \mathbf{1}_n \otimes \mathbf{A}, k) Q(\mathbf{Y}, \mathbf{Y}, h)$, puisque $Q(\mathbf{1}_n \otimes \mathbf{A}, \mathbf{1}_n \otimes \mathbf{A}, k) = \sigma^2(\mathbf{A})$ est une constante indépendante de n , et puisque

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Y}}(0)}{n} \right] = \sigma^2 \left(\frac{\delta_0}{2} - \frac{g_0^2}{3s} \right)$$

on voit aisément que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{Q(\mathbf{1}_n \otimes \mathbf{A}, \mathbf{1}_n \otimes \mathbf{A}, k) Q(\mathbf{Y}, \mathbf{Y}, h)}{n^2} \right] = 0$$

En ce qui concerne le terme $Q(\mathbf{1}_n \otimes \mathbf{1}_n \otimes \mathbf{A}, k) Q(\mathbf{D}, \mathbf{D}, h)$, il suffit d'appliquer un argument analogue. Pour le produit $Q(\mathbf{Y}, \mathbf{Y}, k) Q(\mathbf{D}, \mathbf{D}, h)$, nous avons

$$|Q(\mathbf{Y}, \mathbf{Y}, k) Q(\mathbf{D}, \mathbf{D}, h)| \leq \sigma^2(\mathbf{Y}) \sigma^2(\mathbf{D}) = c_{\mathbf{Y}}(0) c_{\mathbf{D}}(0),$$

et donc

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}[|Q(\mathbf{Y}, \mathbf{Y}, k) Q(\mathbf{D}, \mathbf{D}, h)|] \\ \leq \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \mathbb{E}[|c_{\mathbf{Y}}(0)|^2 / n] \mathbb{E}[|c_{\mathbf{D}}(0)|^2] \right\}^{1/2} = 0. \end{aligned}$$

Et, pour donner un dernier exemple, pour le terme $Q(\mathbf{Y}, \mathbf{D}, k) Q(\mathbf{Y}, \mathbf{D}, h)$, nous avons

$$|Q(\mathbf{Y}, \mathbf{D}, k) Q(\mathbf{Y}, \mathbf{D}, h)| \leq \sigma^2(\mathbf{Y}) \sigma^2(\mathbf{D}) = c_{\mathbf{Y}}(0) c_{\mathbf{D}}(0),$$

et donc

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}[|Q(\mathbf{Y}, \mathbf{D}, k)Q(\mathbf{Y}, \mathbf{D}, h)|] \\ \leq \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \mathbb{E}[|c_{\mathbf{Y}}(0)|^2/n] \mathbb{E}[|c_{\mathbf{D}}(0)|^2] \right\}^{1/2} = 0. \end{aligned}$$

Pour chacun des 35 produits qui composent la somme $S(h, k, n)$, nous pourrions démontrer que leur espérance, une fois divisée par n^2 , converge vers 0 lorsque $n \rightarrow \infty$. Sachant de plus que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Z}}(k)}{n} \right] = \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Y}}(k)}{n} \right],$$

on conclut donc que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Z}}(k)}{n} \right] \mathbb{E} \left[\frac{c_{\mathbf{Z}}(h)}{n} \right] = \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Y}}(k)}{n} \right] \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Y}}(h)}{n} \right],$$

et par conséquent,

$$\lim_{n \rightarrow \infty} \text{Cov} \left[\frac{c_{\mathbf{Z}}(ks)}{n}, \frac{c_{\mathbf{Z}}(hs)}{n} \right] = \frac{\sigma^4 \{15sh_0 - 14g_0^4\}}{45s^2}.$$

Nous allons donc résumer le tout dans le théorème 3.2.

Théorème 3.2. *Soit Z_1, \dots, Z_{ns} une série chronologique de longueur ns générée par un processus stochastique tel que*

$$\nabla_s Z_t = \omega_t, \quad t \in \mathbb{N}$$

où ω_t peut s'exprimer sous la forme

$$\omega_t = \sum_{j=0}^{\infty} \Psi_j e_{t-j}, \quad t \in \mathbb{Z}$$

et où

- i) $\{e_t\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, de moyenne nulle et de variance σ^2 ,
- ii) $\sum_{0 \leq j} |\Psi_j| < \infty$, $\Psi_0 = 1$.

Alors :

- a) $\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{c_{\mathbf{Z}}(ms+r)}{n} \right] = \sigma^2 \left(\frac{\delta_r}{2} - \frac{g_0^2}{3s} \right), \quad r = 0, \dots, s-1,$
- b) $\lim_{n \rightarrow \infty} \text{Cov} \left[\frac{c_{\mathbf{Z}}(ks)}{n}, \frac{c_{\mathbf{Z}}(hs)}{n} \right] = \frac{\sigma^4 (15sh_0 - 14g_0^4)}{45s^2}.$

Les constantes δ_r , g_0^2 et h_0 sont liées à la matrice \mathbf{H}_0 apparaissant dans la représentation du théorème 3.1, à savoir : δ_r est l'élément de la r -ième diagonale

supérieure (et inférieure) de \mathbf{H}_0^2 , g_0^2 est la somme des éléments d'une ligne de \mathbf{H}_0^2 , tandis que h_0 est la somme des carrés des éléments d'une ligne de \mathbf{H}_0^2 .

Notons que le théorème 3.2 nous indique que les deux premiers moments asymptotiques des covariances échantillonnales du processus $\{Z_t\}$ sont les mêmes que ceux de la transformation linéaire de la marche aléatoire des sections 2.4 et 2.5. Par conséquent, les résultats de la section 2.7 s'appliquent également au coefficient de corrélation du délai ks , c'est-à-dire, $r_Z(ks)$.

3.3.3. Autocovariances avec données non centrées

Si nous considérons les autocovariances échantillonnales calculées à partir des données non centrées, il faut remplacer la matrice \mathbf{A} de (3.11) par \mathbf{I}_{ns} . Ainsi, dans l'équation (3.12) où nous explicitons T_3 , le terme \bar{d} disparaît, et seule la première sommation demeure dans (3.13), ce qui est également vrai pour l'équation (3.14). Par conséquent, $E[|T_3|/n]$ tend vers zéro quand n tend vers l'infini.

En ce qui concerne la discussion relative au terme T_5 , les termes \bar{y} et \bar{d} disparaissent et le fait que $E[|T_5|]/n$ tend vers zéro quand n tend vers l'infini demeure vrai. On conclut donc que seul le terme T_2 contribue à l'espérance, et nous avons

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left[\frac{c_Z(ms+r)}{ns} \right] &= \lim_{n \rightarrow \infty} \frac{\sigma^2 \delta_r (n-m)(s(n-m+1) - 2r)}{2n^2 s^2} \\ &= \frac{\sigma^2 \delta_r}{2s}, \quad r = 0, \dots, s-1 \end{aligned} \quad (3.16)$$

En ce qui a trait à la covariance asymptotique entre deux autocovariances échantillonnales, nous devons remplacer la matrice \mathbf{A} apparaissant dans la définition de $Q(\mathbf{u}, \mathbf{v}, k)$ (cf. équation (3.15)), par \mathbf{I}_{ns} . Aussi, dans (3.10), le terme $\bar{\omega}$ disparaît. Avec ces nouvelles définitions, le lemme 3.3 et le corollaire 3.1 demeurent valides. Une discussion analogue à celle que nous avons faite permet alors de conclure, en utilisant (2.46), que

$$\lim_{n \rightarrow \infty} \text{Cov} \left[\frac{c_Z(ks)}{n}, \frac{c_Z(hs)}{n} \right] = \frac{\sigma^4 h_0}{3s} \quad (3.17)$$

Cela nous indique donc que le théorème 3.2 demeure valide pour des autocovariances échantillonnales basées sur des données non centrées si nous remplaçons la conclusion a) par (3.16) et la conclusion b) par (3.17).

Chapitre 4

Les résultats de Dickey

Résumé

Ce chapitre présente les travaux fondamentaux de DICKEY (1976) et de ses collaborateurs, qui ont contribué à la compréhension des processus intégrés d'ordre un. L'objectif est d'exposer ces résultats en fournissant des détails suffisants pour permettre au lecteur de suivre les démonstrations et d'en saisir les implications. Nous nous appuyons notamment sur des notions d'ordre en probabilité et de convergence en loi, dont une présentation détaillée est disponible dans les ouvrages de FULLER (1976) et BISHOP ET COLLAB. (1975).

Nous étudions en particulier trois variables fondamentales associées à la marche aléatoire : la somme normalisée des carrés des observations, le dernier terme de la série normalisée, et la moyenne des observations normalisée. En diagonalisation la forme quadratique associée, nous exprimons ces variables en termes de variables canoniques, ce qui permet d'établir leur distribution asymptotique et de mieux comprendre leur comportement pour de grandes valeurs de n .

4.1. Introduction

Le but de ce chapitre est de présenter une partie des travaux de DICKEY (1976) qui ont été publiés dans DICKEY et FULLER (1979). Nous présenterons ces résultats en donnant suffisamment de détails afin que le lecteur puisse suivre aisément ce que nous présenterons par la suite. Cependant, certains résultats concernant l'ordre en probabilité d'une suite de variables aléatoires seront utilisés. Ces résultats sont détaillés dans (FULLER, 1976, chap. 5), et des informations supplémentaires peuvent être trouvées dans (BISHOP ET COLLAB., 1975, chap. 14).

4.2. Variables de base

Nous nous intéressons au modèle suivant

$$X_t = X_{t-1} + e_t, \quad t \geq 1 \quad (4.1)$$

où $\mathbf{X}_0 = 0$ et où $\{e_t\}_{t \geq 1}$ est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$. Trois variables fondamentales sont étudiées :

$$\begin{aligned} \Gamma_n &= \sum_{t=1}^n X_t^2/n^2 \\ T_n &= X_n/\sqrt{n} \\ W_n &= \bar{X}/\sqrt{n}. \end{aligned}$$

Dans un premier temps, nous diagonaliserons la forme quadratique Γ_n , c'est-à-dire que nous l'exprimerons sous forme canonique (et nous appellerons « *variables canoniques* » les variables y apparaissant). Par la suite, nous écrirons les variables T_n et W_n en termes des variables canoniques, ce qui nous permettra d'obtenir la distribution asymptotique du vecteur $(\Gamma_n, T_n, W_n)'$.

4.2.1. Diagonalisation de Γ_n

Considerons

$$n^2 \Gamma_n = \sum_{t=1}^n X_t^2$$

Puisque

$$X_t = \sum_{j=1}^t e_j$$

nous avons

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{T}_n \mathbf{e} \quad (4.2)$$

et par conséquent,

$$\sum_{t=1}^n X_t^2 = \mathbf{e}' \mathbf{T}_n' \mathbf{T}_n \mathbf{e}$$

et on se convainc aisément que

$$\mathbf{T}'_n \mathbf{T}_n = \begin{bmatrix} n & n-1 & n-2 & \dots & 2 & 1 \\ n-1 & n-1 & n-2 & \dots & 2 & 1 \\ n-2 & n-2 & n-2 & \dots & 2 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 2 & 2 & \dots & 2 & 1 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix} = \mathbf{A}_n$$

et

$$-\mathbf{A}_n^{-1} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 1 & -2 & 1 & \dots & 0 & 0 \\ 0 & 1 & -2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -2 & 1 \\ 0 & 0 & 0 & \dots & 1 & -2 \end{bmatrix}$$

Dickey met en évidence les valeurs propres de la matrice \mathbf{A}_n (de fait, il obtient celles de \mathbf{A}_{n-1} , qu'il dénote par \mathbf{A}_n), en utilisant un résultat de RUTHERFORD (1946) montrant que le déterminant de la matrice $m \times m$

$$\mathbf{R}_m(x, a, b) = \begin{bmatrix} x+b & 1 & 0 & \dots & 0 & 0 \\ 1 & x & 1 & \dots & 0 & 0 \\ 0 & 1 & x & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 & \vdots \\ 0 & 0 & 0 & \dots & x & 1 \\ 0 & 0 & 0 & \dots & 1 & x+a \end{bmatrix}$$

est

$$|\mathbf{R}_m(x, a, b)| = \frac{\sin(m+1)\theta + (a+b)\sin(m\theta) + ab\sin((m-1)\theta)}{\sin(\theta)}$$

où $x = 2 \cos(\theta)$.

Pour trouver les valeurs propres de \mathbf{A}_n on doit chercher λ tel que

$$|\mathbf{A}_n - \lambda \mathbf{I}_n| = 0$$

Puisque les valeurs propres de \mathbf{A}_n sont différentes de zéro, nous pouvons écrire

$$|-\mathbf{A}_n^{-1} + \lambda^{-1} \mathbf{I}_n| = 0$$

et ainsi

$$|-\mathbf{A}_n^{-1} + \lambda^{-1} \mathbf{I}_n| = |\mathbf{R}_n(-2 + \lambda^{-1}, 0, 1)|$$

Par conséquent, $|\mathbf{R}_n(-2 + \lambda^{-1}, 0, 1)| = 0$, si et seulement si

$$\frac{\sin((n+1)\theta) + \sin(n\theta)}{\sin(\theta)} = 0$$

avec $-2 + \lambda^{-1} = 2 \cos(\theta)$. Or,

$$\sin(p) + \sin(q) = 2 \sin\left(\frac{p+q}{2}\right) \cos\left(\frac{p-q}{2}\right).$$

Cela nous indique que

$$\sin((n+1)\theta) + \sin(n\theta) = 2 \sin\left(\frac{(2n+1)\theta}{2}\right) \cos\left(\frac{\theta}{2}\right).$$

On cherche donc θ tel que

$$2 \sin\left(\frac{(2n+1)\theta}{2}\right) \cos\left(\frac{\theta}{2}\right) / \sin(\theta) = 0.$$

Avec $\theta \in (0, \pi)$, on a $\sin(\theta) \neq 0$, $\cos\left(\frac{\theta}{2}\right) \neq 0$ et $\sin\left(\frac{(2n+1)\theta}{2}\right) = 0$ si et seulement si

$$\theta = \theta_k = \frac{2k\pi}{2n+1}, \quad k = 1, \dots, n.$$

Dénotons par $\lambda_{n,1}$ la valeur propre associée à $\theta_{n-i+1} = 2(n-i+1)\pi/(2n+1)$. Dans l'énoncé du résultat de Rutherford, x et θ doivent satisfaire $x = 2 \cos(\theta)$. Dans notre cas, cela donne

$$\begin{aligned} -2 + \lambda_{n,i}^{-1} = 2 \cos(\theta_{n-i+1}) &\Leftrightarrow \lambda_{n,i}^{-1} = 2(1 + \cos(\theta_{n-i+1})) \\ &\Leftrightarrow \lambda_{n,i}^{-1} = 2\left(1 + 2 \cos^2(\theta_{n-i+1}/2) - 1\right) \\ &\Leftrightarrow \lambda_{n,i}^{-1} = 4 \cos^2(\theta_{n-i+1}/2) \\ &\Leftrightarrow \lambda_{n,i} = \frac{1}{4 \cos^2\left(\frac{(n-i+1)\pi}{2n+1}\right)} \end{aligned}$$

$i = 1, \dots, n$.

Puisque $\text{tr}(\mathbf{A}_n) = n(n+1)/2 = O(n^2)$ les valeurs propres de \mathbf{A}_n doivent aussi être d'ordre $O(n^2)$. Nous diviserons $\lambda_{n,i}$ par n^2 afin de normaliser les valeurs propres.

Évaluons donc

$$\lim_{n \rightarrow \infty} \frac{\sqrt{\lambda_{n,i}}}{n} = \lim_{n \rightarrow \infty} \frac{1}{\cos\left(\frac{(n-i+1)\pi}{2n+1}\right) \cdot (2n)} = \lim_{n \rightarrow \infty} \frac{1}{\cos(x_n) \cdot (2n)}$$

où $x_n = ((n - i + 1)\pi)/(2n + 1)$. Évidemment, $\lim_{n \rightarrow \infty} x_n = \frac{\pi}{2}$. Ainsi, pour n très grand, x_n est très près de $\pi/2$. Réécrivons l'expression sous une forme plus simple :

$$\frac{1}{\cos(x_n) \cdot (2n)} = \frac{1}{\cos\left(\frac{(n-i+1)\pi}{2n+1}\right)} \cdot \frac{1}{2n}.$$

Notons que

$$x_n = \frac{\pi}{2} - \frac{(2i-1)\pi}{2(2n+1)}.$$

Puisque $\cos(x)$ est proche de $\sin(x)$ pour des arguments proches de $\frac{\pi}{2}$, pour n grand, nous avons que

$$\cos(x_n) \approx \sin\left(\frac{(2i-1)\pi}{2(2n+1)}\right).$$

Pour des petites valeurs de $\varepsilon = \frac{(2i-1)\pi}{2(2n+1)}$, nous savons que $\sin(\varepsilon) \approx \varepsilon$. Par conséquent :

$$\cos(x_n) \approx \frac{(2i-1)\pi}{2(2n+1)}.$$

Ainsi,

$$\frac{1}{\cos(x_n) \cdot (2n)} \approx \frac{2(2n+1)}{(2i-1)\pi \cdot 2n}.$$

Quand $n \rightarrow \infty$, $2n + 1 \approx 2n$, donc :

$$\lim_{n \rightarrow \infty} \frac{2(2n+1)}{(2i-1)\pi \cdot 2n} = \frac{2}{(2i-1)\pi}.$$

Finalement, nous avons montré que :

$$\lim_{n \rightarrow \infty} \frac{\lambda_{n,i}}{n^2} = \frac{4}{\pi^2(2i-1)^2} = \gamma_i^2$$

où $\gamma_i = 2(-1)^{i+1}/(\pi(2i-1))$.

Soit $\mathbf{x}_{n,i}$, le i -ième vecteur propre de \mathbf{A}_n et soit $x_{n,i,t}$, la t -ième composante de $\mathbf{x}_{n,i}$. Nous avons

$$\begin{aligned} \mathbf{A}_n \mathbf{x}_{n,i} &= \lambda_{n,i} \mathbf{x}_{n,i} \\ \Rightarrow \mathbf{x}_{n,i} &= \lambda_{n,i}^{-1} \mathbf{A}_n \mathbf{x}_{n,i}. \end{aligned}$$

En reprenant le raisonnement de Dickey avec $(n-1)$ remplacé par n , on trouve

$$x_{n,i,t} = \cos(t\theta_i) + (4\lambda_{n,i} - 1)^{-1/2} \sin(t\theta_i), \quad t = 1, 2, \dots, n$$

où $\theta_i = \arccos(1 - 1/(2\lambda_{n,i}))$. Dickey montre de plus que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n x_{n,i,t}^2 = \frac{1}{2},$$

c'est-à-dire,

$$\lim_{n \rightarrow \infty} \frac{\|\mathbf{x}_{n,i}\|}{\sqrt{n}} = \frac{1}{\sqrt{2}}. \quad (4.3)$$

Cela signifie que $\|\mathbf{x}_{n,i}\| = O(n^{1/2})$. Maintenant, puisque $\{\mathbf{x}_{n,1}, \mathbf{x}_{n,2}, \dots, \mathbf{x}_{n,n}\}$ sont les vecteurs propres d'une matrice symétrique, en posant

$$\boldsymbol{\xi}_i = \mathbf{x}_{n,i} / \|\mathbf{x}_{n,i}\|,$$

les vecteurs $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n\}$ fournissent une base orthonormale de \mathbb{R}^n (cf. (SEARLE, 1982, section 11.6)). D'après l'équation (3.1) de DICKEY et FULLER (1979), $\xi_{i,t}$, la t -ième composante de $\boldsymbol{\xi}_i$, est donnée par

$$\xi_{i,t} = \frac{2}{(2n+1)^{1/2}} \cos \left[\frac{(2t-1)(2i-1)\pi}{2(2n+1)} \right]$$

Soit $\mathbf{E} = [\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n]$, alors \mathbf{E} est orthogonale et, pour tout $\mathbf{e} \in \mathbb{R}^n$, il existe $\mathbf{Z} \in \mathbb{R}^n$ tel que

$$\mathbf{e} = \mathbf{E}\mathbf{Z}$$

$$\text{et } \mathbf{Z} = \mathbf{E}^{-1}\mathbf{e} = \mathbf{E}'\mathbf{e}$$

Posons $\mathbf{Z} = (Z_{n,1}, \dots, Z_{n,n})'$, l'indice n étant introduit pour indiquer que \mathbf{Z} dépend de n . On a donc

$$Z_{n,i} = \frac{1}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n x_{n,i,t} e_t$$

Cette relation est valide en particulier pour le vecteur \mathbf{e} de (4.2). Sachant que $\mathbf{E}'\mathbf{A}_n\mathbf{E} = \mathbf{D}_1 = \text{Diag}\{\lambda_{n,1}, \lambda_{n,2}, \dots, \lambda_{n,n}\}$, nous voyons que

$$\begin{aligned} \sum_{t=1}^n X_t^2 &= \mathbf{e}'\mathbf{A}_n\mathbf{e} \\ &= \mathbf{Z}'\mathbf{E}'\mathbf{A}_n\mathbf{E}\mathbf{Z} \\ &= \sum_{i=1}^n \lambda_{n,i} Z_{n,i}^2 \end{aligned}$$

De plus,

$$\mathbf{Z} = \mathbf{E}'\mathbf{e} \Rightarrow \text{E}[\mathbf{Z}] = \mathbf{0}_n \quad \text{et} \quad \text{Var}[\mathbf{Z}] = \sigma^2 \mathbf{I}_n \quad (4.4)$$

On peut donc résumer ainsi : il existe un vecteur $\mathbf{Z} : \mathcal{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$, tel que

$$\Gamma_n = \sum_{i=1}^n \frac{\lambda_{n,i}}{n^2} Z_{n,i}^2$$

et

$$\lim_{n \rightarrow \infty} \frac{\lambda_{n,i}}{n^2} = \frac{4}{\pi^2(2i-1)^2} = \gamma_1^2$$

4.2.2. Structure de T_n

Puisque $T_n = X_n/\sqrt{n}$, on a que

$$\begin{aligned} T_n &= \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{i=1}^n \frac{x_{n,i,t}}{\|\mathbf{x}_{n,i}\|} Z_{n,i} \\ &= \sum_{i=1}^n \left[\sum_{t=1}^n \frac{x_{n,i,t}}{\sqrt{n} \|\mathbf{x}_{n,i}\|} \right] Z_{n,i} \end{aligned}$$

par (DICKY, 1976, p. 31, équation (4.25)),

$$\begin{aligned} \mathbf{t}_n &= \sum_{i=1}^n \left[\frac{x_{n,i,t}}{\sqrt{n} \|\mathbf{x}_{n,i}\|} \cos\left(\frac{(n+1)\theta_i}{2}\right) \sin\left(\frac{n\theta_i}{2}\right) \csc\left(\frac{\theta_i}{2}\right) \right] Z_{n,i} \\ &\quad + O_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Nous dénoterons par $\alpha_{n,i}$ le terme entre crochets. Cela nous permet d'écrire

$$T_n = \sum_{i=1}^n \alpha_{n,i} Z_{n,i} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

De plus, par (DICKY, 1976, p. 31, équation (4.27)),

$$\lim_{n \rightarrow \infty} \alpha_{n,i} = \frac{2\sqrt{2}(-1)^{i+1}}{\pi(2i-1)} = \sqrt{2}\gamma_i$$

4.2.3. Structure de W_n

Puisque $W_n = \bar{X}/\sqrt{n}$, on a

$$\begin{aligned}
W_n &= n^{-3/2} \sum_{i=1}^n X_i \\
&= n^{-3/2} \sum_{i=1}^n \sum_{j=1}^i e_j \\
&= n^{-3/2} \sum_{t=1}^n (n-t+1) e_t \\
&= n^{-3/2} \sum_{t=1}^n (n-t+1) \sum_{i=1}^n \frac{x_{n,i,t}}{\|\mathbf{x}_{n,i}\|} Z_{n,i} \\
&= \sum_{i=1}^n \left[\frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n (n-t+1) x_{n,i,t} \right] Z_{n,i}
\end{aligned}$$

Dans un premier temps, (DICKY, 1976, p. 32) montre que

$$\begin{aligned}
\frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n (n-t+1) x_{n,i,t} &= \frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} \left[\sum_{t=1}^n (n+1) \cos(t\theta_i) - \sum_{t=1}^n t \cos(t\theta_i) \right] \\
&\quad + O\left(\frac{1}{n}\right)
\end{aligned}$$

et en utilisant les identités que l'on retrouve dans JOLLEY (1961), à savoir,

$$\begin{aligned}
\sum_{t=1}^n \cos(t\theta_i) &= \sin\left(\frac{n\theta_i}{2}\right) \cos\left(\frac{(n+1)\theta_i}{2}\right) \csc\left(\frac{\theta_i}{2}\right) = a_{n,i}, \\
\sum_{t=1}^n t \cos(t\theta_i) &= \frac{(n+1) \sin\left(\left(n+\frac{1}{2}\right)\theta_i\right) \csc\left(\frac{1}{2}\theta_i\right)}{2} - \frac{1 - \cos((n+1)\theta_i)}{4 \sin^2\left(\frac{1}{2}\theta_i\right)} \\
&= b_{n,i} - c_{n,i}
\end{aligned}$$

Il nous convainc que

$$\begin{aligned}
&\frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n (n-t+1) x_{n,i,t} \\
&= \frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} [(n+1)a_{n,i} - b_{n,i} + c_{n,i}] + O\left(\frac{1}{n}\right) \\
&= \frac{n^{-1/2}(n+1)}{\|\mathbf{x}_{n,i}\|} \left[\frac{a_{n,i}}{n} - \frac{b_{n,i}}{n(n+1)} + \frac{c_{n,i}}{n(n+1)} \right] + O\left(\frac{1}{n}\right) \\
&= \beta_{n,i} + O\left(\frac{1}{n}\right)
\end{aligned}$$

Il démontre par la suite (page 33, équation (4.36) et suivantes) que :

- (1) $\frac{a_{n,i}}{n}$ tend vers γ_i
- (2) $\frac{b_{n,i}}{n(n+1)}$ tend vers γ_1
- (3) $\frac{c_{n,i}}{n(n+1)}$ tend vers γ_i^2

ce qui lui permet d'obtenir que

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n (n-t+1)x_{n,i,t} \\ &= \lim_{n \rightarrow \infty} \frac{n^{-1/2}}{\|\mathbf{x}_{n,i}\|} (n+1) \left[\frac{a_{n,i}}{n} - \frac{b_{n,i}}{n(n+1)} + \frac{c_{n,i}}{n(n+1)} \right] \end{aligned}$$

et en utilisant (4.3),

$$\lim_{n \rightarrow \infty} \frac{n+1}{\sqrt{n} \|\mathbf{x}_{n,i}\|} \gamma_i^2 = \sqrt{2} \gamma_i^2.$$

Nous résumons comme suit :

$$W_n = \sum_{i=1}^n \beta_{n,i} Z_{n,i} + O_p \left(\frac{1}{\sqrt{n}} \right)$$

où $\beta_{n,i}$ tend vers $\sqrt{2} \gamma_i^2$ lorsque $n \rightarrow \infty$.

4.2.4. Distribution asymptotique de Γ_n , T_n et W_n

La distribution asymptotique des variables Γ_n , T_n et W_n est donnée par le théorème 4.1 (voir DICKEY et FULLER (1979)).

Théorème 4.1. *Sous les hypothèses du modèle (4.1), le vecteur $\mathbf{X}_n = (\Gamma_n, T_n, W_n)'$ converge en loi vers le vecteur $X = (\Gamma, T, W)'$, où*

$$\begin{aligned} \Gamma &= \sum_{i=1}^{\infty} \gamma_i^2 \varepsilon_i^2, \\ T &= \sqrt{2} \sum_{i=1}^{\infty} \gamma_i \varepsilon_i, \\ W &= \sqrt{2} \sum_{i=1}^{\infty} \gamma_i^2 \varepsilon_i, \end{aligned}$$

et $\{\varepsilon_i\}$, $i = 1, 2, \dots$ est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$.

Chapitre 5

Distributions asymptotiques

Résumé

Ce chapitre traite des distributions asymptotiques dans le cadre de la marche aléatoire saisonnière et de modèles saisonniers plus généraux. Il s'appuie sur des résultats fondamentaux permettant d'obtenir la distribution asymptotique de certaines quantités clés, telles que $ns(1 - r(ds))$ et $ns(1 - r'(ds))$. Pour établir ces résultats, deux théorèmes de convergence et un corollaire issus de la littérature sont d'abord énoncés. Ces théorèmes concernent la convergence en loi de suites de vecteurs aléatoires et l'effet des transformations continues sur ces suites.

Par la suite, le chapitre explore les implications de ces résultats dans le contexte spécifique des processus saisonniers. Des démonstrations détaillées et des formulations rigoureuses sont présentées pour justifier l'application des théorèmes aux modèles étudiés. L'accent est mis sur la compréhension des distributions limites et sur leur rôle dans l'analyse des processus aléatoires sous des hypothèses raisonnables de stationnarité et d'ergodicité.

5.1. Introduction

Nous disposons des résultats de base permettant d'obtenir la distribution asymptotique de $ns(1 - r(ds))$ et celle de $ns(1 - r'(ds))$ dans le contexte de la marche aléatoire saisonnière et du modèle saisonnier plus général. Mais auparavant, nous énonçons deux théorèmes de convergence et un corollaire que l'on peut retrouver dans (SERFLING, 1980, pp. 19, 24).

Théorème 5.1. *Soit $\{\mathbf{X}_n\}_{n \geq 1}$ une suite de vecteurs aléatoires de dimension s . Supposons que la suite $\{\mathbf{X}_n\}$ converge en loi vers \mathbf{X} , un vecteur aléatoire de dimension*

s. Soit g une fonction mesurable au sens de Borel définie sur \mathbb{R}^s . Supposons que g soit continue en \mathbf{X} , sauf possiblement sur un ensemble A tel que $\int_A dF_{\mathbf{X}} = 0$. Alors, $g(\mathbf{X}_n)$ converge en loi vers $g(\mathbf{X})$.

Théorème 5.2. *Soient $\{X_n\}_{n \geq 1}$ et $\{Y_n\}_{n \geq 1}$ deux suites de variables aléatoires. Supposons que la suite $\{X_n\}$ converge en loi vers X , tandis que la suite $\{Y_n\}$ converge en probabilité vers une constante finie c . Alors,*

- (1) $X_n + Y_n$ converge en loi vers $X + c$,
- (2) $X_n Y_n$ converge en loi vers cX ,
- (3) X_n / Y_n converge en loi vers X/c , si $c \neq 0$.

Corollaire 5.1. *La convergence en probabilité vers une constante est équivalente à la convergence en loi vers cette constante.*

5.2. Distribution asymptotique de $ns(1 - r'(ds))$ dans le contexte de la marche aléatoire saisonnière

Les résultats de Dickey peuvent être exploités afin d'obtenir la distribution asymptotique de $1 - r'(ds)$, où $r'(ds)$ est le coefficient d'autocorrélation de délai ds , lorsque les données proviennent du modèle

$$\nabla_s X_t = e_t, \quad t \geq 1 \quad (5.1)$$

où $\{e_t\}_{t \geq 1}$, est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$. Le coefficient d'autocorrélation de délai ds est calculé sans centrer les données, c'est-à-dire,

$$r'(ds) = \frac{\sum_{t=1}^{(n-d)s} X_t X_{t+ds}}{\sum_{t=1}^{ns} X_t^2} \quad (5.2)$$

De plus, nous supposons que les valeurs initiales $X_{-s+1}, X_{-s+2}, \dots, X_0$ sont nulles. Si nous écrivons $r'(ds)$ sous la forme

$$r'(ds) = \frac{\sum_{i=1}^s \sum_{t=1}^{n-d} X_{(t-1)s+i} X_{(t+d-1)s+i}}{\sum_{i=1}^s \sum_{t=1}^n X_{(t-1)s+i}^2}$$

nous voyons que le numérateur et le dénominateur de $r'(ds)$ peuvent s'exprimer comme la somme de s sommations indépendantes puisque

$$X_{ts+i} = X_{(t-1)s+i} + e_{ts+i}$$

Il sera plus commode de considérer $1 - r'(ds)$:

$$\begin{aligned}
1 - r'(ds) &= \frac{\sum_{i=1}^s \sum_{t=1}^n X_{(t-1)s+i}^2 - \sum_{i=1}^s \sum_{t=1}^{n-d} X_{(t-1)s+i} X_{(t+d-1)s+i}}{\sum_{i=1}^s \sum_{t=1}^n X_{(t-1)s+i}^2} \\
&= \frac{\sum_{i=1}^s \left(\sum_{t=1}^n X_{(t-1)s+i}^2 - \sum_{t=1}^{n-d} X_{(t-1)s+i} X_{(t+d-1)s+i} \right)}{\sum_{i=1}^s \sum_{t=1}^n X_{(t-1)s+i}^2} \tag{5.3}
\end{aligned}$$

Encore sous cette forme, le numérateur et le dénominateur sont formés de s sommes indépendantes entre elles. Nous allons nous intéresser à l'un des s termes intervenant au numérateur ainsi qu'à l'un de ceux apparaissant au dénominateur, en fixant la valeur de i . Puisque s est aussi fixé, nous nous contenterons d'écrire X_t et e_t à la place de $X_{(t-1)s+i}$ et $e_{(t-1)s+i}$ respectivement, c'est-à-dire,

$$\begin{aligned}
X_t &\leftarrow X_{(t-1)s+1}, \\
e_t &\leftarrow e_{(t-1)s+1}.
\end{aligned}$$

Nous inspirant de la décomposition du numérateur suggérée par l'article de HASZA (1980), nous serons en mesure d'appliquer les résultats de Dickey puisque les expressions qui en découleront seront en termes des variables Γ_n et T_n , à un terme $O_p(n^{-1/2})$ près.

5.2.1. Dénominateur de $1 - r'(ds)$

Chercher la distribution d'un des s termes intervenant au dénominateur de $1 - r'(ds)$ équivaut à chercher la distribution de

$$\sum_{t=1}^n X_t^2,$$

où X_t est une marche aléatoire. Il s'agit effectivement de $n^2 \Gamma_n$ dont il est question à la section 4.2.1. Nous devrions peut-être utiliser la notation $n^2 \Gamma_{n,r}$, $r = 1, \dots, s$, puisque nous avons effectivement une somme de s termes indépendants ayant cette structure. Au niveau des variables canoniques, nous aurions des variables $Z_{n,i,r}$ où le n nous indiquerait que cette variable est introduite lorsque nous avons n observations d'une période de longueur s , et le r spécifierait une des s positions à l'intérieur d'une période. Cette notation ne sera pas utilisée pour l'instant, car il n'y a pas de confusion possible pour le moment.

5.2.2. Numérateur de $ns(1 - r'(ds))$

Remarquons premièrement que

$$\begin{aligned} \sum_{t=1}^n X_t^2 - \sum_{t=1}^{n-d} X_{t+d} X_t \\ = \sum_{t=n-d+1}^n X_t^2 - \sum_{t=1}^{n-d} X_{t+d} (X_{t+d} - X_t) + \sum_{t=1}^{n-d} (X_{t+d} - X_t)^2 \end{aligned} \quad (5.4)$$

Il s'agit d'une autre forme pour l'un des termes du numérateur de $1 - r'(ds)$.

Lemme 5.1.

$$\sum_{t=n-d+1}^n X_t^2 = dX_n^2 + O_p(n^{1/2})$$

Preuve. Nous avons

$$X_n = \sum_{t=1}^n e_t \Rightarrow E[X_n] = 0 \text{ et } \text{Var}[X_n] = n\sigma^2.$$

Par conséquent, $X_n = O_p(n^{1/2})$ et pour j fixé, $X_{n-j} = O_p(n^{1/2})$. Ainsi,

$$\begin{aligned} dX_n^2 - \sum_{t=n-d+1}^n X_t^2 &= \sum_{t=n-d+1}^n (X_n^2 - X_t^2) \\ &= \sum_{j=1}^d (X_n^2 - X_{n-j+1}^2) \\ &= \sum_{j=1}^d (X_n - X_{n-j+1})(X_n + X_{n-j+1}) \end{aligned}$$

Mais, puisque

$$X_n - X_{n-j+1} = \sum_{t=n-j+2}^n e_t = O_p(1),$$

et que

$$X_n + X_{n-j+1} = O_p(n^{1/2}) + O_p(n^{1/2}) = O_p(n^{1/2}),$$

le résultat suit. □

Lemme 5.2.

$$\sum_{t=1}^{n-d} X_{t+d} (X_{t+d} - X_t) = \frac{d}{2} (X_n^2 + n\sigma^2) + O_p(n^{1/2})$$

Preuve. Nous avons

$$\begin{aligned}
& \sum_{t=1}^{n-d} X_{t+d} (X_{t+d} - X_t) \\
&= \sum_{t=1}^{n-d} X_{t+d} \sum_{j=1}^d e_{t+j} \\
&= \sum_{j=1}^d \sum_{t=1}^{n-d} X_{t+d} e_{t+j}.
\end{aligned}$$

Fixons j , $1 \leq j \leq d$. Alors,

$$\begin{aligned}
\sum_{t=1}^{n-d} X_{t+d} e_{t+j} &= \sum_{t=1}^{n-d} (X_{t+j} + X_{t+d} - X_{t+j}) e_{t+j} \\
&= \sum_{t=1}^{n-d} X_{t+j} e_{t+j} + \sum_{t=1}^{n-d} (X_{t+d} - X_{t+j}) e_{t+j}
\end{aligned} \tag{5.5}$$

Examinons le second terme de (5.5) :

$$\begin{aligned}
\sum_{t=1}^{n-d} (X_{t+d} - X_{t+j}) e_{t+j} &= \sum_{t=1}^{n-d} \sum_{i=t+j+1}^{t+d} e_i e_{t+j} \\
&= \sum_{t=1}^{n-d} \sum_{i=j+1}^d e_{t+i} e_{t+j} \\
&= \sum_{t=1}^{n-d} \sum_{i=1}^{d-j} e_{t+j+i} e_{t+j} \\
&= \sum_{t=1}^{d-j} \sum_{i=1}^{n-d} e_{t+j+i} e_{t+j} \\
&= \sum_{i=1}^{d-j} \sum_{t=j+1}^{n-d+j} e_{t+i} e_t.
\end{aligned}$$

Or :

$$\sum_{t=j+1}^{n-d+j} e_t e_{t+i} = \sum_{t=1}^{n-i} e_t e_{t+i} + O_p(1)$$

puisque la différence entre les deux sommes est constituée d'un nombre fini, fixe (avec i), et indépendant de n , de termes $O_p(1)$. Maintenant,

$$Q = \sum_{t=1}^{n-i} e_t e_{t+i} = \mathbf{e}' \mathbf{P}_n(i) \mathbf{e} = \frac{1}{2} \mathbf{e}' \{ \mathbf{P}_n(i) + \mathbf{P}'_n(i) \} \mathbf{e}$$

et

$$\text{Var}(Q) = 2 \text{tr} \left(\frac{\sigma^4}{4} \{ \mathbf{P}_n(i) + \mathbf{P}'_n(i) \}^2 \right) = \frac{\sigma^4}{2} \text{tr} \left(\{ \mathbf{P}_n(i) + \mathbf{P}'_n(i) \}^2 \right).$$

Mais, $\text{tr}\{\mathbf{C}'\mathbf{C}\} = \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq n} c_{i,j}^2$, pour toute matrice \mathbf{C} de dimension $n \times n$. Or, la matrice $\{\mathbf{P}_n(i) + \mathbf{P}'_n(i)\}$ n'est formée que de « 0 » et de « 1. » Il s'agit donc de compter le nombre de « 1 » apparaissant dans la matrice $\{\mathbf{P}_n(i) + \mathbf{P}'_n(i)\}$. On voit aisément qu'il y en a $2(n-i)$. D'où $\text{Var}(Q) = \sigma^4(n-i) = O(n)$. Ainsi, étant donné que $E(Q) = 0$, on conclut que $Q = O_p(n^{1/2})$. Finalement,

$$\sum_{t=j+1}^{n-d+j} e_t e_{t+i} = O_p(n^{1/2}).$$

Par conséquent,

$$\sum_{i=1}^{d-j} \sum_{t=j+1}^{n-d+j} e_t e_{t+i} = O_p(n^{1/2}).$$

Examinons maintenant le premier terme de (5.5) :

$$\begin{aligned} \sum_{t=1}^{n-d} X_{t+j} e_{t+j} &= \sum_{t=j+1}^{n-d+j} X_t e_t \\ &= \sum_{t=1}^n X_t e_t - \sum_{t=1}^j X_t e_t - \sum_{t=n-d+j+1}^n X_t e_t \\ &= \sum_{t=1}^n X_t e_t + O_p(1) - \sum_{k=0}^{d-j-1} X_{n-k} e_{n-k} \\ &= \sum_{t=1}^n X_t e_t + O_p(1) + O_p(n^{1/2}) \\ &= \sum_{t=1}^n X_t e_t + O_p(n^{1/2}) \\ &= \frac{1}{2} \left[\left(\sum_{t=1}^n e_t \right)^2 + \sum_{t=1}^n e_t^2 \right] + O_p(n^{1/2}). \end{aligned}$$

Or,

$$\text{Var} \left[\sum_{t=1}^n e_t^2 \right] = 2n\sigma^4 = O(n)$$

et

$$E \left[\sum_{t=1}^n e_t^2 \right] = n\sigma^2.$$

Par conséquent,

$$\sum_{t=1}^n e_t^2 = n\sigma^2 + O_p(n^{1/2}).$$

Donc,

$$\sum_{t=1}^{n-d} X_{t+j} e_{t+j} = \frac{1}{2} (X_n^2 + n\sigma^2) + O_p(n^{1/2}).$$

Puisque nous devons sommer pour $j \in \{1, 2, \dots, d\}$, la démonstration est terminée.
□

Lemme 5.3.

$$\sum_{t=1}^{n-d} (X_{t+d} - X_t)^2 = nd\sigma^2 + O_p(n^{1/2})$$

Preuve. Nous avons

$$Q = \sum_{t=1}^{n-d} (X_{t+d} - X_t)^2 = \sum_{t=1}^{n-d} \left(\sum_{j=1}^d e_{t+j} \right)^2 \quad (5.6)$$

Puisque $\frac{1}{\sigma\sqrt{d}} \sum_{j=1}^d e_{t+j}$ est distribué selon une loi $\mathcal{N}(0, 1)$, le carré de cette expression suit une loi du χ^2 à un degré de liberté. Ainsi,

$$\mathbb{E} \left[\left(\sum_{j=1}^d e_{t+j} \right)^2 \right] = d\sigma^2.$$

Par conséquent,

$$\mathbb{E}[Q] = d(n-d)\sigma^2.$$

Pour obtenir la variance de Q , nous allons expliciter la matrice de la forme quadratique (5.6). Pour ce faire, notons $\mathbf{P}(k, d)$ la matrice $n \times n$ dont tous les éléments sont nuls, sauf ceux apparaissant sur la diagonale aux positions $(k+i, k+i)$ pour $i = 1, \dots, d$, qui sont égaux à 1. Cette matrice annule les éléments du vecteur \mathbf{e} , sauf les éléments e_{k+1}, \dots, e_{k+d} . On a

$$\begin{aligned} \sum_{t=1}^{n-d} \left(\sum_{j=1}^d e_{t+j} \right)^2 &= \sum_{t=1}^{n-d} (\mathbf{1}'_n \mathbf{P}(t, d) \mathbf{e})^2 \\ &= \sum_{t=1}^{n-d} \mathbf{e}' \mathbf{P}'(t, d) \mathbf{1}_n \mathbf{1}'_n \mathbf{P}(t, d) \mathbf{e} \\ &= \mathbf{e}' \sum_{t=1}^{n-d} \{ \mathbf{P}'(t, d) \mathbf{1}_n \mathbf{1}'_n \mathbf{P}(t, d) \} \mathbf{e} \end{aligned}$$

La matrice $\mathbf{P}'(t, d) \mathbf{1}_n \mathbf{1}'_n \mathbf{P}(t, d)$ contient des éléments nuls, sauf ceux en position $(t+i, t+j)$ pour $i = 1, \dots, d$ et $j = 1, \dots, d$, qui sont égaux à 1. On peut vérifier que

$$\sum_{t=1}^{n-d} \mathbf{P}'(t, d) \mathbf{1}_n \mathbf{1}'_n \mathbf{P}(t, d) = \mathbf{P},$$

dont la forme explicite est donnée dans la figure 5.1.

Pour obtenir la variance de Q , on doit évaluer :

$$2\sigma^4 \operatorname{tr}(\mathbf{P}^2) = 2\sigma^4 \operatorname{tr}(\mathbf{P}'\mathbf{P}),$$

ce qui revient à calculer la somme des carrés de tous les éléments de \mathbf{P} . Bien qu'il soit possible d'explicitier cette somme de carrés, il n'est pas nécessaire de le faire. Dans cette matrice, il y a $2d - 1$ diagonales contenant des éléments non nuls. Sur chacune d'elles, il y a tout au plus $n - 1$ éléments non nuls. Le carré d'un quelconque de ces éléments ne peut être supérieur à 1, et on conclut que...

$$\operatorname{Var}[Q] = 2\sigma^4 \operatorname{tr}[\mathbf{P}'\mathbf{P}] \leq 2\sigma^4(2d - 1)(n - 1)d^2 = O(n).$$

Ainsi,

$$\sum_{t=1}^{n-d} (X_{t+d} - X_t)^2 = dn\sigma^2 + O_p(n^{1/2}),$$

ce qui termine la démonstration du lemme. □

En utilisant les résultats des trois derniers lemmes, (5.4) devient :

$$\begin{aligned} & \sum_{t=1}^n X_t^2 - \sum_{t=1}^{n-d} X_{t+d}X_t \\ &= \sum_{t=n-d+1}^n X_t^2 - \sum_{t=1}^{n-d} X_{t+d}(X_{t+d} - X_t) + \sum_{t=1}^{n-d} (X_{t+d} - X_t)^2 \\ &= dX_n^2 - \frac{d}{2}X_n^2 - \frac{dn}{2}\sigma^2 + nd\sigma^2 + O_p(n^{1/2}) \\ &= \frac{d}{2}(X_n^2 + n\sigma^2) + O_p(n^{1/2}). \end{aligned}$$

$$= \frac{\frac{ds}{2} \sum_{r=1}^s (T_{n,r}^2 + \sigma^2) + O_p(n^{-1/2})}{\sum_{r=1}^s \Gamma_{n,r}} \quad (5.7)$$

Posons

$$\boldsymbol{\chi}_n = (T_{n,1}, \Gamma_{n,1}, \dots, T_{n,s}, \Gamma_{n,s})'$$

et

$$\boldsymbol{\chi} = (T_1, \Gamma_1, \dots, T_s, \Gamma_s)'$$

où

$$T_r = \sqrt{2} \sum_{t=1}^{\infty} \gamma_t \varepsilon_{t,r}, \quad r = 1, 2, \dots, s, \quad (5.8)$$

$$\Gamma_r = \sum_{t=1}^{\infty} \gamma_t^2 \varepsilon_{t,r}^2, \quad r = 1, 2, \dots, s, \quad (5.9)$$

et où $\{\varepsilon_{t,r}\}$ pour $r = 1, 2, \dots, s$ et $t = 1, 2, \dots$, est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$, et où

$$\gamma_i = \frac{2(-1)^{i+1}}{\pi(2i-1)}.$$

En utilisant une généralisation évidente du théorème 4.1, nous voyons que $\boldsymbol{\chi}_n$ converge en loi vers $\boldsymbol{\chi}$. Cela nous amène au théorème suivant.

Théorème 5.3. *Sous les hypothèses du modèle (5.1), si $r'(ds)$ est défini par (5.2), alors, pour $d \geq 1$, la variable $ns(1 - r'(ds))$ dans l'expression (5.7) converge en loi vers*

$$R'(d, s) = \frac{\frac{ds}{2} \sum_{r=1}^s (T_r^2 + \sigma^2)}{\sum_{r=1}^s \Gamma_r} \quad (5.10)$$

Preuve. Écrivons l'équation (5.7) sous la forme

$$ns(1 - r'(ds)) = \frac{g(\boldsymbol{\chi}_n)}{h(\boldsymbol{\chi}_n)} + \frac{O_p(n^{-1/2})}{h(\boldsymbol{\chi}_n)},$$

où

$$g(\boldsymbol{\chi}_n) = \frac{ds}{2} \sum_{r=1}^s (T_{n,r}^2 + \sigma^2)$$

et

$$h(\boldsymbol{\chi}_n) = \sum_{r=1}^s \Gamma_{n,r}.$$

Posons

$$V(\boldsymbol{\chi}_n) = \frac{O_p(n^{-1/2})}{h(\boldsymbol{\chi}_n)}.$$

D'une part, $h(\boldsymbol{\chi})$ est positif avec probabilité 1. Le quotient $g(\boldsymbol{\chi}_n)/h(\boldsymbol{\chi}_n)$ converge donc en loi vers $g(\boldsymbol{\chi})/h(\boldsymbol{\chi})$ selon le théorème 5.1. D'autre part, on constate que $V(\boldsymbol{\chi}_n)$ converge en probabilité vers 0. En effet, si $Y_n = O_p(n^{-1/2})$, alors $Y_n = O_p(1)$ (cf. (BISHOP ET COLLAB., 1975, p. 484)). En appliquant le théorème 5.2, on a que $V(\boldsymbol{\chi}_n)$ converge en loi vers 0. Par le corollaire 5.1, $V(\boldsymbol{\chi}_n)$ converge vers 0 en probabilité. Finalement, par le théorème 5.2, on conclut que $g(\boldsymbol{\chi}_n)/h(\boldsymbol{\chi}_n) + V(\boldsymbol{\chi}_n)$ converge en loi vers $g(\boldsymbol{\chi})/h(\boldsymbol{\chi})$. Ce qui termine la démonstration. \square

Corollaire 5.2. *Dans le contexte du théorème 5.3, la variable $r'(ds)$ converge en probabilité vers 1.*

Preuve. La variable aléatoire $n(1 - r'(ds))$ converge en loi vers $R'(d, s)$, ce qui implique que $1 - r'(ds) = R'(d, s)/n$ converge en loi vers 0. Par le corollaire 5.1, $r'(ds)$ converge en probabilité vers 1. \square

5.3. Distribution asymptotique de $ns(1 - r(ds))$ dans le contexte du modèle plus général

Les résultats de DICKEY (1976) nous permettent aussi d'obtenir la distribution asymptotique d'un coefficient d'autocorrélation échantillonnal dont le délai est un multiple de la période s , lorsque le modèle est défini par (3.1). Pour ce faire, nous nous sommes inspirés de HASZA (1980). Cependant, au niveau du détail, notre travail sera plus ardu. Nous ferons appel au théorème 3.1 et nous devons au préalable démontrer trois lemmes qui nous conduiront à un théorème où nous donnerons des approximations intéressantes du numérateur et du dénominateur de $ns(1 - r(ds))$. L'intérêt de ces approximations est qu'on y retrouve des expressions ne faisant intervenir que les variables aléatoires de base, à savoir : $\Gamma_{n,r}$, $T_{n,r}$ et $W_{n,r}$ pour $r = 1, \dots, s$.

5.3.1. Numérateur de $ns(1 - r(ds))$

Considérons le modèle (3.1). Si nous appliquons le théorème 3.1, nous pouvons écrire

$$\mathbf{Z}_\tau = \mathbf{Z}_0 + \mathbf{H}_0 \mathbf{X}_\tau - \mathbf{D}_\tau + \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{e}_{-j}.$$

Dans cette représentation, \mathbf{Z}_0 est un vecteur de valeurs initiales, \mathbf{X}_τ est une marche aléatoire multivariée et \mathbf{D}_τ est un processus stationnaire au sens strict. Comme précédemment, on dénote $(\mathbf{Z}'_1, \dots, \mathbf{Z}'_n)'$ par \mathbf{Z} , et $(\mathbf{D}'_1, \dots, \mathbf{D}'_n)'$ par \mathbf{D} .

Dans un premier temps, nous nous intéressons à $ns(c(0) - c(ds))$, c'est-à-dire, le numérateur de $ns(1 - r(ds))$.

$$\begin{aligned} ns(c(0) - c(ds)) &= \sum_{t=1}^{ns} (Z_t - \bar{Z})^2 - \sum_{t=1}^{(n-d)s} (Z_t - \bar{Z})(Z_{t+ds} - \bar{Z}) \\ &= \sum_{t=(n-d)s+1}^{ns} (Z_t - \bar{Z})^2 + \bar{Z} \sum_{t=(n-d)s+1}^{ns} Z_t - \bar{Z} \sum_{t=1}^{ds} Z_t \\ &\quad - \sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) + \sum_{t=1}^{(n-d)s} (Z_{t+ds} - Z_t)^2. \end{aligned} \quad (5.11)$$

Nous allons maintenant démontrer les lemmes 5.4 à 5.6 qui établissent l'ordre en probabilité des différents termes de (5.11). Il s'agit d'une généralisation de trois lemmes de HASZA (1980).

Lemme 5.4. *Sous les hypothèses du modèle (3.1) :*

$$\begin{aligned} a) \quad & \sum_{t=(n-d)s+1}^{ns} (Z_t - \bar{Z})^2 = d(\mathbf{X}_n - \bar{\mathbf{X}})' \mathbf{H}_0^2 (\mathbf{X}_n - \bar{\mathbf{X}}) + O_p(n^{1/2}), \\ b) \quad & \bar{Z} \sum_{t=(n-d)s+1}^{ns} Z_t = dc_0^2 \bar{X} \mathbf{1}'_s \mathbf{X}_n + O_p(n^{1/2}), \\ c) \quad & \bar{Z} \sum_{t=1}^{ds} Z_t = O_p(n^{1/2}), \\ d) \quad & \sum_{t=(n-d)s+1}^{ds} Z_t^2 = d\mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + O_p(n^{1/2}), \end{aligned}$$

où $\bar{Z} = \frac{1}{ns} \sum_{t=1}^{ns} Z_t$, $\bar{X} = \frac{1}{ns} \sum_{t=1}^{ns} X_t$, $\bar{\mathbf{X}} = \bar{X} \mathbf{1}_s$, $c_0 = \sum_{j=0}^{\infty} \Psi_j$, et \mathbf{H}_0 est défini en (3.7).

Preuve. En posant

$$\mathbf{A} = \mathbf{Z}_0 + \sum_{j=0}^{\infty} \mathbf{H}_{j+1} \mathbf{a}_{-j} = (\lambda_1, \lambda_2, \dots, \lambda_s)',$$

nous pouvons écrire

$$\mathbf{Z}_\tau = \mathbf{A} + \mathbf{H}_0 \mathbf{X}_\tau - \mathbf{D}_\tau.$$

Définissons la fonction $\eta(t)$ comme suit :

$$\begin{aligned}\eta : \mathbb{N} &\longrightarrow \mathbb{N} \\ t &\longmapsto t - s \lfloor (t-1)/s \rfloor\end{aligned}$$

où $\lfloor x \rfloor$ désigne la partie entière de x . Cette fonction donne $\{i\}$ comme image à l'ensemble des indices $\{i, s+i, 2s+i, 3s+i, \dots\}$. Elle permet donc de trouver la position d'une observation à l'intérieur d'une période de longueur s . Définissons aussi

$$\begin{aligned}\zeta : \mathbb{N} &\longrightarrow \mathbb{N} \\ t &\longmapsto \lfloor (t-1)/s \rfloor + 1\end{aligned}$$

Cette fonction permet de savoir à quelle période appartient une observation donnée, puisque les indices $\{ks+1, ks+2, \dots, ks+s\}$ auront pour image $\{k+1\}$. À l'aide de ces deux fonctions, nous pouvons écrire

$$Z_t = \lambda_{\eta(t)} + \left(\mathbf{P}^{\eta(t)} \mathbf{a}\right)' \mathbf{X}_{\zeta(t)} - \mathbf{D}_t$$

où \mathbf{P} est la matrice définie dans la démonstration du lemme 2.2, et \mathbf{a} est la s -ième ligne de la matrice \mathbf{H}_0 .

On remarque que

$$\begin{aligned}\sum_{t=1}^{ns} Z_t &= (\mathbf{1}_n \otimes \mathbf{1}_s)' \mathbf{Z} \\ &= (\mathbf{1}_n \otimes \mathbf{1}_s)' (\mathbf{1}_n \otimes \mathbf{A} + (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} - \mathbf{D}) \\ &= n \otimes \mathbf{1}'_s \mathbf{A} + (\mathbf{1}'_n \otimes \mathbf{1}'_s \mathbf{H}_0) \mathbf{X} - (\mathbf{1}_n \otimes \mathbf{1}_s)' \mathbf{D} \\ &= ns\bar{\Lambda} + c_0 ns\bar{X} - ns\bar{D}.\end{aligned}$$

Par conséquent,

$$\bar{Z} = \bar{\Lambda} + c_0 \bar{X} - \bar{D} \tag{5.12}$$

Ainsi,

$$Z_t - \bar{Z} = (\lambda_{\eta(t)} - \bar{\Lambda}) + \left(\mathbf{a}' \left(\mathbf{P}^{\eta(t)}\right)' \mathbf{X}_{\zeta(t)} - c_0 \bar{X}\right) - (D_t - \bar{D}).$$

Pour simplifier l'écriture, nous poserons

$$V_t = \mathbf{a}' \left(\mathbf{P}^{\eta(t)}\right)' \mathbf{X}_{\zeta(t)}.$$

Notons que

$$\mathbf{v}_t = \mathbf{H}_0 \mathbf{X}_t$$

et que

$$\mathbf{v} = (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X}.$$

Démonstration de la partie a) : On vérifie aisément que

$$\bar{V} = c_0 \bar{X},$$

et par conséquent,

$$\begin{aligned} (Z_t - \bar{Z})^2 &= [(\lambda_{\eta(t)} - \bar{\Lambda}) + (V_t - \bar{V}) - (D_t - \bar{D})]^2 \\ &= (\lambda_{\eta(t)} - \bar{\Lambda})^2 + (V_t - \bar{V})^2 + (D_t - \bar{D})^2 + 2(\lambda_{\eta(t)} - \bar{\Lambda})(V_t - \bar{V}) \\ &\quad - 2(\lambda_{\eta(t)} - \bar{\Lambda})(D_t - \bar{D}) - 2(V_t - \bar{V})(D_t - \bar{D}). \end{aligned} \quad (5.13)$$

Nous devons sommer l'équation (5.13) pour les valeurs de $t \in \{(n-d)s + 1, \dots, ns\}$. Nous allons donc considérer chacun des termes de cette relation.

— Pour le *premier* terme, on voit que

$$\sum_{t=(n-d)s+1}^{ns} (\lambda_{\eta(t)} - \bar{\Lambda})^2 = O_p(1) \quad (5.14)$$

puisque'il s'agit d'une constante indépendante de n .

— Pour le *troisième* terme, nous avons

$$\sum_{t=(n-d)s+1}^{ns} (D_t - \bar{D})^2 = O_p(1)$$

En effet, nous savons que \bar{D} converge en probabilité vers 0 (voir la discussion relative au terme T_3 à la section 3.3.1). Donc ce terme est $O_p(1)$. De plus, \mathbf{D}_τ est un processus stationnaire. Chacune de ses composantes $(\mathbf{D}_\tau)_i$, $i = 1, \dots, s$, a une variance σ_i^2 indépendante de τ . Par conséquent, $\text{Var}[(\mathbf{D}_\tau)_i] = \mathbf{O}(1)$, $i = 1, \dots, s$. Ainsi,

$$\begin{aligned} D_{ns} - \bar{D} &= O_p(1) \\ \Rightarrow D_{ns-k} - \bar{D} &= O_p(1), \quad k = 0, 1, \dots, ds - 1 \\ \Rightarrow (D_{ns-k} - \bar{D})^2 &= O_p(1) \\ \Rightarrow \sum_{t=(n-d)s+1}^{ns} (D_t - \bar{D})^2 &= O_p(1) \end{aligned} \quad (5.15)$$

car il s'agit d'une somme de ds quantités d'ordre $O_p(1)$.

— En ce qui concerne le *sixième* terme :

$$\sum_{t=(n-d)s+1}^{ns} (V_t - \bar{V})(D_t - \bar{D}) = O_p(n^{1/2}) \quad (5.16)$$

Pour s'en convaincre, on note que :

a) premièrement.

$$\mathbb{E}[V_t] = \mathbb{E}\left[\mathbf{a}' (\mathbf{P}^{\eta(t)})' \mathbf{X}_{\zeta(t)}\right]$$

$$= \mathbf{a}'(\mathbf{P}^{\eta(t)})' \mathbf{E}[\mathbf{X}_{\zeta(t)}] = 0$$

b) deuxièmement

$$\begin{aligned} \text{Var}[V_t] &= \text{Var}\left[\mathbf{a}'(\mathbf{P}^{\eta(t)})' \mathbf{X}_{\zeta(t)}\right] \\ &= \mathbf{a}'(\mathbf{P}^{\eta(t)})' \left(\zeta(t)\sigma^2 \mathbf{I}_s\right) \mathbf{P}^{\eta(t)} \mathbf{a} \\ &= \zeta(t)\sigma^2 \mathbf{a}' \mathbf{a} \end{aligned}$$

puisque la matrice $\mathbf{P}^{\eta(t)}$ est orthogonale.

On peut donc affirmer que

$$V_{ns} = O_p\left(n^{1/2}\right) \quad (5.17)$$

et il en est de même pour V_{ns-k} , $k = 0, \dots, ds - 1$.

c) troisièmement, $\bar{V} = c_0 \bar{X} \Rightarrow \mathbf{E}[\bar{V}] = 0$ et puisque, selon (2.3),

$$\bar{X} = \frac{\mathbf{1}'_n}{n} \otimes \frac{\mathbf{1}'_s}{s} \mathbf{X}, \quad \Sigma_X = \sigma^2 (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s),$$

la variance de \bar{V} sera

$$\begin{aligned} &c_0^2 \frac{\sigma^2}{n^2 s^2} (\mathbf{1}'_n \otimes \mathbf{1}'_s) (\mathbf{T}_n \mathbf{T}'_n \otimes \mathbf{I}_s) (\mathbf{1}_n \otimes \mathbf{1}_s) \\ &= c_0^2 \frac{\sigma^2}{n^2 s^2} (\mathbf{1}'_n \mathbf{T}_n \mathbf{T}'_n \mathbf{1}_n) \otimes (\mathbf{1}'_s \mathbf{1}_s) \end{aligned}$$

et en utilisant (2.14) :

$$\begin{aligned} &= c_0^2 \frac{\sigma^2}{n^2 s^2} \frac{n(n+1)(2n+1)}{6} s \\ &= c_0^2 \sigma^2 \frac{(n+1)(2n+1)}{6ns} = O(n). \end{aligned}$$

On conclut que

$$\bar{V} = O_p\left(n^{1/2}\right). \quad (5.18)$$

Il en découle que pour k fixé,

$$(V_{ns-k} - \bar{V}) = O_p\left(n^{1/2}\right)$$

et que

$$(V_{ns-k} - \bar{V}) (D_{ns-k} - \bar{D}) = O_p\left(n^{1/2}\right).$$

Enfin,

$$\sum_{t=(n-d)s+1}^{ns} (V_t - \bar{V}) (\mathbf{D}_t - \bar{D}) = O_p\left(n^{1/2}\right).$$

— Pour le *second* terme, intéressons-nous maintenant à

$$\sum_{t=(n-d)s+1}^{ns} (V_t - \bar{V})^2.$$

Nous avons

$$\begin{aligned} \sum_{t=(n-d)s+1}^{ns} (V_t - \bar{V})^2 &= \sum_{k=0}^{ds-1} (V_{ns-k} - \bar{V})^2 \\ &= \sum_{r=0}^{s-1} \sum_{\nu=0}^{d-1} (V_{(n-\nu)s-r} - \bar{V})^2. \end{aligned}$$

Pour l'instant, maintenons r fixé. Alors,

$$\begin{aligned} d(V_{ns-r} - \bar{V})^2 - \sum_{\nu=0}^{d-1} (V_{(n-\nu)s-r} - \bar{V})^2 \\ &= \sum_{\nu=0}^{d-1} \left\{ (V_{ns-r} - \bar{V})^2 - (V_{(n-\nu)s-r} - \bar{V})^2 \right\} \\ &= \sum_{\nu=0}^{d-1} \left\{ V_{ns-r}^2 - 2\bar{V}V_{ns-r} + \bar{V}^2 - V_{(n-\nu)s-r}^2 + 2\bar{V}V_{(n-\nu)s-r} - \bar{V}^2 \right\} \\ &= \sum_{\nu=0}^{d-1} \left\{ V_{ns-r}^2 - V_{(n-\nu)s-r}^2 \right\} - 2\bar{V} \sum_{\nu=0}^{d-1} (V_{ns-r} - V_{(n-\nu)s-r}) \end{aligned} \quad (5.19)$$

On remarque alors que

$$\begin{aligned} V_{ns-r} - V_{(n-\nu)s-r} &= \mathbf{a}' \left(\mathbf{P}^{\eta(ns-r)} \right)' \mathbf{X}_{\zeta(ns-r)} - \mathbf{a}' \left(\mathbf{P}^{\eta((n-\nu)s-r)} \right)' \mathbf{X}_{\zeta((n-\nu)s-r)} \\ &= \mathbf{a}' \left(\mathbf{P}^{s-r} \right)' (\mathbf{X}_n - \mathbf{X}_{n-\nu}) \\ &= \mathbf{a}' \mathbf{P}^r \sum_{j=0}^{\nu-1} \mathbf{e}_{n-j} \quad \text{puisque } \left(\mathbf{P}^{s-r} \right)' = \mathbf{P}^r \\ &= O_p(1) \end{aligned} \quad (5.20)$$

et par conséquent,

$$\sum_{\nu=0}^{d-1} (V_{ns-r} - V_{(n-\nu)s-r}) = O_p(1),$$

et tenant compte de la relation (5.18),

$$\bar{V} \sum_{\nu=0}^{d-1} (V_{ns-r} - V_{(n-\nu)s-r}) = O_p(n^{1/2}). \quad (5.21)$$

De plus, on note que

$$\begin{aligned} V_{ns-r}^2 - V_{(n-\nu)s-r}^2 &= (V_{ns-r} - V_{(n-\nu)s-r}) (V_{ns-r} + V_{(n-\nu)s-r}) \\ &= O_p(1) \cdot O_p(n^{1/2}) = O_p(n^{1/2}), \end{aligned}$$

par (5.17) et (5.20) ce qui signifie que

$$\sum_{\nu=0}^{d-1} \left(V_{ns-r}^2 - V_{(n-\nu)s-r}^2 \right) = O_p \left(n^{1/2} \right). \quad (5.22)$$

Les équations (5.19), (5.21) et (5.22) nous permettent de voir que

$$\sum_{\nu=0}^{d-1} \left(V_{(n-\nu)s-r} - \bar{V} \right)^2 = d \left(V_{ns-r} - \bar{V} \right)^2 + O_p \left(n^{1/2} \right) \quad (5.23)$$

et en faisant la somme pour $r \in \{0, 1, \dots, s-1\}$, nous obtenons

$$\begin{aligned} \sum_{r=0}^{s-1} \sum_{\nu=0}^{d-1} \left(V_{(n-\nu)s-r} - \bar{V} \right)^2 &= d \sum_{r=0}^{s-1} \left(V_{ns-r} - \bar{V} \right)^2 + O_p \left(n^{1/2} \right) \\ &= d \left\| \mathbf{V}_n - \bar{V} \mathbf{1}_s \right\|^2 + O_p \left(n^{1/2} \right) \\ &= d \left\| \mathbf{H}_0 \left(\mathbf{X}_n - \bar{\mathbf{X}} \right) \right\|^2 + O_p \left(n^{1/2} \right) \\ &= d \left(\mathbf{X}_n - \bar{\mathbf{X}} \right)' \mathbf{H}_0^2 \left(\mathbf{X}_n - \bar{\mathbf{X}} \right) + O_p \left(n^{1/2} \right). \end{aligned}$$

— On peut se convaincre sans difficulté que le *quatrième* terme est d'ordre $O_p \left(n^{1/2} \right)$ tandis que le cinquième terme est d'ordre $l_p(1)$.

Compte tenu de cette dernière remarque et des équations (5.13), (5.14), (5.15), (5.16), (5.23), on constate que

$$\begin{aligned} \sum_{t=(n-d)s+1}^{ns} \left(Z_t - \bar{Z} \right)^2 &= O_p(1) + d \left(\mathbf{X}_n - \bar{\mathbf{X}} \right)' \mathbf{H}_0^2 \left(\mathbf{X}_n - \bar{\mathbf{X}} \right) + O_p \left(n^{1/2} \right) + O_p(1) \\ &\quad + 2O_p \left(n^{1/2} \right) + 2O_p(1) + 2O_p \left(n^{1/2} \right). \end{aligned}$$

Ce qui termine la démonstration de la partie a).

Démonstration de la partie b) : Concentrons-nous maintenant sur l'expression

$$\bar{Z} \sum_{t=(n-d)s+1}^{ns} Z_t.$$

Nous avons

$$\begin{aligned}
\sum_{t=(n-d)s+1}^{ns} Z_t &= \sum_{t=(n-d)s+1}^{ns} \lambda_{\eta(t)} + \sum_{t=(n-d)s+1}^{ns} V_t - \sum_{t=(n-d)s+1}^{ns} D_t \\
&= ds\bar{A} - \sum_{r=0}^{s-1} \sum_{\nu=0}^{d-1} D_{(n-\nu)s-r} + d \sum_{r=0}^{s-1} V_{ns-r} \\
&\quad + \sum_{r=0}^{s-1} \sum_{\nu=0}^{d-1} (V_{(n-\nu)s-r} - V_{ns-r}) \\
&= d \sum_{r=0}^{s-1} V_{ns-r} + O_p(1).
\end{aligned}$$

De plus, selon (5.12), sachant que \bar{A} et \bar{D} sont d'ordre $O_p(1)$, on a

$$\bar{Z} = \bar{V} + O_p(1).$$

Donc,

$$\begin{aligned}
\bar{Z} \sum_{t=(n-d)s+1}^{ns} Z_t &= (\bar{V} + O_p(1)) \left(d \sum_{r=0}^{s-1} V_{ns-r} + O_p(1) \right) \\
&= d\bar{V} \mathbf{1}'_s \mathbf{V}_n + \bar{V} O_p(1) + d \mathbf{1}'_s \mathbf{V}_n O_p(1) + O_p(1) \\
&= d\bar{V} \mathbf{1}'_s \mathbf{V}_n + O_p(n^{1/2}) \\
&= dc_0^2 \bar{X} \mathbf{1}'_s \mathbf{X}_n + O_p(n^{1/2}).
\end{aligned}$$

Démonstration de la partie c) : Cette affirmation devient évidente à la lumière de l'équation (5.18) et du fait que la sommation impliquée est d'ordre $O_p(1)$, puisqu'elle est formée d'un nombre fixe (et indépendant de n) de termes d'ordre $O_p(1)$.

Démonstration de la partie d) : Puisque

$$Z_t = \lambda_{\eta(t)} + V_t - D_t,$$

on a

$$\sum_{t=(n-d)s+1}^{ns} Z_t^2 = \sum_{t=(n-d)s+1}^{ns} \left(\lambda_{\eta(t)}^2 + V_t^2 + D_t^2 + 2\lambda_{\eta(t)}V_t - 2\lambda_{\eta(t)}D_t - 2V_tD_t \right).$$

Nous allons déterminer l'ordre de chacun de ces six termes.

$$(1) \quad \sum_{t=(n-d)s+1}^{ns} \lambda_{\eta(t)}^2 = O_p(1),$$

puisque'il s'agit d'une constante indépendante de n .

$$(2) \quad \sum_{t=(n-d)s+1}^{ns} D_t^2 = \sum_{r=1}^s \sum_{\nu=n-d}^n D_{\nu s+r}^2 = O_p(1),$$

puisque \mathbf{D}_τ est un processus stationnaire, sa matrice de covariances est indépendante de n .

$$(3) \quad \sum_{t=(n-d)s+1}^{ns} V_t D_t = \sum_{r=1}^s \sum_{\nu=n-d}^n V_{vs+r} D_{vs+r} = O_p(n^{1/2}),$$

car $V_{vs+r} = O_p(n^{1/2})$ et $D_{vs+r} = O_p(1)$, donc $V_{vs+r} D_{vs+r} = O_p(n^{1/2})$, et nous additionnons ds termes de cet ordre.

$$(4) \quad \sum_{t=(n-d)s+1}^{ns} V_t^2 = d\mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + O_p(n^{1/2}).$$

En effet, pour r fixé, nous avons

$$\begin{aligned} dV_{ns-r}^2 - \sum_{\nu=0}^{d-1} V_{(n-\nu)s-r}^2 &= \sum_{\nu=0}^{d-1} (V_{ns-r}^2 - V_{(n-\nu)s-r}^2) \\ &= \sum_{\nu=0}^{d-1} (V_{ns-r} - V_{(n-\nu)s-r}) (V_{ns-r} + V_{(n-\nu)s-r}). \end{aligned}$$

Or, $V_{ns-r} - V_{(n-\nu)s-r} = O_p(1)$, tandis que $V_{ns-r} + V_{(n-\nu)s-r} = O_p(n^{1/2})$. On conclut que

$$\sum_{\nu=0}^{d-1} V_{(n-\nu)s-r}^2 = dV_{ns-r}^2 + O_p(n^{1/2}),$$

et en sommant sur r ,

$$\sum_{r=0}^{s-1} \sum_{\nu=0}^{d-1} V_{(n-\nu)s-r}^2 = d \sum_{r=0}^{s-1} V_{ns-r}^2 + O_p(n^{1/2}) = d\mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + O_p(n^{1/2}).$$

$$(5) \quad \sum_{t=(n-d)s+1}^{ns} \lambda_{\eta(t)} V_t = O_p(n^{1/2}),$$

car il s'agit d'une combinaison linéaire de ds termes d'ordre $O_p(n^{1/2})$.

$$(6) \quad \sum_{t=(n-d)s+1}^{ns} \lambda_{\eta(t)} D_t = O_p(1),$$

car il s'agit d'une combinaison linéaire de ds termes d'ordre $O_p(1)$.

Ceci termine la démonstration du lemme 5.4. □

En appliquant le lemme 5.4 à l'équation (5.11), nous obtenons

$$\begin{aligned} ns(c(0) - c(ds)) &= d(\mathbf{X}_n - \bar{\mathbf{X}})' \mathbf{H}_0^2 (\mathbf{X}_n - \bar{\mathbf{X}}) + dc_0^2 \bar{\mathbf{X}}' \mathbf{1}'_s \mathbf{X}_n + O_p(n^{1/2}) \\ &\quad - \sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) + \sum_{t=1}^{(n-d)s} (Z_{t+ds} - Z_t)^2. \end{aligned} \quad (5.24)$$

Nous allons maintenant entreprendre la démonstration d'un lemme analogue au lemme 2 de HASZA (1980). Il s'agit de trouver une bonne approximation pour la première sommation de l'équation (5.24). Réécrivons donc cette sommation sous une autre forme.

$$\begin{aligned}
\sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) &= \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} Z_{(\nu+d)s+r} (Z_{(\nu+d)s+r} - Z_{\nu s+r}) \\
&= \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} (\lambda_r + V_{(\nu+d)s+r} - D_{(\nu+d)s+r}) \left(\sum_{k=1}^d \omega_{(\nu+k)s+r} \right) \\
&= \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \sum_{k=1}^d \lambda_r \omega_{(\nu+k)s+r} + \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \sum_{k=1}^d V_{(\nu+d)s+r} \omega_{(\nu+k)s+r} \\
&\quad - \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \sum_{k=1}^d D_{(\nu+d)s+r} \omega_{(\nu+k)s+r}.
\end{aligned}$$

Pour k et r fixés, puisque ω_t est de moyenne nulle et stationnaire, nous avons

$$\frac{1}{n} \sum_{\nu=0}^{n-d-1} \omega_{(\nu+k)s+r} \xrightarrow{P} 0,$$

et par conséquent,

$$\sum_{r=1}^s \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} \lambda_r \omega_{(\nu+k)s+r} = o_p(n).$$

Introduisons quelques notations. Soit $\omega_{\nu,r} = \omega_{\nu s+r}$ pour $\nu = 0, 1, 2, \dots, n-1$ et $r = 1, 2, \dots, s$, et soit $D_{\nu,r} = D_{\nu s+r}$ pour $\nu = 0, 1, 2, \dots, n-1$ et $r = 1, 2, \dots, s$. Puisque \mathbf{D}_τ est un processus multivarié stationnaire, chacune de ses composantes est également stationnaire. Nous dénoterons $E[\omega_{0,u} D_{h,\nu}]$ par $\gamma_{\omega,u,\mathbf{D},\nu}(h)$. Il s'agit de la covariance croisée entre la u -ième composante de $\boldsymbol{\omega}_\tau$ et la ν -ième composante de \mathbf{D}_τ . Nous savons que

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^{n-h-1} \omega_{ts+r} D_{(t+h)s+r} = \gamma_{\omega,r,\mathbf{D},r}(h).$$

Ainsi,

$$\sum_{t=1}^{n-h-1} \omega_{ts+r} D_{(t+h)s+r} = n \gamma_{\omega,r,\mathbf{D},r}(h) + o_p(n).$$

Il s'ensuit que

$$\begin{aligned}
\sum_{\nu=0}^{n-d-1} \sum_{k=1}^d D_{(\nu+d)s+r} \omega_{(\nu+k)s+r} &= \sum_{\nu=d}^{n-1} \sum_{k=1}^d D_{\nu s+r} \omega_{(\nu-d+k)s+r} \\
&= \sum_{h=0}^{d-1} \sum_{\nu=d}^{n-1} D_{\nu s+r} \omega_{(\nu-h)s+r} \\
&= n \sum_{h=0}^{d-1} \gamma_{\omega, r, \mathbf{D}, r}(h) + o_p(n).
\end{aligned}$$

On peut alors écrire :

$$\begin{aligned}
\sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) &= \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \sum_{k=1}^d V_{(\nu+d)s+r} \omega_{(\nu+k)s+r} \\
&\quad - n \sum_{r=1}^s \sum_{h=0}^{d-1} \gamma_{\omega, r, \mathbf{D}, r}(h) + o_p(n). \tag{5.25}
\end{aligned}$$

D'autre part, considérons $\mathbf{\Gamma}_{\omega, \mathbf{D}}(h)$, la matrice des covariances croisées de délai h entre le vecteur $\boldsymbol{\omega}_\tau$ et le vecteur $\mathbf{D}_{\tau+h}$. De fait, $\mathbf{\Gamma}_{\omega, \mathbf{D}}(h) = \mathbb{E}[\boldsymbol{\omega}_0 \mathbf{D}'_h]$. Dénotons par $\mathbf{\Gamma}_{\omega, \mathbf{D}}(h, i, j)$ l'élément (i, j) de la matrice $\mathbf{\Gamma}_{\omega, \mathbf{D}}(h)$. On a

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\omega}_0 \mathbf{D}'_h] &= \mathbb{E} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j e_{-j} e'_{h-i} \mathbf{H}'_{i+1} \right] \quad \text{par (3.5) et (3.8)} \\
&= \sigma^2 \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{H}'_{j+h+1} \\
&= \sigma^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\Psi}'_{i+j+h+1} \\
&= \sum_{i=0}^{\infty} \mathbf{\Gamma}_{\omega}(i+h+1).
\end{aligned}$$

Il faut aussi remarquer que

$$\sum_{r=1}^s \gamma_{\omega, r, \mathbf{D}, r}(h) = \text{tr}[\mathbf{\Gamma}_{\omega, \mathbf{D}}(h)] = \sum_{i=0}^{\infty} \text{tr}[\mathbf{\Gamma}_{\omega}(i+h+1)].$$

D'autre part, nous avons

$$\begin{aligned}
\sigma^2 \mathbf{H}_0^2 &= \sigma^2 \mathbf{H}_0 \mathbf{H}'_0 \\
&= \sigma^2 \left[\sum_{j=0}^{\infty} \Psi_j \Psi'_j + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \Psi_j \Psi'_{i+j+1} + \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \Psi_{i+j+1} \Psi'_i \right] \\
&= \Gamma_{\omega}(0) + \sigma^2 \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \Psi_j \Psi'_{i+j+1} + \sigma^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Psi_{i+j+1} \Psi'_i \\
&= \Gamma_{\omega}(0) + \sum_{i=0}^{\infty} \Gamma_{\omega}(i+1) + \sum_{i=0}^{\infty} \Gamma_{\omega}(-i-1).
\end{aligned}$$

Ainsi,

$$\frac{d}{2} \sigma^2 \mathbf{H}_0^2 = \frac{d}{2} \left(\Gamma_{\omega}(0) + \sum_{i=1}^{\infty} \Gamma_{\omega}(i) + \sum_{i=1}^{\infty} \Gamma'_{\omega}(i) \right). \quad (5.26)$$

Par ailleurs,

$$\sum_{h=0}^{d-1} \sum_{i=0}^{\infty} \Gamma_{\omega}(i+h+1) = d \sum_{i=0}^{\infty} \Gamma_{\omega}(i) - \sum_{h=0}^{d-1} (d-h) \Gamma_{\omega}(h). \quad (5.27)$$

Au niveau des traces, en dénotant par δ_0 l'élément apparaissant sur la diagonale de \mathbf{H}_0^2 (voir démonstration du lemme 2.2), nous déduisons de l'équation (5.26) que

$$\begin{aligned}
\text{tr} \left(\frac{d}{2} \sigma^2 \mathbf{H}_0^2 \right) &= \frac{d \sigma^2 s \delta_0}{2} = \frac{d}{2} \left(s \gamma_{\omega}(0) + 2s \sum_{i=1}^{\infty} \gamma_{\omega}(is) \right) \\
\implies \frac{d \sigma^2 s \delta_0}{2} &= \frac{ds}{2} \left(-\gamma_{\omega}(0) + 2 \sum_{i=0}^{\infty} \gamma_{\omega}(is) \right) \\
\implies ds \sum_{i=0}^{\infty} \gamma_{\omega}(is) &= \frac{\sigma^2 ds \delta_0}{2} + \frac{ds \gamma_{\omega}(0)}{2} = \frac{ds}{2} (\sigma^2 \delta_0 + \gamma_{\omega}(0)). \quad (5.28)
\end{aligned}$$

Donc,

$$\begin{aligned}
\sum_{r=1}^s \sum_{h=0}^{d-1} \gamma_{\omega,r,D,r}(h) &= \text{tr} \left(\sum_{h=0}^{d-1} \sum_{i=0}^{\infty} \Gamma_{\omega}(i+h+1) \right) \\
&= ds \sum_{i=0}^{\infty} \gamma_{\omega}(is) - s \sum_{h=0}^{d-1} (d-h) \gamma_{\omega}(hs) \quad \text{par (5.27)} \\
&= \frac{ds}{2} (\sigma^2 \delta_0 + \gamma_{\omega}(0)) - s \sum_{h=0}^{d-1} (d-h) \gamma_{\omega}(hs), \quad \text{par (5.28)} \\
&= \frac{\sigma^2 ds \delta_0}{2} + \frac{s}{2} \left(d \gamma_{\omega}(0) - 2 \sum_{h=0}^{d-1} (d-h) \gamma_{\omega}(hs) \right) \\
&= \frac{\sigma^2 ds \delta_0}{2} - \frac{s}{2} \sum_{h=-d+1}^{d-1} (d-|h|) \gamma_{\omega}(hs).
\end{aligned}$$

Cela nous permet d'écrire (5.25) sous la forme

$$\sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) = \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \sum_{k=1}^d V_{(\nu+d)s+r} \omega_{(\nu+k)s+r} - \frac{n\sigma^2 ds \delta_0}{2} + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n). \quad (5.29)$$

Il reste donc à trouver une forme plus commode pour la triple sommation de (5.29).

Nous avons

$$\begin{aligned} & \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \sum_{k=1}^d V_{(\nu+d)s+r} \omega_{(\nu+k)s+r} \\ &= \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} \mathbf{V}'_{\nu+d+1} \boldsymbol{\omega}_{\nu+k+1} \\ &= \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} \mathbf{V}'_{\nu+d+1} \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1} \\ &= \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} (\mathbf{V}_{\nu+d+1} - \mathbf{V}_{\nu+k-j+1} + \mathbf{V}_{\nu+k-j+1})' \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1} \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} (\mathbf{V}_{\nu+d+1} - \mathbf{V}_{\nu+k-j+1})' \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1} \\ & \quad + \sum_{j=0}^{\infty} \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} \mathbf{V}'_{\nu+k-j+1} \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1}. \end{aligned} \quad (5.30)$$

Nous devons donc examiner chacune des deux triples sommations de (5.30). En ce qui concerne la première, un calcul direct montre que

$$\begin{aligned} & \sum_{j=0}^{\infty} \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} (\mathbf{V}_{\nu+d+1} - \mathbf{V}_{\nu+k-j+1})' \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1} \\ &= \sum_{j=0}^{\infty} \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} \sum_{i=\nu+k-j+2}^{\nu+d+1} \mathbf{e}'_i \mathbf{H}_0 \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1}, \quad \text{puisque } \mathbf{V}_\tau = \mathbf{H}_0^2 \mathbf{X}_\tau, \\ &= \sum_{k=1}^d \text{tr} \left(\mathbf{H}_0 \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \sum_{\nu=0}^{n-d-1} \sum_{i=\nu+k-j+2}^{\nu+d+1} \mathbf{e}_{\nu+k-j+1} \mathbf{e}'_i \right). \end{aligned} \quad (5.31)$$

Or,

$$\begin{aligned} \frac{1}{n} \sum_{\nu=0}^{n-d-1} \sum_{i=\nu+k-j+2}^{\nu+d+1} \mathbf{e}_{\nu+k-j+1} \mathbf{e}'_i &= \frac{1}{n} \sum_{\nu=0}^{n-d-1} \sum_{i=1}^{d-k+j} \mathbf{e}_{\nu+k-j+1} \mathbf{e}'_{i+\nu+k-j+1} \\ &= \frac{1}{n} \sum_{\nu=0}^{n-d-1} \sum_{i=1}^{d-m} \mathbf{e}_{\nu+m+1} \mathbf{e}'_{i+\nu+m+1}, \end{aligned}$$

et

$$\frac{1}{n} \sum_{\nu=0}^{n-d-1} \mathbf{e}_{\nu+m+1} \mathbf{e}'_{\nu+m+i+1} \xrightarrow{P} \mathbf{0}_{s \times s}$$

puisque les \mathbf{e}_i sont non corrélés. Ainsi, (5.31) devient :

$$\sum_{k=1}^d \operatorname{tr} \left(\mathbf{H}_0 \sum_{j=0}^{\infty} \Psi_j \mathbf{o}_p(n) \right) = o_p(n),$$

car $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$. Cela nous permet de conclure que la première triple sommation de (5.30) est $o_p(n)$. En ce qui concerne la seconde triple sommation, maintenons k fixé pour l'instant. On remarque que

$$\begin{aligned} & \operatorname{tr} \left[\sum_{j=0}^{\infty} \sum_{\nu=0}^{n-d-1} \mathbf{V}'_{\nu+k-j+1} \Psi_j \mathbf{e}_{\nu+k-j+1} \right] \\ &= \operatorname{tr} \left[\sum_{j=0}^{\infty} \Psi_j \sum_{\nu=0}^{n-d-1} \mathbf{e}_{\nu+k-j+1} \mathbf{V}'_{\nu+k-j+1} \right] \\ &= \operatorname{tr} \left[\sum_{j=0}^{\infty} \Psi_j \left(\sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{V}'_{\nu} + o_p(n) \right) \right] \\ &= \operatorname{tr} \left[\mathbf{H}_0 \left(\sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{V}'_{\nu} + o_p(n) \right) \right], & \text{car } \mathbf{H}_0 = \sum_{j=0}^{\infty} \Psi_j, \\ &= \operatorname{tr} \left[\mathbf{H}_0 \sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{X}'_{\nu} \mathbf{H}_0 \right] + o_p(n), & \text{car } \mathbf{V}_t = \mathbf{H}_0 \mathbf{X}_t, \\ &= \frac{1}{2} \operatorname{tr} \left[\mathbf{H}_0 \left(\sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{X}'_{\nu} + \sum_{\nu=1}^n \mathbf{X}_{\nu} \mathbf{e}'_{\nu} \right) \mathbf{H}_0 \right] + o_p(n). \end{aligned} \quad (5.32)$$

Un calcul assez simple nous montre que :

$$\begin{aligned} a) \quad & \sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{X}'_{\nu} = \sum_{\nu=1}^n \mathbf{e}_{\nu} \sum_{j=1}^n \mathbf{e}'_j = \sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{e}'_{\nu} + \sum_{j=1}^{n-1} \sum_{j=1+1}^n \mathbf{e}_j \mathbf{e}'_i \\ b) \quad & \mathbf{X}_n \mathbf{X}'_n = \sum_{j=1}^n \sum_{j=1}^n \mathbf{e}_i \mathbf{e}'_j \\ &= \sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{e}'_{\nu} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{e}_i \mathbf{e}'_j + \sum_{j=1}^{n-1} \sum_{i=j+1}^n \mathbf{e}_i \mathbf{e}'_j \\ &= \sum_{\nu=1}^n \mathbf{e}_{\nu} \mathbf{e}'_{\nu} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{e}_i \mathbf{e}'_j + \sum_{i=1}^{n-1} \sum_{j=1+1}^n \mathbf{e}_j \mathbf{e}'_i. \end{aligned}$$

En continuant le développement de (5.32), nous obtenons

$$\begin{aligned}
& \frac{1}{2} \operatorname{tr} \left[\mathbf{H}_0 \left(\sum_{\nu=1}^n \mathbf{e}_\nu \mathbf{e}'_\nu + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{e}_j \mathbf{e}'_i \right. \right. \\
& \quad \left. \left. + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{e}_i \mathbf{e}'_j + \sum_{\nu=1}^n \mathbf{e}_\nu \mathbf{e}'_\nu \right) \mathbf{H}_0 \right] + o_p(n) \\
&= \frac{1}{2} \operatorname{tr} \left[\mathbf{H}_0 \left(\mathbf{X}_n \mathbf{X}'_n + \sum_{\nu=1}^n \mathbf{e}_\nu \mathbf{e}'_\nu \right) \mathbf{H}_0 \right] + o_p(n), \\
&= \frac{1}{2} \operatorname{tr} \left[\mathbf{H}_0 \left(\mathbf{X}_n \mathbf{X}'_n + n\sigma^2 \mathbf{I}_s \right) \mathbf{H}_0 \right] + o_p(n), \\
&= \frac{1}{2} \operatorname{tr} [\mathbf{H}_0 \mathbf{X}_n \mathbf{X}'_n \mathbf{H}_0] + \frac{ns}{2} \sigma^2 \delta_0 + o_p(n), \\
&= \frac{1}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{ns}{2} \sigma^2 \delta_0 + o_p(n).
\end{aligned}$$

Ainsi, en sommant cette expression pour $k \in \{1, 2, \dots, d\}$, d étant fixe par rapport à n , on trouve

$$\sum_{j=0}^{\infty} \sum_{k=1}^d \sum_{\nu=0}^{n-d-1} \mathbf{V}'_{\nu+k-j+1} \boldsymbol{\Psi}_j \mathbf{e}_{\nu+k-j+1} = \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{dns}{2} \sigma^2 \delta_0 + o_p(n)$$

En substituant cette dernière expression dans l'équation (5.29), nous obtenons

$$\begin{aligned}
& \sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) \\
&= \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{n\sigma^2 ds \delta_0}{2} - \frac{n\sigma^2 ds \delta_0}{2} + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n) \\
&= \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n).
\end{aligned}$$

Nous avons donc démontré l'analogie du lemme 2 de HASZA (1980).

Lemme 5.5.

$$\sum_{t=1}^{(n-d)s} Z_{t+ds} (Z_{t+ds} - Z_t) = \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n) \quad (5.33)$$

La démonstration du lemme 5.6 (généralisation du lemme 3 de HASZA (1980)) est assez immédiate.

Lemme 5.6.

$$\sum_{t=1}^{(n-d)s} (Z_{t+ds} - Z_t)^2 = ns \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n) \quad (5.34)$$

Preuve. On a

$$\begin{aligned} & \sum_{t=1}^{(n-d)s} (Z_{t+ds} - Z_t)^2 \\ &= \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} (Z_{(\nu+d)s+r} - Z_{\nu s+r})^2 \\ &= \sum_{r=1}^s \sum_{\nu=0}^{n-d-1} \left(\sum_{h=1}^d \omega_{(\nu+h)s+r} \right)^2 \end{aligned}$$

Considérons alors

$$\begin{aligned} & \frac{1}{n} \sum_{\nu=0}^{n-d-1} \left(\sum_{h=1}^d \omega_{(\nu+h)s+r} \right)^2 \\ &= \frac{1}{n} \sum_{\nu=0}^{n-d-1} \sum_{h=1}^d \omega_{(\nu+h)s+r}^2 + \frac{2}{n} \sum_{\nu=0}^{n-d-1} \sum_{h=1}^{d-1} \sum_{k=h+1}^d \omega_{(\nu+h)s+r} \omega_{(\nu+k)s+r}. \end{aligned}$$

Or, le premier terme tend en probabilité vers $d\gamma_\omega(0)$. Le second terme tend en probabilité vers

$$2 \sum_{h=1}^{d-1} (d-h) \gamma_\omega(hs).$$

Donc,

$$\begin{aligned} & \frac{1}{n} \sum_{\nu=0}^{n-d-1} \left(\sum_{h=1}^d \omega_{(\nu+h)s+r} \right)^2 \xrightarrow{P} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) \\ & \Rightarrow \sum_{t=1}^{(n-d)s} (Z_{t+ds} - Z_t)^2 = ns \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n). \end{aligned}$$

Cela termine la démonstration du lemme. \square

Les équations (5.24), (5.33) et (5.34) nous indiquent que

$$\begin{aligned} ns(c(0) - c(ds)) &= d (\mathbf{X}_n - \bar{\mathbf{X}})' \mathbf{H}_0^2 (\mathbf{X}_n - \bar{\mathbf{X}}) \\ &+ dc_0^2 \bar{X}' \mathbf{1}'_s \mathbf{X}_n - \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n \\ &+ \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n^{1/2}) + o_p(n). \end{aligned}$$

Selon (FULLER, 1976, p. 227)), le terme $O_p(n^{1/2}) + o_p(n)$ est d'ordre $o_p(n)$, ainsi,

$$\begin{aligned}
ns(c(0) - c(ds)) &= d\mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + d\bar{\mathbf{X}}' \mathbf{H}_0^2 \bar{\mathbf{X}} - 2d\mathbf{X}'_n \mathbf{H}_0^2 \bar{\mathbf{X}} + dc_0^2 \bar{\mathbf{X}} \mathbf{1}'_s \mathbf{X}_n \\
&\quad - \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n) \\
&= \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + dsg_0^2 \bar{\mathbf{X}}^2 - 2dg_0^2 \bar{\mathbf{X}} \mathbf{1}'_s \mathbf{X}_n + dc_0^2 \bar{\mathbf{X}} \mathbf{1}'_s \mathbf{X}_n \\
&\quad + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n) \\
&= \frac{d}{2} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + d(c_0^2 - 2g_0^2) \bar{\mathbf{X}} \mathbf{1}'_s \mathbf{X}_n + dsg_0^2 \bar{\mathbf{X}}^2 \\
&\quad + \frac{ns}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(n).
\end{aligned}$$

5.3.2. Dénominateur de $1 - r(ds)$

Nous allons maintenant trouver une expression plus simple pour $ns c(0)$. De manière plus précise, nous avons

$$\begin{aligned}
ns c(0) &= \sum_{t=1}^{ns} (Z_t - \bar{Z})^2 \\
&= \sum_{t=1}^{ns} (\lambda_{\eta(t)} - \bar{\Lambda})^2 + \sum_{t=1}^{ns} (V_t - \bar{V})^2 + \sum_{t=1}^{ns} (D_t - \bar{D})^2 \\
&\quad + 2 \sum_{t=1}^{ns} (\lambda_{\eta(t)} - \bar{\Lambda}) (V_t - \bar{V}) - 2 \sum_{t=1}^{ns} (\lambda_{\eta(t)} - \bar{\Lambda}) (D_t - \bar{D}) \\
&\quad - 2 \sum_{t=1}^{ns} (V_t - \bar{V}) (D_t - \bar{D}).
\end{aligned} \tag{5.35}$$

Examinons chacun des six termes.

(1) Pour le premier :

$$\sum_{t=1}^{ns} (\lambda_{\eta(t)} - \bar{\Lambda})^2 = \sum_{\nu=0}^{n-1} \sum_{r=1}^s (\lambda_r - \bar{\Lambda})^2 = ns\sigma^2(\Lambda) = O(n). \tag{5.36}$$

(2) Pour le second, développons :

$$\begin{aligned}
& \sum_{t=1}^{ns} (D_t - \bar{D})^2 \\
&= \sum_{\nu=0}^{n-1} \sum_{r=1}^s (D_{\nu s+r} - \bar{D})^2 \\
&= \sum_{\nu=0}^{n-1} \sum_{r=1}^s (D_{\nu s+r} - \bar{D}_r)^2 + \sum_{\nu=0}^{n-1} \sum_{r=1}^s (\bar{D}_r - \bar{D})^2 \\
&\quad + 2 \sum_{\nu=0}^{n-1} \sum_{r=1}^s (D_{\nu s+r} - \bar{D}_r) (\bar{D}_r - \bar{D}) \quad \text{où } \bar{D}_r = \frac{1}{n} \sum_{\nu=0}^{n-1} D_{\nu s+r} \\
&= \sum_{\nu=0}^{n-1} \sum_{r=1}^s (D_{\nu s+r} - \bar{D}_r)^2 + n \sum_{r=1}^s (\bar{D}_r - \bar{D})^2 \\
&= \sum_{r=1}^s \Gamma_{\mathbf{D}}(0, r, r) + o_p(n) = o_p(n)
\end{aligned}$$

où $\Gamma_{\mathbf{D}}(0, r, r)$ désigne l'élément situé à la position r, r dans la matrice des covariances au délai 0 associée au processus \mathbf{D} , notée $\mathbf{\Gamma}_{\mathbf{D}}(0)$. Par conséquent,

$$\sum_{t=1}^{ns} (D_t - \bar{D})^2 = \text{tr}(\mathbf{\Gamma}_{\mathbf{D}}(0)) + o_p(n) \quad (5.37)$$

(3) Développons :

$$\begin{aligned}
& \sum_{t=1}^{ns} (\lambda_{\eta(t)} - \bar{\Lambda}) (V_t - \bar{V}) \\
&= \sum_{r=1}^s (\lambda_r - \bar{\Lambda}) \sum_{\nu=0}^{n-1} (v_{\nu s+r} - \bar{V}) \\
&= \sum_{r=1}^s (\lambda_r - \bar{\Lambda}) \sum_{\nu=0}^{n-1} V_{\nu s+r} - \sum_{r=1}^s (\lambda_r - \bar{\Lambda}) \sum_{\nu=0}^{n-1} \bar{V} \\
&= \sum_{r=1}^s (\lambda_r - \bar{\Lambda}) \sum_{\nu=0}^{n-1} V_{\nu s+r} + O_p(n^{3/2}), \quad \text{car } \bar{V} = O_p(n^{1/2}).
\end{aligned}$$

Examinons maintenant $\sum_{\nu=0}^{n-1} V_{\nu s+r}$. Nous avons

$$\begin{aligned}
\sum_{\nu=0}^{n-1} V_{\nu s+r} &= \sum_{\nu=1}^n \mathbf{a}' \mathbf{P}^r \mathbf{X}_{\nu} \\
&= \mathbf{a}' \mathbf{P}^r (\mathbf{1}'_n \otimes \mathbf{I}_s) \mathbf{X} \\
&= \mathbf{a}' \mathbf{P}^r (\mathbf{1}'_n \otimes \mathbf{I}_s) (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e} \\
&= \mathbf{a}' \mathbf{P}^r (\mathbf{1}'_n \mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}
\end{aligned}$$

dont la variance est

$$\begin{aligned}
& \sigma^2 \mathbf{a}' (\mathbf{1}'_n \mathbf{T}_n \otimes \mathbf{I}_s) (\mathbf{T}'_n \mathbf{1}_n \otimes \mathbf{I}_s) \mathbf{a} \\
&= \sigma^2 \mathbf{a}' (\mathbf{1}'_n \mathbf{T}_n \mathbf{T}'_n \mathbf{1}_n \otimes \mathbf{I}_s) \mathbf{a} \\
&= \sigma^2 \mathbf{a}' \left(\frac{n(2n+1)(n+1)}{6} \otimes \mathbf{I}_s \right) \mathbf{a} \\
&= \sigma^2 \frac{n(2n+1)(n+1)}{6} \mathbf{a}' \mathbf{a} = O(n^3) \\
&\Rightarrow \sum_{\nu=0}^{n-1} V_{\nu s+r} = O_p(n^{3/2}), \quad \text{pour } r = 1, 2, \dots, s.
\end{aligned}$$

Cela nous permet de conclure que

$$\sum_{r=1}^s (\lambda_r - \bar{\lambda}) \sum_{\nu=0}^{n-1} (V_{\nu s+r} - \bar{V}) = O_p(n^{3/2}) \quad (5.38)$$

$$(4) \sum_{t=1}^{ns} (\lambda_{\eta(t)} - \bar{\lambda}) (D_t - \bar{D}) = \sum_{r=1}^s (\lambda_r - \bar{\lambda}) \sum_{\nu=0}^{n-1} (D_{\nu s+r} - \bar{D})$$

Or,

$$\begin{aligned}
\sum_{\nu=0}^{n-1} (D_{\nu s+r} - \bar{D}) &= \sum_{\nu=0}^{n-1} (D_{\nu s+r} - \bar{D}_r) + \sum_{\nu=0}^{n-1} (\bar{D}_r - \bar{D}) \\
&= 0 + n(\bar{D}_r - \bar{D}) = o_p(n), \\
&\Rightarrow \sum_{\nu=0}^{n-1} (\lambda_{\eta(t)} - \bar{\lambda}) (D_t - \bar{D}) = o_p(n). \quad (5.39)
\end{aligned}$$

(5) On peut aussi montrer que

$$\sum_{t=1}^{ns} (V_t - \bar{V}) (D_t - \bar{D}) = O_p(n^{3/2})$$

En effet, $(D_t - \bar{D}) = O_p(1)$ et la sommation devient

$$\begin{aligned}
\sum_{t=1}^{ns} (V_t - \bar{V}) (D_t - \bar{D}) &= O_p(1) \sum_{t=1}^{ns} (V_t - \bar{V}) \\
&\text{(La variance de } D_t \text{ ne dépend pas de } n \text{)} \\
&= O_p(1) O_p(n^{3/2}) = O_p(n^{3/2}). \quad (5.40)
\end{aligned}$$

Des équations (5.36) à (5.40), on conclut que

$$\sum_{t=1}^{ns} (Z_t - \bar{Z})^2 = \sum_{t=1}^{ns} (V_t - \bar{V})^2 + O_p(n^{3/2}) \quad (5.41)$$

Or,

$$\begin{aligned}
\sum_{t=1}^{ns} (V_t - \bar{V})^2 &= \sum_{\nu=0}^{n-1} \sum_{r=1}^s V_{\nu s+r}^2 - \frac{1}{ns} \left(\sum_{\nu=0}^{n-1} \sum_{r=1}^s V_{\nu s+r} \right)^2 \\
&= \sum_{\nu=1}^n \|\mathbf{V}_\nu\|^2 - ns\bar{V}^2 \\
&= \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu - nsc_0^2 \bar{X}^2
\end{aligned} \tag{5.42}$$

Et par conséquent de (5.41) et (5.42), il découle

$$\frac{c(0)}{ns} = \frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu - \frac{c_0^2}{ns} \bar{X}^2 + O_p(n^{-1/2}) \tag{5.43}$$

Nous avons donc démontré un théorème qui est une généralisation de celui de HASZA (1980).

Théorème 5.4. *Considérons le modèle suivant*

$$\nabla_s Z_t = \omega_t, \quad t \in \mathbb{N}$$

où ω_t admet la représentation

$$\omega_t = \sum_{j=0}^{\infty} \Psi_j e_{t-j}, \quad t \in \mathbb{Z}$$

et où

- a) $\{e_t\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, de moyenne nulle et de variance σ^2 ,
- b) $\sum_{j=0}^{\infty} j |\Psi_j| < \infty$, $\Psi_0 = 1$.

Pour $d \geq 1$, on a

$$\begin{aligned}
c(0) - c(ds) &= \frac{d}{2ns} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{d}{ns} (c_0^2 - 2g_0^2) \bar{X} \mathbf{1}'_s \mathbf{X}_n + \frac{d}{n} g_0^2 \bar{X}^2 \\
&\quad + \frac{1}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(1)
\end{aligned} \tag{5.44}$$

$$\frac{c(0)}{ns} = \frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu - \frac{c_0^2}{ns} \bar{X}^2 + o_p(n^{-1/2}). \tag{5.45}$$

5.3.3. Distribution asymptotique

Supposons maintenant que le bruit blanc est gaussien. Puisque $ns(1 - r(ds)) = ns(c(0) - c(ds))/c(0)$, on constate que

$$ns(1 - r(ds)) = \frac{\left[\frac{d}{2ns} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{d}{ns} (c_0^2 - 2g_0^2) \bar{X} \mathbf{1}'_s \mathbf{X}_n + \frac{d}{n} g_0^2 \bar{X}^2 \right] + \frac{1}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(1)}{\frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu - \frac{c_0^2}{ns} \bar{X}^2 + O_p(n^{-1/2})} \quad (5.46)$$

Portons notre attention sur le premier terme du dénominateur de (5.46), à savoir

$$\Gamma_n^* = \frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu$$

Nous diagonaliserons la forme quadratique Γ_n^* . Cette diagonalisation nous fournira les variables canoniques que nous utiliserons afin de représenter tous les termes apparaissant dans l'expression (5.46). Nous généraliserons donc les résultats de DICKEY (1976) dans ce contexte précis. De simples considérations nous indiquent que

$$n^2 s^2 \Gamma_n^* = \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu = \mathbf{e}' (\mathbf{T}'_n \mathbf{T}_n \otimes \mathbf{H}_0^2) \mathbf{e} \quad (5.47)$$

En effet, $\mathbf{X} = (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e}$ en posant $\mathbf{X}_0 = \mathbf{0}$, dans l'équation (2.2). De plus,

$$(\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} = (\mathbf{I}_n \otimes \mathbf{H}_0^2) (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e} = (\mathbf{T}_n \otimes \mathbf{H}_0^2) \mathbf{e}$$

L'équation (5.47) découle du fait que

$$\begin{aligned} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu &= \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{H}_0^2)' \mathbf{X} \\ &= \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{H}_0)' (\mathbf{I}_n \otimes \mathbf{H}_0) \mathbf{X} \\ &= \mathbf{e}' (\mathbf{T}'_n \mathbf{T}_n \otimes \mathbf{H}_0^2) \mathbf{e} \end{aligned}$$

Nous avons déjà vu, en suivant la démarche de Dickey, comment obtenir la diagonalisation de $\mathbf{T}'_n \mathbf{T}_n$. La matrice \mathbf{H}_0^2 est une matrice réelle et symétrique. Nous pouvons donc trouver s vecteurs propres orthonormaux $\zeta_1, \zeta_2, \dots, \zeta_s$, tels que si nous formons la matrice

$$\mathbf{F} = [\zeta_1, \zeta_2, \dots, \zeta_s], \quad s \times s, \quad (5.48)$$

nous aurons

$$\mathbf{F}' \mathbf{H}_0^2 \mathbf{F} = \mathbf{D}_2, \quad (5.49)$$

où $\mathbf{D}_2 = \text{Diag}\{\eta_1, \eta_2, \dots, \eta_s\}$, η_i , $i = 1, \dots, s$, étant les valeurs propres de \mathbf{H}_0^2 (voir (SEARLE, 1982, section 11.6)). La démonstration du lemme suivant est simple et se retrouve dans (GRAHAM, 1981, p. 27).

Lemme 5.7. *Soit \mathbf{A} une matrice $p \times p$ et soit \mathbf{B} une matrice $q \times q$. Dénotons par $\mathbf{u}_1, \dots, \mathbf{u}_p$ les vecteurs propres associés aux valeurs propres $\lambda_1, \dots, \lambda_p$ de la matrice \mathbf{A} et par $\mathbf{v}_1, \dots, \mathbf{v}_q$ les vecteurs propres associés aux valeurs propres μ_1, \dots, μ_q de la matrice \mathbf{B} . Alors, le vecteur de longueur pq formé par $\mathbf{u}_i \otimes \mathbf{v}_j$ est le vecteur propre associé à la valeur propre $\lambda_i \mu_j$ de la matrice $\mathbf{A} \otimes \mathbf{B}$.*

Dans le cas qui nous intéresse, puisque \mathbf{H}_0^2 ne dépend pas de n , ni ses vecteurs propres ζ_1, \dots, ζ_s , ni ses valeurs propres η_1, \dots, η_s , ne dépendent de n . De plus, connaissant la diagonalisation de $\mathbf{T}'_n \mathbf{T}_n$ ainsi que celle de \mathbf{H}_0^2 , nous pouvons diagonaliser la forme quadratique de départ, à savoir

$$\mathbf{e}' (\mathbf{T}'_n \mathbf{T}_n \otimes \mathbf{H}_0^2) \mathbf{e}.$$

Rappelons que $\xi_{n,i}$ désigne le i -ième vecteur propre normalisé de la matrice $\mathbf{T}'_n \mathbf{T}_n$ associé à la valeur propre $\lambda_{n,i}$. En posant

$$\mathbf{E} = [\xi_{n,1}, \xi_{n,2}, \dots, \xi_{n,n}],$$

nous avons

$$\mathbf{E}' \mathbf{T}'_n \mathbf{T}_n \mathbf{E} = \mathbf{D}_1,$$

où $\mathbf{D}_1 = \text{diag}\{\lambda_{n,1}, \lambda_{n,2}, \dots, \lambda_{n,n}\}$. De plus, les vecteurs $\xi_{n,i} \otimes \zeta_j$, $i = 1, \dots, n$ et $j = 1, \dots, s$, sont les vecteurs propres orthonormés de $\mathbf{T}'_n \mathbf{T}_n \otimes \mathbf{H}_0^2$. La matrice $\mathbf{E} \otimes \mathbf{F}$ est une matrice orthogonale. Ses colonnes forment une base de \mathbb{R}^{ns} . Ainsi,

$$\forall \mathbf{e} \in \mathbb{R}^{ns} \exists \mathbf{Z} \in \mathbb{R}^{ns} \text{ tel que } \mathbf{e} = \mathbf{E} \otimes \mathbf{F} \mathbf{Z} \text{ et } \mathbf{Z} = \mathbf{E}' \otimes \mathbf{F}' \mathbf{e}$$

et, par (5.47),

$$\begin{aligned} \sum_{\nu=1}^n \mathbf{X}'_{\nu} \mathbf{H}_0^2 \mathbf{X}_{\nu} &= \mathbf{e}' (\mathbf{T}'_n \mathbf{T}_n \otimes \mathbf{H}_0^2) \mathbf{e} \\ &= \mathbf{Z}' (\mathbf{E}' \otimes \mathbf{F}') (\mathbf{T}'_n \mathbf{T}_n \otimes \mathbf{H}_0^2) (\mathbf{E} \otimes \mathbf{F}) \mathbf{Z} \\ &= \mathbf{Z}' (\mathbf{E}' \mathbf{T}'_n \mathbf{T}_n \mathbf{E}) \otimes (\mathbf{F}' \mathbf{H}_0^2 \mathbf{F}) \mathbf{Z} \\ &= \mathbf{Z}' (\mathbf{D}_1 \otimes \mathbf{D}_2) \mathbf{Z}, \end{aligned}$$

où \mathbf{D}_1 et \mathbf{D}_2 sont les matrices diagonales contenant respectivement les valeurs propres de $\mathbf{T}'_n \mathbf{T}_n$ et de \mathbf{H}_0^2 .

Convenons de dénoter les éléments de \mathbf{Z} de la façon suivante :

$$\mathbf{Z}' = (Z_{n,1,1}, \dots, Z_{n,1,s}, Z_{n,2,1}, \dots, Z_{n,2,s}, \dots, Z_{n,n,1}, \dots, Z_{n,n,s})$$

où l'indice n indique simplement que le vecteur \mathbf{Z} dépend de n .

Il est alors possible d'écrire la forme quadratique de départ sous forme canonique :

$$n^2 s^2 \Gamma_n^* = \sum_{i=1}^n \sum_{r=1}^s \lambda_{n,i} \eta_r Z_{n,ir}^2 = \sum_{r=1}^s \eta_r \sum_{i=1}^n \lambda_{n,i} Z_{n,i,r}^2$$

où les $Z_{n,i,r}$ sont des variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$. Si nous posons

$$\Gamma_{n,r} = \frac{1}{n^2} \sum_{t=1}^n \lambda_{n,t} Z_{n,t,r}^2$$

nous pouvons affirmer que

$$\Gamma_n^* = \frac{1}{s^2} \sum_{r=1}^s \eta_r \Gamma_{n,r}.$$

et puisque $\Gamma_{n,r}$ est une variable ayant une structure identique à celle de Γ_n présentée à la section 4.2.1, la distribution asymptotique de Γ_n^* est connue.

5.3.3.1. Les variables $T_{n,j}^*$. Posons $T_{n,j}^* = X_{n,j}/\sqrt{n}$ pour $j = 1, \dots, s$ et soit $\mathbf{T}_n^* = (T_{n,1}, T_{n,2}, \dots, T_{n,s})'$. Puisque X_t est une marche aléatoire, il est clair que

$$\mathbf{X}_n = \sum_{\nu=1}^n \mathbf{e}_\nu = (\mathbf{1}' \otimes \mathbf{I}_s) \mathbf{e}.$$

Par conséquent, puisque $\mathbf{T}_n^* = \mathbf{X}_n/\sqrt{n}$, on a

$$\begin{aligned} \mathbf{T}_n^* &= \frac{1}{\sqrt{n}} (\mathbf{1}'_n \otimes \mathbf{I}_s) \mathbf{e} \\ &= \frac{1}{\sqrt{n}} (\mathbf{1}'_n \otimes \mathbf{I}_s) (\mathbf{E} \otimes \mathbf{F}) \mathbf{Z} \\ &= \frac{1}{\sqrt{n}} (\mathbf{1}'_n \mathbf{E} \otimes \mathbf{F}) \mathbf{Z}. \end{aligned}$$

Mais puisque \mathbf{E} est une matrice dont les colonnes sont formées des vecteurs propres normalisés $\boldsymbol{\xi}_{n,i}$ de $\mathbf{T}_n' \mathbf{T}_n$, le produit $\mathbf{1}'_n \mathbf{E}$ nous donne une matrice $1 \times s$, dont la i -ième composante est la somme des composantes de $\boldsymbol{\xi}_{n,i}$. On voit ainsi que

$$\frac{1}{\sqrt{n}} (\mathbf{1}'_n \mathbf{E} \otimes \mathbf{F}) = \frac{1}{\sqrt{n}} \left[\frac{1}{\|\mathbf{x}_{n,1}\|} \sum_{t=1}^n x_{n,1,t} \mathbf{F}, \dots, \frac{1}{\|\mathbf{x}_{n,n}\|} \sum_{t=1}^n x_{n,n,t} \mathbf{F} \right],$$

de sorte que

$$\frac{1}{\sqrt{n}} (\mathbf{1}'_n \mathbf{E} \otimes \mathbf{F}) z = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n x_{n,i,t} \mathbf{F} \mathbf{Z}_i,$$

et ainsi, la j -ième composante de \mathbf{T}_n^* est

$$\begin{aligned} T_{n,j}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n x_{n,i,t} \sum_{k=1}^s \zeta_{j,k} Z_{n,i,k} \\ &= \sum_{k=1}^s \zeta_{j,k} \sum_{i=1}^n \left(\frac{1}{\sqrt{n}\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n x_{n,i,t} \right) Z_{n,i,k} \\ &= \sum_{k=1}^s \zeta_{j,k} \sum_{i=1}^n \alpha_{n,i} Z_{n,i,k}, \end{aligned}$$

où $\alpha_{n,i}$ est l'expression entre crochets. Or, nous connaissons le comportement de la variable aléatoire $\sum_{i=1}^n \alpha_{n,i} Z_{n,i,k}$, puisqu'elle a la même structure que la variable T_n de la section 4.2.2. On peut résumer ainsi :

$$T_{n,j}^* = \sum_{r=1}^s \zeta_{jr} T_{n,r}$$

ce qui signifie que $\mathbf{T}_n^* = \mathbf{F}' \boldsymbol{\tau}_n$, où $\boldsymbol{\tau}_n = (T_{n,1}, \dots, T_{n,s})'$ et où

$$T_{n,r} = \sum_{i=1}^n \alpha_{n,i} Z_{n,i,r}.$$

5.3.3.2. Les variables $W_{n,j}^*$. Soit $W_{n,j}^* = \bar{X}_j / \sqrt{n}$, où

$$\bar{X}_j = \frac{1}{n} \sum_{t=1}^n X_{t,j}, \quad j = 1, \dots, s$$

sont les moyennes des observations ayant la même position relative à l'intérieur d'une période. Nous allons exprimer les variables $W_{n,j}^*$ en termes des variables canoniques. Notons premièrement que

$$\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_s \end{bmatrix} = \frac{1}{n} \mathbf{1}'_n \otimes \mathbf{I}_s \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} = \frac{1}{n} (\mathbf{1}'_n \otimes \mathbf{I}_s) (\mathbf{T}_n \otimes \mathbf{I}_s) \mathbf{e},$$

et en posant $\mathbf{n} = (n, n-1, \dots, 1)'$, on a

$$\mathbf{W}_n^* = \left(\frac{1}{n^{3/2}} (n, n-1, \dots, 1) \otimes \mathbf{I}_s \right) \mathbf{e} = n^{-3/2} (\mathbf{n}' \otimes \mathbf{I}_s) \mathbf{e}.$$

Un raisonnement analogue à celui de la section précédente nous indique que

$$\begin{aligned} \mathbf{W}_n^* &= n^{-3/2} (\mathbf{n}' \otimes \mathbf{I}_s) \mathbf{e}, \\ &= n^{-3/2} (\mathbf{n}' \otimes \mathbf{I}_s) (\mathbf{E} \otimes \mathbf{F}) \mathbf{Z}, \\ &= n^{-3/2} (\mathbf{n}' \mathbf{E} \otimes \mathbf{F}) \mathbf{Z}, \end{aligned}$$

et ainsi,

$$\mathbf{W}_n^* = \frac{1}{n^{3/2}} \left[\frac{1}{\|\mathbf{x}_{n,1}\|} \sum_{t=1}^n (n-t+1)x_{n,1,t} \mathbf{F}, \dots, \frac{1}{\|\mathbf{x}_{n,n}\|} \sum_{t=1}^n (n-t+1)x_{n,n,t} \mathbf{F} \right] \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix}.$$

En isolant $W_{n,j}^*$, la j -ième composante de \mathbf{W}_n^* , on constate que

$$\begin{aligned} W_{n,j}^* &= \frac{1}{n^{3/2}} \sum_{i=1}^n \frac{1}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n (n-t+1)x_{n,i,t} \sum_{k=1}^s \zeta_{j,k} Z_{n,i,k} \\ &= \sum_{k=1}^s \zeta_{j,k} \sum_{i=1}^n \left(\frac{n^{-3/2}}{\|\mathbf{x}_{n,i}\|} \sum_{t=1}^n (n-t+1)x_{n,i,t} \right) Z_{n,i,k} \\ &= \sum_{k=1}^s \zeta_{j,k} \sum_{i=1}^n \delta_{n,i} Z_{n,i,k}, \end{aligned}$$

où $\delta_{n,i}$ est l'expression entre parenthèses. Nous connaissons le comportement asymptotique de $\sum_{i=1}^n \delta_{n,i} Z_{n,i,k}$, puisqu'il a la même structure que la variable W_n de la section 4.2.3. On peut résumer ainsi :

$$W_{n,j}^* = \sum_{r=1}^s \zeta_{j,r} W_{n,r}$$

ce qui signifie que $\mathbf{W}_n^* = \mathbf{F}' \mathbf{W}_n$, où $\mathbf{W}_n = (W_{n,1}, \dots, W_{n,s})'$, et où

$$W_{n,r} = \sum_{i=1}^n \delta_{n,i} Z_{n,i,r}.$$

On se donne alors une suite de variables aléatoires $\varepsilon_{t,j}$ pour $1 \leq t < \infty$, $1 \leq j \leq s$, indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$, et on pose :

$$\Gamma_j = \sum_{t=1}^{\infty} \gamma_t^2 \varepsilon_{t,j}^2, \quad j = 1, \dots, s, \quad (5.50)$$

$$T_j = \sqrt{2} \sum_{t=1}^{\infty} \gamma_t \varepsilon_{t,j}, \quad j = 1, \dots, s, \quad (5.51)$$

$$W_j = \sqrt{2} \sum_{t=1}^{\infty} \gamma_t^2 \varepsilon_{t,j}, \quad j = 1, \dots, s$$

Soit

$$\boldsymbol{\chi} = (\Gamma_1, \dots, \Gamma_s, T_1, \dots, T_s, W_1, \dots, W_s)'$$

Le théorème 4.1 nous assure que ce vecteur est bien défini. De plus, la suite des vecteurs

$$\boldsymbol{\chi}_n = (\Gamma_{n,1}, \dots, \Gamma_{n,s}, T_{n,1}, \dots, T_{n,s}, W_{n,1}, \dots, W_{n,s})'$$

converge en loi vers $\boldsymbol{\chi}$, puisque les composantes $\Gamma_{n,r}$, $T_{n,r}$ et $W_{n,r}$ ont des structures identiques à celles des variables Γ_n , T_n et W_n . Elles partagent donc les mêmes propriétés.

Nous avons d'une part que $ns(1 - r(ds))$ est égal au quotient suivant :

$$\frac{\left(\frac{d}{2ns} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n + \frac{d}{ns} (c_0^2 - 2g_0^2) \bar{X} \mathbf{1}'_s \mathbf{X}_n + \frac{d}{n} g_0^2 \bar{X}^2 \right) + \frac{1}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(ns) + o_p(1)}{\frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu - \frac{c_0^2}{ns} \bar{X}^2 + O_p(n^{-1/2})}. \quad (5.52)$$

Pour les termes du numérateur de (5.52), on note que :

a) D'une part

$$\begin{aligned} \frac{1}{n} \mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n &= \mathbf{T}'_n \mathbf{H}_0^2 \mathbf{T}_n^* \\ &= \boldsymbol{\tau}'_n \mathbf{F}' \mathbf{H}_0^2 \mathbf{F} \boldsymbol{\tau}_n \\ &= \sum_{k=1}^s \eta_k T_{n,k}^2, \end{aligned}$$

puisque \mathbf{F} diagonalise \mathbf{H}_0^2 , (cf. équation (5.49)).

b) D'autre part

$$\begin{aligned} \frac{1}{n} \bar{X} \mathbf{1}'_s \mathbf{X}_n &= \frac{1}{s} \mathbf{1}'_s \mathbf{W}_n^* \mathbf{1}'_s \mathbf{T}_n^* \\ &= \frac{1}{s} \mathbf{1}'_s \mathbf{F}' \mathbf{W}_n \mathbf{1}'_s \mathbf{F}' \mathbf{T}_n \\ &= \frac{1}{s} \sum_{j=1}^s \sum_{k=1}^s \zeta_{j,k} W_{n,k} \sum_{i=1}^s \sum_{r=1}^s \zeta_{i,r} T_{n,r} \end{aligned}$$

où $\zeta_{j,k}$, $j = 1, \dots, s$, sont les éléments du k -ième vecteur propre $\boldsymbol{\zeta}_k$ de \mathbf{F} , voir (5.48).

c) De plus

$$\begin{aligned} \frac{\bar{X}^2}{n} &= \frac{1}{s^2} (\mathbf{1}'_s \mathbf{W}_n^*)^2 \\ &= \frac{1}{s^2} \left[\sum_{j=1}^s \sum_{k=1}^s \zeta_{j,k} W_{n,k} \right]^2. \end{aligned}$$

Pour le dénominateur de (5.52),

$$\frac{1}{n^2} \sum_{\nu=1}^s \mathbf{X}'_{\nu} \mathbf{H}_0^2 \mathbf{X}_{\nu} = \sum_{r=1}^s \eta_r \Gamma_{n,r}$$

En substituant ces dernières expressions dans (5.52) et en multipliant le numérateur et le dénominateur par s^2 , nous obtenons

$$\frac{\left[\frac{ds}{2} \sum_{r=1}^s \eta_r T_{n,r}^2 + d(c_0^2 - 2g_0^2) \left(\sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} W_{n,r} \right) \left(\sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} T_{n,r} \right) + dg_0^2 \left(\sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} W_{n,r} \right)^2 + \frac{s^2}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_{\omega}(hs) + o_p(1) \right]}{\sum_{r=1}^s \eta_r \Gamma_{n,r} - \frac{c_0^2}{s} \left(\sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} W_{n,r} \right)^2 + O_p(n^{-1/2})}$$

Le théorème suivant se démontre par un raisonnement analogue à celui utilisé pour prouver le théorème 5.4.

Théorème 5.5. *Considérons le modèle suivant*

$$\nabla_s Z_t = \omega_t, \quad t \in \mathbb{N}$$

où ω_t admet la représentation

$$\omega_t = \sum_{j=0}^{\infty} \Psi_j e_{t-j}, \quad t \in \mathbb{Z}$$

et où :

- a) e_t est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$,
- b) $\sum_{j=0}^{\infty} j |\Psi_j| < \infty$, $\Psi_0 = 1$.

Dénotons par $r(ds)$ le coefficient d'autocorrélation de délai ds , i.e.

$$r(ds) = \frac{\sum_{t=1}^{(n-d)s} (Z_t - \bar{Z})(Z_{t+ds} - \bar{Z})}{\sum_{t=1}^{ns} (Z_t - \bar{Z})^2}$$

Alors

$$ns(1 - r(ds)) \xrightarrow{Loi} R(d, s)$$

où $R(d, s)$ est connu est donné par

$$R(d, s) = \frac{\left[\frac{ds}{2} \sum_{r=1}^s \eta_r T_r^2 + d (c_0^2 - 2g_0^2) \sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} W_r \sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} T_r \right. \\ \left. + dg_0^2 \left(\sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} W_r \right)^2 + \frac{s^2}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) \right]}{\sum_{r=1}^s \eta_r \Gamma_r - \frac{c_0^2}{s} \left(\sum_{r=1}^s \sum_{j=1}^s \zeta_{j,r} W_r \right)^2} \quad (5.53)$$

Si nous appliquons le théorème 5.5 avec $s = 1$, nous obtenons comme corollaire le théorème 2 de HASZA (1980).

Corollaire 5.3 : Si $\nabla Z_t = \omega_t$, et que ω_t satisfait les conditions du théorème 5.5, et si $r(d)$ dénote le coefficient d'autocorrélation échantillonnal de délai d , alors $n(1 - r(d))$ converge en loi vers $R(d, 1) = R_d$, où R_d est donné par

$$R_d = \frac{d \left(\frac{T^2}{2} - WT + W^2 \right) + \frac{c_0^{-2}}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(h)}{\Gamma - W^2}$$

Preuve. Si $s = 1$, alors $\mathbf{H}_0 = c_0$, i.e., il s'agit d'un scalaire, et $\mathbf{H}_0^2 = c_0^2$. Par conséquent, $g_0^2 = c_0^2$. La seule valeur propre de \mathbf{H}_0^2 est $\eta_1 = c_0^2$ et elle est associée au vecteur propre $\zeta_1 = \zeta_{1,1} = 1$. Par simple substitution de ces valeurs dans l'expression (5.53), on obtient le résultat cherché après avoir divisé le numérateur et le dénominateur par c_0^2 et simplifié l'expression. \square

5.4. Distribution asymptotique de $ns(1 - r(ds))$ dans le contexte de la marche aléatoire saisonnière

Nous nous plaçons maintenant dans le contexte de la marche aléatoire saisonnière. Nous pouvons utiliser les résultats démontrés jusqu'ici pour mettre en évidence la distribution asymptotique de $ns(1 - r(ds))$, où $r(ds)$ est le coefficient de corrélation de délai ds , calculé de la façon usuelle, c'est-à-dire, à partir des données centrées. Afin de pouvoir exploiter les résultats de Dickey tels que nous les avons présentés, nous supposons que $\nabla_s Z_t = e_t$, où $\{e_t\}$ est une suite de variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$. Dans ce cas particulier, la matrice \mathbf{H}_0 prend

une forme simple donnant une matrice \mathbf{H}_0^2 encore plus simple.

$$\mathbf{H}_0 = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

et nous constatons que $\mathbf{H}_0^2 = \mathbf{I}_s$. De plus, $g_0^2 = c_0 = 1$, de sorte que les expressions du théorème 5.4 prennent les formes suivantes :

$$c(0) - c(ds) = \frac{d}{2} \left[\frac{\mathbf{X}'_n \mathbf{X}_n}{ns} - 2 \frac{\bar{X} \mathbf{1}'_s \mathbf{X}_n}{ns} + 2 \frac{\bar{X}^2}{n} + \sigma^2 \right] + o_p(1),$$

$$\frac{c(0)}{ns} = \frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_{\nu} \mathbf{X}_{\nu} - \frac{\bar{X}^2}{ns} + o_p(n^{-1/2}).$$

Rappelons maintenant certaines notations :

$$\begin{aligned} X_{j,r} &= X_{(j-1)s+r}, & j &= 1, \dots, n, \quad r = 0, \dots, s-1, \\ \bar{X}_r &= \frac{1}{n} \sum_{i=1}^n X_{i,r}, & \bar{X} &= \frac{1}{s} \sum_{r=1}^s \bar{X}_r, \\ T_r &= \sqrt{2} \sum_{i=1}^{\infty} \gamma_i \varepsilon_{i,r}, & T_{n,r} &= X_{n,r} / \sqrt{n}, \\ W_r &= \sqrt{2} \sum_{i=1}^{\infty} \gamma_i^2 \varepsilon_{i,r}, & W_{n,r} &= \bar{X}_r / \sqrt{n}, \\ \Gamma_r &= \sum_{i=1}^{\infty} \gamma_i^2 \varepsilon_{i,r}^2, & \Gamma_{n,r} &= \frac{1}{n^2} \sum_{t=1}^n X_{t,r}^2, \end{aligned}$$

où les $\{\varepsilon_{i,r}\}_{i=1}^{\infty}$, $r = 1, \dots, s$, sont des suites de variables aléatoires normales indépendantes de moyenne nulle et de variance σ^2 . Par le théorème 4.1,

$$(T_{n,r}, W_{n,r}, \Gamma_{n,r}) \xrightarrow{Loi} (T_r, W_r, \Gamma_r).$$

À cause de l'indépendance entre $X_{t,u}$ et $X_{t,\nu}$, $u \neq \nu$, le vecteur de longueur $3s$, à savoir,

$$\boldsymbol{\chi}_n = (T_{n,1}, W_{n,1}, \Gamma_{n,1}, \dots, T_{n,s}, W_{n,s}, \Gamma_{n,s})$$

converge en loi vers le vecteur

$$\boldsymbol{\chi} = (T_1, W_1, \Gamma_1, \dots, T_s, W_s, \Gamma_s)'.$$

Nous pouvons exprimer $ns(c(0) - c(ds))/c(0) = ns(1 - r(ds))$ comme le quotient de deux fonctions continues de χ_n . Plus précisément, on voit que :

$$\begin{aligned}
 ns(1 - r(ds)) &= \frac{\left(\frac{d}{2s} \sum_{r=1}^s \frac{X_{n,r}^2}{n} - \frac{d}{s^2} \sum_{r=1}^s \frac{\bar{X}_r}{\sqrt{n}} \sum_{t=1}^s \frac{X_{n,t}}{\sqrt{n}} \right. \\
 &\quad \left. + \frac{d}{s^2} \left(\sum_{t=1}^s \frac{\bar{X}_t}{\sqrt{n}} \right)^2 + \frac{d\sigma^2}{2} + o_p(1) \right)}{\frac{1}{s^2} \sum_{r=1}^s \sum_{t=1}^n \frac{X_{t,r}^2}{n^2} - \frac{1}{s^3} \left(\sum_{t=1}^s \frac{\bar{X}_r}{\sqrt{n}} \right)^2 + O_p(n^{-1/2})} \\
 &= \frac{\left(ds \left[\frac{1}{2} \sum_{r=1}^s T_{n,r}^2 - \frac{1}{s} \sum_{r=1}^s W_{n,r} \sum_{t=1}^s T_{n,t} \right. \right. \\
 &\quad \left. \left. + \frac{1}{s} \left(\sum_{t=1}^s W_{n,t} \right)^2 + \frac{s\sigma^2}{2} \right] + o_p(1) \right)}{\sum_{r=1}^s \Gamma_{n,r} - \frac{1}{s} \left(\sum_{t=1}^s W_{n,t} \right)^2 + O_p(n^{-1/2})} \tag{5.54}
 \end{aligned}$$

qui, par un raisonnement analogue à celui utilisé pour démontrer le théorème 5.4, converge en loi vers :

$$R(d, s) = \frac{ds \left(\frac{1}{2} \sum_{r=1}^s T_r^2 - \frac{1}{s} \sum_{r=1}^s W_r \sum_{t=1}^s T_t + \frac{1}{s} \sum_{r=1}^s \sum_{t=1}^s W_r W_t + \frac{s\sigma^2}{2} \right)}{\sum_{r=1}^s \Gamma_r - \frac{1}{s} \sum_{r=1}^s \sum_{t=1}^s W_r W_t} \tag{5.55}$$

puisque le dénominateur est positif avec probabilité 1.

Sachant que $ns(1 - r(ds))$ tend en loi vers $R(d, s)$, $1 - r(ds)$ converge en loi vers 0, c'est-à-dire que $1 - r(ds)$ tend en probabilité vers 0, donc $r(ds)$ converge en probabilité vers 1.

La comparaison des équations (5.55) et (5.10) nous montre que les distributions asymptotiques sont différentes selon que nous centrons ou non les données. Cela est également vrai pour le modèle saisonnier plus général comme nous le montrerons dans la section suivante.

5.5. Distribution asymptotique de $ns(1 - r'(ds))$ dans le contexte du modèle plus général

Si nous ne centrons pas les données, le numérateur de $(1 - r'(ds))$ est

$$ns(c(0) - c(ns)) = \sum_{t=1}^{ns} Z_t^2 - \sum_{t=1}^{(n-d)s} Z_t Z_{t+ds}$$

Un simple calcul nous montre alors que

$$\begin{aligned} & \sum_{t=1}^{ns} Z_t^2 - \sum_{t=1}^{(n-d)s} Z_t Z_{t+ds} \\ &= \sum_{t=(n-d)s+1}^{ns} Z_t^2 - \sum_{t=1}^{(n-d)s} Z_{t+ds}(Z_{t+ds} - Z_t) + \sum_{t=1}^{(n-d)s} (Z_{t+ds} - Z_t)^2 \end{aligned} \quad (5.56)$$

En appliquant les lemmes 5.4 a), 5.5, et 5.6, et en divisant par ns , l'expression (5.56) devient

$$c'(0) - c'(ds) = \frac{d\mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n}{2ns} + \frac{1}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(1) \quad (5.57)$$

En utilisant le fait que $Z_t = \lambda_{\eta(t)} + V_t - D_t$, on a :

$$Z_t^2 = \lambda_{\eta(t)}^2 + V_t^2 - D_t^2 + 2\lambda_{\eta(t)}V_t - 2\lambda_{\eta(t)}D_t - 2V_tD_t$$

En sommant sur les valeurs de $t = 1, \dots, ns$, par un raisonnement analogue à celui de la section 5.3.2, on montre que, voir (5.41),

$$\sum_{t=1}^{ns} Z_t^2 = \sum_{t=1}^{ns} V_t^2 + O_p(n^{3/2})$$

et par conséquent,

$$\frac{c'(0)}{ns} = \frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu + O_p(n^{-1/2}) \quad (5.58)$$

L'équation (5.58) est l'analogie de (5.43). De (5.57) et (5.58), on obtient

$$ns(1 - r'(ds)) = \frac{\frac{d\mathbf{X}'_n \mathbf{H}_0^2 \mathbf{X}_n}{2ns} + \frac{1}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs) + o_p(1)}{\frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{H}_0^2 \mathbf{X}_\nu + O_p(n^{-1/2})}$$

qui converge en loi vers

$$R'(d, s) = \frac{\frac{ds}{2} \sum_{r=1}^s \eta_r T_r^2 + \frac{s^2}{2} \sum_{h=-d+1}^{d-1} (d - |h|) \gamma_\omega(hs)}{\sum_{r=1}^s \eta_r \Gamma_r} \quad (5.59)$$

si nous supposons que le bruit blanc est gaussien. Cette expression est différente de celle fournie par le théorème 5.5. Par conséquent, le fait de centrer ou non les données a un impact évident sur la distribution asymptotique des coefficients de corrélation échantillonnaires, même si la moyenne théorique de la série est nulle.

Si nous considérons à nouveau la marche aléatoire saisonnière, nous avons $\gamma_\omega(0) = \sigma^2$ et $\gamma_\omega(k) = 0$, pour $k \geq 1$. De plus, $\eta_r = 1$, pour $r = 0, \dots, s$, puisque $\mathbf{H}_0^2 = \mathbf{I}_s$. Si nous substituons ces valeurs dans (5.59), nous retrouvons l'expression du théorème 5.4.

5.6. Distribution asymptotique des autocovariances de la marche aléatoire saisonnière

En appliquant les résultats des sections précédentes, nous pouvons obtenir sans trop de difficultés la distribution asymptotique des autocovariances dont le délai est un multiple de la période. Pour pouvoir appliquer les résultats de Dickey, nous supposons que le bruit blanc est gaussien.

5.6.1. Avec observations centrées

En divisant l'équation (5.44) par ns et en soustrayant le résultat de l'équation (5.45), nous obtenons

$$\begin{aligned} \frac{c(ds)}{ns} &= \frac{1}{n^2 s^2} \sum_{\nu=1}^n \mathbf{X}'_\nu \mathbf{X}_\nu - \frac{\bar{X}^2}{ns} \\ &\quad - \frac{d}{2ns} \left(\frac{\mathbf{X}'_n \mathbf{X}_n}{ns} - \frac{2\bar{X} \mathbf{1}'_s \mathbf{X}_n}{ns} + \frac{2\bar{X}^2}{ns} + \sigma^2 \right) + o_p(n^{-1/2}). \end{aligned}$$

Sachant que

$$Y_n = O_p(n^{-1/2}) \implies Y_n = o_p(1)$$

on obtient

$$\frac{c(ds)}{ns} = \frac{1}{s^2} \sum_{r=1}^s \Gamma_{n,r} - \frac{1}{s^3} \left(\sum_{r=1}^s W_{n,r} \right)^2 - \frac{d}{2s^2} \sum_{r=1}^s \left(\frac{T_{n,r}}{\sqrt{n}} \right)^2$$

$$+ \frac{d}{s^3} \sum_{r=1}^s \frac{W_{n,r}}{\sqrt{n}} \sum_{t=1}^s \frac{W_{n,t}}{\sqrt{n}} - \frac{d}{s^3} \left(\sum_{r=1}^s \frac{W_{n,r}}{\sqrt{n}} \right)^2 + o_p(1) \quad (5.60)$$

Du fait que Y_n converge en loi vers Y , on déduit que $Y_n = O_p(1)$. Par conséquent, puisque $W_{n,r}$ converge en loi vers W_r , on a $W_{n,r} = O_p(1)$ et $W_{n,r}/\sqrt{n} = O_p(n^{-1/2})$. Ainsi, $W_{n,r}/\sqrt{n} = o_p(1)$. Le même argument montre que $T_{n,r}/\sqrt{n} = o_p(1)$. L'équation (5.60) devient alors

$$\frac{c(ds)}{ns} = \frac{1}{s^2} \sum_{r=1}^s \Gamma_{n,r} - \frac{1}{s^3} \left(\sum_{r=1}^s W_{n,r} \right)^2 + o_p(1)$$

qui converge en loi vers

$$C(s) = \frac{1}{s^2} \sum_{r=1}^s \Gamma_r - \frac{1}{s^3} \left(\sum_{r=1}^s W_r \right)^2$$

où Γ_r et W_r sont définis par (5.50) et (5.51). Utilisant la définition de Γ_r et de W_r , il est possible d'obtenir la moyenne de cette distribution. Sachant que

$$\sum_{j=1}^{\infty} \gamma_j^2 = \frac{1}{2} \quad \text{et} \quad \sum_{j=1}^{\infty} \gamma_j^4 = \frac{1}{6},$$

selon (DICKY, 1976, p. 45), on a :

$$\begin{aligned} \text{a) } E[\Gamma_r] &= \sum_{j=1}^{\infty} \gamma_j^2 E[\varepsilon_{i,r}] = \frac{1}{2} \sigma^2 \\ \text{b) } E \left[\sum_{u=1}^s \sum_{v=1}^s W_u W_v \right] &= 2 \sum_{u=1}^s \sum_{v=1}^s \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \gamma_i^2 \gamma_j^2 E[\varepsilon_{i,u} \varepsilon_{j,v}] \\ &= 2 \sum_{u=1}^s \sum_{i=1}^{\infty} \gamma_i^4 \sigma^2 \\ &= \frac{s \sigma^2}{3} \end{aligned}$$

À la lumière des deux dernières équations, on a

$$E[C(s)] = \frac{\sigma^2(3s-2)}{6s^2}$$

qui est l'expression obtenue en (2.20). Ainsi, la distribution limite a une moyenne égale à la limite des moyennes des lois exactes.

5.6.2. Avec observations non centrées

Ce cas est un peu plus simple. Nous avons

$$\begin{aligned} c'(ds) &= \frac{1}{ns} \sum_{t=1}^{(n-d)s} X_t X_{t+ds} \\ &= \frac{1}{ns} \sum_{i=1}^s \sum_{t=1}^{n-d} X_{(t-1)s+i} X_{(t+d-1)s+i} \end{aligned}$$

Chercher la distribution asymptotique de cette statistique est équivalent à chercher celle d'une somme de s termes de la forme

$$\begin{aligned} \sum_{t=1}^{n-d} X_t X_{t+d} &= - \sum_{t=1}^{n-d} X_{t+d} (X_{t+d} - X_t) + \sum_{t=1}^{n-d} X_{t+d}^2 \\ &= - \sum_{t=1}^{n-d} X_{t+d} (X_{t+d} - X_t) + \sum_{t=1}^n X_t^2 + O_p(1), \end{aligned}$$

puisque $\sum_{t=1}^n X_t^2 = O_p(1)$. En appliquant le lemme 5.2, nous obtenons

$$\sum_{t=1}^{n-d} X_t X_{t+d} = -\frac{d}{2} (X_n^2 + n\sigma^2) + \sum_{t=1}^{n-d} X_t^2 + O_p(n^{1/2}).$$

En divisant par $n^2 s^2$, la dernière équation devient

$$\begin{aligned} \frac{1}{n^2 s^2} \sum_{t=1}^{n-d} X_t X_{t+d} &= -\frac{d}{2s^2} \frac{X_n^2/n}{n} - \frac{d\sigma^2}{2ns^2} + \frac{1}{s^2} \sum_{t=1}^{n-d} \frac{X_t^2}{n^2} + O_p(n^{-3/2}) \\ &= -\frac{d}{2ns^2} T_n^2 + \frac{\Gamma_n}{s^2} + O_p(n^{-3/2}) \\ &= O_p(1/n) + \frac{\Gamma_n}{s^2} + O_p(n^{-3/2}) \\ &= \frac{\Gamma_n}{s^2} + O_p(1/n) \\ &= \frac{\Gamma_n}{s^2} + o_p(1). \end{aligned}$$

On conclut donc que

$$\frac{c'(ds)}{ns} = \frac{1}{s^2} \sum_{r=1}^s \Gamma_{n,r} + o_p(1) \quad (5.61)$$

Il découle de (5.61) que $c(ds)/(ns)$ converge en loi vers

$$C'(s) = \frac{1}{s^2} \sum_{r=1}^s \Gamma_r.$$

Sachant que $E[\Gamma_r] = \sigma^2/2$ et $\text{Var}[\Gamma_r] = \sigma^4/3$ ((DICKY, 1976, p. 45)), on a

$$\begin{aligned} E[c'(ds)/(ns)] &= \sigma^2/(2s), \\ \text{Var}[c'(ds)/(ns)] &= \sigma^4/(3s^3). \end{aligned}$$

Il est intéressant de noter que ces moments de la loi limite correspondent aux limites des moments des distributions exactes, c.f. (2.45) et (2.46) avec $h_0 = 1$.

5.7. Quelques comparaisons

5.7.1. Processus autorégressif d'ordre 1

Nous pouvons obtenir une expression assez simple pour la distribution asymptotique de $n(1 - r'(1))$ où

$$r'(1) = \frac{\sum_{t=1}^{n-1} X_t X_{t+1}}{\sum_{t=1}^n X_t^2}.$$

lorsque les X_t proviennent du modèle

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots$$

où les ε_t sont des variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$ et où $\mathbf{X}_0 = 0$, sous l'hypothèse $H_0 : \rho = 1$. Ce modèle est celui considéré par DICKY et FULLER (1979). La statistique $r'(1)$ est légèrement différente de l'estimateur de vraisemblance maximale et des moindres carrés de ρ , à savoir

$$r^*(1) = \frac{\sum_{t=1}^{n-1} X_t X_{t+1}}{\sum_{t=1}^{n-1} X_t^2}.$$

En appliquant le théorème 5.4 avec $s = 1$ et $d = 1$, nous voyons que

$$n(1 - r'(1)) \xrightarrow{\text{Loi}} \frac{(\sigma^2 + T^2)/2}{\Gamma} \quad (5.62)$$

Dickey et Fuller obtiennent une forme sensiblement différente pour la distribution asymptotique de $n(1 - r^*)$. De leur corollaire 1, on conclut que

$$n(1 - r^*) \xrightarrow{\text{Loi}} \frac{(\sigma^2 - T^2)/2}{\Gamma}$$

La comparaison des deux distributions asymptotiques met en évidence l'importance du terme X_n au niveau du dénominateur. Puisque $X_n = O_p(n)$, son impact n'est pas négligeable.

5.7.2. Modèle saisonnier simple

Dans DICKEY ET COLLAB. (1984), on considère, entre autres, le modèle

$$Z_t = \rho Z_{t-s} + \varepsilon_t, \quad t = 1, 2, \dots$$

où $Z_{-s+1}, Z_{-s+2}, \dots, Z_0$ sont des valeurs initiales et où les ε_t sont des variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(0, \sigma^2)$. Sous l'hypothèse $H_0 : \rho = 1$, les auteurs obtiennent la distribution asymptotique de $ns(r^*(s) - 1)$, où

$$r^*(s) = \frac{\sum_{t=1}^n Z_t Z_{t-s}}{\sum_{t=1}^n Z_{t-s}^2}.$$

D'une façon plus explicite

$$ns(1 - r^*(s)) \xrightarrow{\text{Loi}} \frac{\frac{s}{2} \sum_{r=1}^s (\sigma^2 - T_r^2)}{\sum_{r=1}^s \Gamma_r}$$

En posant $d = 1$, dans le théorème 5.4 (ou 5.5), nous obtenons

$$ns(1 - r'(s)) \xrightarrow{\text{Loi}} \frac{\frac{s}{2} \sum_{r=1}^s (\sigma^2 + T_r^2)}{\sum_{r=1}^s \Gamma_r}$$

La différence entre les deux distributions asymptotiques est du même type que celle observée dans le cas du modèle autorégressif d'ordre 1. Maintenant, si nous posons $d = 1$ dans l'équation (5.55), nous obtenons

$$ns(1 - r(s)) \xrightarrow{\text{Loi}} \frac{\frac{s}{2} \sum_{r=1}^s (\sigma^2 + T_r^2) - \sum_{r=1}^s \sum_{t=1}^s W_r T_t + \sum_{r=1}^s \sum_{t=1}^s W_r W_t}{\sum_{r=1}^s \Gamma_r - \frac{1}{s} \sum_{r=1}^s \sum_{t=1}^s W_r W_t}$$

Le coefficient de corrélation $r(s)$ considéré ici est calculé en soustrayant la moyenne générale des observations. Parmi les modèles étudiés par Dickey, Hasza et Fuller, celui où il serait indiqué d'utiliser une telle statistique est (1.5), à savoir,

$$(Z_t - \mu) = \rho(Z_{t-s} - \mu) + \varepsilon_t, \quad t = 1, 2, \dots$$

Ils prennent en considération $r^*(s)$, l'estimateur des moindres carrés de ρ , et mettent en évidence le fait que

$$ns(1 - r^*(s)) \xrightarrow{\text{Loi}} \frac{\frac{s}{2} \sum_{r=1}^s (\sigma^2 - T_r^2) + \sum_{r=1}^s \sum_{t=1}^s W_r T_t - \sum_{r=1}^s \sum_{t=1}^s W_r W_t}{\sum_{r=1}^s \Gamma_r - \frac{1}{s} \sum_{r=1}^s \sum_{t=1}^s W_r W_t}$$

La différence entre les deux distributions asymptotiques est toujours du même type.

Chapitre 6

Résultats numériques

Résumé

Ce chapitre est consacré aux tests asymptotiques appliqués aux modèles de marches aléatoires saisonnières. Il débute par l'introduction des statistiques de test basées sur les distributions asymptotiques établies dans le chapitre précédent. L'objectif principal est d'évaluer la significativité des effets saisonniers et de vérifier certaines hypothèses structurelles des modèles étudiés.

Après un rappel des principes généraux des tests asymptotiques, plusieurs résultats sont présentés, notamment des conditions sous lesquelles les statistiques de test convergent en loi vers des distributions limites bien définies. Des théorèmes et corollaires viennent appuyer ces résultats, permettant de préciser les conditions d'application des tests en pratique. Enfin, des illustrations numériques et des considérations sur l'efficacité des tests sont abordées, soulignant l'importance des distributions asymptotiques pour l'inférence statistique dans les modèles saisonniers.

6.1. Introduction

Afin d'étudier la forme et les moments des distributions mises en évidence au chapitre cinq, nous procéderons par simulation. D'une façon plus précise, nous considérerons le modèle

$$X_t = X_{t-s} + \varepsilon_t \quad (6.1)$$

où ε_t est une suite de variables aléatoires $\mathcal{N}(0, \sigma^2)$. Nous nous intéresserons à la distribution exacte des statistiques $ns(1 - r(ds))$ et $ns(1 - r'(ds))$, de même qu'à leur distribution asymptotique. Sans perdre de généralité, nous supposerons que $\sigma^2 = 1$. On peut se rendre compte que les distributions impliquées ne dépendent pas

vraiment de σ^2 (et ainsi, faire disparaître σ^2 au chapitre 5). En effet, considérons $ns(1 - r'(ds))$ sous la forme (5.7). En divisant le numérateur et le dénominateur par σ^2 , on obtient

$$ns(1 - r'(ds)) = \frac{\frac{ds}{2} \sum_{r=1}^s \left[\frac{T_{n,r}^2}{\sigma^2} + 1 \right] + O_p(n^{-1/2})}{\frac{1}{\sigma^2} \sum_{r=1}^s \Gamma_{n,r}} \quad (6.2)$$

Ainsi, si nous considérons des variables canoniques de variance égale à 1, (voir (4.4)), (6.2) s'écrit

$$ns(1 - r'(ds)) = \frac{\frac{ds}{2} \sum_{r=1}^s [T_{n,r}^2 + 1] + o_p(n^{-1/2})}{\sum_{r=1}^s \Gamma_{n,r}}$$

et si nous supposons que la variance des $\varepsilon_{t,r}$ apparaissant en (5.8) et (5.9) est égale à 1, la variable $ns(1 - r'(ds))$ converge en loi vers

$$R'(d, s) = \frac{\frac{ds}{2} \sum_{r=1}^s (T_r^2 + 1)}{\sum_{r=1}^s \Gamma_r}.$$

Nous pourrions faire un raisonnement analogue relativement aux équations (5.54) et (5.55).

Nous supposerons maintenant que les valeurs initiales sont nulles. Le paramètre s prendra les valeurs 1, 4 et 12. Pour l'étude des distributions exactes, nous prendrons des longueurs de séries (N), égales à 60, 108 et 204, ces longueurs étant des multiples de 12. Pour $s = 1$, nous considérerons les 5 premiers coefficients d'autocorrélation, tandis que pour $s = 4$ et 12, nous nous limiterons aux délais s et $2s$. En ce qui a trait aux distributions asymptotiques, il suffit d'étudier $r(s)$ et $r'(s)$, puisque $R(d, s) = dR(1, s)$ et $R'(d, s) = dR'(1, s)$ selon les équations (5.10) et (5.55). Nous résumons le tout dans le tableau 6.1.

Pour chacune des 12 cellules du tableau 6.1, nous ferons une simulation de 10 000 observations. Pour étudier les distributions exactes, en utilisant l'équation (6.1), nous simulerons 10 000 séries chronologiques de longueur appropriée et nous calculerons pour chacune d'elles les autocorrélations requises. La simulation des distributions asymptotiques n'est pas aussi immédiate. Elle se fera en utilisant les équations (5.10) et (5.55) et la méthode sera exposée à la section 6.2.2.

Tableau 6.1 – Nombre de délais considérés pour les différentes valeurs de N et s

N	s		
	1	4	12
60	5	2	2
108	5	2	2
204	5	2	2
∞	1	1	1

Signalons enfin que tous les programmes reliés à la simulation ont été écrits en Pascal 6000 et exécutés sur un ordinateur Cyber 173. Cet ordinateur utilisant des mots mémoire de 60 bits, un calcul fait en simple précision donne un résultat dont la précision est comparable à celle obtenue en double précision sur un ordinateur ayant des mots mémoire de 32 bits.

6.2. Détails de la mise en œuvre

6.2.1. Générateur aléatoire de variables normales

La quantité de nombres aléatoires normaux nécessaires aux simulations est considérable. DICKEY et FULLER (1981) ont utilisé le générateur « *Super Duper* » de l'Université McGill. Puisque nous ne disposons pas de ce générateur, nous avons mis en œuvre l'algorithme M de (KNUTH, 1981, p. 122), basé sur la méthode de MARSAGLIA (1961) (connue sous le nom anglais de « *Marsaglia's rectangle-wedge-tail method* »¹) et sur les travaux de WALKER (1977). Citons ici Knuth :

The resulting method yields an extremely efficient program since its average running time is very small.

Le générateur uniforme utilisé est celui du système Pascal 6000.²

Pour mettre en œuvre cet algorithme, nous devons construire certaines tables de constantes, et la création de celles-ci est relativement complexe. Nous fournissons à l'annexe II le programme *ConsDef* écrit selon les spécifications de (KNUTH, 1981, exercices 7 et 10, section 3.4.1). Ce programme conserve dans un fichier les constantes calculées avec un nombre suffisant de décimales et imprime un tableau comparable au tableau 6.2.

1. Une traduction française pourrait être : « Méthode rectangle-triangle-queue de Marsaglia. »

2. Nous disposons d'une version en C++ utilisable avec R qui est plus rapide que la fonction `rnorm` de R. On observe une réduction de 50% du temps d'exécution.

Tableau 6.2 – Table de constantes nécessaires à l'exécution de l'algorithme M de Knuth

j	P_j	P_{j+16}	Q_j	Y_j	Y_{j+16}	Z_j	Z_{j+16}	S_{j+1}	D_{j+15}	E_{j+15}
0	0.000	0.067		0.00	0.59	0.20	0.21	0.0		
1	0.849	0.161	0.236	-0.92	0.96	1.32	0.24	0.2	0.505	25.00
2	0.970	0.236	0.206	-5.86	-0.06	6.66	0.26	0.4	0.773	12.50
3	0.855	0.285	0.234	-0.58	0.12	1.38	0.28	0.6	0.876	8.33
4	0.994	0.308	0.201	-33.16	1.31	34.96	0.29	0.8	0.939	6.25
5	0.995	0.304	0.201	-39.51	0.31	41.31	0.29	1.0	0.986	5.00
6	0.933	0.280	0.214	-2.57	1.12	2.97	0.28	1.2	0.995	4.06
7	0.923	0.241	0.217	-1.61	0.54	2.61	0.26	1.4	0.987	3.37
8	0.727	0.197	0.275	0.67	0.75	0.73	0.25	1.6	0.979	2.86
9	1.000	0.152	0.200	0.00	0.56	0.00	0.24	1.8	0.972	2.47
10	0.691	0.112	0.289	0.35	0.17	0.65	0.23	2.0	0.966	2.16
11	0.454	0.079	0.440	-0.17	0.38	0.37	0.22	2.2	0.960	1.92
12	0.287	0.052	0.698	0.92	-0.01	0.28	0.21	2.4	0.954	1.71
13	0.174	0.033	1.150	0.36	0.39	0.24	0.21	2.6	0.948	1.54
14	0.101	0.020	1.974	-0.02	0.20	0.22	0.20	2.8	0.942	1.40
15	0.057	0.086	3.526	0.19	0.78	0.21	0.22	3.0	0.936	1.27

Le lecteur peut vérifier que le tableau 6.2 correspond bien au tableau 1 de (KNUTH, 1981, section 3.4.1). La fonction `Normal` définie dans le programme `SimulI` de l'annexe A est la codification de l'algorithme M.

6.2.2. Simulation des distributions asymptotiques

Pour simuler les différentes distributions asymptotiques mises en évidence au chapitre cinq, nous devons évaluer des combinaisons linéaires infinies de variables aléatoires. Nous utiliserons la méthode suggérée par DICKEY (1976) et reprise par HASZA (1980). Il suggère de remplacer les sommes infinies par des sommes finies. Ces sommes finies ne sont pas obtenues par simple troncation des sommes infinies. Il s'agit plutôt de conserver les 72 premiers termes des séries, où nous remplaçons les γ_1 par des γ_1^* satisfaisant certaines propriétés.

Tableau 6.3 – Série $\sum_{i=1}^{\infty} \gamma_i^k$ pour différentes valeurs de k où $\gamma_i = \frac{2(-1)^{i+1}}{\pi(2i-1)}$

k	1	2	3	4	5	6	7	8
$\sum_{i=1}^{\infty} \gamma_i^k$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{5}{48}$	$\frac{1}{15}$	$\frac{61}{1440}$	$\frac{17}{630}$

6.2.2.1. Calcul des γ_i^* . Constatant que les variables aléatoires T_r , W_r et Γ_r , $r = 1, \dots, s$, sont des sommes infinies et pondérées de variables $\mathcal{N}(0, \sigma^2)$ (ou du

carré de ces variables), Dickey choisit de construire un ensemble fini de poids qui fait en sorte que les deux premiers moments des nouvelles variables sont égaux à ceux des sommes infinies. Pour les statistiques considérées par Dickey, ces deux premiers moments dépendent de $\sum_{i=1}^{\infty} \gamma_i^k$, $k = 2, \dots, 8$. Les valeurs de ces différentes sommes sont données au tableau 6.3 (DICKEY, 1976, p. 45).

On doit donc trouver des γ_i^* , $i = 1, \dots, 72$, dont les sommes correspondantes approcheront ces valeurs théoriques. Voici, dans ses grandes lignes, la procédure de Dickey :

- 1) Pour $i = 1, \dots, 14$, poser $\gamma_i^* = \gamma_i$.
- 2) Fixer pour l'instant $\gamma_{15}^* = \gamma_{15}$.
- 3) Trouver des constantes α et β telles que :

$$3.1) \gamma_i^{*2} = \alpha + \beta(i - 15), i = 16, \dots, 72,$$

$$3.2) \sum_{i=1}^{72} \gamma_i^{*2} = \sum_{i=1}^{\infty} \gamma_i^2 = \frac{1}{2},$$

$$3.3) \sum_{i=1}^{72} \gamma_i^{*4} = \sum_{i=1}^{\infty} \gamma_i^4 = \frac{1}{6},$$

et pour lesquelles on devrait avoir

$$3.4) \alpha + \beta(i - 15) \approx 0, \text{ pour } i = 72.$$

- 4) Évaluer $\delta = \max_{2 \leq k \leq 8} \left| \sum_{i=1}^{72} \gamma_i^{*k} - \sum_{i=1}^{\infty} \gamma_i^k \right|$. Si $\delta > 5 \times 10^{-10}$, assigner à γ_{15}^* une autre valeur et retourner à 3).

Comme on peut le voir dans la détermination des constantes γ_i^* , le point délicat est le troisième. D'un point de vue pratique, nous ne tiendrons compte que des contraintes 3.1), 3.2) et 3.3) et nous vérifierons si la condition 3.4) est satisfaite.

Pour déterminer les constantes α et β , nous utiliserons l'algorithme de Newton-Raphson à deux dimensions. Cet algorithme nécessite un vecteur de valeurs initiales. Aussi, lorsque $\gamma_{15}^* = \gamma_{15}$, ce vecteur de valeurs initiales sera obtenu par régression de γ_i^{*2} sur i , pour $i = 15, \dots, 72$, puis les valeurs finales de α et β calculées pour la dernière itération serviront de valeurs initiales pour la suivante.

Décrivons brièvement l'algorithme de Newton-Raphson tel que présenté dans BURDEN ET COLLAB. (1981), algorithme 9.1). Auparavant, introduisons les notations

suivantes :

$$\mathbf{x} = (\alpha, \beta)',$$

$$\mathbf{F}(\mathbf{x}) = \left(\sum_{i=1}^{72} \gamma_i^{*2} - \frac{1}{2}, \sum_{i=1}^{72} \gamma_i^{*4} - \frac{1}{6} \right)' = (F_1(\mathbf{x}), F_2(\mathbf{x}))',$$

$$\mathbf{J}(\mathbf{x}) = \begin{pmatrix} \frac{\partial F_1(\mathbf{x})}{\partial x_1} & \frac{\partial F_1(\mathbf{x})}{\partial x_2} \\ \frac{\partial F_2(\mathbf{x})}{\partial x_1} & \frac{\partial F_2(\mathbf{x})}{\partial x_2} \end{pmatrix}$$

Algorithme de Newton-Raphson

Début

$\mathbf{x} :=$ Point initial ;

$k := 1$; { compteur du nombre d'itérations }

Répéter

Évaluer $\mathbf{F}(\mathbf{x})$ et $\mathbf{J}(\mathbf{x})$;

Résoudre $\mathbf{J}(\mathbf{x})\mathbf{y} = -\mathbf{F}(\mathbf{x})$;

Poser

$\mathbf{x} := \mathbf{x} + \mathbf{y}$;

$k := k + 1$

jusqu'à ce que $\|\mathbf{y}\|_\infty < \varepsilon$ ou $k =$ maximum d'itérations permis.

Fin

où $\|\mathbf{y}\|_\infty = \max\{|y_1|, |y_2|\}$, et ε est une petite valeur (par exemple $\varepsilon = 10^{-17}$).

Pour résoudre le système $\mathbf{J}(\mathbf{X})\mathbf{Y} = -\mathbf{F}(\mathbf{X})$, nous utiliserons l'algorithme d'élimination de Gauss avec pivot maximal, c.f.(BURDEN ET COLLAB., 1981, algorithme 6.2).

Pour faciliter la programmation, nous écrirons les contraintes sous une forme plus commode. Pour la contrainte 3.2), nous avons

$$F_1(\mathbf{x}) = \sum_{i=1}^{14} \gamma_i^{*2} + \gamma_{15}^{*2} + \sum_{i=16}^{72} \gamma_i^{*2} - 1/2$$

de sorte que si nous convenons de dénoter par S_1 la quantité $\sum_{i=1}^{14} \gamma_i^2$, en utilisant la contrainte 3.1), la dernière équation devient

$$\begin{aligned} F_1(\mathbf{x}) &= S_1 + \gamma_{15}^{*2} + \sum_{i=16}^{72} (\alpha + \beta(i - 15)) - 1/2 \\ &= 57\alpha + \beta \sum_{i=16}^{72} (i - 15) - (1/2 - S_1 - \gamma_{15}^{*2}) \end{aligned}$$

$$= 57 \cdot \alpha + 57 \cdot 58 \cdot \beta / 2 - (1/2 - S_1 - \gamma_{15}^{*2})$$

$$F_1(\mathbf{x}) = 57(\alpha + 29\beta - K_1) = 0,$$

c'est-à-dire

$$\alpha + 29\beta - K_1 = 0$$

où $K_1 = (1/2 - S_1 - \gamma_{15}^{*2}) / 57$. Un travail analogue au niveau de la contrainte 3.3) où nous posons

$$S_2 = \sum_{i=1}^{14} \gamma_i^4, \quad \text{et}$$

$$K_2 = \frac{1/6 - S_2 - \gamma_{15}^{*4}}{57},$$

nous conduit à

$$\alpha^2 + 58\alpha\beta + \frac{3335}{3}\beta^2 - K_2 = 0.$$

Par conséquent,

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} 1 & 29 \\ 2\alpha + 58\beta & 58\alpha + \frac{6670}{3}\beta \end{bmatrix}.$$

Le programme ConsDick de l'annexe B a été utilisé pour obtenir les constantes α , β et γ_{15}^* .

Avec $\gamma_{15}^* = \gamma_{15} = 2,195\,240\,657 \times 10^{-2}$, l'algorithme de Newton-Raphson a produit les valeurs $\alpha = 2,259\,117\,281 \times 10^{-4}$, $\beta = -3,705\,215\,269\,3 \times 10^{-6}$, après 16 itérations. L'écart maximal entre $\sum_{i=1}^{72} \gamma_i^{*k}$ et la valeur théorique, pour $k = 2, \dots, 8$, était inférieur à 3.1×10^{-6} , et $\|y\|_\infty < 1.0 \times 10^{-17}$. Nous avons par la suite examiné le voisinage de γ_{15} et essayé plusieurs valeurs de γ_{15}^* . Les valeurs finalement retenues sont

$$\gamma_{15}^* = 1,973\,346\,8 \times 10^{-2},$$

$$\alpha = 2,414\,766\,208\,86 \times 10^{-4},$$

$$\beta = -4,185\,977\,814\,01 \times 10^{-6}.$$

Cet ensemble de valeurs fait en sorte que l'écart maximal entre $\sum_{i=1}^{72} \gamma_i^{*k}$ et la valeur théorique, pour $k = 2, \dots, 8$, est inférieur à 4.33×10^{-10} , et $\|y\|_\infty < 2.8 \times 10^{-18}$. C'est donc en utilisant ces dernières valeurs que nous avons déterminé l'ensemble des 72 γ_i^* nécessaires à la simulation des distributions asymptotiques. La fonction **QAsymp** définie dans le programme Simul1 de l'annexe B évalue $R(1, s)$ selon l'équation (5.55). Notons que les vecteurs **GammaI** et **GammaI2** contiennent les valeurs γ_i et γ_i^{*2} .[†]

[†]. L'annexe C propose et justifie une méthode alternative, plus simple à programmer.

6.3. Programmes de simulation

L'ensemble des simulations a été réalisé en utilisant quatre variantes d'un même programme. Ces variantes sont expliquées à l'annexe B. Pour chacune des simulations faites, les programmes produisaient : la moyenne, la variance, le minimum, le maximum et une liste des principaux percentiles.

Pour obtenir les percentiles, on peut conserver en mémoire ou sur fichier toutes les valeurs simulées et les trier lorsque la simulation prend fin. Il va de soi que si le nombre de valeurs est très grand, cette solution est peu efficace. Dans notre cas, lorsque $s = 1$, nous devons générer 50 000 valeurs. Nous préférons utiliser un histogramme de 800 classes et compter, au fur et à mesure que la simulation se déroule, le nombre d'observations appartenant à chacune des classes de cet histogramme. La méthode est décrite dans LATOUR (1986). Une simulation préliminaire de quelques centaines d'observations permet d'avoir une idée du minimum et de l'étendue de la distribution. Cette information est requise par le programme de simulation. Les percentiles sont calculés par interpolation linéaire inverse des fréquences cumulées associées à cet histogramme de 800 classes.

Notons finalement que la moyenne, le minimum, l'étendue et l'information relative à l'histogramme (fréquence de chacune des classes, le nombre de classes et le nombre d'observations) sont conservées sur fichier pour ne pas avoir à reprendre les simulations si un traitement ultérieur s'avère nécessaire.

6.4. Résultats des simulations

Nous allons dans un premier temps juger de l'impact de l'utilisation de données centrées ou non centrées dans le calcul des autocorrélations au niveau des distributions de $R(1, s)$ et $R'(1, s)$. Les figures 6.1 à 6.6³ illustrent la forme des densités dans chacun des cas. Au tableau 6.4, nous présentons les deux premiers moments évalués par simulation et au tableau 6.5 figurent les percentiles obtenus par ces mêmes simulations.

3. Rééditée à l'aide de la fonction `density` de R en 2025.

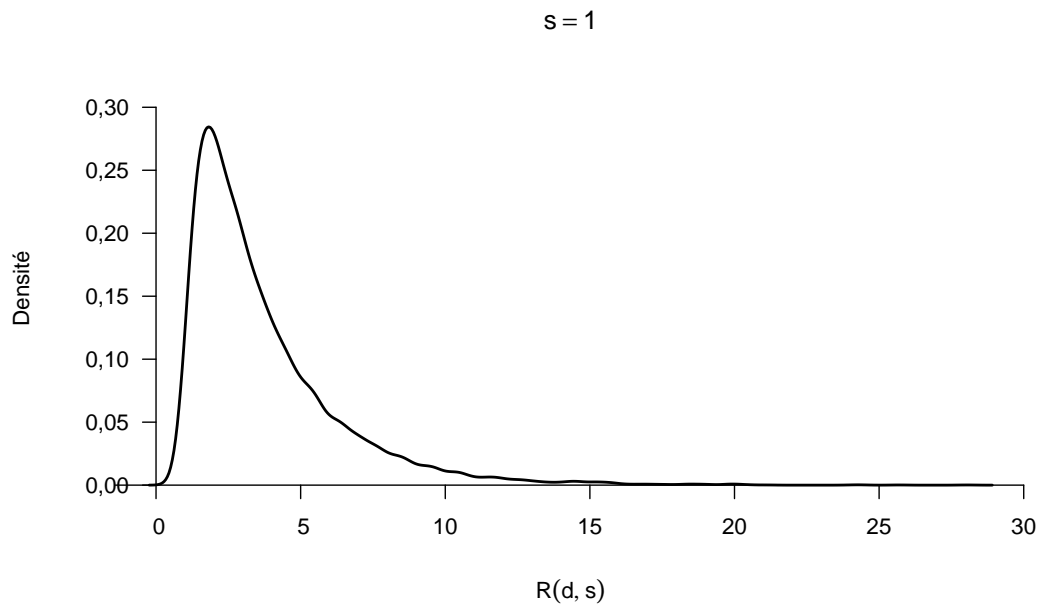


Figure 6.1 – Distribution de $R(d, s)$ lorsque $s = 1$.

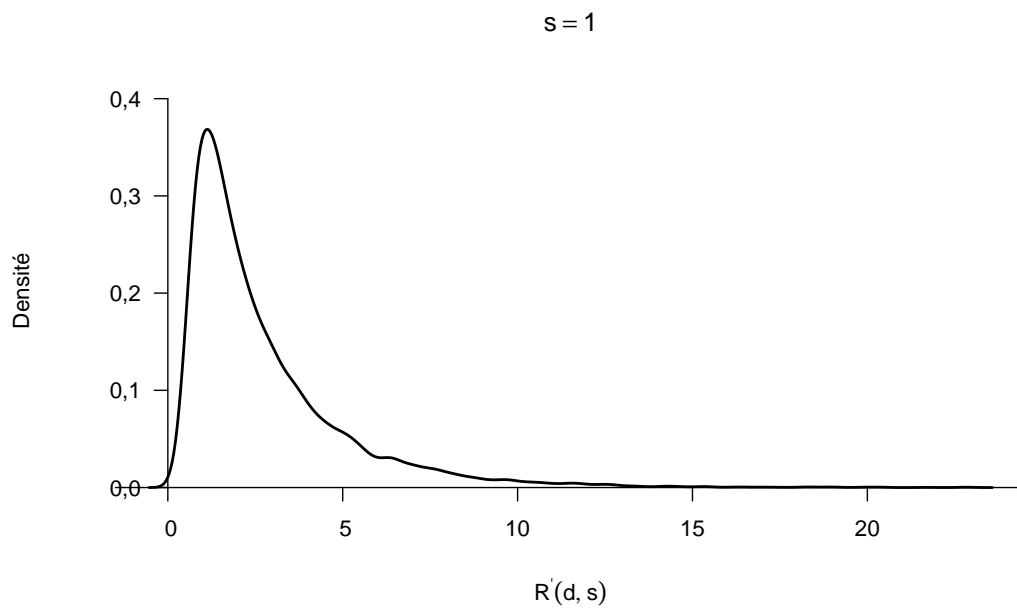


Figure 6.2 – Distribution de $R'(d, s)$ lorsque $s = 1$.

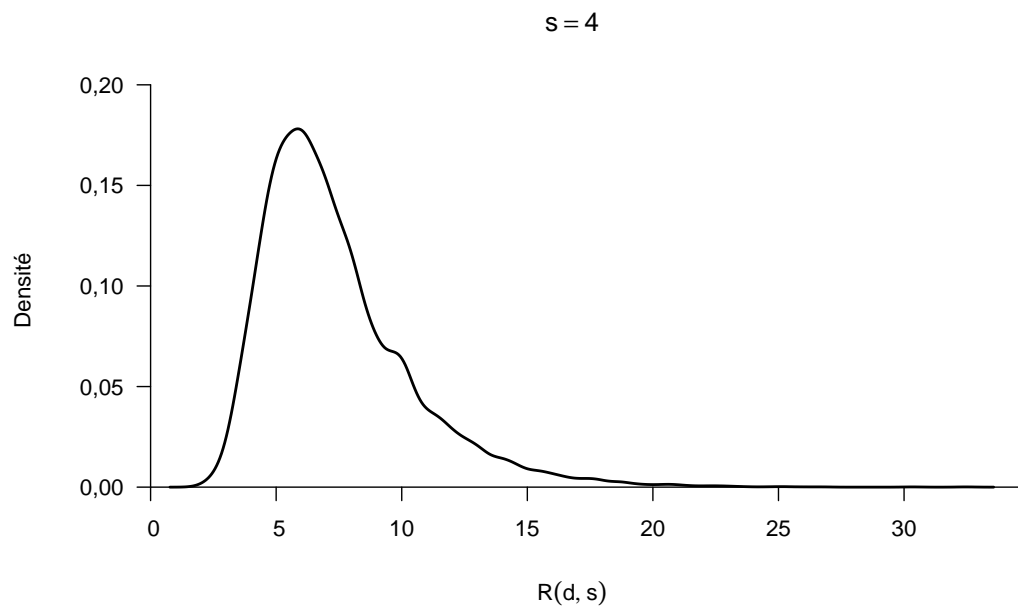


Figure 6.3 – Distribution de $R(d, s)$ lorsque $s = 4$.

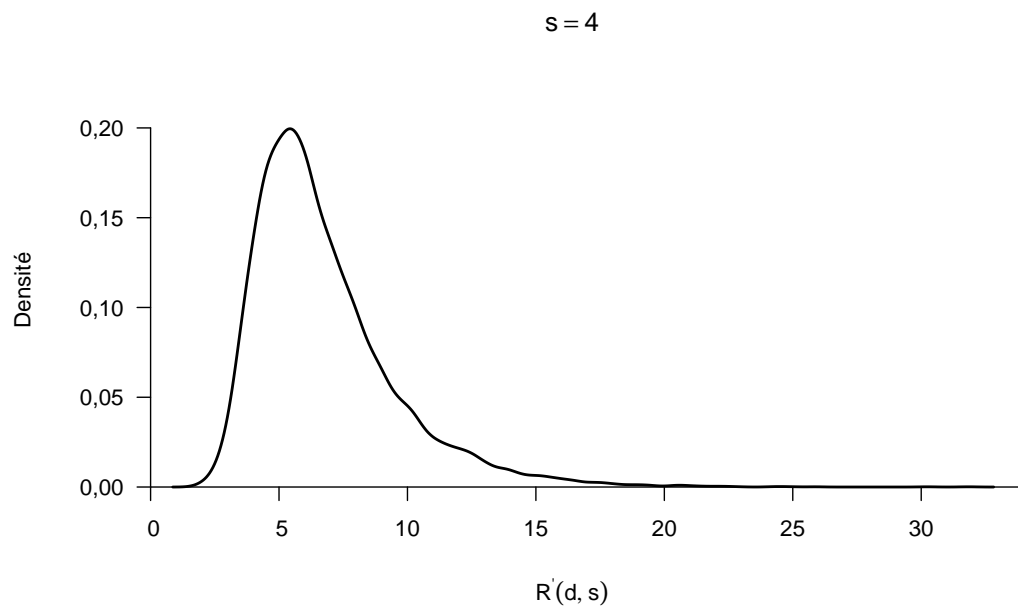


Figure 6.4 – Distribution de $R'(d, s)$ lorsque $s = 4$.

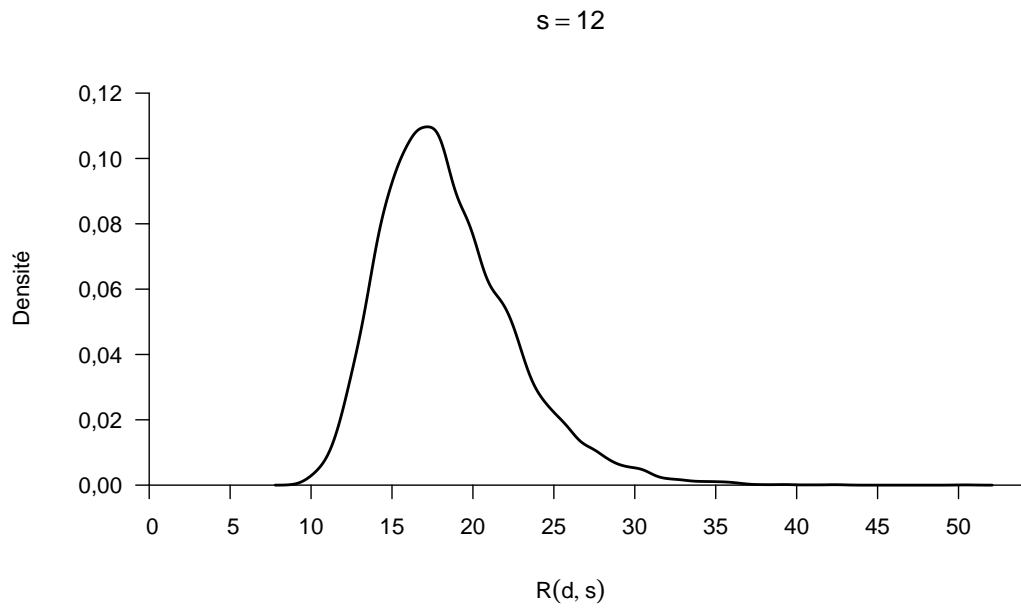


Figure 6.5 – Distribution de $R(d, s)$ lorsque $s = 12$.

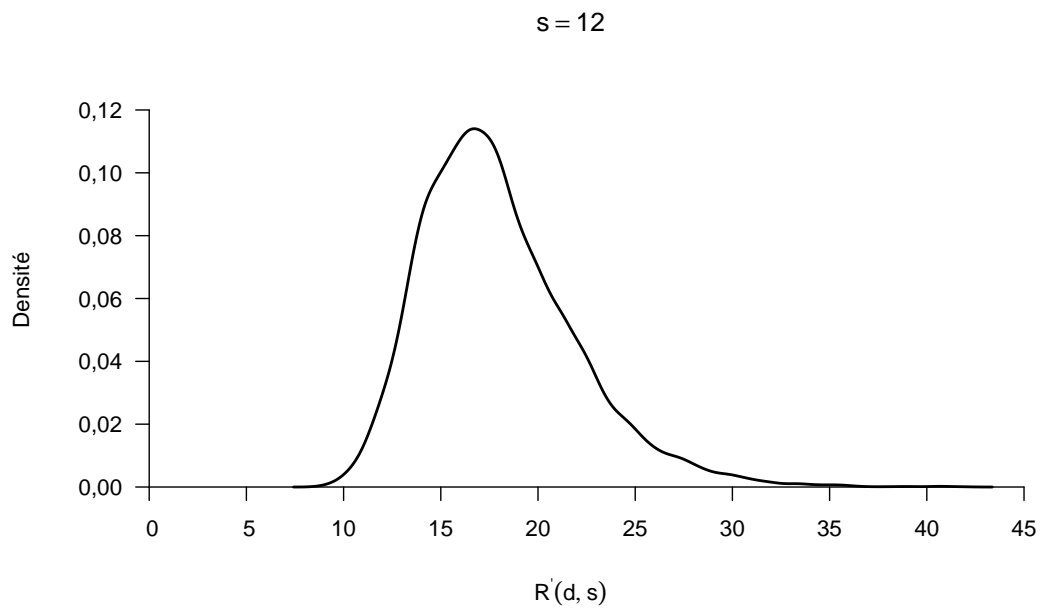


Figure 6.6 – Distribution de $R'(d, s)$ lorsque $s = 12$.

Tableau 6.4 – Les deux premiers moments des distributions asymptotiques

s	Avec données centrées		Avec données non centrées	
	Moyenne	Variance	Moyenne	Variance
1	7.39	19.67	3.82	10.49
4	11.24	23.16	9.50	15.38
12	26.61	36.52	25.39	30.99

Tableau 6.5 – Percentiles des distributions asymptotiques

Avec données centrées									
s	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	1.91	2.21	2.56	3.06	6.28	13.18	16.09	18.85	22.90
4	4.55	5.14	5.67	6.36	10.15	17.57	20.66	23.34	26.94
12	16.40	17.60	18.52	19.85	25.75	34.42	37.81	41.02	45.10
Avec données non centrées									
s	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	0.65	0.78	0.92	1.14	2.78	7.75	10.36	12.63	16.10
4	4.11	4.56	5.00	5.63	8.58	14.63	17.22	19.58	23.03
12	15.82	16.96	17.95	19.10	24.63	32.60	35.54	38.36	42.34

Les expressions analytiques des distributions asymptotiques avaient mis en évidence les différences structurelles induites par le fait de centrer ou non les données dans le calcul des autocorrélations. L'importance de ces différences se reflète dans les chiffres des tableaux 6.4 et 6.5. Nous constatons que pour $s = 1$, la moyenne et la variance de la distribution sont divisées par deux lorsque nous passons d'une situation à l'autre. Cette différence s'atténue lorsque la longueur de la période augmente. Ainsi, pour $s = 4$, la moyenne et la variance décroissent de façon sensible, mais pour $s = 12$, les deux premiers moments sont relativement proches les uns des autres. Il en est de même au niveau des percentiles puisque plus s est grand, moins les différences sont marquées.

Nous allons maintenant dégager les points saillants relatifs à la convergence des distributions exactes vers les distributions asymptotiques en examinant les tableaux 6.6 à 6.10. Pour comparer les percentiles, nous utiliserons l'écart relatif par rapport au percentile de la distribution asymptotique. D'une manière plus précise, dénotons par $x_N(d)$, un percentile de la distribution exacte de $N(1 - r(d))$ et par $x_\infty(d)$, celui de la distribution asymptotique qui lui correspond. L'écart relatif E est défini par

$$E = \left| \frac{x_N(d) - x_\infty(d)}{x_\infty(d)} \right| \quad (6.3)$$

Dans un premier temps, nous nous concentrerons sur les données relatives au cas où $d = 1$. Le tableau 6.6 résume l'information relative aux deux premiers moments. Ainsi, pour $d = 1$, nous constatons que la moyenne des distributions exactes semble converger vers la moyenne des distributions asymptotiques correspondantes avec une vitesse qui décroît lorsque s augmente, qu'il s'agisse des distributions avec données centrées ou de celles avec données non centrées.

Tableau 6.6 – Les deux premiers moments des distributions exactes et asymptotiques de $N(1 - r(ds))$ et de $N(1 - r'(ds))$

		$N(1 - r(ds))$							
		Moyenne				Variance			
		Taille				Taille			
s	d	60	108	204	∞	60	108	204	∞
1	1	7.01	7.16	7.30	7.39	15.89	17.72	18.67	19.67
	2	13.36	13.97	14.39	14.78	49.04	61.23	69.37	78.68
	3	19.14	20.46	21.26	22.17	85.85	121.30	144.88	177.04
	4	24.39	26.58	27.94	29.56	119.55	187.81	239.52	314.74
	5	29.20	32.40	34.42	36.95	147.99	256.52	347.81	491.78
4	1	10.23	10.69	10.84	11.24	16.81	19.10	20.00	23.16
	2	19.10	20.56	21.24	22.48	45.72	61.59	71.58	92.64
12	1	21.58	23.61	24.94	26.61	14.87	22.07	26.94	36.52
	2	37.82	43.85	47.99	53.22	21.36	52.25	81.69	146.08
		$N(1 - r'(ds))$							
		Moyenne				Variance			
		Taille				Taille			
s	d	60	108	204	∞	60	108	204	∞
1	1	3.62	3.70	3.72	3.82	8.68	9.50	9.86	10.49
	2	6.99	7.26	7.36	7.64	29.21	34.42	37.08	41.96
	3	10.14	10.67	10.93	11.46	55.78	70.26	79.35	94.41
	4	13.07	13.96	14.43	15.28	84.42	112.99	134.52	167.84
	5	15.83	17.11	17.85	19.10	113.39	160.24	199.50	262.25
4	1	8.72	9.09	9.34	9.50	10.98	13.12	14.33	15.38
	2	16.50	17.60	18.37	19.00	30.44	42.73	51.50	61.52
12	1	20.85	22.63	23.87	25.39	13.11	18.40	23.80	30.99
	2	36.91	42.30	46.05	50.78	18.67	43.04	71.77	123.96

Tableau 6.7 – Percentiles des distributions exactes et asymptotiques de $N(1-r(ds))$ pour $s = 1$

		$N(1-r(ds))$								
d	N	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	60	1.93 (0.01) [†]	2.25 (0.02)	2.57 (0.00)	3.04 (0.01)	6.04 (0.04)	12.28 (0.07)	15.01 (0.07)	17.36 (0.08)	20.37 (0.11)
	108	1.90 (0.01)	2.24 (0.01)	2.58 (0.01)	3.07 (0.00)	6.08 (0.03)	12.73 (0.03)	15.42 (0.04)	18.01 (0.04)	21.60 (0.06)
	204	1.94 (0.02)	2.24 (0.01)	2.58 (0.01)	3.05 (0.00)	6.18 (0.02)	12.99 (0.01)	15.89 (0.01)	18.46 (0.02)	22.05 (0.04)
	∞	1.91	2.21	2.56	3.06	6.28	13.18	16.09	18.85	22.90
	2	60	4.08 (0.07)	4.64 (0.05)	5.27 (0.03)	6.12 (0.00)	11.71 (0.07)	22.88 (0.13)	27.01 (0.16)	31.00 (0.18)
108	3.88 (0.02)	4.54 (0.03)	5.22 (0.02)	6.17 (0.01)	11.99 (0.05)	24.64 (0.07)	29.39 (0.09)	33.73 (0.11)	39.90 (0.13)	
204	3.97 (0.04)	4.53 (0.02)	5.22 (0.02)	6.11 (0.00)	12.25 (0.02)	25.53 (0.03)	30.92 (0.04)	36.13 (0.04)	42.83 (0.06)	
∞	3.82	4.42	5.12	6.12	12.55	25.36	32.18	37.70	45.80	
3	60	6.29 (0.10)	7.13 (0.08)	8.05 (0.05)	9.23 (0.01)	17.12 (0.09)	31.91 (0.19)	37.18 (0.23)	42.38 (0.25)	48.15 (0.30)
	108	6.01 (0.05)	6.94 (0.05)	7.92 (0.03)	9.26 (0.01)	17.71 (0.06)	35.72 (0.10)	41.91 (0.13)	47.94 (0.15)	56.28 (0.18)
	204	5.98 (0.04)	6.83 (0.03)	7.88 (0.03)	9.24 (0.01)	18.21 (0.03)	37.32 (0.06)	45.05 (0.07)	52.55 (0.07)	62.50 (0.09)
	∞	5.73	6.63	7.68	9.18	18.84	39.54	48.27	56.55	68.70
	4	60	8.71 (0.14)	9.79 (0.11)	10.85 (0.06)	12.36 (0.01)	22.18 (0.12)	39.70 (0.25)	45.86 (0.29)	50.66 (0.33)
108	8.23 (0.08)	9.34 (0.06)	10.65 (0.04)	12.37 (0.01)	23.32 (0.07)	45.63 (0.13)	53.51 (0.17)	61.08 (0.19)	69.88 (0.24)	
204	7.99 (0.05)	9.22 (0.04)	10.58 (0.03)	12.32 (0.01)	24.08 (0.04)	48.52 (0.08)	58.31 (0.09)	68.25 (0.09)	80.31 (0.12)	
∞	7.64	8.84	10.24	12.24	25.12	52.72	64.36	75.40	91.60	
5	60	11.34 (0.19)	12.52 (0.13)	13.77 (0.08)	15.52 (0.01)	25.77 (0.15)	46.54 (0.29)	52.72 (0.34)	57.32 (0.39)	63.40 (0.45)
	108	10.51 (0.10)	11.84 (0.07)	13.38 (0.05)	15.47 (0.01)	28.69 (0.09)	55.01 (0.17)	64.15 (0.20)	71.94 (0.24)	81.92 (0.28)
	204	10.11 (0.06)	12.62 (0.14)	13.21 (0.03)	15.47 (0.01)	29.86 (0.05)	59.48 (0.10)	71.31 (0.11)	83.13 (0.12)	97.42 (0.15)
	∞	9.55	11.05	12.80	15.30	31.40	65.90	80.45	94.25	114.50

† Les chiffres entre parenthèses indiquent l'écart relatif entre le percentile de la distribution exacte et celui de la distribution asymptotique.

Se référant au tableau 6.6, au niveau des statistiques calculées à partir de données centrées, lorsque $s = 1$, on note que pour une longueur de série aussi petite que 60, l'écart entre la moyenne de la distribution exacte et la moyenne de la distribution asymptotique est mineur (7.01 versus 7.39). Il est un peu plus important

Tableau 6.8 – Percentiles des distributions exactes et asymptotiques de $N(1 - r'(ds))$ pour $s = 1$

		$N(1 - r'(ds))$								
d	N	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	60	0.64	0.76	0.90	1.11	2.68	7.36	9.49	11.95	14.55
		(0.02) [†]	(0.03)	(0.02)	(0.03)	(0.04)	(0.05)	(0.08)	(0.05)	(0.10)
	108	0.64	0.79	0.94	1.13	2.69	7.60	9.96	12.45	14.96
		(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.07)
	204	0.64	0.77	0.92	1.13	2.70	7.55	9.93	12.43	15.79
	(0.02)	(0.01)	(0.00)	(0.01)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	
	∞	0.65	0.78	0.92	1.14	2.78	7.75	10.36	12.63	16.10
2	60	1.33	1.56	1.84	2.27	5.29	14.15	17.88	21.59	26.54
		(0.02)	(0.00)	(0.00)	(0.00)	(0.05)	(0.09)	(0.14)	(0.15)	(0.18)
	108	1.34	1.60	1.89	2.28	5.35	14.86	19.17	23.41	20.84
		(0.03)	(0.03)	(0.03)	(0.00)	(0.04)	(0.04)	(0.07)	(0.07)	(0.10)
	204	1.30	1.56	1.85	2.27	5.42	14.84	19.55	24.36	30.42
	(0.00)	(0.00)	(0.01)	(0.00)	(0.03)	(0.04)	(0.06)	(0.04)	(0.06)	
	∞	1.30	1.56	1.84	2.28	5.56	15.50	20.72	25.26	32.20
3	60	2.07	2.43	2.80	3.41	7.83	20.31	25.24	30.14	37.01
		(0.06)	(0.04)	(0.01)	(0.00)	(0.06)	(0.13)	(0.19)	(0.20)	(0.23)
	108	2.04	2.44	2.86	3.43	7.97	21.63	27.61	33.39	41.52
		(0.05)	(0.04)	(0.04)	(0.00)	(0.04)	(0.07)	(0.11)	(0.12)	(0.14)
	204	1.96	2.34	2.79	3.41	8.08	22.12	28.87	35.63	44.53
	(0.01)	(0.00)	(0.01)	(0.00)	(0.03)	(0.05)	(0.07)	(0.06)	(0.08)	
	∞	1.95	2.34	2.76	3.42	8.34	23.25	31.08	37.89	48.30
4	60	2.89	3.32	3.78	4.58	10.20	25.87	31.71	37.47	45.25
		(0.11)	(0.06)	(0.03)	(0.00)	(0.08)	(0.17)	(0.23)	(0.26)	(0.30)
	108	2.77	3.27	3.81	4.57	10.55	27.97	35.57	42.73	52.30
		(0.07)	(0.05)	(0.04)	(0.00)	(0.05)	(0.10)	(0.14)	(0.15)	(0.19)
	204	2.62	3.16	3.72	4.55	10.71	29.27	37.48	46.63	57.70
	(0.01)	(0.01)	(0.01)	(0.00)	(0.04)	(0.06)	(0.10)	(0.08)	(0.10)	
	∞	2.60	3.12	3.68	4.56	11.12	31.00	41.44	50.52	64.40
5	60	3.72	4.25	4.81	5.74	12.58	30.92	37.53	44.35	51.50
		(0.14)	(0.09)	(0.05)	(0.01)	(0.09)	(0.20)	(0.28)	(0.30)	(0.36)
	108	3.53	4.13	4.82	5.73	13.07	34.10	42.88	50.98	62.29
		(0.09)	(0.06)	(0.05)	(0.01)	(0.06)	(0.12)	(0.17)	(0.19)	(0.23)
	204	3.31	3.96	4.64	5.69	13.32	36.16	45.78	57.28	70.64
	(0.02)	(0.02)	(0.01)	(0.00)	(0.04)	(0.07)	(0.12)	(0.09)	(0.12)	
	∞	3.25	3.90	4.60	5.70	13.90	38.75	51.80	63.15	80.50

† Les chiffres entre parenthèses indiquent l'écart relatif entre le percentile de la distribution exacte et celui de la distribution asymptotique.

au niveau des variances (15.89 pour la distribution exacte et 19.67 pour la distribution asymptotique). Pour $s = 4$, les moyennes se ressemblent (10.23 versus 11.24) mais la différence est plus marquée au niveau des variances (16.81 versus 23.16). Finalement, lorsqu'on regarde ces quantités pour $s = 12$, l'écart est plus prononcé au niveau des

Tableau 6.9 – Percentiles des distributions exactes et asymptotiques de $N(1-r(ds))$ et $N(1-r'(ds))$ pour $s = 4$

		$N(1-r(ds))$								
d	N	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	60	4.37	4.89	5.35	5.98	9.35	15.63	18.16	20.50	23.45
		(0.04) [†]	(0.05)	(0.06)	(0.06)	(0.08)	(0.11)	(0.12)	(0.12)	(0.13)
	108	4.46	4.97	5.51	6.18	9.73	16.48	19.11	21.59	24.70
		(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.06)	(0.08)	(0.07)	(0.08)
	204	4.52	5.00	5.48	6.15	9.91	16.77	19.33	21.97	25.46
	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)	(0.05)	(0.06)	(0.06)	(0.05)	
	∞	4.55	5.14	5.67	6.36	10.15	17.57	20.66	23.34	26.94
2	60	9.17	10.01	10.81	11.86	17.69	20.24	32.04	35.68	40.40
		(0.01)	(0.03)	(0.05)	(0.07)	(0.13)	(0.20)	(0.22)	(0.24)	(0.25)
	108	9.11	10.04	11.04	12.34	18.85	31.02	35.71	40.12	45.99
		(0.00)	(0.02)	(0.03)	(0.03)	(0.07)	(0.12)	(0.14)	(0.14)	(0.15)
	204	9.18	10.10	11.02	12.34	19.44	32.56	37.52	41.06	40.35
	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)	(0.07)	(0.09)	(0.10)	(0.10)	
	∞	9.10	10.28	11.34	12.72	20.30	35.14	41.32	45.68	53.86
		$N(1-r'(ds))$								
d	N	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	60	4.05	4.49	4.90	5.38	7.93	13.13	15.10	17.17	20.07
		(0.01)	(0.02)	(0.02)	(0.04)	(0.08)	(0.10)	(0.12)	(0.12)	(0.13)
	108	4.13	4.52	4.97	5.50	8.27	13.76	16.05	18.23	21.53
		(0.00)	(0.01)	(0.01)	(0.02)	(0.04)	(0.06)	(0.07)	(0.07)	(0.07)
	204	4.13	4.57	4.96	5.56	8.49	14.14	16.67	19.20	22.59
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	
	∞	4.11	4.56	5.00	5.63	8.56	14.63	17.22	19.58	23.03
2	60	8.69	9.39	10.00	10.81	15.18	23.92	27.37	30.53	34.55
		(0.06)	(0.03)	(0.00)	(0.04)	(0.12)	(0.18)	(0.21)	(0.22)	(0.25)
	108	8.53	9.30	10.02	11.00	16.09	26.25	30.32	34.14	39.70
		(0.04)	(0.02)	(0.00)	(0.02)	(0.06)	(0.10)	(0.12)	(0.13)	(0.14)
	204	8.40	9.26	10.08	11.17	16.78	27.64	32.41	37.19	43.30
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.06)	(0.06)	(0.05)	(0.06)	
	∞	8.22	9.12	10.00	11.26	17.16	29.26	34.44	39.16	46.06

† Les chiffres entre parenthèses indiquent l'écart relatif entre le percentile de la distribution exacte et celui de la distribution asymptotique.

moyennes et des variances. De fait, l'écart absolu entre les moyennes est de 5.03 (ce qui correspond à un écart relatif de 19%), tandis qu'au niveau des variances, il est de 21.65 (ce qui donne un écart relatif de 59%). Évidemment, pour une longueur de série de 108 ou 204, les différences sont moins importantes, mais demeurent tout de même assez marquées, tout au moins au niveau des variances. Ces dernières remarques

Tableau 6.10 – Percentiles des distributions exactes et asymptotiques de $N(1-r(ds))$ et $N(1-r'(ds))$ pour $s = 12$

		$N(1-r(ds))$								
d	N	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	60	14.88	15.71	16.33	17.16	21.05	26.63	28.74	30.73	33.28
		(0.09) [†]	(0.11)	(0.12)	(0.14)	(0.18)	(0.23)	(0.24)	(0.25)	(0.26)
	108	15.51	16.43	17.26	18.29	22.92	29.76	32.25	34.78	38.11
		(0.05)	(0.07)	(0.07)	(0.08)	(0.11)	(0.14)	(0.15)	(0.15)	(0.15)
	204	15.96	17.00	17.88	19.04	24.18	31.83	34.33	36.87	40.43
		(0.03)	(0.03)	(0.03)	(0.04)	(0.06)	(0.08)	(0.09)	(0.10)	(0.10)
	∞	16.40	17.60	18.52	19.85	25.75	34.92	37.81	41.02	45.10
		$N(1-r'(ds))$								
d	N	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990
1	60	14.47	15.25	15.92	16.75	20.33	25.63	27.51	29.31	31.82
		(0.09)	(0.10)	(0.11)	(0.12)	(0.17)	(0.21)	(0.23)	(0.24)	(0.25)
	108	15.19	16.05	16.85	17.73	22.02	28.26	30.73	32.59	35.08
		(0.04)	(0.05)	(0.06)	(0.07)	(0.11)	(0.13)	(0.14)	(0.15)	(0.17)
	204	15.47	16.40	17.28	18.35	23.13	30.25	32.96	35.56	39.15
		(0.02)	(0.03)	(0.04)	(0.04)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)
	∞	15.82	16.96	17.95	19.10	24.63	32.60	35.54	38.36	42.34
2	60	29.45	30.16	30.92	31.95	36.33	42.70	44.93	46.92	49.45
		(0.07)	(0.11)	(0.14)	(0.16)	(0.26)	(0.35)	(0.37)	(0.39)	(0.42)
	108	30.89	32.16	33.33	34.76	41.42	51.13	54.41	57.81	61.35
		(0.02)	(0.05)	(0.07)	(0.09)	(0.16)	(0.22)	(0.23)	(0.25)	(0.28)
	204	31.43	32.95	34.48	36.48	44.82	57.17	61.62	66.26	72.28
		(0.01)	(0.03)	(0.04)	(0.05)	(0.09)	(0.12)	(0.13)	(0.14)	(0.15)
	∞	31.64	33.92	35.90	38.20	49.26	65.20	71.08	76.72	84.68

[†] Les chiffres entre parenthèses indiquent l'écart relatif entre le percentile de la distribution exacte et celui de la distribution asymptotique.

s'appliquent aussi aux moments des distributions des statistiques basées sur des données non centrées.

Il est aussi intéressant de comparer les distributions au niveau des percentiles. Les tableaux 6.7 à 6.10 ont été produits à cette fin. Cet ensemble de tableaux permet non seulement de se faire une idée de la vitesse de convergence des percentiles des

distributions exactes vers ceux des distributions asymptotiques, mais il permet aussi de constater encore mieux les différences entre les distributions des statistiques calculées avec des données centrées et celles calculées à partir des données non centrées. Comme précédemment, nous ne nous arrêterons pour l'instant que sur l'information reliée au cas où $d = 1$.

Attardons-nous aux distributions des statistiques basées sur des données centrées. Comme on peut le voir, les percentiles des distributions exactes sont plus près de ceux de la distribution asymptotique pour $s = 1$ que pour $s = 4$ ou $s = 12$. De façon plus précise, si $N = 60$ et $s = 1$, l'écart relatif varie de 1% à 11%, tandis que pour $s = 4$, il varie de 4% à 13%, et pour $s = 12$, de 9% à 26%.

Notons au passage que l'écart relatif est plus important pour les grands percentiles que pour les petits. Pour $N = 108$, les écarts relatifs sont tout au plus de 6% pour $s = 1$, inférieurs à 8% pour $s = 4$, mais peuvent atteindre 15% pour $s = 12$. Évidemment, pour $N = 204$, les écarts sont moins importants. Ainsi, pour $s = 1$, aucun des écarts n'est supérieur à 4%. Mais pour $s = 4$ et $s = 12$, ces écarts peuvent aller jusqu'à 6% ou 10%. Un examen similaire des tableaux relatifs aux données non centrées nous mène à des conclusions semblables.

D'autre part, si nous maintenons N , d et s fixés, tout comme pour les deux premiers moments, nous remarquons des différences sensibles au niveau des percentiles lorsque nous utilisons des statistiques basées sur des données centrées comparativement à celles impliquant des données non centrées. Néanmoins, ces différences prennent moins d'importance lorsque s croît.

Pour comprendre ce qui se passe lorsque d varie, considérons

$$ns(1 - r'(ds)) = \frac{\frac{ds}{2} \sum_{r=1}^s [T_{n,r}^2 + 1] + O_p(n^{-1/2})}{\sum_{r=1}^s r_{n,r}}$$

Si nous faisons abstraction du terme $O_p(n^{-1/2})$, on constate que $ns(1 - r'(ds))$ a une distribution qui devrait être assez proche de celle de $dns(1 - r'(s))$. Évidemment, cela est d'autant plus vrai que n est grand. Une réflexion analogue en regard de l'équation (5.54) nous indique qu'il devrait en être de même pour $ns(1 - r(ds))$ relativement à $dns(1 - r(s))$.

Les écarts relatifs associés aux percentiles lorsque $d \neq 1$ devraient aussi refléter cette propriété. En utilisant l'équation (6.3), et le fait que $x_\infty(d) = dx_\infty(1)$, nous

avons

$$E = \left| \frac{x_N(d) - x_\infty(d)}{x_\infty(d)} \right| = \left| \frac{x_N(d) - dx_\infty(1)}{dx_\infty(1)} \right|$$

Mais si nous avons $x_N(d) \approx dx_N(1)$, cela signifierait que l'écart relatif serait à peu près le même, peu importe la valeur de d . Cependant, nous notons des différences appréciables au niveau des écarts relatifs lorsque d varie. Cette propriété asymptotique ne se manifeste pas nécessairement de façon évidente même pour $N = 204$. Nous présentons finalement aux figures 6.7 à 6.24, les densités des distributions exactes.

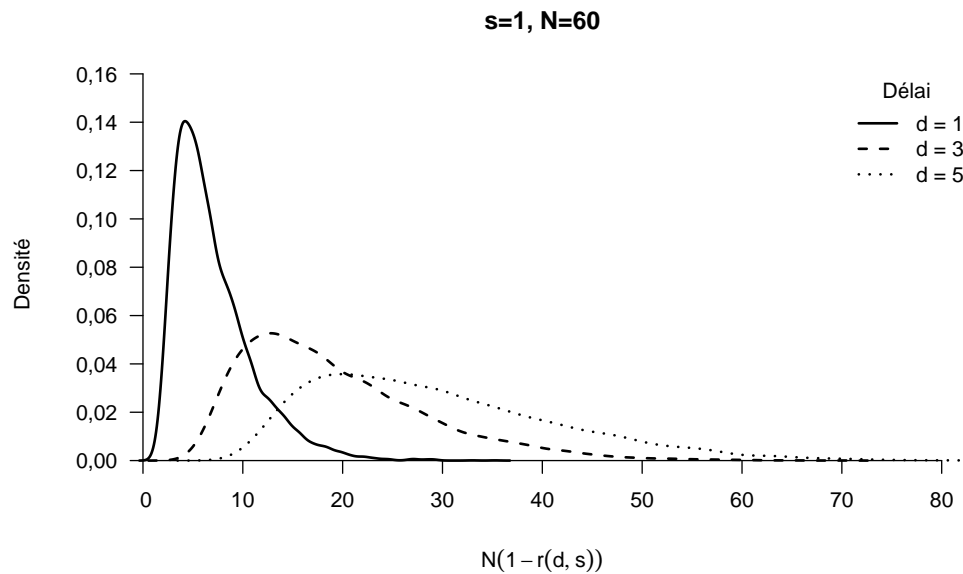


Figure 6.7 – Distribution de $N(1 - r(ds))$ lorsque $s = 1$ et $N = 60$.

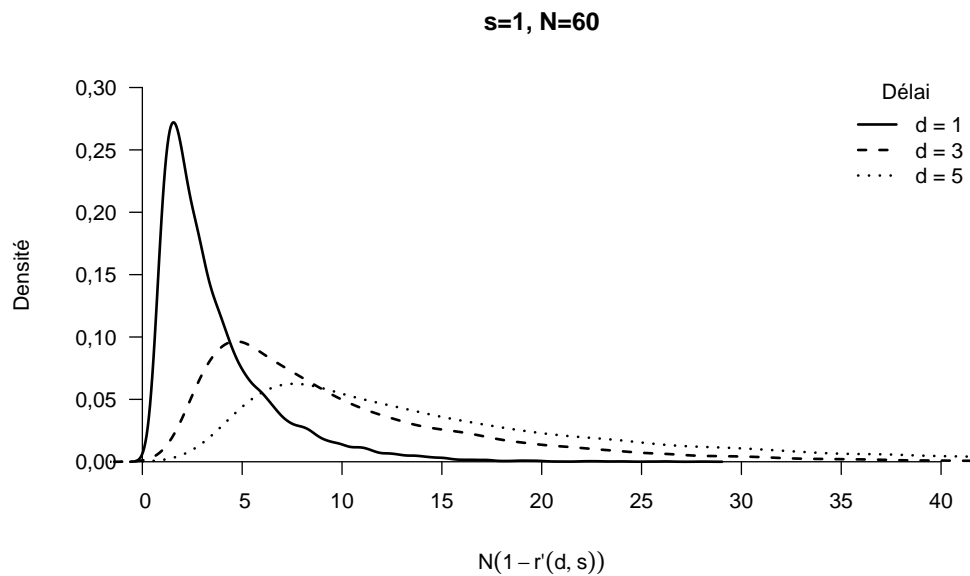


Figure 6.8 – Distribution de $N(1-r'(ds))$ lorsque $s = 1$ et $N = 60$.

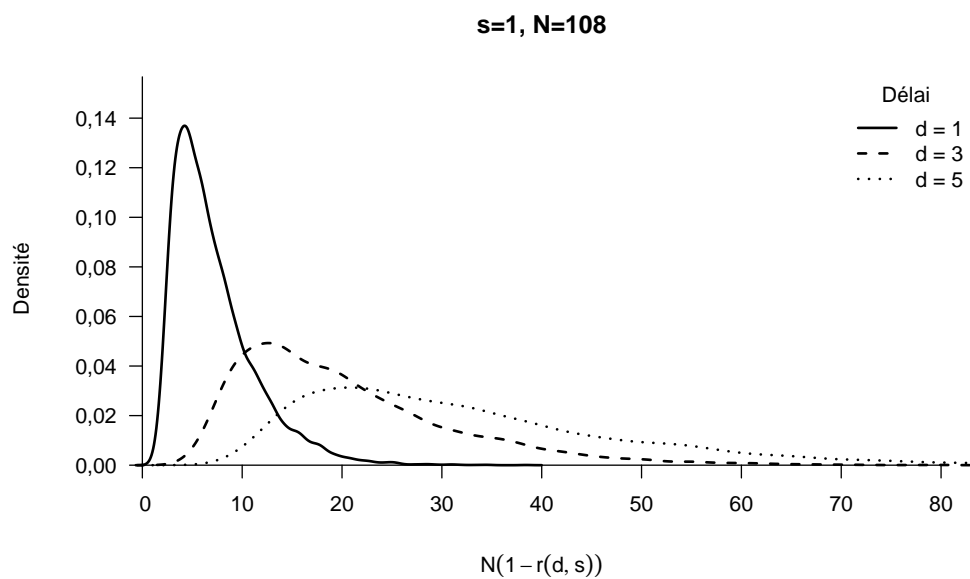


Figure 6.9 – Distribution de $N(1-r(ds))$ lorsque $s = 1$ et $N = 108$.

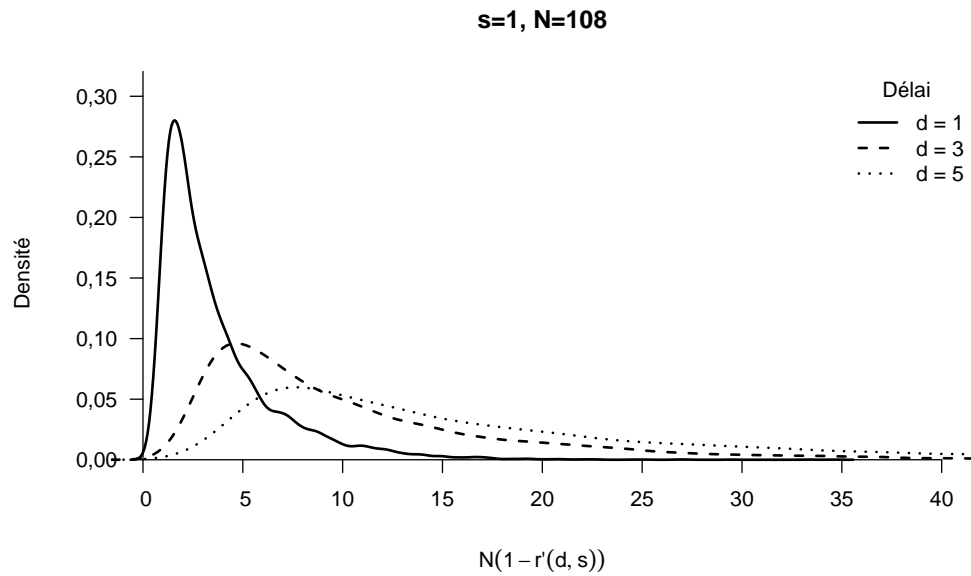


Figure 6.10 – Distribution de $N(1 - r'(ds))$ lorsque $s = 1$ et $N = 108$.

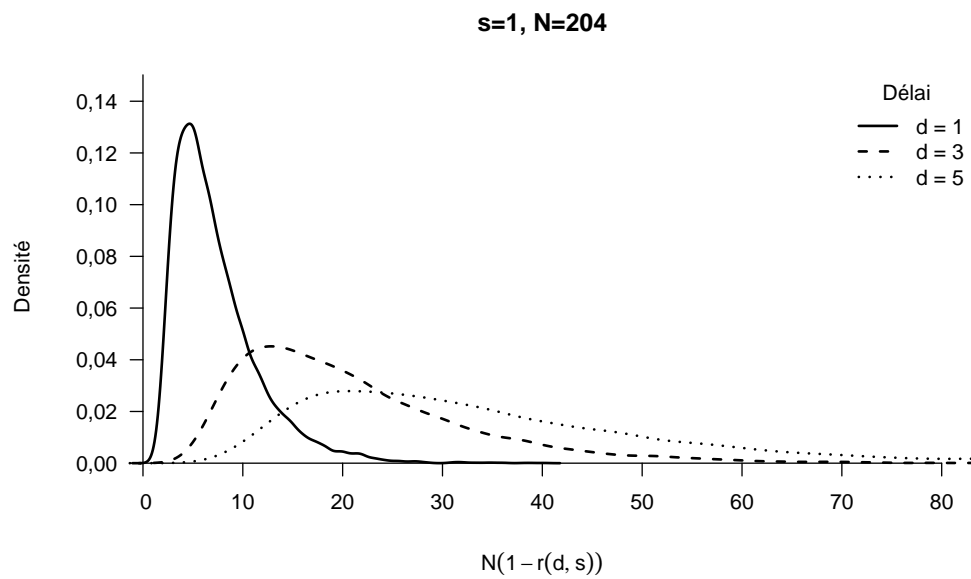


Figure 6.11 – Distribution de $N(1 - r(ds))$ lorsque $s = 1$ et $N = 204$.

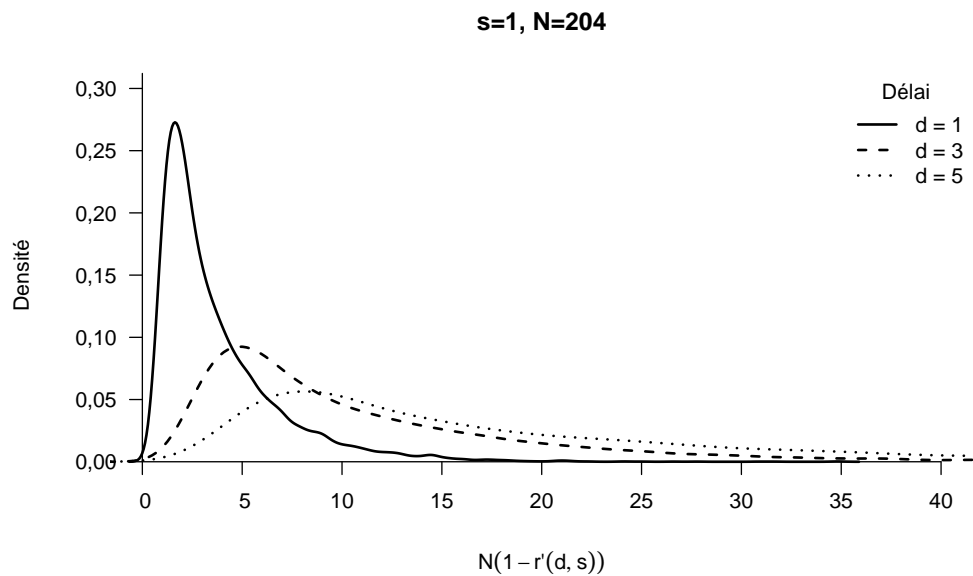


Figure 6.12 – Distribution de $N(1-r'(ds))$ lorsque $s = 1$ et $N = 204$.

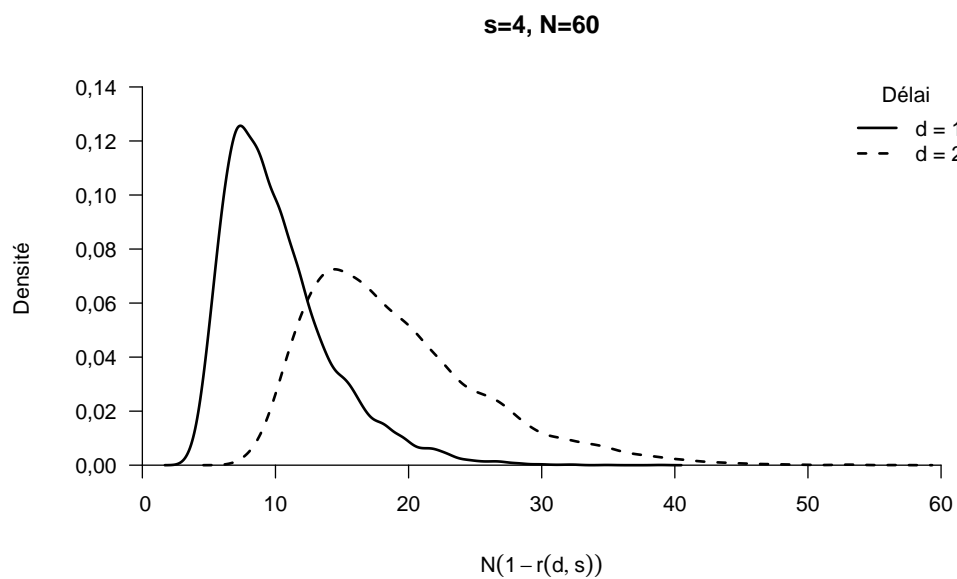


Figure 6.13 – Distribution de $N(1-r(ds))$ lorsque $s = 4$ et $N = 60$.

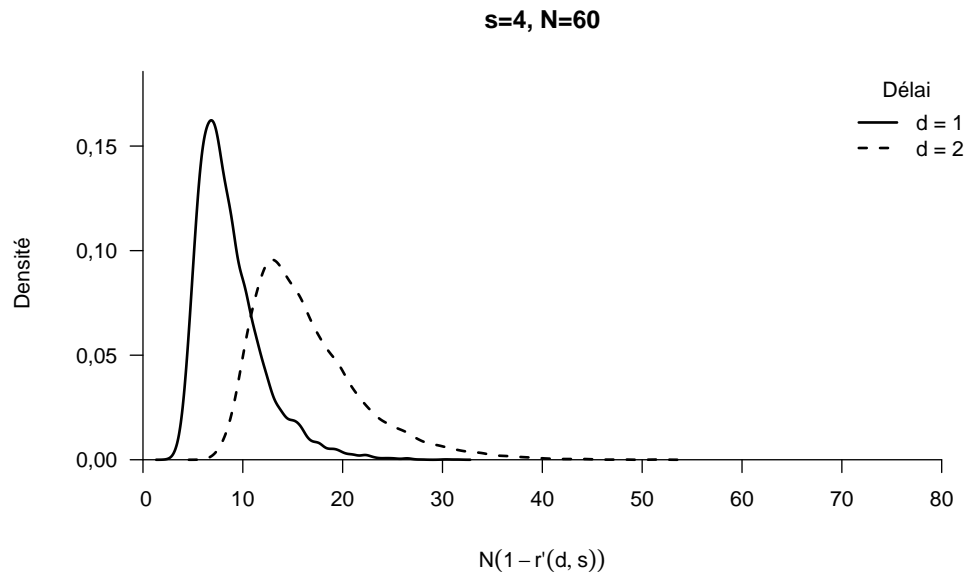


Figure 6.14 – Distribution de $N(1-r'(ds))$ lorsque $s = 4$ et $N = 60$.

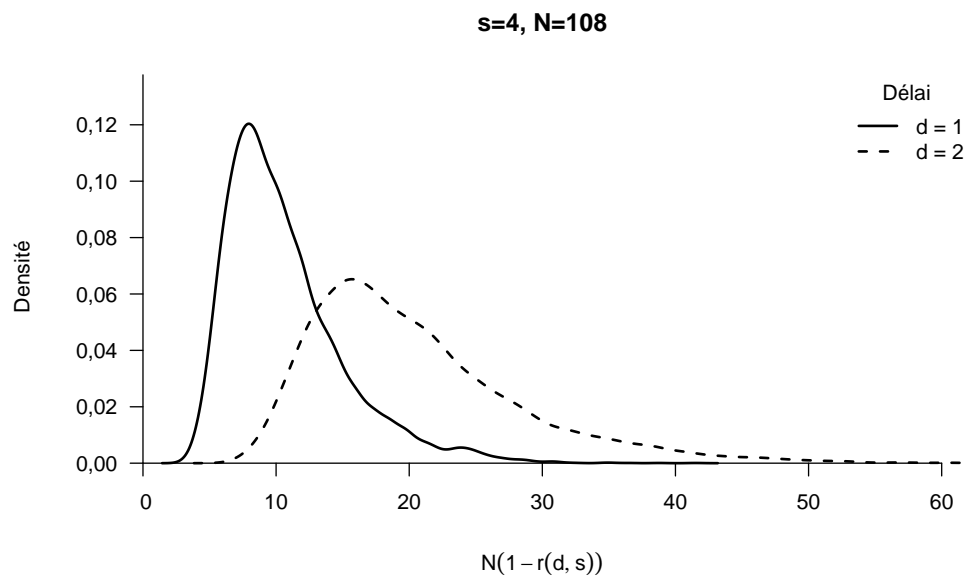


Figure 6.15 – Distribution de $N(1-r(ds))$ lorsque $s = 4$ et $N = 108$.

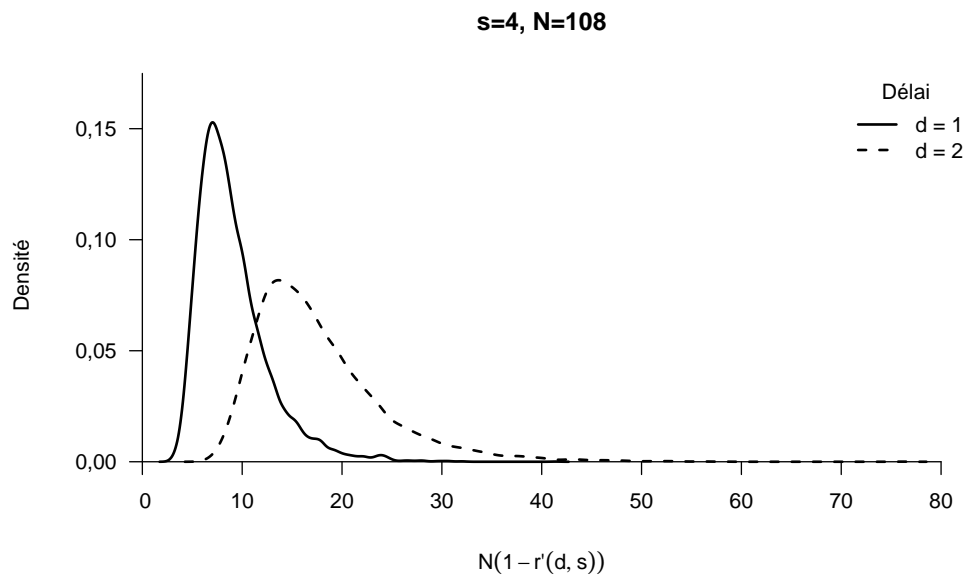


Figure 6.16 – Distribution de $N(1 - r'(ds))$ lorsque $s = 4$ et $N = 108$.

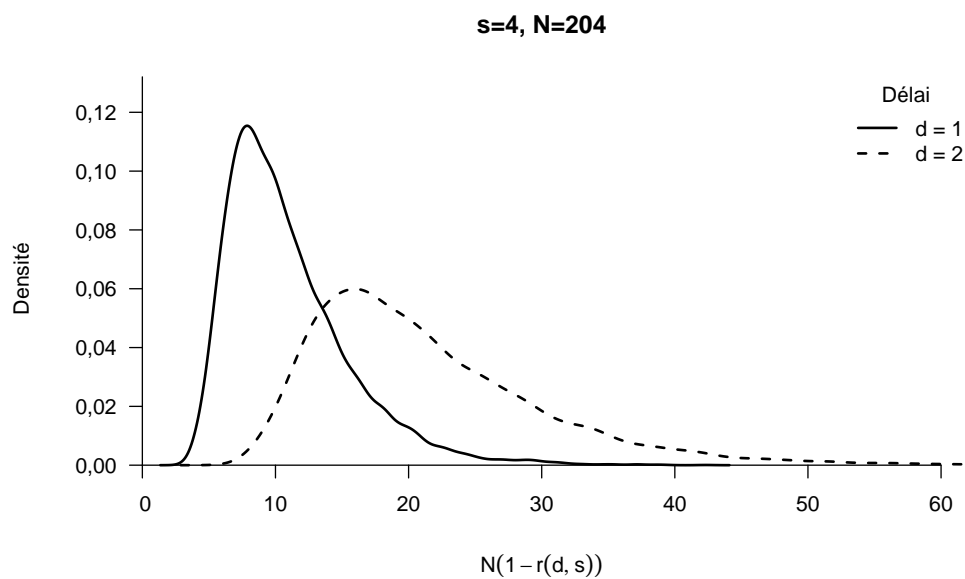


Figure 6.17 – Distribution de $N(1 - r(ds))$ lorsque $s = 4$ et $N = 204$.

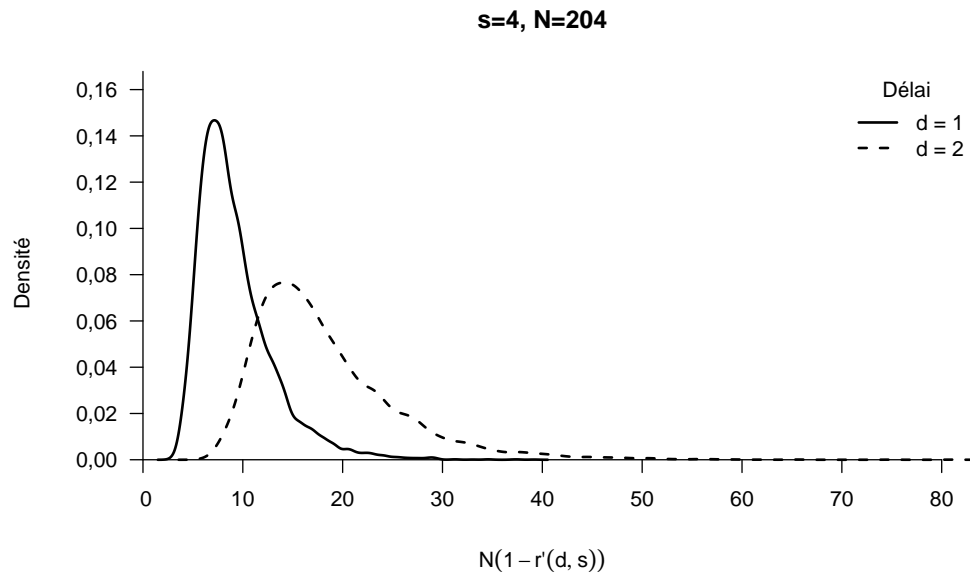


Figure 6.18 – Distribution de $N(1-r'(ds))$ lorsque $s = 4$ et $N = 204$.

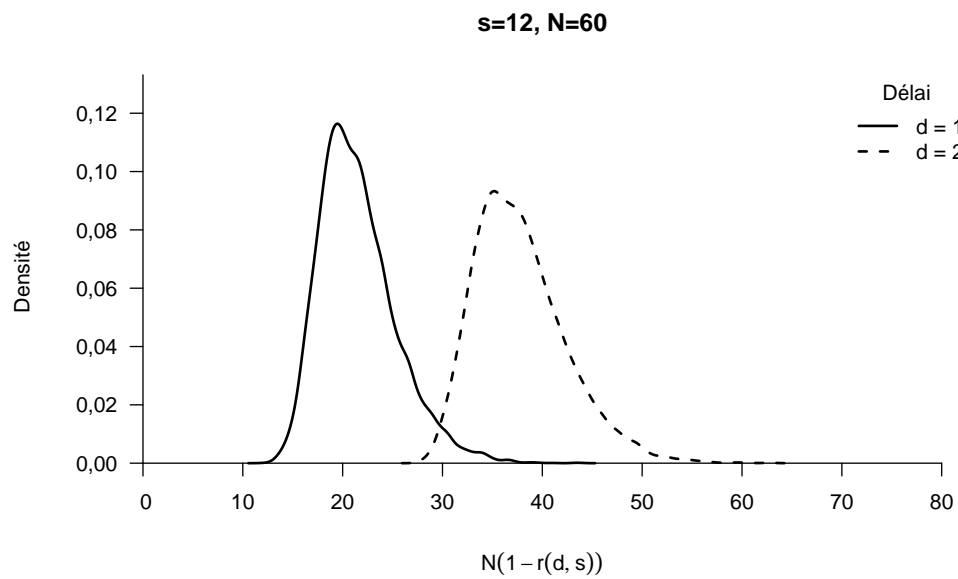


Figure 6.19 – Distribution de $N(1-r(ds))$ lorsque $s = 12$ et $N = 60$.

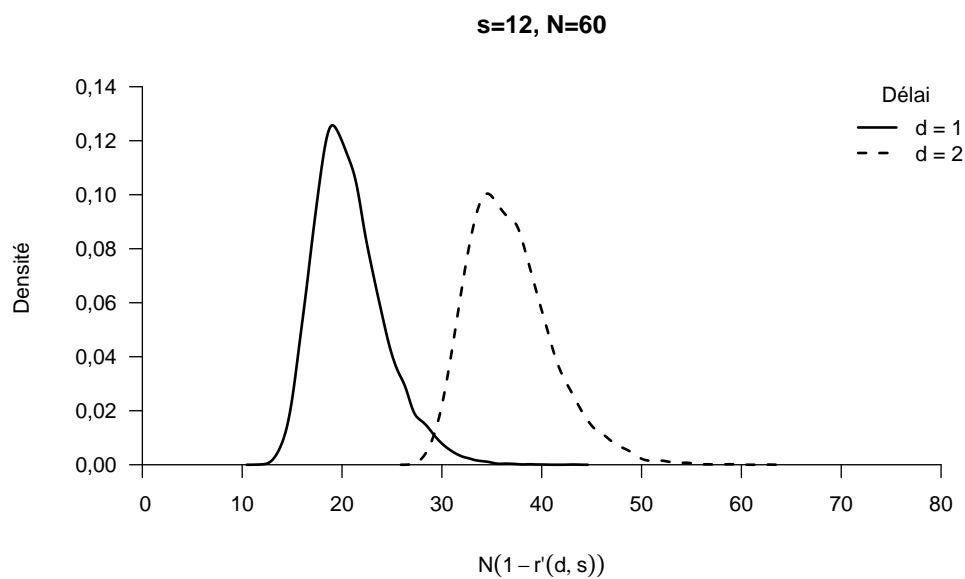


Figure 6.20 – Distribution de $N(1 - r'(ds))$ lorsque $s = 12$ et $N = 60$.

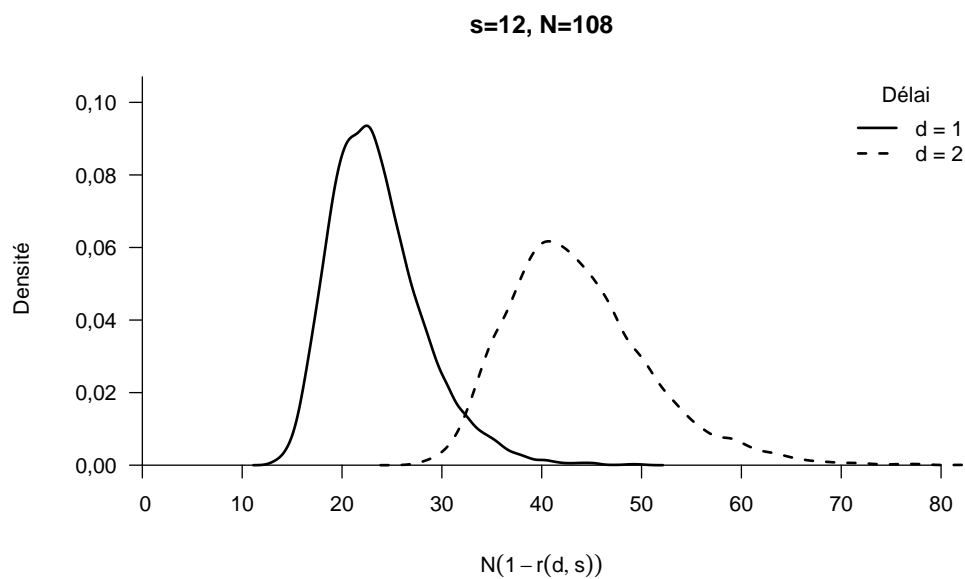


Figure 6.21 – Distribution de $N(1 - r(ds))$ lorsque $s = 12$ et $N = 108$.

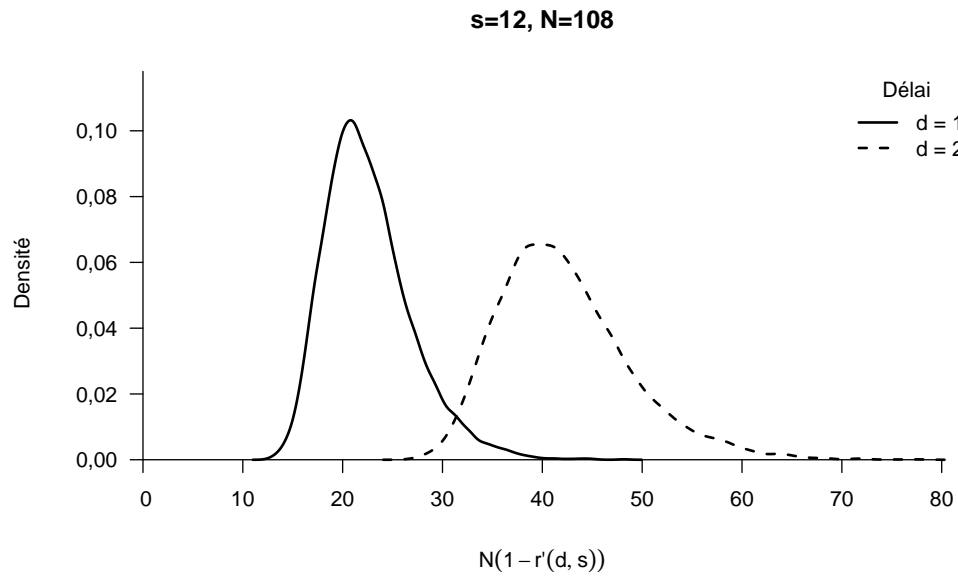


Figure 6.22 – Distribution de $N(1 - r'(ds))$ lorsque $s = 12$ et $N = 108$.

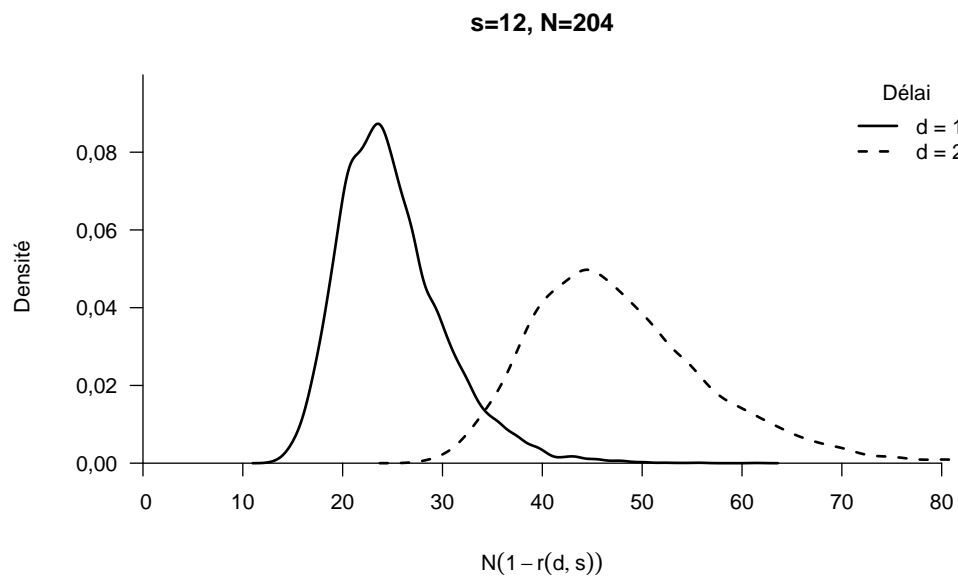


Figure 6.23 – Distribution de $N(1 - r(ds))$ lorsque $s = 12$ et $N = 204$.

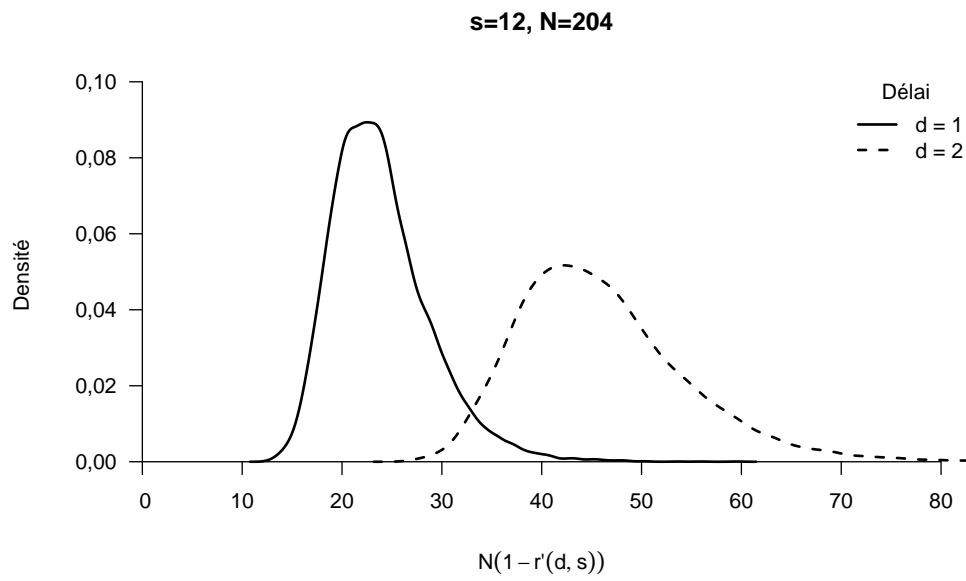


Figure 6.24 – Distribution de $N(1 - r'(ds))$ lorsque $s = 12$ et $N = 204$.

Annexe A

Utilisation de MacSyma

A.1. Description du logiciel

MacSyma est un système écrit en Lisp capable de manipuler symboliquement et numériquement des expressions mathématiques. Il a été conçu par le groupe Mathlab du MIT. Il s'agit d'un logiciel ayant des capacités étonnantes. Il peut tout aussi bien factoriser un polynôme, dériver ou intégrer une fonction, travailler avec le développement en série de Taylor ou de Laurent, évaluer une limite, résoudre des équations différentielles, manipuler des matrices et des tenseurs, etc... Ce système possède des procédures graphiques et son propre éditeur de texte. L'utilisateur a le choix de travailler en mode interactif ou en mode par lot¹. Nous invitons le lecteur à consulter la table des matières des deux manuels de référence de MacSyma pour qu'il puisse juger de la grande portée de ce logiciel.

Les deux manuels de référence sont nécessaires pour utiliser le programme de façon efficace. Néanmoins, pour une introduction au système, nous recommandons la lecture de GONG (1983) où un problème relié aux séries chronologiques est étudié via MacSyma. Le livre de R. H. Rand (1984) explique de façon détaillée comment utiliser le logiciel et traite un ensemble de problèmes variés. De plus, on y trouve un glossaire des principales fonctions de MacSyma.

La grande majorité des calculs faits dans cette thèse ont été vérifiés par le programme. Dans certains cas, bien que les manipulations algébriques nécessaires au développement des calculs soient élémentaires, il eut été fastidieux de les opérer manuellement.

1. Pour les plus jeunes qui ne l'ont pas connu, le mode par lot est un mode d'exécution dans lequel un programme informatique ou un ensemble de commandes est exécuté sans interaction directe avec l'utilisateur. Dans ce mode, les tâches sont regroupées dans un fichier en un « lot » et exécutées automatiquement selon un script ou une liste de commandes prédéfinie.

Afin d'éclairer le lecteur sur le type d'utilisation que nous avons fait de MacSyma, nous allons présenter une séance de travail où certains calculs de la section 2.6 sont exécutés. Les commentaires ayant été insérés dans le fichier de contrôle, cet exemple est auto-documenté. Nous fournissons, de plus, les fichiers de contrôle utilisés pour certains autres calculs de la thèse.

A.2. Exemple de séance MacSyma

Nous présentons ici une séance complète d'utilisation du programme MacSyma². Nous avons demandé le fonctionnement en mode par lot en appelant la fonction Batch. L'argument de cette fonction est le nom d'un fichier contenant les commandes à exécuter. Ce choix est motivé par le fait que nous pouvons insérer des commentaires entre les commandes. Ceux-ci sont délimités par les symboles `'/*'` et `*/'`. La lecture des résultats de la séance devient alors plus compréhensible. Nous avons retouché les commentaires de cet exemple afin de les compléter et de placer les signes diacritiques.


Maxima 5.47.0 <https://maxima.sourceforge.io>

using Lisp SBCL 2.3.9

Distributed under the GNU Public License. See the file COPYING.

Dedicated to the memory of William Schelter.

The function `bug_report()` provides bug reporting information.

2. La présentation de cette séance diffère de celle de la première édition. Nous avons en effet utilisé wxMaxima 23.10.0  dans un environnement Sequoia 15.3.1 sur un MacIntosh d'Apple. Il s'agit en fait de la version 5.47.0 de Maxima

```

-> batch(prog16 );

/*
Le but de ce programme est d'évaluer la covariance entre deux autocovari-
ances échantillonales lorsque nous ne centrons pas les données. Il nous
donne aussi la variance d'une autocovariance, de même que la variance de
la variance.
*/

/*
SimpSum est une variable logique indiquant à MacSyma s'il est nécessaire
d'essayer de simplifier les sommations. Si SimpSum est vraie, alors le
programme essaie de simplifier les sommations.
*/

simpsum :true ;

```

true (simpsum)

```

-> /*
Dans le calcul de la covariance de la section 2.5, interviennent 6 som-
mations. Elles seront liées aux variables T1 à T6. La première, T1, est
la contribution des blocs de la diagonale principale. Nous utilisons ici
deux fonctions de MacSyma : sum et ratSmp.

La fonction sum se compose de quatre paramètres essentiels : le pre-
mier
définit l'expression à évaluer, le deuxième indique l'indice de sommation,
et les deux derniers spécifient respectivement la borne inférieure et la
borne supérieure de cet indice.

La fonction ratsimp est appelée avec un unique argument, qui consiste en
une expression, dans le but de la simplifier.
*/

```

```

t1 :ratsimp(sum(2*i^2+i*h,i,1,n-k));

```

$$\begin{aligned}
 & \frac{1}{6} \left(4n^3 + (-12k + 3h + 6)n^2 \right. \\
 & \quad \left. + (12k^2 - 6hk - 12k + 3h + 2)n \right. \\
 & \quad \left. - 4k^3 + (3h + 6)k^2 + (-3h - 2)k \right)
 \end{aligned}
 \tag{t1}$$

->

/*

La contribution des blocs de la partie triangulaire supérieure est la somme de deux doubles sommations, T2 et T3. Pour chacune, nous procéderons en deux étapes.

*/

```
ratsimp(sum (2*i^2+(nu+h)*i,i,1,n-k-nu));
```

$$\begin{aligned}
 & -\frac{1}{6} \left(\nu^3 + (-6n + 6k - 3h - 3)\nu^2 \right. && (\% \text{ o3}) \\
 & \quad + \left(9n^2 + (-18k + 6h + 9)n + 9k^2 + (-6h - 9)k + 3h + 2 \right) \nu \\
 & \quad - 4n^3 + (12k - 3h - 6)n^2 + \left(-12k^2 + (6h + 12)k - 3h - 2 \right) n \\
 & \quad \left. + 4k^3 + (-3h - 6)k^2 + (3h + 2)k \right)
 \end{aligned}$$

->

/*

Pour référer ou résultat de la dernière expression on peut utiliser son numéro, i.e. %o3, ou encore le symbole %.

*/

```
t2 :ratsimp (sum(% ,nu,1,h));
```

$$\begin{aligned}
 t_2 = \frac{1}{24} \left(16hn^3 + n^2(-48hk - 6h(h-1)) \right. &&& (\text{t2}) \\
 & \quad + n(48hk^2 + 12h(h-1)k - 4h^3 - 6h^2 - 6h) \\
 & \quad - 16hk^3 + 6h(h-1)k^2 + (4h^3 + 6h^2 + 6h)k \\
 & \quad \left. + 3h^4 + 2h^3 - 3h^2 - 2h \right)
 \end{aligned}$$

->

/*

La contribution des blocs de la partie triangulaire inferieure est la somme de trois doubles sommations, T4, T5 et T6.

*/

```
ratsimp(sum(i^2+h*i,i,1,n-k-nu));
ratsimp(sum(% ,nu,h+1,n-k-1));
t3 :ratsimp(2*%);
ratsimp(sum(j^2+k*j,j,1,n-k-nu));
t4 :ratsimp(2*sum(% ,nu,k,n-h-1));
ratsimp(sum(2*j^2+(nu+k)*j,j,1,n-h-nu));
t5 :ratsimp(sum(% ,nu,k-h+1,k-1));
ratsimp(sum(2*j^2+(2*nu+h)*j,j,1,n-k));
t6 :ratsimp(sum(% ,nu,1,k-h));
```

$$\begin{aligned}
& -\frac{1}{6} \left(2\nu^3 + (-6n + 6k - 3h - 3)\nu^2 \right. & (\%o5) \\
& \quad + \left(6n^2 + (-12k + 6h + 6)n + 6k^2 + (-6h - 6)k + 3h + 1 \right) \nu \\
& \quad - 2n^3 + (6k - 3h - 3)n^2 + \left(-6k^2 + (6h + 6)k - 3h - 1 \right) n \\
& \quad \left. + 2k^3 + (-3h - 3)k^2 + (3h + 1)k \right)
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{12} \left(n^4 + (-4k - 2h)n^3 \right. & (\%o6) \\
& \quad + \left(6k^2 + 6hk - 1 \right) n^2 \\
& \quad + \left(-4k^3 - 6hk^2 + 2k + 2h^3 \right) n \\
& \quad \left. + k^4 + 2hk^3 - k^2 - 2h^3k - h^4 + h^2 \right)
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{6} \left(n^4 + (-4k - 2h)n^3 \right. & (t3) \\
& \quad + \left(6k^2 + 6hk - 1 \right) n^2 \\
& \quad + \left(-4k^3 - 6hk^2 + 2k + 2h^3 \right) n \\
& \quad \left. + k^4 + 2hk^3 - k^2 - 2h^3k - h^4 + h^2 \right)
\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{6} \left(2\nu^3 + (-6n + 3k - 3)\nu^2 \right. & (\%o8) \\
& \quad \left. + (6n^2 + (6 - 6k)n - 3k + 1)\nu \right)
\end{aligned}$$

$$- 2n^3 + (3k - 3)n^2 + (3k - 1)n - k^3 + k)$$

$$\begin{aligned} & \frac{1}{6} \left(n^4 + (4 - 6k)n^3 \right. \\ & \quad + (12k^2 - 18k + 5)n^2 \\ & \quad + (-8k^3 + 24k^2 - 16k + 2)n \\ & \quad + k^4 + (-2h - 10)k^3 + 11k^2 \\ & \quad \left. + (2h^3 + 6h^2 + 6h - 2)k - h^4 - 4h^3 - 5h^2 - 2h \right) \end{aligned} \quad (\text{t4})$$

$$\begin{aligned} & -\frac{1}{6} \left(\nu^3 + (-6n - 3k + 6h - 3)\nu^2 \right. \\ & \quad + (9n^2 + (6k - 18h + 9)n + (3 - 6h)k + 9h^2 - 9h + 2)\nu \\ & \quad - 4n^3 + (-3k + 12h - 6)n^2 \\ & \quad + ((6h - 3)k - 12h^2 + 12h - 2)n \\ & \quad \left. + (3h - 3h^2)k + 4h^3 - 6h^2 + 2h \right) \end{aligned} \quad (\%o10)$$

$$\begin{aligned} & \frac{1}{24} \left((16h - 16)n^3 \right. \\ & \quad + ((24 - 24h)k - 30h^2 + 54h - 24)n^2 \\ & \quad + ((36h^2 - 60h + 24)k + 20h^3 - 54h^2 + 42h - 8)n \\ & \quad + (8h - 8)k^3 + (6h - 6h^2)k^2 \\ & \quad \left. + (-12h^3 + 30h^2 - 26h + 8)k - 5h^4 + 18h^3 - 19h^2 + 6h \right) \end{aligned} \quad (\text{t5})$$

$$\begin{aligned} & \frac{1}{6} \left((6n^2 + (6 - 12k)n + 6k^2 - 6k)\nu \right. \\ & \quad + 4n^3 + (-12k + 3h + 6)n^2 \\ & \quad + (12k^2 + (-6h - 12)k + 3h + 2)n \\ & \quad \left. - 4k^3 + (3h + 6)k^2 + (-3h - 2)k \right) \end{aligned} \quad (\%o12)$$

$$\begin{aligned} & \frac{1}{6} \left((4k - 4h)n^3 \right. \\ & \quad + (-9k^2 + (9h + 9)k - 9h)n^2 \\ & \quad \left. + (6k^3 + (-6h - 15)k^2 + (15h + 5)k - 5h)n \right) \end{aligned} \quad (\text{t6})$$

$$-k^4 + (h+6)k^3 + (-6h-5)k^2 + 5hk)$$

-> /*
 Pour avoir des expressions plus interessantes, on peut indiquer
 l'importance des variables via la fonction RatVars. Celle-ci a comme
 argument la liste des variables dont on veut indiquer l'importance
 selon l'ordre croissant de leur importance. Par exemple, si nous avons
 un polynôme en h et en k, il peut être exprimé comme un polynome en k
 dont les coefficients sont des polynomes en h, ou bien comme un polynome
 en h dont les coefficients sont des polynomes en k.
 */
 ratvars(n, h[0]);

$$[n, h_0] \quad (\% \text{ o14})$$

-> /*
 On additionne le tout et on multiplie par h[0]/(n^2*s)
 */
 ratsimp(h[0]*(t1+t2+t3+t4+t5+t6)/(n^2*s));
 /*
 On obtient la variance de l'autocovariance en substituant h à k dans la
 dernière expression via la fonction RatSubst. On demande aussi de
 simplifier l'expression.
 */
 ratsimp(ratsubst(h,k,%));

$$\frac{h_0}{12n^2s} \left(4n^4 + (-12k + 4h + 8)n^3 \right. \quad (\% \text{ o15})$$

$$\begin{aligned} &+ (18k^2 + (-6h - 30)k - 18h^2 + 18h + 8)n^2 \\ &+ (-12k^3 + 42k^2 + (24h^2 - 18h - 30)k + 12h^3 - 30h^2 + 14h + 4)n \\ &+ 2k^4 + (-2h - 20)k^3 + (22 - 6h^2)k^2 \\ &+ (-4h^3 + 30h^2 + 6h - 4)k - 5h^4 + 2h^3 - 19h^2 - 2h) \end{aligned}$$

$$\begin{aligned} &\frac{h_0}{12n^2s} \left(4n^4 + (8 - 8h)n^3 \right. \quad (\% \text{ o16}) \\ &\quad + (-6h^2 - 12h + 8)n^2 \\ &\quad + (24h^3 - 6h^2 - 16h + 4)n \\ &\quad \left. - 15h^4 + 12h^3 + 9h^2 - 6h) \right) \end{aligned}$$

```

-> /*
    Finalement, la variance de la variance est obtenue en remplaçant h par 0
    dans la dernière expression. On peut aussi demander la factorisation de
    l'expression finale.
*/
factor (ratsimp(ratsubst(0,h,%)));

```

$$\frac{h_0 (n + 1) (n^2 + n + 1)}{3ns} \quad (\% \text{ o17})$$

A.3. Autres programmes MacSyma

Les fichiers `Proph` et `Progi` contiennent les commandes permettant d'obtenir la covariance entre deux autocovariances échantillonnales calculées à partir des données centrées. En effet, ils effectuent, étape par étape, tous les calculs décrits à la section 2.5.

On notera que la plupart des commandes sont terminées par le symbole `$`. L'effet de ce symbole est que MacSyma n'imprime pas le résultat de l'évaluation de la commande. On termine ainsi une commande lorsque le résultat produit est connu ou lorsque sa forme explicite ne nous intéresse pas vraiment.

Cela permet d'obtenir une liste plus courte de la séance.

A.3.1. ProgA

```

(% i1) /* Premier programme MacSyma */
writefile(sortie1)$
      NIL/ · Startsdribblingto$SORTIE1(2024/10/19, 09 : 11 : 27). · /

(% i2) /* Permettre les simplifications */
simpsum :true$
(% i3) /* Définition de r(i,j) */
r(i, j) := ((2 * n + 1) * (n - 3 * (i + j) + 1) + 3 * (i^2 + j^2))/(6 * n*s)$

```

```

(% i12) /* Contribution des blocs diagonaux */
t1 : ratsimp(h[0] *s* sum(2 * i^2 + h * i, i, 1, n - k))$
temp : (i + h) * r(i, i) + i * (r(i + h, i + k) + r(i, i + k) + r(i + h, i))$
ratsimp(%)$
t2 : ratsimp(s*g[0]^4 * sum(%, i, 1, n - k))$
temp : ratsimp(r(i, i) * r(i + h, i + k) + r(i + h, i) * r(i, i + k))$
t3 : ratsimp(s^2*g[0]^4 * sum*(%, i, 1, n - k))$
res1 : ratsimp(t1 + t2 + t3)$
ratvars(n)$
rat(res1)$
(% i18) /* Contribution de la partie triangulaire supérieure */
2 * i^2 + (l + h) * i$
t1 : ratsimp(sum(%, i, 1, n - k - l))$
ratvars(l)$
rat(t1)$
ratsimp(sum(%, l, 1, h))$
t1 : %*s * h[0]$
(% i25) /* Calcul du second terme de la contribution triangulaire supérieure */
t2 : i^2 + l * h$
ratvars(i)$
rat(t2)$
t2 : ratsimp(sum(%, i, 1, n - k - l))$
ratvars(l)$
rat(t2)$
t2 : 2 * s * h[0] * ratsimp(sum(%, l, h + 1, n - k - 1))$
(% i32) /* Calcul du troisième terme de la contribution triangulaire */
t3 : ratsimp(i * (r(i, i + l) + r(i, i + l + k) + r(i + h, i + l + k) + r(i + h, i +
l)))$
ratvars(t1)$
rat(t3)$
t3 : ratsimp(sum(\%, i, 1, n - k - l))$
ratvars(l)$
rat(t3)$
t3 : ratsimp(sum(%, l, 1, n - k - 1))$

```

```

(% i39) /* Calcul du quatrième terme de la contribution triangulaire */
t4 : ratsimp(l * r(i, i + l + k) + h * r(i, i + l))$
ratvars(i)$
rat(t4)$
t4 : ratsimp(sum(%, i, 1, n - k - l))$
ratvars(l)$
rat(t4)$
t4 : ratsimp(sum(%, l, 1, h))$
(% i46) /* Calcul du cinquième terme de la contribution triangulaire */
t5 : ratsimp(r(i, i + l + k) + r(i, i + l))$
ratvars(i)$
rat(t5)$
t5 : ratsimp(sum(%, i, 1, n - k - l))$
ratvars(l)$
rat(t5)$
t5 : h * ratsimp(sum(%, l, h + 1, n - k - l))$
(% i53) /* Calcul du sixième terme de la contribution triangulaire */
t6 : ratsimp(r(i, i + l) * r(i + h, i + l + k) + r(i, i + l + k) * r(i + h, i + l))$
ratvars(i)$
rat(t6)$
t6 : ratsimp(sum(%, i, 1, n - k - l))$
ratvars(l)$
rat(t6)$
t6 : ratsimp(sum(%, l, 1, n - k - l))$
(% i55) /* Calcul final et mise en forme */
ratvars(s, g[0], h[0], n)$
res2 : ratsimp(t1 + t2 + s[0] * g[0]^4 * (t3 + t4 + t5) + s^2 * g[0]^4 * t6)$
(% i57) /* Sauvegarde et fin du programme */
save("sortie1", valeur1, res1, res2)$
quit();

```

A.3.2. ProgB

```

(% i1) /* Second programme MacSyma */

      /* Écriture dans le fichier de sortie */
      /* writefile(sortie2); */

      /* Permettre les simplifications */
      simpsum :true;
(% i2) /* Définition de r(i,j) */
r(i, j) := ((2 * n + 1) * (n - 3 * (i + j) + 1) + 3 * (i^2 + j^2)) / (6 * n * s);
(% i9) /* contribution de la partie triangulaire inférieure */
      /*
      *****
      */
      j*k+j^2;
      ratvars(j);
      rat(%th(2));
      t1 : ratsimp(sum(%, j, 1, n - h - 1));
      ratvars(1);
      rat(%th(2));
      t1 : 2 * s * h[0] * ratsimp(sum(%, 1, k, n - h - 1));
(% i16) /*
      *****
      */
      ratsimp(j * r(1 + h + j, j + k) + (k + j) * r(1 + h + j, j) + j * r(1 + j, j + k) +
      (k + j) * r(1 + j, j));
      ratvars(j);
      rat(%th(2));
      t2 : ratsimp(sum(%, j, 1, n - h - 1));
      ratvars(1);

      rat(%th(2));
      t2 : s*g[0]^4 * ratsimp(sum(%, 1, k, n - h - 1));

```

```

(% i20) /*
*****
*/
ratsimp(r(l + j, j) * r(l + h + j, j + k) + r(l + j, j + k) * r(l + h + j, j));
ratvars(j);
rat(%th(2));
t3 : s^2 * g[0]^4 * ratsimp(sum(%, l, k, n - h - 1));
(% i27) /*
*****
*/
j*(l+k)+2*j^2;
ratvars(j);
rat(%th(2));
t4 :ratsimp(sum(%,j,1,n-h-1));
ratvars(l);
rat(%th(2));
t4 : s*h[0]*ratsimp(sum(%,l,k-h+1,k-1));
(% i34) /*
*****
*/
ratsimp(j*r(l+h+j,j+k)+(l+j)*r(l+h+j,j)+j*r(l+j,j+k)+(k+j)*r(l+j,j));
ratvars(j);
rat(%th(2));
t5 : ratsimp(sum(%, j, 1, n - h - 1));
ratvars(j);
rat(%th(2));
t5 : s*g[0]^4*ratsimp(sum(%,l,k-h+1,k-1));
(% i41) /*
*****
*/
ratsimp(r(l+j, j) * r(l+h+j, j+k) + r(l+j, j + k) * r(l + h + j, j));
ratuars(j);
rat(%th(2));
t6 : ratsimp(sum(%, j, 1, n-1-1));
ratvars(l);
rat(%th(2));
t6 :s^2*rg[0]^4*ratsimp(sum(%, l, k-h+1, k-1));

```

```

(% i45) /*
*****
*/
j*(2*i+h)+2*j^2;
ratvars(j)$
rat(%th(2));
t7 : s*h[0] * ratsimp(sum(%, 1, 1, k - h));
(% i49) /*
*****
*/
ratsimp(j*r(1+h+j,j+k) + (1+j)*r(1+h+j,j) + j*r(1+j,j+k)+(1+h+j)*r(1+j,j));
ratvars(j);
rat(%th(2));
t8 : s*g[0]^4*ratsimp(sum(%, 1, 1, n - h));
(% i56) /*
*****
*/
ratsimp(r(1+j,j)*r(1+h+j,j+k) + r(1+j,j+k)*r(1+h+j,j));
ratavars(j);
rat(%th(2));
t9 :ratsimp(sum(%,j, 1, n-k));
ratvars(1);
rat(%th(2));
t9 : s^2 * g[0]^4 * ratsimp(sum(%, 1, 1, k - h));
(% i60) /*
*****
Résultats finaux
*****
*/
res3 : ratsimp(t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 + t9);
resultat : ratsimp(res1 + res2 + res3);
ratvars(h[0], g[0], n);
result2 : rat(resultat)$
save(valeur2,res1,res2,res3,result2)
quit();

```

ProgF.

Le fichier ProgF contient les commandes permettant le calcul des expressions de base conduisant à l'évaluation des expressions (2.47) et (2.48). ProgF est chargé de trouver $a_1 = E[c(ks)]$, $b_1 = E[c(0)]$, $a_2 = E[c(ks)c(0)]$, $a_3 = E[c^2(0)]$ et $a_4 = E[c^2(ks)]$.

```
(% i2) writefile(sortie6)$
/*
*****
*
* Sixieme programme Macsym *
*
*****
*/
ratvars(h,k,g0,h0,n,sigma)$
/*
Evaluation de la covariance
*/
(% i3) resultat :-((56*g0^4-60*s*h0)*n^6
+((-52*k-52*h+120)*g0^4+(60*k+60*h-120)*s*h0)*n^5
+((-60*k^2+(65*h-90)*k-60*h^2-90*h+110)*g0^4
+(90*k^2+(-90*h+90)*k+90*h^2+90*h-120)*s*h0)*n^4
+((50*k^3-90*k^2+(30*h^2+90*h-90)*k+40*h^3-90*h^2-80*h+60)*g0^4
+(-120*k^3+90*k^2+(-180*h^2-90*h+150)*k-60*h^3+90*h^2
+90*h-60)*s*h0)*n^3
+((15*k^4+(-10*h+60)*k^3+(90*h^2-45)*k^2+(-10*h^3+90*h^2
+55*h-60)*k+15*h^4+30*h^3-45*h^2-30*h+14)*g0^4
+(30*k^4+(30*h-60)*k^3+(90*h^2-30)*k^2+(60*h^3-90*h^2
-90*h+60)*k+15*h^4-30*h^3-15*h^2+30*h)*s*h0)*n^2
+(-9*k^5+(-60*h^2+25)*k^3+(-30*h^4+60*h^2-14)*k
-3*h^5+5*h^3)*g0^4*n
+(-3*h*k^5+(-20*h^3+15*h)*k^3+(-3*h^5+15*h^3-14*h)*k)*g0^4)
/(180*n^2)$
(% i5) rat(ratsimp(resultat*sigma^4/(n^2*s^2)))$
cov(h,k) :=”(%)$
```

```

(% i8) /*
      Evaluation de la variance
      */
      ratsimp(ratsubst(k,h,%th(2)))$
      rat(%)$
      var(k) :="(%)$
(% i9) /*
      Evaluation de l'espérance
      */
      mu(m) :="(rat(-((2*g0^2-3*d0*s)*n^3+((-m+3)*g0^2+(6*m-3)*d0*s)*n^2
      +((-3*m^2-3*m+1)*g0^2+(-3*m^2+3*m)*d0*s)*n
      +(2*m^3-m) *g0^2) *sigma^2/(6*s *n^2)))$
(% i13) /*
      Évaluation du second moment (non centrée)
      */
      rat(mu(k)^2)$
      rat(ratsimp(var(k)+%))$
      mu2(k) :="(%)$
      rat(ratsimp(cov(h,k)+mu(h)*mu(k)))$
(% i21) muhk (h,k) :="(%)$
      /*
      Evaluation de l'approximation de l'espérance de la corrélation
      */
      a1 :mu(k)$
      b1 :mu(0)$
      a2 :muhk (0,k)$
      a3 :mu2(0)$
      a4 :mu2(k)$
      save(valeur6, a1, a2, a3, a4, b1)$
      quit( );

```

Les expressions de base étant disponibles dans le fichier valeur6, on peut obtenir les formes explicites des équations (2.47) et (2.48). Notons qu'il n'est pas souhaitable de travailler avec ces formules directement. Il vaut mieux chercher une forme plus commode pour le calcul. Nous donnons ici les deux séances de travail (Séance A et Séance B) qui mettent en évidence les formes explicites.

Séance A.

(% i1) a1/b1*(1-a2/(a1*b1)+o3/b1^ 2);

$$\frac{a_1 \left(-\frac{a_2}{a_1 b_1} + \frac{a_3}{b_1^2} + 1 \right)}{b_1} \quad (\% \text{ o1})$$

(% i2) ratsimp(%o1);

$$\frac{a_1 b^2 - a_2 b_1 + a_1 a_3}{b_1^3} \quad (\% \text{ o2})$$

```
(% i8) ratvars(n)$
loadfile(valeur6)$
rat(a1*b1^2)$
rot(a2*b1)$
rat(al*a3)$
rat(d5-d6+d7)$
rat(b1^3)$
ratsimp(d8/d9)$
ratvars(h,k,h0,g0,n,sigma)*
rat(%th(2));
```

$$\begin{aligned}
& (\%o12)/R/ \left((40g_0^6 - 180d_0sg_0^4 + 270d_0^2s^2g_0^2 - 135d_0^3s^3) n^6 \right. \\
& + \left((-68k + 140)g_0^6 - 540d_0sg_0^4 + (60skh_0 - 405d_0^2s^2k + 675d_0^2s^2) \right. \\
& g_0^2 + 180d_0s^2kh_0 + 270d_0^3s^3k - 270d_0^3s^3) n^5 \\
& + \left((-12k^2 - 100k + 190) g_0^6 + (468d_0sk^2 + 28800sk - 50500s) g_0^4 \right. \\
& + \left(45d_0^2s^2k^2 - g_0d_0^2s^2k + 540d_0^2s^2) g_0^2 \right. \\
& + \left(-450d_0s^2k^2 + 270d_0s^2k) h_0 - 130d_0^3s^3k^2 + 405d_0^3s^3k \right. \\
& - 135d_0^3s^3) n^4 + \left((28k^3 - 168k^2 - 141k + 125) g_0^6 \right. \\
& + \left(-270d_0sk^3 + 612d_0sk^2 + 432d_0sk - 270d_0s) g_0^4 \right. \\
& + \left((-120sk^3 + g_0sk^2 + g_0sk) h_0 + g_0d_0^2s^2k^3 + 135d_0^2s^2k^2 \right. \\
& - 630d_0^2s^2k + 135d_0^2s^2) g_0^2 + \left(360d_0s^2k^3 - 450d_0s^2k^2 + 90d_0s^2k) \right. \\
& h_0 - 135d_0^3s^3k^2 + 135d_0^3s^3k) n^3 + \left(130k^4 + 122k^3 - 117k^2 - 114k \right. \\
& + 40) g_0^6 + \left(-45d_0sk^4 - 360d_0sk^3 + 208d_0sk^2 + 243d_0sk - 45d_0s) \right. \\
& g_0^4 + \left((60sk^4 - 120sk^3 - 150sk^2 + 30sk) h_0 + 90d_0^2s^2k^3 \right. \\
& + g_0d_0^2s^2k^2 - 135d_0^2s^2k) g_0^2 + \left(-g_0d_0s^2k^4 + 180d_0s^2k^3 \right. \\
& - 90d_0s^2k^2) h_0) n^2 + \left((-18k^5 + 15k^4 + 68k^3 - 18k^2 - 40k + 5) g_0^6 \right. \\
& + \left(27d_0sk^5 - 135d_0sk^3 + 27d_0sk^2 + 45d_0sk) g_0^4 \right. \\
& + \left(30sk^4 + 60sk^3 - 30sk^2) h_0g_0^2) n + \left(-9k^5 + 7k^3 - 5k) g_0^6) \right. \\
& / \left(40g_0^6 - 180d_0sg_0^4 + 270d_0^2s^2g_0^2 - 135d_0^3s^3) n^6 \right. \\
& + \left(140g_0^6 - 540d_0sg_0^4 + 675g_0^2s^2g_0^2 - 270d_0^3s^3) n^5 \right. \\
& + \left(190g_0^6 - 585d_0sg_0^4 + 540d_0^2s^2g_0^2 - 135d_0^3s^3) n^4 \right. \\
& + \left(125g_0^6 - 270d_0sg_0^4 + 135d_0^2s^2g_0^2) n^3 + \left(40g_0^6 - 45d_0sg_0^4) n^2 \right. \\
& \left. + 5g_0^6n) \right)
\end{aligned}$$

168

```
(% i9) quit();
```

Séance B.

(% i1) (a1/b1)^2*(a4/a1^2-2*a2/(a1*b1)+a3/b1^2-(a2/(a1*b1)-a3/b1^2)^2);

$$\frac{a_1^2 \left(- \left(\frac{2a_2}{a_1 b_1} \right) + \frac{a_3}{b_1^2} - \left(\frac{a_2}{a_1 b_1} - \frac{a_3}{b_1^2} \right)^2 + \frac{a_4}{a_1^2} \right)}{b_1^2} \quad (\% \text{ o1})$$

(% i2) ratsimp(%);

$$\frac{a_4 b_1^4 - 2a_1 a_2 b_1^3 + (a_1^2 a_3 - a_2^2) b_1^2 + 2a_1 a_2 a_3 b_1 - a_1^2 a_3^2}{b_1^6} \quad (\% \text{ o2})$$

```
(% i3) loadfile(valeur6)$
      rat(a4*b1^4)$
      rat (2*a1*a2*b1^3)$
      rat(a1^2*a3)$
      rat(%-a2^2)$
      rat(%*b1^2)$
      rat(2*a1*a2*a3*b1)$
      rat(a1^2*a3^2)$
      rat(d4-d5+d8+d9-d10)$
      rat(b1^6)$
      ratsimp(d11/d12)$
      ratuars(h,k,g0,h0,n,sigma)$
      rat(d13);
```

$$\begin{aligned}
& (d15)/R/ - \left(\left(\left(3600s^2k^2g_0^4 + 21600d_0s^3k^2g_0^2 + 32400d_0^2s^4k^2 \right) h_0^2 \right. \right. \\
& + \left(-9360sk^2g_0^8 - 28080d_0s^2k^2g_0^6 - 89100d_0^2s^3k^2g_0^4 \right. \\
& + 48600d_0^3s^4k^2g_0^2 - 36450d_0^4s^5k^2 \left. \right) h_0 + 4464k^2g_0^{12} \\
& + 12960d_0sk^2g_0^{10} + 45360d_0^2s^2k^2g_0^8 - 38800d_0^3s^3k^6g_0^6 \\
& + 32805d_0^4s^4k^2g_0^4 \left. \right) n^{10} \\
& + \dots^\dagger \\
& + \left(81k^{10} + 234k^8 - 431k^6 + 400k^4 - 140k^2 \right) 900^{12} \\
& / \left(\left(1600g_0^{12} - 14400d_0sg_0^{10} + 54000d_0^2s^2g_0^8 - 108000d_0^3s^3g_0^6 \right. \right. \\
& + 121500d_0^4s^4g_0^4 - 72900d_0^5s^5g_0^2 + 18225d_0^6s^6 \left. \right) n^{12} \\
& + \dots^\dagger \\
& + \left(400g_0^{12} - 450d_0sg_0^{10} \right) n^3 + 25g_0^{12}n^2
\end{aligned} \tag{A.1}$$

† Les points de suspension n'apparaissent pas sur la liste originale produite par MacSyma. Le programme écrit la formule au long. Cependant, celle-ci s'écrit sur un peu plus de sept pages et c'est pourquoi nous préférons ne donner que les coefficients des termes dominants.

Annexe B

Programmes reliés à la simulation

B.1. Description du logiciel

L'ensemble des simulations a été réalisé à l'aide de quatre variantes d'un même programme, dont les spécificités sont ici détaillées. Pour chaque simulation effectuée, les programmes généraient divers indicateurs statistiques, notamment la moyenne, la variance, le minimum, le maximum ainsi qu'une série de percentiles clés. Par ailleurs, des programmes préliminaires ont dû être développés afin de compléter et d'optimiser les programmes de simulation.

B.2. Programmes préliminaires

B.2.1. Programme ConsDef

Ce programme a été conçu conformément aux spécifications des exercices 7 et 10 de la section 3.4.1 du livre de KNUTH (1981), dans le but de calculer les constantes nécessaires à la simulation de la loi normale à l'aide de l'algorithme M décrit dans cet ouvrage.¹ Ce programme permet ainsi d'obtenir, avec une précision acceptable, les valeurs figurant dans le tableau 6.2.

```
1 Program CreerTable1;  
2  
3 Const  
4   Pi = 3.14159265358979323846264338327950288;  
5   Debug = False;  
6
```

1. Nous présentons ici une version légèrement modifiée par rapport à celle de la thèse originale. Cette version est compatible avec le compilateur *Free Pascal Compiler version 3.2.2 [2021/05/16]*.

```

7     NPoints = 40;           { Nombre de point utilisés dans la procédure
8                             d'intégration}
9     h = 0.2;               { Largeur des intervalles bornés dans la ↘
→partition
10                             de (0,3)}
11     nDec = 3;             { Nombre de décimales dans l'affichage des
12                             constantes du tableau}
13     tolerance = 1.0e-16; { Ce que l'on considère comme nul}
14     maxIter = 100;       { Nombre d'itérations maximales autorisé dans
15                             NewtonRaphson}
16
17 Type
18     probVal = Record      { Structure d'information pour la méthode de
19                             l'alias de Walker}
20         p: extended;    { Probabilité qu'on garde la valeur choisie en
21                             premier}
22         valeur: integer { Valeur retournée si on ne garde pas la valeur
23                             choisie enpremier}
24     End;
25
26     fonctionDeMasse = Array[1..32] Of probVal;
27     Table = Array[1..40] Of extended;
28
29     { Type pour une fonction prenant un "extended" et retournant un
30         "extended" }
31     TFunction = function(x: extended): extended;
32     TFunction2 = function(x, b:extended): extended;
33
34 Var
35     w, xGauss: table;     { Tableaux de constantes pour l'intégration
36                             numérique}
37     PP: fonctionDeMasse; { Tableau pour la «mise en boîte» de la ↘
→fonction
38                             de masse}
39
40     p: Array[0..32] Of extended; { Tous les tableaux de ces sept ↘
→ }
41     y, z: Array[0..31] Of extended; { déclarations sont utilisés pour ↘
→ }
42     q: Array[1..15] Of extended; { construire le tableau 1 page 123 de ↘
→ }

```

```

43   s: Array[1..16] Of extended;      { Knuth (1986)           ↘
→ }
44   a, b: Array[1..15] Of extended;
45   d, e: Array[16..30] Of extended;
46   x: Array[6..15] Of extended;
47
48   j: integer;
49   total: extended;
50   sortie: Text;
51
52 Procedure GetXW;
53 {↘
→=====↘
→}
54 { Cette procédure lit dans un fichier les tableaux x[] et w[] nécessaires↘
→ }
55 { à la procédure d'intégration Integre           ↘
→ }
56 {↘
→=====↘
→}
57   Const
58       Debug = False;
59   Var
60       i: integer;
61       Entree: Text;
62       fileName: String;
63   Begin
64       fileName := 'ConstGauss.txt';
65       Assign(Entree, fileName); { Associer le fichier à Entree }
66       Reset(Entree);           { Ouvrir le fichier en lecture }
67
68       { Lire les valeurs de xGauss }
69       For i := 1 To 20 Do
70           Begin
71               Readln(Entree, xGauss[20 + i]);
72               xGauss[21 - i] := -xGauss[20 + i];
73           End;
74
75       { Lire les valeurs de W }
76       For i := 1 To 20 Do
77           Begin

```

```

78         ReadLn(Entree, W[20 + i]);
79         W[21 - i] := W[20 + i];
80     End;
81
82     { Si Debug est activé, afficher les tableaux }
83     If Debug Then
84         For i := 1 To 40 Do
85             Writeln(i : 2, xGauss[i] : 20 : 15, w[i] : 20 : 15);
86
87         Close(Entree); { Fermer le fichier }
88     End;
89
90     Function Integre(f: TFunction; a, b: extended): extended;
91     {↘
92     →=====↘
93     →}
94     { Cette fonction retourne la valeur de l'intégrale de la fonction f entre↘
95     → }
96     { a et b. Elle est basée sur la quadrature de Gauss avec 40 points ↘
97     → }
98     {↘
99     →=====↘
100    →}
101     Const
102         Debug = False;
103     Var
104         i: Integer;
105         t1, t2, somme, xi, fi: extended;
106     Begin
107         If Debug Then
108             Writeln('Integre: entree');
109         somme := 0;
110         If b > a Then
111             Begin
112                 t1 := (b - a) / 2;
113                 t2 := (b + a) / 2;
114                 For i := 1 To NPoints Do
115                     Begin
116                         xi := xGauss[i] * t1 + t2;
117                         fi := f(xi);
118                         Somme := Somme + w[i] * fi;
119                     End;
120                 If Debug Then

```

```

114         Writeln(xi, fi);
115         End;
116         Somme := Somme * t1
117     End;
118     Integre := Somme;
119     If Debug Then
120         Writeln('Integre: Sortie')
121     End;
122
123     Function NormDens (x: extended): extended;
124     {↘
125     →=====↘
126     →}
127     { Cette fonction retourne la valeur de la densité d'une loi N(0,1) ↘
128     → }
129     {↘
130     →=====↘
131     →}
132     Begin
133         NormDens := exp(-sqr(x) / 2) / (sqrt(2 * pi))
134     End;
135
136     Function NewtonRaphson (x, a, tolerance: extended;
137         F: TFunction2;
138         FPrime: TFunction): extended;
139     {↘
140     →=====↘
141     →}
142     { Cette fonction retourne la valeur de x solution de F(x)=a. On doit ↘
143     → }
144     { donner en entrée la fonction F et sa dérivée FPrime. Le nombre maximal ↘
145     → }
146     { d'itérations est maxIter. Le paramètre tolérance est un petit nombre ↘
147     → }
148     {réel considéré comme nulle. ↘
149     → }
150     {↘
151     →=====↘
152     →}
153     Var
154         nouveauX, ancienX, distance: extended;
155         iter: integer;

```

```

143 Begin
144     iter := 0;
145     ancienX := x;
146     distance := 2 * tolerance;
147     While (distance > tolerance) And (iter < maxIter) Do
148         Begin
149             nouveauX := ancienX - F(ancienX, a) / FPrime(ancienX);
150             distance := abs(nouveauX - ancienX);
151             ancienX := nouveauX;
152             iter := iter + 1
153         End;
154     NewtonRaphson := ancienX;
155 End;
156
157 Function FjPlus15 (x: extended): extended;
158 {↘
159     →=====↘
160     →}
159 { Cette fonction retourne la différence deux hauteurs de la fonction de ↘
160     → }
160 { densité normale (multiplié par 2) divisée par la probabilité d'être ↘
161     → }
161 { dans la zone délimité par les extrémités de l'intervalle ↘
162     → }
162 {↘
163     →=====↘
164     →}
163 Begin
164     FjPlus15 := sqrt(2/pi) * (exp(-sqr(x)/2) - exp(-sqr(j)/50))/p[j+15];
165 End;
166
167 Function FjPlus15Prime (x: extended): extended;
168 {↘
169     →=====↘
170     →}
169 { Cette fonction retourne la dérivée de la fonction FjPlus15 ↘
170     → }
170 {↘
171     →=====↘
172     →}
171 Begin
172     FjPlus15Prime := -sqr(2 / pi) * x * exp(-sqr(x) / 2) / p[j + 15]

```

```

173 End;
174
175 Function AAnnuler (x, a: extended): extended;
176 Begin
177     AAnnuler := FjPlus15Prime(x) - a;
178 End;
179
180 Function AAnnulerPrime (x: extended): extended;
181 Begin
182     AAnnulerPrime := sqrt(2/pi)*exp(-sqr(x)/2)*(sqr(x)-1)/p[j+15];
183 End;
184
185 Procedure QuickSort (Var a: fonctionDeMasse;
186                     n: integer);
187     Procedure Sort (l, r: integer);
188         Var
189             i, j: integer;
190             x: extended;
191             w: probVal;
192         Begin
193             i := l;
194             j := r;
195             x := a[(l + r) Div 2].p;
196             Repeat
197                 While a[i].p < x Do
198                     i := i + 1;
199                 While a[j].p > x Do
200                     j := j - 1;
201                 If i <= j Then
202                     Begin
203                         w := a[i];
204                         a[i] := a[j];
205                         a[j] := w;
206                         i := i + 1;
207                         j := j - 1
208                     End
209                 Until i > j;
210                 If l < j Then
211                     sort(l, j);
212                 If r > i Then
213                     sort(i, r)
214             End;

```

```
215 Begin
216     Sort(1, n)
217 End;
218
219 Procedure MettreEnboite (Var p: fonctionDeMasse;
220                         k: integer);
221     Var
222         i, j, n: integer;
223         a: fonctionDeMasse;
224         w: probVal;
225         temp: extended;
226
227     Begin
228         For i := 1 To k Do
229             p[i].valeur := i;
230         a := p;
231         QuickSort(a, k);
232         n := k;
233     Repeat
234         temp := k * a[1].p;
235         j := a[1].valeur;
236         If (abs(temp - 1) > 1.0e-12) Then
237             Begin
238                 p[j].p := temp;
239                 p[j].valeur := a[n].valeur;
240             End
241         Else
242             Begin
243                 p[j].p := 1;
244                 p[j].valeur := 0;
245             End;
246         a[1].p := a[n].p - (1 / k - a[1].p);
247         a[1].valeur := a[n].valeur;
248         n := n - 1;
249         i := 1;
250         While (i < n) And (a[i].p > a[i + 1].p) Do
251             Begin
252                 w := a[i];
253                 a[i] := a[i + 1];
254                 a[i + 1] := w;
255                 i := 1 + i;
256             End;
```

```

257     Until n = 0;
258 End;
259
260 Begin { Programme principal }
261
262 { Calcul des bornes des intervalles}
263 For j := 1 To 16 Do
264     s[j] := (j - 1) / 5;
265 If debug Then
266     For j := 1 To 16 Do
267         writeln(j : 2, ' ', s[j] : 4 : 1);
268
269 { Calcul des surfaces des 31 composantes de la décomposition de la ↘
    →surface}
270 GetXW;
271 total := 0;
272 For j := 1 To 15 Do
273     Begin
274         p[j] := sqrt(2 / 25 / pi) * exp(-sqr(j) / 50);
275         p[j + 15] := 2 * (Integre(@NormDens, s[j], s[j + 1])) - p[j];
276         total := total + p[j] + p[j + 15];
277     End;
278 p[31] := 1 - total;
279 {composante ayant une probabilité nulle ↘
    → }
280 p[32] := 0;
281
282 {↘
    →=====↘
    →}
283 { Copier les différentes probabilités dans une structure appropriée afin ↘
    → }
284 { de faire la mise en boîte. ↘
    → }
285 {↘
    →=====↘
    →}
286 For j := 1 To 31 Do
287     pp[j].p := p[j];
288 pp[32].p := 0;
289 MettreEnboite(pp, 32);

```

```

290 {↘
→=====↘
→}
291 { Copier les résultats de la mise en boîte dans les tableaux appropriés. ↘
→ }
292 {↘
→=====↘
→}
293 p[0] := 0;
294 y[0] := pp[32].valeur;
295 For j := 1 To 31 Do
296   Begin
297     p[j] := pp[j].p;
298     y[j] := pp[j].valeur
299   End;
300 If debug Then
301   For j := 0 To 31 Do
302     writeln(j : 3, p[j] : 16 : 12, y[j] : 3);
303
304 For j := 1 To 5 Do
305   Begin
306     a[j] := FjPlus15(s[j]);
307     b[j] := -h * FjPlus15Prime(s[j] + h);
308   End;
309 For j := 6 To 15 Do
310   b[j] := FjPlus15(s[j]);
311
312 For j := 6 To 15 Do
313   Begin
314     x[j] := NewtonRaphson(j / 5, -b[j] / h, tolerance, @AAnnuler,
315                           @AAnnulerPrime);
316     a[j] := FjPlus15(x[j]) + (x[j] - s[j]) * b[j] / h;
317   End;
318 If debug Then
319   For j := 1 To 15 Do
320     If j < 6 Then
321       writeln(j : 3, a[j] : 16 : 12, b[j] : 16 : 12)
322     Else
323       writeln(j : 3, a[j] : 16 : 12, b[j] : 16 : 12, x[j] : 16 : 12);
324
325 For j := 1 To 15 Do
326   d[j + 15] := a[j] / b[j];

```

```

327
328   For j := 1 To 5 Do
329     e[j + 15] := 25 / j;
330   For j := 6 To 15 Do
331     e[j + 15] := 1 / (exp((2 * j - 1) / 50) - 1);
332
333   For j := 0 To 31 Do
334     If abs(p[j] - 1) > tolerance Then
335       Begin
336         z[j] := 1 / (5 - 5 * p[j]);
337         y[j] := y[j] / 5 - z[j]
338       End
339     Else
340       Begin
341         z[j] := 0;
342         y[j] := 0
343       End;
344   For j := 1 To 15 Do
345     q[j] := 1 / (5 * P[j]);
346
347   Writeln('Table 1 of Knuth (1981), page 123');
348   Writeln('j':3,'p[j]':8,'p[j+16]':8,'q[j]':8,'y[j]':8,'y[j+16]':8,'z[j]'\
→:8,'z[j+16]':8,'s[j+1]':8,'d[j+15]':8,'e[j+15]':8);
349   Writeln(0:3,p[0]:8:nDec,p[16]:8:nDec,'':8,y[0]:8:nDec,y[16]:8:nDec,z\
→[0]:8:nDec,z[16]:8:nDec,s[1]:8:nDec,'':8,'':8);
350   Forj:=1 To 5 Do
351     Writeln(j:3,p[j]:8:nDec,p[j+16]:8:nDec,q[j]:8:nDec,y[j]:8:nDec,y[j\
→+16]:8:nDec,z[j]:8:nDec,z[j+16]:8:nDec,s[j+1]:8:nDec,d[j+15]:8:nDec,e[j\
→+15]:8:nDec);
352
353 {↘
→=====↘
→}
354 { Sauvegarde des résultats dans un fichier approprié. Il sera lu lorsque ↘
→ }
355 { nous voudrions simuler des observations selon une loi normale ↘
→ }
356 {↘
→=====↘
→}
357 { Assigner la variable de fichier à un fichier physique }
358 Assign(Sortie, 'NormData.txt');

```

```

359 { Ouvrir le fichier pour l'écriture }
360 ReWrite(Sortie);
361 For j := 0 To 31 Do
362   Begin
363     Writeln(Sortie,p[j] : 24 : 20);
364   End;
365 For j := 1 To 15 Do
366   Begin
367     Writeln(Sortie,q[j] : 24 : 20);
368   End;
369 For j := 0 To 31 Do
370   Begin
371     Writeln(Sortie,y[j] : 24 : 20);
372   End;
373 For j := 0 To 31 Do
374   Begin
375     Writeln(Sortie,z[j] : 24 : 20);
376   End;
377 For j := 1 To 16 Do
378   Begin
379     Writeln(Sortie,s[j] : 24 : 20);
380   End;
381 For j := 1 To 15 Do
382   Begin
383     Writeln(Sortie,d[j + 15] : 24 : 20);
384   End;
385 For j := 1 To 15 Do
386   Begin
387     Writeln(Sortie,e[j + 15] : 24 : 20);
388   End;
389 close(sortie)
390 End.

```

Programmes/Pascal/Consdef/Consdef.p

B.2.2. Le program ConsDick

Le programme ConsDick permet de trouver les constantes γ_i^* , $i = 1, \dots, 72$, selon la méthode proposée par DICKEY (1976) et reprise par HASZA (1980).

```

1 {$f-}
2 Program ConsDick(input, output);
3 {$i'math '}

```

```

4 Const
5 unsurpi2 = 0.1013211836423377714438795;
6 unsurpi4 = 0.01026598225468433518915278;
7   pi = 3.14159256358979;
8   epsilon = 1.0e-17;
9   dimension = 2;
10  dimension1 = 3;
11  maxiter = 1000;
12 Type
13  matrix = array [integer, integer] of real;
14  vector = array [integer] of real;
15  mataug = array [1..dimension, 1..dimension1] of real;
16  vectreal = array [1..dimension] of real;
17 Var
18  i, niter: integer;
19  s1, s2 : real;
20  k1, k2 : real;
21  a : mataug;
22  x, y : vectreal;
23  gamma15 : real;
24  gauche, droite: real;
25  succes : boolean;
26  sommetheorique : array [2..8] of real;
27  sommeobtenue : array [2..8] of real;
28  erreur : real;
29 Procedure VerifieGamma15(Var erreur:real);forward;
30 Procedure Resoud (Var a:dynamic matrix; Var x:dynamic vector;
31                  Var succes: boolean; n:integer);
32 Const
33  epsilon = 1.0e-20;
34  Var
35      indice : array[1..100] of integer;
36      pivot,xmax : real;
37      i,j,k,imax : integer;
38  Begin
39  If n > 100 then
40      Begin
41          succes := false;
42          WriteLn(' *** On ne peut résoudre un systeme ayant plus de 100
43                  équations ')
44      End
45  else

```

```

46     Begin
47         succes := true;
48         For i:= 1 to n do
49             indice[i] := i;
50     For i:=1 to n - 1 do
51         Begin
52             imax := 1;
53             xmax := a[indice[i], i];
54             For j := 1 + 1 to n do
55                 If a[indice[j],i] > xmax then
56                     Begin
57                         imax := j;
58                         xmax := a[indice[i],i]
59                     End;
60             if abs(xmax) > epsilon then
61                 Begin
62                     If indice[i] <> indice[imax] then
63                         Begin
64                             j := indice[i];
65                             indice[i] := indice[imax];
66                             indice[imax] := j
67                         End;
68                     For j := i+1 to n do
69                         Begin
70                             pivot := a[indice[j],i]/a[indice[i],i];
71                             For k := 1 to n+1 do
72                                 a[indice[j],k] := a[indice[j],k]
73                                     - pivot*a[indice[i],k]
74                             End
75                         End
76                     else
77                         succes := false
78                     End;
79     If (abs(a[indice[n],n]) > epsilon) and succes then
80         Begin
81             x[n] := a[indice[n],n+1]/a[indice[n],n];
82             For i:= n-1 downto 1 do
83                 Begin
84                     x[i] := a[indice[i],n+1];
85                     For j:=i+1 to n do
86                         x[i] := x[i] - a[indice[i],j]*x[j];
87                     x[i] := x[i]/a[indice[i],i]

```

```

88             End
89         End
90     else
91         succes := false;
92     If not succes then
93         Begin
94             Writeln;
95             WriteLn ( ' *** resoud *** tentative de division par
96                     zero ');
97             Writeln
98         End
99     End
100 End;
101
102 Procedure FaireA;
103     Begin
104         a[1,1]:= 1;
105         a[1,2]:= 29;
106         a[2,1]:= 2*x[1]+58*x x[2];
107         a[2,2]:= 58*x[1]+(2223+1/3)*x[2];
108         a[1,3]:= -(x[1]+29*x[2]-k1);
109         a[2,3]:= -(sqr(x[1]) + 58*x[1]*x[2] + (1111+2/3)*sqr(x[2]) - \
→k2)
110     End;
111
112 Function Gamma(i:integer): real;
113     const
114         pisur2 = 1.5707963267948966192;
115     Begin
116         Gamma := (2 * (i mod 2) - 1 ) / ((2*i-1)*pisur2)
117     End;
118
119 Procedure Trouve(Var alpha, beta, gamma15: real);
120     Var
121         s1, s2, t1, t2 : real;
122         i : integer;
123     Begin
124         s1:=0;
125         s2:= 0;
126         For i:= 72 downto 1 do
127             Begin
128                 s1 := s1 + i*sqr(Gamma(i));

```

```

129         s2 := s2 + sqr(Gamma(i))
130     End;
131     { remarque: Les sommes sont effectuées dans l'ordre inverse
132       afin de garantir une meilleure précision arithmétique }
133     t1 := 15*sqr(gamma15) + s1;
134     t2 := sqr(gamma15) + s2;
135     beta := (t1-43.5*t2)/(16254.5);
136     alpha:= s2/58 - beta*43.5
137 End;
138
139 Procedure Resume (Var erreur:real; imprime:boolean);
140     Var
141         i, j : integer;
142         temp : array[2..8] of real;
143         gammai, delta, maxdelta : real;
144     Begin
145         If imprime then
146             Begin
147                 WriteLn(' x[1]=' ,x[1]:20, ' x[2]=' ,x[2]:20);
148                 WriteLn(' y[1]=' ,y[1]:20, ' y[2]=' ,y[2]:20);
149                 WriteLn(' nombre d''iterations: ',niter:4);
150                 WriteLn(' valeur actuelle de gamma15: ',gamma15:20);
151                 WriteLn(' contrainte (3) de dickey:',x[1]+x[2]*(72-15)\
→:20);
152             End;
153         For j := 2 to 8 do
154             sommeobtenue[j]:= 0;
155         For i:= 72 downto 16 do
156             Begin
157                 gammai := (2*(i mod 2) - 1)*sqrt(abs(x[1]+x[2]*(i-15)));
158                 For J := 2 to 8 do
159                     temp[j]:= poweri(gammai,j);
160                 For j := 2 to 8 do
161                     sommeobtenue[j]:= sommeobtenue[j] + temp[j]
162             End;
163         For j := 2 to 8 do
164             sommeobtenue[j] := sommeobtenue[j] + poweri(gamma15,j);
165         For i := 14 downto 1 do
166             Begin
167                 Gammai := Gamma(i);
168                 For j := 2 to 8 do
169                     sommeobtenue[j] := sommeobtenue[j] + poweri(gammai,j)

```

```

170     End;
171     maxdelta := 0;
172     If imprime then
173         Writeln ('k':3,'obtenue':15,' theorique':20,'difference':20);
174     For J := 2 to 8 do
175         Begin
176             delta := abs(sommetheorique[j]-sommeobtenue[j]);
177             if delta > maxdelta then maxdelta := delta;
178             If imprime then
179                 WriteLn(j:3, sommeobtenue[j]:20, sommetheorique[j]:20,
180                     delta:20);
181         End;
182     erreur := maxdelta
183 End;
184
185 Procedure inltialise;
186     Var i: integer;
187     Begin
188         s1 := 0;
189         s2 := 0;
190         For i:= 14 downto 1 do
191             Begin
192                 s1 := s1 + sqr(1/(2*i-1));
193                 s2 := s2 + sqr(sqr(1/(2*i-1)))
194             End;
195         s1 := s1 * 4 * unsurpi2;
196         s2 := s2 * 16 * unsurpi4;
197         sommetheorique[2]:= 1/2;
198         sommetheorique[3]:= 1/4;
199         sommetheorique[4]:= 1/6;
200         sommetheorique[5]:= 5/48;
201         sommetheorique[6]:= 1/15;
202         sommetheorique[7]:= 61/1440;
203         sommetheorique[8]:= 17/630;
204         gamma15 := 2/((2*15-1)*pi);
205         trouve(x[1],x[2],gamma15);
206         VerifieGamma15(erreur);
207         resume (erreur, true);
208         gauche := 0.019733398;
209         droite := 0.019733501;
210     End;
211

```

```

212 Procedure VerifieGamma15 {Var erreur: real};
213   Begin
214     k1 := (1/2-s1-sqr(gamma15))/57;
215     k2 := (1/6-s2-sqr(sqr(gamma15)))/57;
216     niter := 1;
217     repeat
218       faireA;
219       resoud(a, y, succes, 2);
220       x[1] := x[1] + y[1];
221       x[2] := x[2] + y[2];
222       niter := niter + 1;
223       until (abs(y[1])<epsilon) and (abs(y[2])<epsilon) or (niter=
→maxiter)
224           or not succes;
225     resume(erreur, false)
226   End;
227
228 Begin
229   inltialise;
230   For i := 0 to 25 do
231     Begin
232       gamma15 := gauche + i*(droite-gauche)/25;
233       VerifieGamma15(erreur);
234       resume (erreur, true);
235       WriteLn; Writeln (' maximum des differences: ', erreur:20);
236       Writeln(' gamma15 =', gamma15:20, ' erreur =', erreur:20)
237     End
238 End.

```

Programmes/Pascal/consDickorg.pas

B.3. Programmes de simulation

B.3.1. Loi asymptotique avec données centrées

Le programme Simul1 a été écrit afin d'obtenir un histogramme, la moyenne et la variance, ainsi qu'un ensemble de percentiles des asymptotiques de $ns(1 - r(ds))$. Il imprime l'information pertinente et conserve aussi cette information dans un fichier pour des manipulations ultérieures.

```

1 {$L-}
2 {$a+}

```

```

3 Program Simul1(Input, Output, BinHist, CarHist);
4 {$i'random'}
5 Type
6   VectHist = Array[0..800] of Integer;
7   InfHisto = Record
8     Min, Range : Real;
9     NClass, NObs : Integer;
10    Histo: VectHist;
11    end;
12   Vecteur = Array[Integer] of Real;
13   VectPYZ = Array[0..31] of Real;
14   vectQ = Array[1..15] of Real;
15   vectS = Array[1..16] of Real;
16   vectDE = Array[16..30] of Real;
17   VectGamma = Array[1..72] of Real;
18   Vect15 = Array[1..15] of Real;
19 Var
20   NP, NY, NZ : VectPYZ;
21   NQ : VectQ;
22   ND, NE : VectDE;
23   NS : VectS;
24   Seed1, Seed2 : Integer;
25   GammaI, GammaI2 : VectGamma;
26   i, NClass, Numero, S : Integer;
27   Histo : VectHist;
28   Pourcent : Vect15;
29   Min, Range, Facteur, x, Somme, Somme2, Moyenne, Variance : Real;
30   NObs : integer;
31   MinObs, MaxObs : Real;
32   BinHist : file of InfHisto;
33   CarHist : text;
34
35 Value Histo = VectHist(801 of 0);
36
37 {
38   Les valeurs suivantes sont les constantes requises par l algo-
39   rithme M de D. Knuth figurant a la page 123 de Knuth(1981).
40   Elles ont été calculées selon les specifications données aux
41   exercices 7 et 10 de la section 3.4.1 de ce livre par le pro-
42   gramme ConsDef
43 }
44

```

45 Value NP = VectPYZ

46 (0.00000000000000, 0.84873739496400, 0.96998897933580,
 47 0.85503104286931, 0.99427926423630, 0.99515870951199,
 48 0.93274375463479, 0.92340337100404, 0.72737066777509,
 49 0.99999999999974, 0.69108437136881, 0.45407478843176,
 50 0.28664998777399, 0.17386200619118, 0.10131778025214,
 51 0.05672765967281, 0.06727492121681, 0.16051226307567,
 52 0.23553408391936, 0.28540215132048, 0.30758398387906,
 53 0.30389585363882, 0.27952114309042, 0.24148483829247,
 54 0.19705221929318, 0.15244760717495, 0.11211301538775,
 55 0.07852835391591, 0.05246425946492, 0.03346991210839,
 56 0.02040720747344, 0.08639347400094);

57

58 Value NQ = VectQ

59 (0.23564414763236, -0.20618790961620, 0.23390963599268,
 60 0.20115073017601, 0.20097296852085, 0.21442116230337,
 61 0.21659006917262, 0.27496297123376, 0.20000000000005,
 62 0.28940026469397, 0.44045607705009, 0.69771501318776,
 63 1.15033758312945, 1.97398718647932, 3.52561697686031);

64

65 Value NY = VectPYZ

66 (0.00000000000000, -0.92220385833230, -5.86421852951332,
 67 -0.57960570289337, -33.16053799044107, -39.51129922804694,
 68 -2.57370153379787, -1.61108096559386, 0.66640322092802,
 69 0.00000000000000, 0.35257403166619, -0.16635054722143,
 70 0.91963272366688, 0.35790969366035, -0.02254807718165,
 71 0.18797215666196, 0.58557454436530, 0.96175947401834,
 72 -0.06162055860565, 0.12012230323731, 1.31115630582833,
 73 0.31268667045659, 1.12240684361248, 0.53632695811937,
 74 0.75091779963083, 0.56402638740319, 0.17474610680622,
 75 0.38295588274482, -0.01107383230427, -0.39307421206485,
 76 0.19583353253941, 0.78108737809059);

77

78 Value NZ = VectPYZ

79 (0.20000000000000, 1.32220385833230, 6.66421852951333,
 80 1.37960570289337, 34.96053799044103, 41.31129922804689,
 81 2.97370153379788, 2.61108096559386, 0.73359677907198,
 82 0.00000000000000, 0.64742596833381, 0.36635054722143,
 83 0.28036727633311, 0.24209030633965, 0.22254807718165,
 84 0.21202784333804, 0.21442545563470, 0.23824052598165,
 85 0.26162055860565, 0.27987769676269, 0.28884369417166,
 86 0.28731332954341, 0.27759315638752, 0.26367304188063,

```

87         0.24908220036917,  0.23597361259681,  0.22525389319378,
88         0.21704411725518,  0.21107383230427,  0.20692578793515,
89         0.20416646746059,  0.21891262190941);
90
91 Value NS = VectS
92         (0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2,
93         2.4, 2.6, 2.8, 3.0);
94
95 Value ND = VectDE
96         (0.50503350066879, 0.772956883181701, 0.87642431729707,
97         0.93921124285770, 0.98608681560905, 0.99515450131766,
98         0.98674801424353, 0.97921135862260, 0.97227391617338,
99         0.96575234004515, 0.95953097292161, 0.95353409608020,
100        0.94771026493741, 0.94202340198953, 0.93644752494987);
101
102 Value NE = VectDE
103        (25.00000000000000,12.50000000000000, 8.333333333333331,
104        6.250000000000000, 5.000000000000000, 4.06377310692091,
105        3.36779614093342, 2.85829591351012, 2.46945536484918,
106        2.16316966400260, 1.91584991123493, 1.71211186543338,
107        1.54149408253681, 1.39663465926647, 1.27220242787081);
108
109 {$L+}
110 Function Normal : Real;
111     Var
112         U V, Temp, X, F: Real;
113         Signe : Boolean;
114         j : Integer;
115     Begin
116         U := Random;
117         Signe := U >= 0.5;
118         U:= 2*U - Trunc(2*U);
119         j:= Trunc(32*U);
120         F:= 32*U - j;
121         If(F < NP[j]) and (j > 15)
122             then Begin
123                 if j < 31
124                     then Repeat
125                         U := Random;
126                         V := Random;
127                         if U > V
128                             then Begin

```

```

129             Temp := U;
130             U := V;
131             V := Temp
132             End;
133             X := NS[j - 15] + U / 5
134             until (V <= ND[j]) or
135                 (V <= NE[j]*Exp(Sqr(NS[j - 14]) - Sqr(X)) / 2) ↘
→-1)
136             else Repeat
137                 U := Random;
138                 V := Random;
139                 X := Sqrt(9 - 2*Ln(V))
140                 until U*X < 3
141             End
142             else
143                 |f F >= NP[j] then X := NY[j] + F*NZ[j]
144                 else X := NS[j] + F*NQ[j];
145                 If Signe then x := -x;
146                 Normal := X
147             End;
148
149 Function QAsymp(S: Integer): Real;
150     Const
151         Rac2 = 1.414213562;
152     Var T, Tr, T2, W, G : Real;
153         X : Real;
154         i, j : Integer;
155     Begin
156         T:= 0 ;
157         T2 := 0;
158         W := 0;
159         G := 0;
160         For i:= 1 to S do
161             Begin
162                 Tr := 0;
163                 For j := 72 downto 1 do
164                     Begin
165                         X := Normal;
166                         Tr := Tr + GammaI[j]*X;
167                         W:= W + GammaI2[j]*X;
168                         G:= G + GammaI2[j]*Sqr(X)
169                     End;

```

```

170         T := T + Tr;
171         T2 := T2 + Sqr(Tr);
172     End;
173     T := T*Rac2;
174     T2 := 2*T2;
175     W:= W*Rac2;
176     QAsymp := (S*(T2 + S) / 2 + Sqr(W) - T*W) / (G - Sqr(W) / S)
177 End;
178
179 Procedure SetGamma(Var GammaI, GammaI2: VectGamma);
180 Const
181     Gamma 15 = 1.9733468e-02;
182 Var i : Integer;
183 Function Gamma1(i: Integer): Real;
184 Const
185     PiSur2 = 1.5707963267948966192;
186 Begin
187     Gamma1 := (2*(i mod 2) -1) / ((2*i - 1)*PiSur2)
188 End;
189 Function Gamma2(i: Integer): Real;
190 Const
191     Alpha = 2.41476620886e-04;
192     Beta = -4.18597781401e-06;
193 Begin
194     Gamma2 := (2*(i mod 2) -1)*Sqrt(Abs(Alpha + Beta*(i - 15)))
195 End;
196 Begin
197     For i := 1 to 14 do
198         Begin
199             GammaI[i] := Gamma1(i);
200             GammaI2[i] := Sqr(GammaI[i])
201         End;
202     GammaI[15]:= Gamma15;
203     GammaI2[15] := Sqr(Gamma15);
204     For i := 16 to 72 do
205         Begin
206             GammaI[i]:= Gamma2(i);
207             GammaI2[i] := Sqr(GammaI[i])
208         End
209     End;
210 Procedure Cumul(Var Histo: VectHist);
211     Var

```

```

212     i : Integer;
213     Begin
214         For i:= 1 to NClass do Histo[i] := Histo[i] + Histo[i - 1]
215     End;
216
217 Procedure DetPour(Pourcent: Real; Var Numero: Integer; Var Percentile: \
→Real;
218                 Var Histo: VectHist);
219     Var Nombre, Temp: Real;
220     Begin
221         Percentile:= 0.0;
222         Nombre := Pourcent*Histo[NClass];
223         While Nombre > Histo[Numero] do
224             Numero := Numero + 1;
225         If(Numero <> 0) and (Numero <> NClass)
226             then Percentile:= Min + (Numero + (Nombre - Histo[Numero - \
→1])
227                                     / (Histo[Numero] - Histo[Numero - 1]))
228                                     *Range / NClass
229             else WriteLn('*** Attention: Classe limite*** ', Pourcent\
→);
230     End;
231 Procedure PerctPr(Var Histo: VectHist);
232     Const
233         MxLignes = 60;
234     var
235         NLignes, i : Integer;
236         Pourcent, Percentile : Real;
237         Numero : integer;
238     Procedure Entete;
239     Begin
240         WriteLn; WriteLn;
241         WriteLn('1', 'Pourcent': 10, 'Percentile': 15);
242         WriteLn; WriteLn;
243         NLignes := 5
244     End;
245     Begin
246         Entete;
247         Numero := 1;
248         For i := 1 to 99 do
249             Begin
250                 Pourcent := i / 100;

```

```

251         DetPour(Pourcent, Numero, Percentile, Histo);
252         WriteLn(Pourcent: 8: 3, Percentile: 15: 3);
253         NLignes := Nlignes + 1;
254         If NLignes = MxLignes then Entete;
255         If i = 2
256             then Begin
257                 DetPour(0.025, Numero, Percentile, Histo);
258                 WriteLn('0.025': 8, Percentile: 15: 3);
259                 Nlignes := Nlignes + 1;
260                 If NLignes = MxLignes then Entete
261                     End;
262         If i = 97
263             then Begin
264                 DetPour(0.975, Numero, Percentile, Histo);
265                 WriteLn('0.975': 8, Percentile: 15: 3);
266                 Nlignes := NLignes + 1;
267                 If NLignes = MxLignes then Entete
268                     End
269             End
270     End;
271
272 Procedure Resume(NClass, NObs: Integer; Min, Range: Real;
273                 Var Histo: VectHist);
274     Var Bloc, Reste, i, j, MClass : Integer;
275         Delta, Sup, Inf : Real;
276     Begin
277         Rewrite(CarHist);
278         Bloc := Trunc(NClass / 50);
279         Reste := NClass - 50*Bloc;
280         MClass := Trunc(NClass / Bloc);
281         Delta := Range / NClass;
282         Sup := Min + Delta;
283         WriteLn(CarHist, Min: 9: 4, Sup: 14: 4, Histo[0]: 14, Histo[0]: ↘
→14);
284         If Reste = 0 then MClass := MClass - 1;
285         For i := 2 to MClass do
286             Begin
287                 j := Bloc*(i - 1);
288                 Inf:= Min + (1 + j - Bloc)*Delta;
289                 Sup := Min + (1 + j)*Delta;
290                 WriteLn(CarHist, Inf: 9: 4, Sup: 14: 4, Histo[j]: 14,
291                     Histo[j] - Histo[j - Bloc]: 14)

```

```

292         End;
293         WriteLn(CarHist, Sup: 9: 4, Min + Range: 14: 4, Histo[Bloc*(\
→MClass - 1]): 14,
294             NObs - Histo[Bloc*(MClass - 1]): 14)
295     End;
296
297 Procedure HistPr;
298     Const
299         MxLign = 55;
300     Var
301         NLigne, i: Integer;
302         Delta, Inf, Sup: Real;
303         Cumulatif: Real;
304     Begin
305         WriteLn('1 Borne', 'Borne': 14);
306         WriteLn(' inferieure superieure Frequence');
307         WriteLn; WriteLn;
308         Delta := Range / NClass;
309         Sup := Min + Delta;
310         Cumulatif:= Histo[0];
311         WriteLn(' Moins de ', Sup: 8: 4, Histo[0]: 12, Cumulatif / NObs: 12: \
→4);
312         NLigne := 5;
313         For i := 2 to NClass do
314             Begin
315                 Cumulatif := Cumulatif + Histo[i - 1];
316                 Inf := Min + (i - 1)*Delta;
317                 Sup := Min + i*Delta;
318                 WriteLn(Inf: 9: 4, Sup: 14: 4, Histo[i - 1]: 12, Cumulatif / \
→NObs: 12: 4);
319                 Nligne := Nligne + 1;
320                 If NLigne = MxLign
321                     then Begin
322                         WriteLn('1 Borne', 'Borne': 14);
323                         WriteLn(' inferieure superieure Frequence',
324                             'Cumulatif': 12);
325                         WriteLn; WriteLn;
326                         Nligne := 4
327                     End
328                 End;
329         Cumulatif:= Cumulatif + Histo[NClass];
330         WriteLn (Sup:9:4, ' et plus ', Histo[NClass]:8,Cumulatif/NObs:12:4)

```

```

331 End;
332
333 { P R O G R A M M E   P R I N C I P A L }
334 Begin
335   ReadLn(S, Min, Range, NObs, NClass);
336   ReadLn(Seed1, Seed2);
337   SetRandom(Seed1, Seed2);
338   SetGamma(GammaI, GammaI2);
339   X := OAsymp(S);
340   MinObs := X;
341   MaxObs := X;
342   Facteur := NClass / Range;
343   Somme := 0.0;
344   Somme2 := 0.0;
345   For i := 1 to NObs do
346     Begin
347       X := OAsymp(S);
348       Numero := Trunc((X - Min)*Facteur);
349       If Numero < 1
350         then Numero := 0
351         else if Numero > NClass then Numero := NClass;
352       Histo[Numero] := Histo[Numero] + 1;
353       If X < MinObs
354         then MinObs := X
355         else if X > MaxObs then MaxObs := X;
356       Somme := Somme + X;
357       Somme2 := Somme2 + Sqr(X)
358     End;
359   { HistPr; }
360   Cumul(Histo);
361   PerctPr(Histo);
362   WriteLn('1', 'Resume de la simulation');
363   WriteLn; WriteLn;
364   WriteLn(' Nombre d''observations:', NObs);
365   WriteLn(' Periode (S): ', S: 2);
366   WriteLn(' Minimum observe: ', MinObs);
367   WriteLn(' Maximum observe: ', MaxObs);
368   WriteLn(' Nombre de classes de l''histogramme: ', NClass:6);
369   GetRandom(Seed1, Seed2);
370   WriteLn(' Seed1 = ', Seed1);
371   WriteLn(' Seed2 = ', Seed2);
372   Moyenne := Somme / NObs;

```

```

373   Variance := Somme2 / (NObs - 1) - NObs*Sqr(Moyenne) / (NObs - 1);
374   WriteLn(' Moyenne de la variable: ', Moyenne);
375   WriteLn(' Variance de la variable: ', Variance);
376   Resume(NClass, NObs, Min, Range, Histo);
377   Rewrite(BinHist);
378   BinHist^.Histo := Histo;
379   BinHist^.Min := Min;
380   BinHist^.Range := Range;
381   BinHist^.NClass := NClass;
382   BinHist^.NObs := NObs;
383   Put(BinHist)
384 End.

```

Programmes/Pascal/Simul1.tex

B.3.2. Loi asymptotique avec données non centrées

Pour obtenir un programme simulant la loi asymptotique quand les données ne sont pas centrées, il suffit de remplacer la fonction QAsymp du programme Simul1 par la fonction UAsymp suivante :

```

1 Function UAsymp(S: Integer): Real;
2 Const
3   Rac2 = 1.414213562;
4 Var
5   Tr, T2, G, X: Real;
6   i, j: Integer;
7 Begin
8   T2 := 0;
9   G := 0;
10  For i := 1 to S do
11  Begin
12    Tr := 0;
13    For j := 72 downto 1 do
14    Begin
15      X := Normal;
16      Tr := Tr + GammaI[j] * X;
17      G := G + GammaI2[j] * Sqr(X)
18    End;
19    T2 := T2 + Sqr(Tr);
20  End;
21  T2 := 2 * T2;

```

```

22   UAsymp := (S * (T2 + S) / 2) / G;
23 End;

```

Programmes/Pascal/UAsymp.tex

B.3.3. Loi exacte avec données centrées

Pour obtenir un programme simulant les lois exactes quand les données sont centrées, il suffit de faire quelques modifications mineures au programme Simul1. L'essentiel de ces modifications consiste à remplacer la fonction QAsymp par une procédure qui simule une série de la taille voulue selon le modèle de la marche aléatoire saisonnière et qui calculera les coefficients d'autocorrélation voulus. Il s'agit de la procédure Gencorr.

```

1  Procedure GenCorr(Var Coeff: Vect5; Longueur, Nombre, S: Integer);
2  Var
3     i, K, d: Integer;
4     Moyenne, Variance, Temp: Real;
5     Serie: Array [-11..204] of Real;
6  Begin
7     { Quelques initialisations }
8     Moyenne := 0.0;
9     For i := 0 downto -S + 1 do
10        Serie[i] := 0;
11
12     { On génère la série en additionnant les valeurs pour le calcul de la
→ moyenne }
13     For i := 1 to Longueur do
14     Begin
15        Serie[i] := Serie[i - S] + Normal;
16        Moyenne := Moyenne + Serie[i];
17     End;
18     Moyenne := Moyenne / Longueur;
19
20     { On centre la série et on calcule la somme des carrés des écarts }
21     Variance := 0.0;
22     For i := 1 to Longueur do
23     Begin
24        Serie[i] := Serie[i] - Moyenne;
25        Variance := Variance + Sqr(Serie[i]);
26     End;
27

```

```

28   { On calcule alors les coefficients de corrélation qui nous inté\
→ressent }
29   For d := 1 to Nombre do
30   Begin
31     Temp := 0.0;
32     k := d * S;
33     For i := 1 to Longueur - K do
34       Temp := Temp + Serie[i] * Serie[i + K];
35     Coeff[d] := (Variance - Temp) / Variance;
36   End
37 End;

```

Programmes/Pascal/Gencorr.tex

B.3.4. Loi exacte avec données non centrées

Il s'agit de faire quelques modifications mineures à la procédure Gencorr. Elle devient :

```

1 Procedure GenCorr(Var Coeff: Vect5; Longueur, Nombre, S: Integer);
2 Var
3   i, K, d: Integer;
4   Variance, Temp: Real;
5   Serie: Array [-11..204] of Real;
6 Begin
7   { Quelques initialisations }
8   For i := 0 downto -s + 1 do
9     Serie[i] := 0;
10
11   { On génère la série selon une marche aléatoire saisonnière }
12   For i := 1 to Longueur do
13     Serie[i] := Serie[i - s] + Normal;
14
15   Variance := 0.0;
16
17   { Calcul de la variance }
18   For i := 1 to Longueur do
19     Variance := Variance + Sqr(Serie[i]);
20
21   { Calcul des coefficients de corrélation }
22   For d := 1 to Nombre do
23   Begin
24     Temp := 0.0;

```

```
25     K := d * S;  
26     For i := 1 to Longueur - K do  
27         Temp := Temp + Serie[i] * Serie[i + K];  
28     Coeff[d] := (Variance - Temp) / Variance;  
29 End;  
30 End;
```

Programmes/Pascal/Gencorr2.tex

Annexe C

Autre approche pour la simulation des lois asymptotiques

C.1. Description du logiciel

Comme établi au théorème 4.1 du chapitre 4, pour les distributions asymptotiques, nous considérons une suite de variables aléatoires indépendantes doublement indicées

$$\{\varepsilon_{i,r}\}_{i=1}^{\infty}, \quad r = 1, \dots, s,$$

avec $\varepsilon_{i,r} \sim \mathcal{N}(0,1)$ i.i.d. À partir de cette suite, nous définissons :

$$T_r = \sqrt{2} \sum_{i=1}^{\infty} \gamma_i \varepsilon_{i,r}, \quad W_r = \sqrt{2} \sum_{i=1}^{\infty} \gamma_i^2 \varepsilon_{i,r}, \quad \Gamma_r = \sqrt{2} \sum_{i=1}^{\infty} \gamma_i^2 \varepsilon_{i,r}^2.$$

Les coefficients γ_i sont donnés par :

$$\gamma_i = \frac{2(-1)^{i+1}}{\pi(2i-1)}.$$

Nous détaillons ci-après les moments et les distributions de ces variables, puis nous proposons une méthode directe pour simuler les distributions asymptotiques visées.

C.2. Moments et distributions

Pour un indice fixe r , les composantes du vecteur $(T_r, W_r, \Gamma_r)'$ sont dépendantes, car elles se construisent à partir des mêmes $\varepsilon_{i,r}$.

C.2.1. La variable T_r

Par linéarité de l'espérance et puisque $E[\varepsilon_{i,r}] = 0$, on a :

$$E[T_r] = \sqrt{2} \sum_{i=1}^{\infty} \gamma_i E[\varepsilon_{i,r}] = 0.$$

L'interversion somme/espérance est justifiée par la convergence en L^2 (et donc en L^1) de la série, conséquence de la convergence de

$$\sum_{i=1}^{\infty} \gamma_i^2 = \frac{4}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{(2i-1)^2},$$

puisque

$$\sum_{i=1}^{\infty} \frac{1}{(2i-1)^2} = \frac{\pi^2}{8}.$$

La variance s'obtient par :

$$\text{Var}(T_r) = 2 \sum_{i=1}^{\infty} \gamma_i^2 \text{Var}(\varepsilon_{i,r}) = 2 \sum_{i=1}^{\infty} \gamma_i^2.$$

En substituant l'expression de γ_i :

$$\sum_{i=1}^{\infty} \gamma_i^2 = \frac{4}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{(2i-1)^2} = \frac{4}{\pi^2} \times \frac{\pi^2}{8} = \frac{1}{2},$$

d'où

$$\text{Var}(T_r) = 2 \times \frac{1}{2} = 1.$$

Ainsi, $T_r \sim \mathcal{N}(0,1)$.

C.2.2. La variable W_r

De même,

$$E[W_r] = \sqrt{2} \sum_{i=1}^{\infty} \gamma_i^2 E[\varepsilon_{i,r}] = 0.$$

La variance s'écrit :

$$\text{Var}(W_r) = 2 \sum_{i=1}^{\infty} \gamma_i^4,$$

et en utilisant

$$\sum_{i=1}^{\infty} \gamma_i^4 = \frac{16}{\pi^4} \sum_{i=1}^{\infty} \frac{1}{(2i-1)^4},$$

avec

$$\sum_{i=1}^{\infty} \frac{1}{(2i-1)^4} = \frac{\pi^4}{96},$$

on obtient :

$$\sum_{i=1}^{\infty} \gamma_i^4 = \frac{16}{\pi^4} \times \frac{\pi^4}{96} = \frac{1}{6},$$

ce qui donne :

$$\text{Var}(W_r) = 2 \times \frac{1}{6} = \frac{1}{3}.$$

Ainsi, $W_r \sim \mathcal{N}(0, 1/3)$.

C.2.2.1. Covariance et corrélation entre T_r et W_r . Étant donné que T_r et W_r sont construites à partir des mêmes $\varepsilon_{i,r}$ (avec $E[\varepsilon_{i,r}\varepsilon_{j,r}] = \delta_{ij}$), on calcule :

$$\text{Cov}(W_r, T_r) = 2 \sum_{i=1}^{\infty} \gamma_i^3.$$

D'après le tableau 6.3, on a :

$$\sum_{i=1}^{\infty} \gamma_i^3 = \frac{1}{4},$$

et donc :

$$\text{Cov}(W_r, T_r) = \frac{1}{2}.$$

La corrélation est alors :

$$\rho(W_r, T_r) = \frac{\text{Cov}(W_r, T_r)}{\sqrt{\text{Var}(W_r) \text{Var}(T_r)}} = \frac{\frac{1}{2}}{\sqrt{\frac{1}{3} \times 1}} = \frac{\sqrt{3}}{2}.$$

Le vecteur $(T_r, W_r)'$ suit ainsi une loi binormale de paramètres :

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}.$$

C.2.3. La variable Γ_r

Puisque $\varepsilon_{i,r} \sim \mathcal{N}(0,1)$, on a $E[\varepsilon_{i,r}^2] = 1$ et $\text{Var}(\varepsilon_{i,r}^2) = 2$. Ainsi,

$$E[\Gamma_r] = \sqrt{2} \sum_{i=1}^{\infty} \gamma_i^2 = \sqrt{2} \times \frac{1}{2} = \frac{\sqrt{2}}{2},$$

et

$$\text{Var}(\Gamma_r) = 2 \sum_{i=1}^{\infty} \gamma_i^4 \text{Var}(\varepsilon_{i,r}^2) = 4 \sum_{i=1}^{\infty} \gamma_i^4 = 4 \times \frac{1}{6} = \frac{2}{3}.$$

On peut montrer que $\text{Cov}(T_r, \Gamma_r) = \text{Cov}(W_r, \Gamma_r) = 0$. La distribution exacte de Γ_r étant plus complexe, nous ne l'aborderons pas ici.

C.3. Simulation du triplet $(T_r, W_r, \Gamma_r)'$

L'approximation numérique repose sur la convergence de la série :

$$\sum_{i=1}^{\infty} \gamma_i^2 = \sum_{i=1}^{\infty} \left(\frac{2}{\pi(2i-1)} \right)^2.$$

Cette série de Dirichlet converge, ce qui garantit qu'une troncature après un nombre fini de termes (par exemple $N = 100$) fournit une approximation adéquate. Une étude numérique a montré que la somme tronquée avec $N = 100$ termes capture environ 99.8% de la variance de T_r et 99.9999% pour W_r et Γ_r , assurant ainsi un compromis optimal entre précision et complexité de calcul.

Le script C.1 présente le code R ayant permis de produire les graphiques des densités asymptotiques de la seconde édition.

Nous précisons que cette annexe résulte d'une « *discussion* » avec ChatGPT 4o. Le script C.1 a été intégralement généré par cette intelligence artificielle.

```

1 # Chargement des bibliothèques nécessaires
2 library(reshape2)
3
4 # Définition du séparateur décimal
5 options(OutDec = ",")
6
7 # Récupération des fichiers CSV
8 csv_files <- list.files(pattern = "data_ns_1_minus_r.*_s=\\d+_n=\\d+\\.\\.",
9   →csv")
10
11 # Boucle pour traiter chaque fichier
12 for (data_file in csv_files) {
13   # Lecture du fichier CSV
14   data <- read.csv(data_file)
15
16   # Vérification du nombre de colonnes
17   num_cols <- ncol(data)
18   if (num_cols < 1) {
19     warning(paste("Fichier ignoré (trop peu de colonnes) :", data_file))
20     next
21   }
22

```

```

23 # Extraction de s et n depuis le nom du fichier
24 matches <- regmatches(data_file, gregexpr("s=\\d+|n=\\d+", data_file))
25 params <- as.numeric(gsub("[^0-9]", "", unlist(matches))) # Conversion ↘
  → en nombres
26 s_value <- params[1]
27 n_value <- params[2]
28 N_value <- s_value * n_value # Calcul de N
29
30 # Création du titre sous la forme "s=4, N=108"
31 title_text <- paste0("s=", s_value, ", N=", format(N_value, decimal.↘
  →mark = ", "))
32
33 # Détermination du type de fichier et ajustement de l'étiquette de l'↘
  →axe des X
34 if (grepl("r_prime", data_file)) {
35   x_label <- expression(N * (1 - r * "' ' (d,s)) # Expression mathé↘
  →matique pour r'
36   output_file <- gsub("data_", "densite_", data_file)
37   output_file <- gsub("\\.csv", ".pdf", output_file)
38 } else {
39   x_label <- expression(N * (1 - r(d,s))) # Expression mathématique ↘
  →pour r
40   output_file <- gsub("data_", "densite_", data_file)
41   output_file <- gsub("\\.csv", ".pdf", output_file)
42 }
43
44 # Définition du fichier PDF avec la taille demandée
45 pdf(output_file, width = 8, height = 5)
46
47 # Estimation de densité pour chaque colonne
48 dens_list <- list()
49 for (i in 1:min(3, num_cols)) {
50   dens_list[[i]] <- density(data[[i]], adjust = 1)
51 }
52
53 # Définition des limites des axes
54 x_min <- min(unlist(lapply(dens_list, function(d) d$x)))
55 x_max <- max(unlist(lapply(dens_list, function(d) d$x)))
56 y_min <- 0
57 y_max <- max(unlist(lapply(dens_list, function(d) d$y))) * 1.1
58
59 # Forcer l'affichage de l'axe des y si x_min > 0

```

```

60  if (x_min > 0) {
61    x_min <- 0
62  }
63
64  # Générer des repères "propres" sur l'axe des x avec virgule comme sé\
→parateur
65  x_ticks <- pretty(c(x_min, x_max), n = 8)
66  x_ticks_labels <- format(x_ticks, decimal.mark = ",")
67
68  # Création du graphique vide avec axes personnalisés
69  plot(dens_list[[1]]$x, dens_list[[1]]$y, type = "n", xlim = c(x_min, x_\
→max), ylim = c(y_min, y_max),
70        main = title_text, xlab = x_label, ylab = "Densité", axes = FALSE)
71
72  # Ajout des axes avec des repères plus propres et la virgule comme sé\
→parateur
73  axis(1, at = x_ticks, labels = x_ticks_labels, pos = 0)
74  axis(2, pos = 0, las = 1)
75
76  # Définition des styles de ligne
77  line_styles <- c("solid", "dashed", "dotted")
78
79  # Définition de la légende avec d en italique et "d="
80  d_values <- c(1, 2, 3) # À adapter si les valeurs réelles sont diff\
→rentes
81  legend_labels <- lapply(d_values, function(d) bquote(italic(d) == .(d))\
→)
82
83  # Ajout des courbes de densité
84  for (i in 1:length(dens_list)) {
85    lines(dens_list[[i]]$x, dens_list[[i]]$y, col = "black", lwd = 2, lty\
→ = line_styles[i])
86  }
87
88  # Ajout d'une légende avec "d=" et d en italique
89  legend("topright", title="Délai", legend = legend_labels,
90        lty = line_styles[1:length(dens_list)], col = "black", lwd = 2, \
→bty="n")
91
92  # Fermeture du fichier PDF
93  dev.off()
94

```

```
95  cat("Graphique généré et enregistré dans", output_file, "\n")  
96 }
```

Script C.1 – Production des graphiques des distributions asymptotiques de la seconde édition

Remerciements

Je tiens à exprimer ma profonde gratitude envers mon directeur de recherche, le professeur Roch Roy, pour sa patience, sa disponibilité et la grande liberté qu'il m'a accordée dans mon travail.

Mes remerciements vont également au Conseil National de la Recherche du Canada, qui m'a généreusement octroyé une bourse d'études.

Un remerciement tout particulier à ma compagne, Louise Laforest, pour son soutien précieux et quotidien durant toutes ces années, ainsi que pour son engagement dans la révision et la mise en page de ce travail.

Ajout lors de la réédition de 2025

Lors de la rédaction de cette thèse en 1986, les normes et directives de présentation recommandaient des formules standard pour exprimer la gratitude d'un étudiant envers son directeur de recherche. Aujourd'hui, à l'occasion de cette réédition, je souhaite exprimer plus explicitement toute ma reconnaissance envers Roch Roy pour son engagement et son influence déterminante dans mon parcours.

J'ai énormément appris grâce à lui. Dès mes premiers pas en mathématiques, il m'a initié à la statistique, et son cours a rapidement orienté mon choix de carrière. Le jour de ma soutenance, j'ai pleinement mesuré l'immense dette intellectuelle et humaine que j'avais envers lui. Passionné par la recherche, il avait également un profond souci de pédagogie : toujours bienveillant, il prenait le temps de répondre aux questions et savait éveiller la curiosité. Par ailleurs, dans ses examens, il n'hésitait pas à proposer des questions bonus sous forme de problèmes ouverts, une manière subtile d'observer nos réactions face à une situation inédite, jamais abordée en cours.

Roch Roy a été un véritable modèle pour moi. Dans la préface, j'évoque certains éléments qui ont guidé mon attitude tout au long de ma carrière. Il reste, sans conteste, une source d'inspiration.

212

Mille mercis !

Grenoble, 5 mars 2025

Liste des ouvrages cités

- ANDERSON, O. 1979, « Formulse for the expected values of sampled variance and covariances from series generated by general autoregressive integrated moving average processes of order (p, d, q) », *Sankhya, B*, vol. 41, p. 177–195.
- ANDERSON, O. et J. G. DE GOOIJER. 1983, « Autoregressive moving average processes of order (p, d, q) , $d=0$ or 1 », *Sankhya, B*, vol. 45, p. 249–256.
- ANDERSON, T. 1971, *The Statistical Analysis of Time Series*, John Wiley & Sons, New York.
- BISHOP, Y., S. FIENBERG et P. HOLLAND. 1975, *Discrete Multivariate Analysis, Theory and Practice*, Massachusetts Institute of Technology Press, Cambridge, Massachusetts.
- BOX, G. et G. JENKINS. 1976, *Time Series Analysis : Forecasting and Control*, revised edition éd., Holden-Day, San Francisco.
- BRILLINGER, D. 1981, *Time Series : Data Analysis and Theory*, augmented edition éd., Holden-Day, San Francisco.
- BURDEN, R. L., J. D. FAIRES et A. C. REYNOLDS. 1981, *Numerical Analysis*, 2^e éd., Prindle, Weber & Smith, Boston, Massachusetts.
- DAVIES, N., J. PETRUCCELLI et M. PATE. 1984, *Exact Moments of the Sample Cross Correlations of Multivariate Autoregressive Moving Average Time Series*, rapport technique, Worcester Polytechnic Institute.
- DE GOOIJER, J. 1980, « Exact moments of the sample autocorrelations from series generated by general arima processes of order (p,d,q) , $d=0$ or 1 », *Journal of Econometrics*, vol. 14, p. 9.
- DE GOOIJER, J. 1981, « An investigation of the moments of the sample autocovariances and autocorrelations for general arima processes », *Journal of Statistical Computation and Simulation*, vol. 12, p. 175–192.
- DICKEY, D. 1976, *Estimation and Hypothesis Testing in Nonstationary Time Series*, thèse de doctorat, Iowa State University.

- DICKEY, D. et W. FULLER. 1979, « Distribution of the estimators for autoregressive time series with a unit root », *Journal of the American Statistical Association*, vol. 74, p. 427–431.
- DICKEY, D. et W. FULLER. 1981, « Likelihood ratio statistics for autoregressive time series with unit root », *Econometrica*, vol. 49, p. 1057–1072.
- DICKEY, D., D. HASZA et W. FULLER. 1984, « Testing for unit roots in seasonal time series », *Journal of the American Statistical Association*, vol. 79, p. 355–367.
- EVANS, G. B. A. et N. E. SAVIN. 1981, « The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model », *The Annals of Statistics*, vol. 9, p. 1114–1118.
- FERRY, L. 2025, *IA : grand remplacement ou complémentarité ?*, Éditions de l'Observatoire, ISBN 979-10-329-1475-5.
- FULLER, W. 1976, *Introduction to Statistical Time Series*, John Wiley & Sons, New York.
- GODOLPHIN, E. J. 1977, « On the autocorrelation structure for seasonal moving average models and its applications for the cramer-wold factorization », *Journal of Applied Probability*, vol. 14, p. 785–794.
- GONG, G. 1983, « Letting macsyma help », dans *Computer Science and Statistics : Proceedings of the Fifteenth Symposium on the Interface*, sous la direction de J. Gentle, Elsevier, p. 237–244.
- GRAHAM, A. 1981, *Matrix Calculus with Applications*, Halsted Press, John Wiley & Sons, New York.
- HASZA, D. 1980, « The asymptotic distribution of the sample autocorrelations for an integrated arma process », *Journal of the American Statistical Association*, vol. 75, p. 349–352.
- HASZA, D. et W. FULLER. 1979, « Estimation for autoregressive processes with unit roots », *The Annals of Statistics*, vol. 7, p. 1106–1120.
- HASZA, D. et W. FULLER. 1982, « Testing for nonstationary parameter specifications in seasonal time series models », *Annals of Statistics*, vol. 10, p. 1209–1216.
- JOLLEY, L. 1961, *Summation of Series*, 2^e éd., Dover, New York.
- JORDÁN, K. 1965, *Calculus of finite differences*, vol. 33, American Mathematical Soc.
- KENDALL, M. 1954, « Note on bias in the estimation of autocorrelation », *Biometrika*, vol. 41, p. 403–404.

- KNUTH, D. 1981, *The Art of Computer Programming, Vol. 2*, 2^e éd., Addison-Wesley, Reading, Massachusetts.
- LATOUR, A. 1978, *Problèmes d'estimation dans un processus de ramification*, mémoire de maîtrise, Université du Québec à Montréal, <https://membres-ljk.imag.fr/Alain.Latour/Pubpdf/master.pdf>.
- LATOUR, A. 1986, « The polar method for the normal distribution », *Byte*, vol. 11, p. 131–132.
- LATOUR, A. 2025, *Problèmes d'estimation dans un processus de ramification (édition révisée)*, mémoire de maîtrise, Université du Québec à Montréal, <https://membres-ljk.imag.fr/Alain.Latour/Pubpdf/master.pdf>. Édition révisée de l'édition originale de 1978.
- MARRIOTT, F. et J. POPE. 1954, « Bias in the estimation of autocorrelation », *Biometrika*, vol. 41, p. 390–402.
- MARSAGLIA, G. 1961, « Expressing a random variable in terms of uniform random variables », *The Annals of Mathematical Statistics*, vol. 32, p. 894–899.
- MÉLARD, G. 1985, *Analyse de données chronologiques*, Les presses de l'Université de Montréal.
- OPENAI. 2024, « Définition de la recherche reproductible », Consultation via une conversation interactive, <https://chat.openai.com/>. Fourni par le modèle IA ChatGPT en décembre 2024.
- PEÑA, D. 1984, « The autocorrelation function of seasonal arma models », *Journal of Time Series Analysis*, vol. 5, p. 269–272.
- PRIESTLEY, M. 1981, *Spectral Analysis and Time Series*, vol. 2, Academic Press, New York.
- RAO, M. 1978, « Asymptotic distribution of an estimator of the boundary parameter of an unstable process », *The Annals of Statistics*, vol. 6, p. 185–190.
- ROY, R. 1977, « On the asymptotic behaviour of the sample autocovariance function for an integrated moving average process », *Biometrika*, vol. 64, p. 419–421.
- ROY, R. 1980, *Analyse et prévision de séries chronologiques*, Notes de cours, Département d'informatique et de recherche opérationnelle, Université de Montréal.
- ROY, R. et P. LEFRANÇOIS. 1981, « Sur les moments des autocovariances échantillonnées d'un processus arima », *Les Annales des Sciences Mathématiques du Québec*, vol. V, p. 87–95.
- RUTHERFORD, D. 1946, « Some continuant determinants arising in physics and chemistry », *Proceedings of the Royal Society of Edinburgh, Section A*, vol. 62, p.

229–236.

- SAID, S. et D. DICKEY. 1981, « Testing arima(p, 1, q) versus arma(p+1, q) », dans *Proceedings of the Business, Economic and Statistical Section*, American Statistical Association, p. 318–322.
- SAID, S. et D. DICKEY. 1984, « Testing for unit roots in autoregressive-moving average models of unknown order », *Biometrika*, vol. 71, p. 599–607.
- SAMARANAYAKE, V. et D. HASZA. 1982, « Asymptotic properties of unstable multiple time series », dans *Proceedings of the Business, Economic and Statistical Section*, American Statistical Association.
- SEARLE, S. 1982, *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York.
- SERFLING, R. 1980, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.
- TIAO, G. et G. BOX. 1981, « Modeling multiple time series with applications », *Journal of the American Statistical Association*, vol. 76, p. 802–816.
- WALKER, A. 1977, « An efficient method for generating discrete random variables with general distributions », *ACM Transactions on Mathematical Software*, vol. 3, p. 253–256.
- WICHERN, D. 1973, « The behavior of the sample autocorrelation function for an integrated moving average process », *Biometrika*, vol. 60, p. 235–239.
- YAJIMA, H. 1985, « Asymptotic properties of the sample autocorrelations and partial autocorrelations of multiplicative ARIMA process », *Journal of Time Series Analysis*, vol. 6, p. 187–201.