

**République Algérienne Démocratique et
Populaire**
**Ministère de l'Enseignement Supérieur et de
la Recherche Scientifique**
Université Mentouri Constantine
Faculté des Sciences
Département de Mathématiques

n° d'ordre : /T.E/2005
n° d'ordre : /MAT/2005

**Thèse présentée pour l'obtention du
diplôme de Doctorat d'Etat
en Mathématique**

Intitulée :
**"Statistique des processus linéaires :
Tests et Estimation"**
par

Meghlaoui Dakhmouche

Spécialité : Statistique théorique

Thèse soutenue le / /2005 devant la Commission d'Examen

| | | |
|----------------------|----------------------|------------|
| DENCHE Mohamed | Professeur | Président |
| BOUSHABA Mahmoud | Maitre de conférence | Rapporteur |
| RAHMANI Fouad Lazhar | Maitre de conférence | Examineur |
| KHALDI Khaled | Maitre de conférence | Examineur |
| BOUTABIA Hacène | Maitre de conférence | Examineur |
| YAROU Mostepha Fateh | Maitre de conférence | Examineur |

ABSTRACT :

In this contribution, we have concentrated on the analysis of the uni-dimensional time series for the simple reason that it remains always much to learn on the behaviour of the simple time series. And the findings that are already known must be summarised and presented clearly to the users. Thus, more effort will be required to describe and explain the current state of knowledge in this field at various technical levels in order to motivate the experts with the use of more sophisticated methods. Because only the practical use of new methods can justify their development.

Until very recently, cleavages between schools led most researchers to deal with the random processes and the time series through a given mode of representation. Contrary to that fact, we propose an approach in which all the modes of representation of a random process are used according to the possibilities of answers to the put questions. For instance, the beauty of the representation "state-space" lies in the simplicity of the structure of the equation of state. In addition, it is clear according to the definition of this equation that the stationarity of the processes is not strictly necessary. Moreover, the use of Gaussian spaces in formalism enables us to tackle the significant problem of the reversal of time and to state a result that seems interesting to us for the determination of the order of the processes. Besides, the method of Box and Jenkins is in so far as the most popular due to the fact that it is more accessible to a very broad range of users. The mathematical level that it requires is elementary. As for the spectral representation of a random process historically called periodogram, it is primarily based on the Fourier's transform of this process. The advantage of this form of parameterization is to replace the study of the random processes within the elegant framework of the harmonic analysis.

Next, we have developed an approach for the study of the likelihood of Gaussian linear processes. It is known that even in discrete time the problems of calculation of the likelihood are never simple. Thus, we have concentrated on the study of some types of likelihood approximation, the one of Whittle, then the one of Box and Jenkins and finally that of Toeplitz and Szegö. The two first have been widely popularised by Box and Jenkins. However, both approximations are unsuccessful when they are used in asymptotic calculations. For this reason, we have introduced the third type of approximation summarised into what is known as Toeplitz-Szegö's method.

In addition, it is noted that the sum of the squares in the approximation of Box and Jenkins is not directly computable because the innovations depend

on values previous to the time series occurrence. But an iterative procedure allows us to calculate this sum of squares in the rational case. As it seems that the convergence of this algorithm was not established, we have devoted a chapter to solve this problem in the case of the moving average models and the mixed models.

A drafting of the work of Rosanov and Ibragimov on the calculation of likelihood in particular in continuous time starting from the equations of Wiener-Hopf has been achieved in our contribution. Then, proceeding from the results of the approximation of the likelihood, we have proposed their application to handle the problems of tests. The goal is to locate the natural framework of latter developments and, thus, to attempt at obtaining approximations easy to handle by applying the theorem of Szegö to the continuous case. By calculating the Laplace's transform of the approximation of Szegö of the likelihood, one can lead the formula of Chernoff and note that it has the asymptotic properties of the lemma of Neyman and Pearson.

Key-words :

Time series - Random processes - State-space representation of a process - Internal representation of a process - Reversal of time - Dual representation of a random process - *ARMA* Models - Approximation of the likelihood functions - Toeplitz-Szegö's method of approximation - Algorithm of the returns - Approximation of the likelihood ratio - Test of the likelihood ratio.

RÉSUMÉ :

Dans cette contribution, nous nous sommes limités essentiellement à l'analyse des séries chronologiques unidimensionnelles pour la simple raison qu'il reste toujours beaucoup à apprendre sur le comportement des séries chronologiques simples. Et les résultats qui ont déjà été établis doivent être résumés et présentés clairement aux utilisateurs. Plus d'effort sera donc exigé pour décrire et expliquer l'état actuel des connaissances dans ce domaine à divers niveaux techniques, afin de motiver les praticiens à l'utilisation de méthodes plus sophistiquées. Car seule l'utilisation pratique des nouvelles méthodologies peut justifier leur développement.

Jusqu'à très récemment, les clivages entre écoles ont conduit la plupart des chercheurs à étudier les processus aléatoires et les séries chronologiques à travers un mode de représentation donné. Contrairement à cela, nous proposons une démarche dans laquelle tous les modes de représentation d'un processus aléatoire sont utilisés en fonction des possibilités de réponses aux questions posées. Par exemple, la beauté de la représentation état-espace réside dans la simplicité de la structure de l'équation d'état. Par ailleurs, il est clair d'après la définition de cette équation, que la stationnarité des processus n'est pas strictement nécessaire. De plus, l'utilisation des espaces gaussiens dans le formalisme nous permet d'aborder le problème important du retournement du temps et d'énoncer un résultat qui nous paraît intéressant pour la détermination de l'ordre des processus. La méthode de Box et Jenkins est de loin la plus populaire. L'avantage de cette méthode, d'aspect heuristique, est d'être accessible à un éventail d'utilisateurs très large. Le niveau mathématique qu'elle requiert est élémentaire. Quant à la représentation spectrale d'un processus aléatoire, historiquement appelée périodogramme, elle est essentiellement basée sur la transformée de Fourier de ce processus. L'avantage de cette forme de paramétrisation est de replacer l'étude des processus aléatoires dans le cadre élégant de l'analyse harmonique.

Par suite, nous avons abordé l'étude de la vraisemblance des processus linéaires gaussiens. On sait que même en temps discret les problèmes de calculs de la vraisemblance ne sont jamais simples. Ainsi, nous nous sommes intéressés à l'étude de quelques types d'approximation de la vraisemblance, celle de Whittle, celle de Box et Jenkins et celle de Toeplitz et Szegö. Les deux premières ont été largement popularisées par Box et Jenkins. Cependant, sur la ln -vraisemblance les deux approximations se prêtent mal aux calculs asymptotiques. On a donc introduit le troisième type d'approximation résumé dans ce qu'on a appelé la méthode de Toeplitz-Szegö.

Par ailleurs, on constate que la somme des carrés dans l'approximation de Box et Jenkins n'est pas directement calculable car les innovations dépendent des observations avant l'instant origine. Mais une procédure itérative permet dans le cas rationnel de calculer cette somme de carrés. Comme il semble que la convergence de cet algorithme n'a pas été établie, nous avons consacré le chapitre quatre à répondre à ce problème dans le cas des modèles en moyennes mobiles et des modèles mixtes.

Une rédaction des travaux de Rosanov et Ibragimov sur le calcul de vraisemblance a été proposée, notamment en temps continu à partir des équations de Wiener-Hopf. Et, à partir des résultats de l'approximation des vraisemblances on a procédé une extension aux problèmes de tests. Le but est donc de situer le cadre naturel de développements ultérieurs; et ainsi, en appliquant le théorème de Szegö au cas continu tenter d'obtenir des approximations faciles à manier. En calculant la transformée de Laplace de l'approximation de Szegö de la vraisemblance, on aboutit à la formule de Chernoff et on constate qu'elle possède les propriétés asymptotiques du lemme de Neyman et Pearson :

Mots-clés :

Séries chronologiques - Processus aléatoires - Représentation externe d'un processus - Représentation interne d'un processus - Retournement du temps - Représentation duale d'un processus aléatoire - Modèles *ARMA* - Approximation des fonctions de vraisemblance - Approximation de Toeplitz-Szegö - Algorithme d'aller-retour - Approximation du rapport de vraisemblances - Test du rapport de vraisemblances.

Table des matières

| | | |
|----------|-------------------------------------------------------------------------------------------|-----------|
| 1 | Introduction | 7 |
| 2 | Paramétrisation des processus aléatoires | 10 |
| 2.1 | Représentation externe d'un processus aléatoire | 10 |
| 2.1.1 | Introduction et notations | 10 |
| 2.1.2 | Le modèle autorégressif d'ordre p ou $AR(p)$ | 12 |
| 2.1.3 | Le modèle en moyennes mobiles d'ordre q ou $MA(q)$ | 13 |
| 2.1.4 | Le modèle mixte autorégressif-moyennes mobiles d'ordre (p, q) ou $ARMA(p, q)$ | 13 |
| 2.1.5 | Le modèle général $ARIMA(p, d, q)$ | 14 |
| 2.1.6 | Principe de prévision | 15 |
| 2.2 | Représentation interne d'un processus aléatoire | 16 |
| 2.2.1 | Introduction | 16 |
| 2.2.2 | Quelques rappels | 17 |
| 2.2.3 | Propriétés des représentations internes | 19 |
| 2.2.4 | Représentation état-espace d'un processus $ARMA$ | 23 |
| 2.3 | Dualité des représentations et principe du retournement du temps | 25 |
| 2.3.1 | Représentation duale d'un processus aléatoire | 25 |
| 2.3.2 | Ordre d'un processus aléatoire | 27 |
| 3 | Approximation de la vraisemblance d'un processus gaussien stationnaire | 29 |
| 3.0.3 | Introduction | 29 |
| 3.1 | Les différents types d'approximations | 30 |
| 3.2 | Méthode d'approximation de Toeplitz et Szegö | 34 |

| | | |
|----------|---------------------------------------------------------------------------------------------------------------------|-----------|
| 4 | Détermination de la vraisemblance inconditionnelle d'un processus aléatoire par la méthode de l'aller-retour | 43 |
| 4.1 | Position du problème | 43 |
| 4.2 | Approximation de la vraisemblance exacte d'un processus MA | 44 |
| 4.2.1 | La vraisemblance exacte d'un processus $MA(q)$ | 44 |
| 4.2.2 | Détermination des valeurs initiales par l'algorithme de l'aller-retour | 45 |
| 4.3 | Approximation de la vraisemblance exacte d'un processus $ARMA$ | 52 |
| 4.3.1 | La vraisemblance exacte d'un processus $ARMA(p, q)$ | 52 |
| 4.3.2 | Détermination des valeurs initiales par l'algorithme de l'aller-retour | 54 |
| 5 | Rapport des vraisemblances de deux processus gaussiens | 61 |
| 5.1 | Equations de Wiener- Hopf | 61 |
| 5.2 | Le rapport de vraisemblances de deux processus gaussiens | 62 |
| 5.2.1 | Condition d'existence | 62 |
| 5.3 | L'équation de Wiener-Hopf du rapport de vraisemblances | 67 |
| 5.4 | Solution dans le cas rationnel de l'équation de vraisemblance | 69 |
| 5.5 | Le lemme de Neyman-Pearson | 75 |
| 5.5.1 | Cas discret | 75 |
| 5.5.2 | Cas continu | 77 |

Remerciements

En premier lieu, je tiens à exprimer ma plus profonde gratitude à l'égard de ma femme pour sa disponibilité de tous les instants et pour son infinie patience pour m'avoir supporté durant tout ce temps, et à mes adorables enfants pour leur foi en moi. Je tiens aussi à exprimer mon plus profond respect et mon admiration à l'égard de Monsieur Didier Dacunha Castelle pour sa fraîcheur d'esprit et ses conseils avisés.

Je remercie vivement Monsieur Boushaba Mahmoud pour tout l'intérêt qu'il a porté à l'aboutissement de cette thèse et pour avoir accepté d'en être le rapporteur.

Je remercie Monsieur Denche Mohamed d'avoir accepté d'être le président de mon jury de soutenance.

J'exprime mes plus sincères remerciements à Messieurs Khaldi Khaled et Rahmani Fouad Lazhar qui m'ont fait l'honneur d'être les examinateurs de cette thèse. Je suis très honoré que Messieurs Boutabia Hacène et Yarou Mostepha Fateh aient accepté de faire partie de ce jury.

Pour finir, je remercie tous les amis et collègues qui m'ont soutenu et encouragé.

Chapitre 1

Introduction

Le travail que nous présentons a été orienté sur les aspects les plus sensibles dans l'étude des processus aléatoires et des séries chronologiques. Dans la vaste littérature qui traite de la statistique des processus linéaires (voir [6], [9], [14], [15], [19], [22], [26], [32], [34]), l'un des problèmes les plus fréquemment rencontrés est celui d'exprimer explicitement la fonction de vraisemblance. On sait que même en temps discret les problèmes de calculs de cette fonction ne sont jamais simples. De même, le rapport de vraisemblances, qui joue un rôle essentiel pour la construction et la détermination de tests d'ajustement et de discrimination de modèles, a fait l'objet d'une attention particulière. Enfin, on constate que beaucoup d'efforts restent à faire concernant les problèmes d'identification et de paramétrisation des modèles de séries chronologiques. Dans ce but, notre contribution a été subdivisée en cinq chapitres qui, nous l'espérons, répondent aux préoccupations du moment.

Après le chapitre introductif, le second chapitre est consacré à la présentation de deux types de paramétrisation des processus aléatoires. Le premier type que l'on a intitulé représentation externe des processus, est un résumé des travaux de Box et Jenkins, Brockwell et Hamilton principalement ([6], [9], [22], [38]). Cette forme d'expression des modèles stochastiques est basée sur le théorème de décomposition de Wold et sur l'idée de G.U. Yule ([38]) qui stipule qu'un phénomène dont les réalisations successives sont fortement dépendantes entre elles, peut être considéré comme étant le résultat du passage à travers un filtre linéaire d'une suite de variables aléatoires ε_t indépendantes et centrées. Dans notre exposé on a adopté les notations de Box et Jenkins

([6], [22]) pour décrire les modèles *ARMA* à l'aide de l'équation simple :

$$P(B)X_t = Q(B)\varepsilon_t$$

Le deuxième type de paramétrisation qu'on a appelé représentation interne d'un processus, est en fait la représentation de Kalman-Bucy ou représentation état-espace ([9], [22], [27], [35]). Cette technique a été développée dans le cadre du contrôle des systèmes linéaires ([13], [24]). Elle est résumée par le système des équations suivantes ou filtre de Kalman-Bucy :

$$\begin{aligned} Y_t &= HX_{t-1} + W_t \\ \text{où} \quad X_t &= FX_{t-1} + V_t \end{aligned}$$

Dans le troisième chapitre nous avons abordé l'étude de la vraisemblance des processus linéaires gaussiens. Ainsi, nous nous sommes intéressés à l'étude de quelques types d'approximation de la vraisemblance. La première méthode exposée est due à Whittle ([38]). Avec cette technique, assez intuitive quoique peu générale, on calcule l'approximation de la vraisemblance L_N^W au moyen du périodogramme. En effet, il a été démontré que la forme quadratique dans l'exponentielle est approchée par $\frac{N}{4\pi} \int \frac{I_N(\lambda)}{f(\lambda, \theta)} d\lambda$, ie. :

$$\ln L_N^W = C \int_{-\pi}^{+\pi} \ln f(\lambda) d\lambda - \frac{N}{4\pi} \int \frac{I_N(\lambda)}{f(\lambda, \theta)} d\lambda$$

Elle permet aussi, sous des hypothèses assez générales, de montrer que la somme des carrés des innovations empiriques $\sum_{j=1}^N \varepsilon_j^2$, est une bonne approximation de la forme quadratique de la vraisemblance. Le deuxième genre d'approximation consiste à considérer tout processus comme limite de processus autorégressif. Ce type d'approximation a été largement popularisé par Box et Jenkins ([6]). Cependant, sur la ln-vraisemblance les deux approximations précédentes se prêtent mal aux calculs asymptotiques. On a donc introduit un troisième type d'approximation résumé dans ce qu'on va appeler la méthode de Toeplitz-Szegö. Il est à rappeler que la matrice d'autocovariances d'un processus n'est autre que la transformée de Toeplitz de sa densité spectrale.

Par ailleurs, on constate que la méthode de Box et Jenkins est une application particulière de la méthode de Whittle dans le cas d'un spectre rationnel

et de ce fait la forme quadratique de la vraisemblance est approchée par L_N^{BJ} telle que :

$$L_N^{BJ} = \sum_{j=1-q}^N \varepsilon_j^2$$

Les innovations ne sont pas calculables en général puisqu'elles dépendent de l'observation avant l'instant origine. Mais une procédure itérative permet dans le cas rationnel de calculer cette vraisemblance. Comme il semble que la convergence de cet algorithme n'a pas été établie, nous avons consacré le chapitre quatre à répondre à ce problème dans le cas des modèles en moyennes mobiles et des modèles mixtes.

Dans le dernier chapitre, on adopte la représentation spectrale d'un processus aléatoire qui est essentiellement la transformée de Fourier de ce processus. A l'origine, cette représentation était appelée périodogramme. L'avantage de cette forme de paramétrisation est de remplacer l'étude des processus aléatoires dans le cadre élégant de l'analyse harmonique. Le dernier volet de notre travail est une extension des résultats du chapitre trois aux problèmes de tests. Une rédaction des résultats bien connus de Rosanov et Ibrahimov ([26]) sur le calcul de vraisemblances y a été résumée, notamment en temps continu à partir des équations de Wiener-Hopf ([17]). Le but est alors de situer le cadre naturel de développements ultérieurs, i.e. l'application de la forme due à Kac du théorème de Szegö, au cas continu pour obtenir des approximations faciles à manier. Un résumé du travail de Pisarenko ([32]) sur le rapport de vraisemblances dans le cas où les densités spectrales sont rationnelles, a été entrepris, mais on constate que l'auteur ne propose pas de statistique pour construire un test. Finalement, pour terminer cette dernière partie on a essayé de déterminer le test de Neyman et Pearson en situation exponentielle. Pour cela, on a considéré la transformée de Laplace de l'approximation de Szegö de la vraisemblance. Le résultat obtenu n'est autre que la formule de Chernoff ([4]) et ainsi il vérifie les propriétés asymptotiques du lemme de Neyman et Pearson. Il est, dans ce cas, équivalent aux tests classiques type porte-manteau.

Chapitre 2

Paramétrisation des processus aléatoires

2.1 Représentation externe d'un processus aléatoire

2.1.1 Introduction et notations

Une classe importante de modèles stochastiques utilisés pour décrire les séries chronologiques est la classe des modèles stationnaires *ARMA* ([6], [9], [22]). La propriété de stationnarité suppose que le processus stochastique fluctue autour d'une valeur moyenne constante. Mais, en économie, dans l'industrie ou dans le monde des affaires où les prévisions jouent un rôle important, la plupart des séries chronologiques qu'on y étudie sont mieux représentées par des modèles non-stationnaires. En particulier, ces séries ne possèdent pas une valeur moyenne fixe. Les processus stochastiques non-stationnaires seront modélisés par les processus *ARIMA* ([6], [22]). Cette vaste classe de processus fournit un ensemble de modèles stationnaires et non-stationnaires qui ont la particularité de représenter assez fidèlement les séries chronologiques rencontrées dans la pratique.

Par ailleurs, ces modèles de représentation des séries chronologiques sont définis en fonction d'équations linéaires de différence à coefficients constants. Pour faciliter les notations, nous allons introduire et utiliser l'opérateur de retard noté B , défini tel que :

$$BX_t = X_{t-1} \quad \text{et} \quad B^m X_t = X_{t-m}$$

L'opération inverse ou translation positive est réalisée par l'opérateur d'avance F défini tel que :

$$FX_t = X_{t+1} \quad \text{et} \quad F^m X_t = X_{t+m}$$

Remarque 1 L'opérateur de retard B et l'opérateur d'avance F sont inverses l'un de l'autre, i.e., $F = B^{-1}$.

On définit un autre opérateur aussi utile que les précédents, appelé opérateur de différence et noté ∇ , tel que :

$$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t$$

Remarque 2 L'opérateur ∇ admet pour inverse l'opérateur somme S , défini tel que :

$$SX_t = \nabla^{-1} X_t = (1 - B)^{-1} X_t = \left(\sum_{j=0}^{\infty} B^j \right) X_t = \sum_{j=0}^{\infty} X_{t-j}$$

La paramétrisation des modèles stochastiques que l'on va utiliser est basée sur l'idée de G.U. Yule ([37]) qui stipule qu'une série chronologique dont les valeurs successives sont fortement dépendantes peut être considérée comme le résultat du passage à travers un filtre linéaire d'une suite de variables aléatoires ε_t indépendantes et centrées. Ces variables sont des tirages aléatoires d'une distribution donnée, en général une loi normale de moyenne nulle et de variance σ_ε^2 . La suite des variables aléatoires $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ est appelée processus de bruit blanc et est notée ε_t . Cette opération de transformation du processus ε_t en un processus X_t par filtrage linéaire n'est autre que le produit de convolution d'une suite $(\psi_j)_{j \geq 0}$ de poids et de la suite des valeurs $(\varepsilon_{t-j})_{j \geq 0}$ du processus des innovations. En d'autres termes, la valeur d'un processus à un instant t est le résultat de la somme pondérée des innovations antérieures à cet instant, i.e. :

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \mu + \Psi(B) X_t$$

où $\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ avec $\psi_0 = 1$

Remarque 3 La constante μ est un niveau constant autour duquel évolue le processus X_t . L'opérateur $\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ est le filtre linéaire qui transforme ε_t en X_t . L'opérateur $\Psi(B)$ est appelé fonction de transfert. Si la série $\sum_{j=0}^{\infty} \psi_j B^j$ est convergente, le filtre $\Psi(B)$ est dit stable et le processus X_t est dit stationnaire, dans le cas contraire on dit que X_t est non-stationnaire.

2.1.2 Le modèle autorégressif d'ordre p ou $AR(p)$

Definition 4 Soient $Y_t, Y_{t-1}, Y_{t-2}, \dots$, les réalisations du processus aux instants $t, t-1, t-2, \dots$. Supposons que $E(Y_t) = \mu$ et posons $X_t = Y_t - \mu$. Si le processus X_t peut s'écrire sous la forme

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.1)$$

où ε_t est un processus de bruit blanc ou processus d'innovation vérifiant les conditions $E(\varepsilon_t) = 0$ et $Var(\varepsilon_t) = \sigma_\varepsilon^2$, alors X_t est appelé processus autorégressif d'ordre p et est noté $AR(p)$.

Remarque 5 La raison principale de cette appellation réside dans le fait que l'expression du processus X_t est une régression multiple de sa valeur courante X_t sur ses p valeurs passées $X_{t-1}, X_{t-2}, \dots, X_{t-p}$. De plus, si on considère $\Phi(B)$ l'opérateur autorégressif d'ordre p défini tel que

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

alors la relation 2.1 peut s'écrire succinctement sous la forme :

$$\Phi(B) X_t = \varepsilon_t$$

Par ailleurs, il est aisé de constater que le modèle $AR(p)$ est un cas particulier des modèles à filtre linéaire. Par exemple, on peut remplacer dans l'expression 2.1 X_{t-1} par :

$$X_{t-1} = \phi_1 X_{t-2} + \phi_2 X_{t-3} + \dots + \phi_p X_{t-p-1} + \varepsilon_{t-1}$$

Et ainsi de suite, on remplace X_{t-2}, X_{t-3}, \dots etc. Finalement, X_t s'exprimera comme une série en ε_t , i.e.

$$\begin{aligned} X_t &= \Psi(B) \varepsilon_t = \Phi^{-1}(B) \varepsilon_t \\ \text{avec } \Psi(B) &= \Phi^{-1}(B) \end{aligned}$$

Un processus autorégressif peut être stationnaire ou non-stationnaire. Pour qu'un processus $AR(p)$ soit stationnaire il faut que les coefficients ψ_j soient définis de telle sorte que la série $\Phi^{-1}(B) = \sum_{j=0}^{\infty} \psi_j B^j$ soit convergente. La condition nécessaire pour qu'il en soit ainsi, est que l'opérateur autorégressif $\Phi(B)$, considéré comme polynôme en B de degré p , ait toutes ses racines à l'extérieur du cercle unité.

2.1.3 Le modèle en moyennes mobiles d'ordre q ou $MA(q)$

Definition 6 Soient $X_t, X_{t-1}, X_{t-2}, \dots$ etc., les réalisations d'un processus aléatoire centré aux instant $t, t-1, t-2, \dots$. Si le processus X_t peut s'écrire sous la forme

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.2)$$

où ε_t est un processus d'innovation vérifiant les conditions $E(\varepsilon_t) = 0$ et $Var(\varepsilon_t) = \sigma_\varepsilon^2$, alors X_t est appelé processus en moyennes mobiles d'ordre q et est noté $MA(q)$.

Remarque 7 On définit l'opérateur des moyennes mobiles d'ordre q , noté $\Theta(B)$, tel que :

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

Ainsi, un processus $MA(q)$ peut s'écrire succinctement sous la forme $X_t = \Theta(B) \varepsilon_t$. Il est évident que le modèle $MA(q)$ est un modèle à filtre linéaire.

2.1.4 Le modèle mixte autorégressif-moyennes mobiles d'ordre (p, q) ou $ARMA(p, q)$

Definition 8 Pour permettre une meilleure flexibilité dans l'ajustement des séries chronologiques, il est souvent utile de combiner les formes autorégressives et les formes moyennes mobiles. Ceci nous amène à définir les modèles mixtes autorégressifs-moyennes mobiles d'ordre (p, q) , notés $ARMA(p, q)$, tels que :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.3)$$

L'expression 2.3 peut s'écrire de façon succincte sous la forme :

$$\Phi(B) X_t = \Theta(B) \varepsilon_t$$

Par ailleurs, on peut montrer que les modèles $ARMA(p, q)$ sont des modèles à filtre linéaire. En effet, on peut remplacer dans l'expression 2.3 X_t par :

$$X_{t-1} = \phi_1 X_{t-2} + \phi_2 X_{t-3} + \dots + \phi_p X_{t-p-1} + \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3} - \dots - \theta_q \varepsilon_{t-q-1}$$

On procédera de la même manière pour $X_t, X_{t-1}, X_{t-2}, \dots etc$ Finalement, X_t peut s'exprimer comme une série en ε_t , i.e.

$$X_t = \Psi(B) \varepsilon_t = \Phi^{-1}(B) \Theta(B) \varepsilon_t$$

Il est établi dans ([2], [6], [9], [22], [30]) que les représentations adéquates des séries chronologiques stationnaires que l'on rencontre dans la pratique sont obtenues à l'aide de modèles AR , de modèles MA ou de modèles $ARMA$ avec des paramètres p et q le plus souvent inférieurs ou égaux à 2.

2.1.5 Le modèle général $ARIMA(p, d, q)$

Certaines séries chronologiques rencontrées dans la pratique, surtout dans les domaines de l'industrie, de l'économie et dans le monde des affaires, présentent des allures non-stationnaires ([2], [6]). Généralement, ces séries n'admettent pas de moyenne fixe. Néanmoins, elles se caractérisent par un comportement homogène. En d'autres termes, bien que le niveau moyen autour duquel la série fluctue change d'une période à l'autre, la série chronologique garde la même allure générale moyennant un changement de temps et d'échelle. Ce genre de comportement peut être représenté par un opérateur généralisé $\Lambda(B)$ qui, considéré comme polynôme en B , admet une ou plusieurs racines sur le cercle unité. Ce type d'opérateur peut s'écrire sous la forme :

$$\Lambda(B) = \Phi(B) (1 - B)^d$$

où $\Phi(B)$ est un opérateur autorégressif stationnaire

Definition 9 *Le modèle général pour représenter des séries chronologiques stationnaires ou non-stationnaires à comportement homogène peut être présenté sous la forme suivante*

$$\Lambda(B) X_t = \Phi(B) (1 - B)^d X_t = \Theta(B) \varepsilon_t \quad (2.4)$$

Ce modèle est alors appelé modèle autorégressif-moyennes mobiles intégré ou modèle $ARIMA(p, d, q)$.

Remarque 10 *L'expression 2.4 peut s'écrire sous la forme suivante :*

$$\begin{aligned}\Phi(B)Z_t &= \Theta(B)\varepsilon_t \\ \text{où } Z_t &= \nabla^d X_t.\end{aligned}\tag{2.5}$$

Ainsi, dans l'expression 2.5 $\Phi(B)$ est un opérateur autorégressif stationnaire, donc le processus Z_t est un processus stationnaire. Autrement dit, la $d^{\text{ème}}$ différence du processus est stationnaire. Dans la pratique d est généralement égal à 0, 1 ou au plus 2.

L'introduction du mot "intégré" dans la dénomination *ARIMA* peut s'expliquer par le fait que X_t peut s'écrire sous la forme suivante :

$$X_t = S^d Z_t$$

où S est l'opérateur somme défini tel que :

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j}$$

Par conséquent, le processus X_t est le résultat d'une "sommation" infinie d'un processus stationnaire Z_t . On aurait pu utilisé le mot "sommé" au lieu de "intégré" dans l'abréviation *ARIMA*.

2.1.6 Principe de prévision

Pour introduire le principe de prévision des valeurs futures d'un processus aléatoire, nous allons procéder comme si la forme et les paramètres du modèle sont exactement connus. Il est clair que les erreurs dues à l'estimation des paramètres ne va pas affecter sérieusement les prévisions, sauf si la taille de l'échantillon utilisé est petite. Les calculs des prévisions seront faits en utilisant l'équation de différence du modèle ([6], [9], [22]). Cette approche nous paraît comme la plus simple et la plus élégante. Cependant, pour approfondir l'étude des prévisions, il est possible qu'on ait recours à d'autres formes du modèle.

Supposons qu'on veuille faire la prévision des valeurs futures $\{X_t, t \geq N + 1\}$ d'un processus aléatoire centré et stationnaire à partir d'une observation $\{X_1, X_2, \dots, X_N\}$ du phénomène. Supposons que le processus $\{X_t, t \in \mathbb{Z}\}$ vérifie l'équation de différence :

$$P(B)X_t = Q(B)\varepsilon_t\tag{2.6}$$

De plus, $\{X_t, t \in \mathbb{Z}\}$ engendre l'espace de Hilbert $L^2(\cdot, \mathcal{A}, P)$ muni du produit scalaire $\langle X, Y \rangle = E(XY)$. Alors, le meilleur prévision d'une valeur future $X_{N+\ell}$ du processus est sa projection orthogonale, notée $\widehat{X}_N(\ell)$, sur le sous-espace de Hilbert \mathcal{H}_N engendré par $\{X_1, X_2, \dots, X_N\}$. En d'autres termes, $\widehat{X}_N(\ell)$ est l'élément de \mathcal{H}_N qui rend minimum la quantité $E \left\{ \left(X_{N+\ell} - \widehat{X}_N(\ell) \right)^2 \right\}$ ([6], [9], [21]). La projection sur l'espace de Hilbert \mathcal{H}_N de l'élément $X_{N+\ell}$ est obtenue en calculant l'espérance conditionnelle de $X_{N+\ell}$ par rapport à \mathcal{H}_N ([6], [9], [35]), i.e.

$$\widehat{X}_N(\ell) = E^{\mathcal{H}_N}(X_{N+\ell})$$

A l'aide du modèle 2.6, le calcul des espérances conditionnelles est fait selon la règle suivante :

$$\begin{cases} E^{\mathcal{H}} X_{N-j} = X_{N-j} & \text{si } j \geq 0 \\ E^{\mathcal{H}} X_{N+j} = \widehat{X}_N(j) & \text{si } j \geq 0 \\ E^{\mathcal{H}} \varepsilon_{N-j} = \varepsilon_{N-j} & \text{si } j \geq 0 \\ E^{\mathcal{H}} \varepsilon_{N+j} = 0 & \text{si } j \geq 0 \end{cases}$$

Il a été établi ([6], [9]) que l'opérateur autorégressif $P(B)$ détermine le comportement de la fonction éventuelle de prévision alors que l'opérateur des moyennes mobiles $Q(B)$ fixe les q valeurs initiales de la fonction éventuelle de prévision. Pour un processus en moyennes mobiles $MA(q)$ les prévisions chutent à zéro dès que $\ell > q$. Alors que pour un processus $AR(p)$ les prévisions décroissent de façon exponentielle si les racines de l'équation $P(B) = 0$ sont réelles, elles décroissent de façon oscillante si les racines de cette équation sont complexes.

2.2 Représentation interne d'un processus aléatoire

2.2.1 Introduction

Les représentations internes, appelées aussi représentations "état-espace" ou représentations de Kalman-Bucy (voir [9], [16], [22], [27], [35]) ont eu un profond impact sur l'étude des processus aléatoires et l'analyse des séries chronologiques. A l'origine, ces techniques ont été développées dans le cadre

du contrôle des systèmes linéaires. En effet, la démarche de Kalman-Bucy ([26]) suppose à priori que le processus $(Y_t, t \in \mathbb{Z})$ est modélisé tel que :

$$Y_t = HX_{t-1} + W_t \quad (2.7)$$

où

$$X_t = FX_{t-1} + V_t \quad (2.8)$$

où $(W \ V)'$ est un bruit blanc.

L'équation 2.8 est interprétée comme décrivant l'évolution de l'"état" X_t du système à l'instant t . Elle est donc appelée équation d'état. Le processus X_t ainsi représenté, est alors markovien. La relation 2.7 est appelée équation d'observation et donc elle définit la suite de réalisations de Y_t . Par ailleurs, en employant le langage suggestif de l'automaticien, on peut dire que Y_t représente la sortie d'un système dynamique excité par un bruit blanc. Il a été établi dans ([6], [16], [22], [35]) que cette représentation est rigoureusement équivalente à celle de Box et Jenkins dans le cas où la densité spectrale $f(\lambda)$ est rationnelle en $e^{i\lambda}$. Dans ce paragraphe, nous allons résumer quelques résultats que l'on peut trouver dans ([16], [27], [31], [35]), sur les propriétés de la fonction de covariance, du spectre et des espaces gaussiens associés au processus Y_t .

2.2.2 Quelques rappels

Les processus que nous considérons par la suite sont des processus gaussiens stationnaires indexés par \mathbb{Z} et à valeurs dans des espaces vectoriels réels de dimension finie ([4], [21], [31]). Les espaces gaussiens réels associés seront toujours notés en lettres rondes. Ainsi, l'espace gaussien engendré par le processus X_t sera noté $\mathcal{H} = \mathcal{H}(X_t, t \in \mathbb{Z})$, si $T \subset \mathbb{Z}$, on notera $\mathcal{H}_T = \mathcal{H}(X_t, t \in T)$. Plus précisément, nous définirons le «passé avant l'instant t » (resp. le «futur après l'instant t ») par $\mathcal{H}_t^- = \mathcal{H}_{]-\infty, t]}$ (resp. par $\mathcal{H}_t^+ = \mathcal{H}_{[t, +\infty[}$). Le «passé infini» $\mathcal{H}_{-\infty}$ sera défini par $\mathcal{H}_{-\infty} = \bigcap_{t \in \mathbb{Z}} \mathcal{H}_t^-$. Nous dirons, comme Rosanov ([26], [34]) que le processus X_t est régulier si $\mathcal{H}_{-\infty} = \{0\}$ et qu'il est singulier si $\mathcal{H}_t^- = \mathcal{H}_{-\infty} \forall t \in \mathbb{Z}$.

Soit \mathbb{H} une famille croissante d'espaces gaussiens $(\mathcal{H}_t, t \in \mathbb{Z})$, nous dirons que le processus X_t est :

- (i) adapté à \mathbb{H} si $\forall t \in \mathbb{Z}, X_t \in \mathcal{H}_t$.
- (ii) innovant par rapport à \mathbb{H} si $\forall t \in \mathbb{Z}, E^{\mathcal{H}_t}(X_{t+1}) = 0$.

(iii) mesurable par rapport à \mathbb{H} si $\mathcal{H}_t \subset \mathcal{H}_\infty$ où $\mathcal{H}_\infty = \bigvee_{t \in \mathbb{Z}} \mathcal{H}_t$ (on rappelle que si $(H_i, i \in I)$ est une famille d'espaces de Hilbert, $\bigvee_{i \in I} H_i$ désigne l'espace de Hilbert engendré par les $H_i, \forall i \in I$).

(iv) markovien par rapport à \mathbb{H} si X_t est adapté à \mathbb{H} et si :

$$E^{\mathcal{H}_{t-1}}(X_t) = E^{X_{t-1}}(X_t) = FX_{t-1}$$

F est donc définie comme la matrice de transition de X_t .

Nous rappelons quelques résultats établis dans ([9], [14], [17], [31], [33], [34]). Soit $\sigma(F)$ le spectre de F et \mathcal{D} (resp. $\delta\mathcal{D}$) le disque unité (resp. le cercle unité) du plan complexe. On dira que le processus gaussien markovien X_t est purement non déterministe si X_t est régulier, et de plus, si $X_t = FX_{t-1} + V_t$ où le processus $\{V_t, t \in \mathbb{Z}\}$ est adapté et innovant par rapport à $\mathcal{H}^- = (\mathcal{H}_t^-, t \in \mathbb{Z})$. Si le processus X_t est non dégénéré (i.e. s'il est de covariance définie positive), alors la matrice F est asymptotiquement stable, i.e. $\sigma(F) \subset \mathcal{D}^\circ$. Par ailleurs, on dira que X_t est déterministe si X_t est singulier, et si $X_t = FX_{t-1}$. Si X_t est non dégénéré alors F est oscillatoire, i.e. $\sigma(F) \subset \partial\mathcal{D}$. Si les processus gaussiens sont centrés et stationnaires, alors ils sont définis à une équivalence près par leur fonction de covariance $\Lambda_X(t)$, avec $\Lambda_X(t) = E(X_{t+s}X_s)$ ou leur spectre $S_X(\lambda)$ défini par le théorème de Bochner tel que :

$$\Lambda_X(t) = \int_{-\pi}^{+\pi} e^{i\lambda t} S_X(d\lambda)$$

La mesure $S_X(d\lambda)$ est une mesure positive bornée en ce sens que $\forall \Delta \in \mathcal{B}[-\pi, +\pi]$, $S_X(\Delta)$ est hermitienne semi-définie positive et $\int_{-\pi}^{+\pi} tr(S_X(d\lambda)) < \infty$. On notera en lettres minuscules les mesures aléatoires associées aux processus considérés, soit :

$$X_t = \int_{-\pi}^{+\pi} e^{i\lambda t} x(d\lambda)$$

On sait que $\forall \Delta \in \mathcal{B}[-\pi, +\pi]$, $x(\Delta)$ est \mathcal{H} -mesurable et $\forall \Delta_1, \Delta_2 \in \mathcal{B}[-\pi, +\pi]$, alors :

$$E \left\{ x(\Delta_1) \overline{x(\Delta_2)} \right\} = S_X(\Delta_1 \cap \Delta_2)$$

où $\overline{x(\Delta_2)}$ désigne le complexe conjugué de $x(\Delta_2)$.

2.2.3 Propriétés des représentations internes

Considérons un processus aléatoire gaussien centré et stationnaire $(Y_t, t \in \mathbb{Z})$ défini sur un espace de probabilité (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{R}^p .

Definition 11 Soit $\mathbb{H} = (\mathcal{H}_t, t \in \mathbb{Z})$ une famille croissante d'espaces gaussiens. On dira que cette famille est admissible si Y_t est adapté à \mathbb{H} . De plus, on dira que le processus gaussien centré (X_t, W_t) stationnairement corrélé avec Y_t , est une \mathbb{H} – représentation d'ordre q de Y_t si et seulement si :

- (i) $(X_t, t \in \mathbb{Z})$ est un processus à valeurs dans \mathbb{R}^q , markovien par rapport à \mathbb{H} .
- (ii) $(W_t, t \in \mathbb{Z})$ est un processus à valeurs dans \mathbb{R}^p , adapté et innovant par rapport à \mathbb{H} .
- (iii) il existe une matrice H de dimension (p, q) telle que :

$$Y_t = HX_{t-1} + W_t \quad \forall t \in \mathbb{Z}$$

Par ailleurs, on dira qu'une représentation de Y_t est d'ordre minimal, noté $O(Y_t)$, s'il n'existe pas de représentation d'ordre strictement inférieur.

Proposition 12 Le processus W_t est un bruit blanc et le vecteur $(X_t \ Y_t)'$ est markovien par rapport à \mathbb{H} .

Démonstration : Nous rappelons qu'un bruit blanc est par définition un processus gaussien totalement «décorrélé». D'après le point (ii) de la définition 11, on a :

$$E(W_{t+s}W_s) = E\{E^{\mathcal{H}_s}(W_{t+s})W_s\} = 0 \quad \forall t \geq 1$$

Donc, W est un bruit blanc.

On déduit aussi à partir du point (iii) de la définition 11 que :

$$W_t = Y_t - E^{\mathcal{H}_{t-1}}Y_t \tag{2.9}$$

Réciproquement, si W_t est défini par la relation 2.9 il est clair que c'est un bruit blanc adapté et innovant par rapport à \mathbb{H} .

Par ailleurs, il est évident que $(X_t \ Y_t)'$ est adapté à \mathbb{H} . De plus, si F désigne la matrice de transition de X_t , on a :

$$E^{\mathcal{H}_{t-1}}X_t = FX_{t-1} \tag{2.10}$$

Par ailleurs, W_t étant innovant par rapport à \mathbb{H} , il vient :

$$E^{\mathcal{H}_{t-1}} Y_t = H X_{t-1} \quad (2.11)$$

D'après les relations 2.10 et 2.11 on peut affirmer que le vecteur $\begin{pmatrix} X_t & Y_t \end{pmatrix}'$ est markovien par rapport à \mathbb{H} , de matrice de transition $\mathbb{F} = \begin{pmatrix} F & 0 \\ H & 0 \end{pmatrix}$. On obtient alors le modèle :

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \mathbb{F} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} V_t \\ W_t \end{bmatrix}$$

Ou bien de façon explicite :

$$\begin{cases} X_t = F X_{t-1} + V_t \\ Y_t = H X_{t-1} + W_t \end{cases}$$

On retrouve le modèle ou la représentation état-espace de Kalman-Bucy. ■

Remarque 13 *On déduit de ce qui précède que :*

$$E^{\mathcal{H}_t} (Y_{t+k}) = H F^{k-1} X_t \quad \forall k \geq 1 \quad (2.12)$$

Proposition 14 *Les représentations minimales sont non dégénérées.*

Démonstration : Soit X_t une \mathcal{H} -représentation de Y_t d'ordre minimal q et de covariance Γ . Si Γ était non définie positive, il existerait une matrice M de changement de base orthogonale sur \mathbb{R}^q telle que :

$$M X_t = \begin{pmatrix} X_t^1 \\ 0 \end{pmatrix} \quad \text{avec} \quad \text{Cov}(X_t^1) > 0$$

X_t^1 étant évidemment markovien par rapport à \mathbb{H} , il est clair que (X_t^1, W_t) serait une représentation de Y_t d'ordre strictement inférieur à q , ce qui contredit l'hypothèse de minimalité de q . ■

Principe de décomposition des représentations minimales.

Soit (X_t, W_t) une \mathcal{H} -représentation minimale de Y_t . Nous savons qu'il existe une base \mathcal{B}_X de \mathbb{R}^q telle que X_t s'écrive dans \mathcal{B}_X sous la forme :

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}$$

où X_{1t} est purement non déterministe, de matrice de transition F_1 asymptotiquement stable, X_{2t} déterministe, de matrice de transition F_2 oscillatoire et $\mathcal{H}_{1t} \perp \mathcal{H}_{2t}$.

Soit $\mathbb{H} = (\mathbb{H}_1, \mathbb{H}_2)$ la décomposition de \mathbb{H} dans \mathcal{B}_X . Définissons alors les processus Y_{1t} et Y_{2t} tels que :

$$Y_{1t} = H_1 X_{1(t-1)} + W_t \quad (2.13)$$

$$Y_{2t} = H_2 X_{2(t-1)} \quad (2.14)$$

Proposition 15 *Si $Y_t = Y_{1t} + Y_{2t}$, alors :*

$$Y_{2t} = E^{\mathcal{H}_{-\infty}}(Y_t) \quad (2.15)$$

où $\mathcal{H}_{-\infty} = \bigcap_{t \in \mathbb{Z}} \mathcal{H}_t$.

Démonstration : X_{2t} étant déterministe et adapté à \mathbb{H} , on a alors :

$$\mathcal{H}_{2t} = \mathcal{H}_{-\infty} \subset \bigcap_{t \in \mathbb{Z}} \mathcal{H}_t = \mathcal{H}_{-\infty}$$

Donc Y_{2t} est $\mathcal{H}_{-\infty}$ - mesurable

De plus, soit V_t le bruit blanc sur la dynamique de X_{1t} , on a alors :

$$X_{1t} = F_1 X_{1(t-1)} + V_t \quad (2.16)$$

Donc,

$$X_{1t} = \sum_{k=0}^{\infty} F_1^k V_{t-k}$$

où la série converge p.s. dans L^2 .

Sachant que le vecteur $(V_t \ W_t)'$ est innovant et adapté à \mathbb{H} , il est clair que :

$$V_t \perp \mathcal{H}_{-\infty} \quad \text{et} \quad W_t \perp \mathcal{H}_{-\infty} \quad \forall t \in \mathbb{Z}$$

On en déduit que X_{1t} et W_t , donc aussi Y_{1t} , sont orthogonaux à $\mathcal{H}_{-\infty}$. La relation 2.15 découle alors immédiatement de la somme des relations 2.13 et 2.14. ■

Proposition 16 Y_{1t} est régulier de densité spectrale $f(\lambda)$ et Y_{2t} est singulier de spectre $S_s(\lambda)$.

Démonstration : La somme des relations 2.13 et 2.14 s'écrit en termes de mesures aléatoires telle que :

$$y(d\lambda) = y_1(d\lambda) + y_2(d\lambda) \quad (2.17)$$

De même que pour chacune des équations 2.13 et 2.14 :

$$y_1(d\lambda) = H_1 e^{-i\lambda} x_1(d\lambda) + w(d\lambda)$$

$$y_2(d\lambda) = H_2 e^{-i\lambda} x_2(d\lambda)$$

D'après la relation 2.16, on a

$$x_1(d\lambda) = (I - e^{-i\lambda} F_1)^{-1} v(d\lambda)$$

D'où

$$\begin{aligned} y_1(d\lambda) &= H_1 e^{-i\lambda} (I - e^{-i\lambda} F_1)^{-1} v(d\lambda) + w(d\lambda) \\ &= H_1 (e^{i\lambda} I - F_1)^{-1} v(d\lambda) + w(d\lambda) \end{aligned}$$

En adoptant les mêmes notations que dans ([16], [35], [36]), on en déduit que Y_{1t} est régulier de densité spectrale $f_1(\lambda)$ définie telle que :

$$f_1(\lambda) = \begin{bmatrix} H_1 (e^{i\lambda} I - F_1)^{-1} & I \end{bmatrix} \begin{pmatrix} Q & S \\ S' & R \end{pmatrix} \begin{bmatrix} (e^{-i\lambda} I - F_1')^{-1} H_1' \\ I \end{bmatrix} \quad (2.18)$$

De même, on a :

$$X_{2t} = F_2 X_{2(t-1)}$$

D'où

$$(I - F_2 e^{-i\lambda}) x_2(d\lambda) = 0$$

On sait que $S_2(d\lambda) = E(x_2(d\lambda) \overline{x_2(d\lambda)})$, il vient alors :

$$(I - F_2 e^{-i\lambda}) S_{X_2}(d\lambda) = 0 \quad (2.19)$$

Si on pose :

$$S_2(d\lambda) = E(y_2(d\lambda) \overline{y_2(d\lambda)})$$

On aura d'après la relation 2.14 :

$$S_2(d\lambda) = H_2 S_{X_2}(d\lambda) H_2'$$

De la relation 2.19 on peut déduire que Y_{2t} est singulier de mesure spectrale $S_2(d\lambda)$ concentrée sur les λ tels que $e^{i\lambda} \in \sigma(F_2)$.

D'après la relation 2.17, sachant que les espaces gaussiens engendrés par Y_{1t} et Y_{2t} sont orthogonaux, on a :

$$S(\lambda) = \int_{-\pi}^{+\pi} f_1(u) du + S_2(\lambda)$$

Si l'on remarque que $f_1(\lambda)$ et $S_2(\lambda)$ forment la décomposition de Lebesgue unique de $S(\lambda)$ en sa densité spectrale et sa fonction de saut, on déduit que :

$$\begin{aligned} f_1(\lambda) &= f(\lambda) && \text{p.p. pour la mesure de Lebesgue} \\ S_2(\lambda) &= S_s(\lambda) \end{aligned}$$

■

Proposition 17 *Si Y_t est régulier, alors ses représentations minimales sont purement non déterministes.*

Démonstration : Si Y_t est régulier, on a $S_s(\lambda) = 0$.

Soit (X_t, W_t) une \mathbb{H} – représentation minimale de Y_t et (X_{1t}, X_{2t}) la décomposition de X_t dans \mathcal{B}_X . D'après la proposition précédente, il vient que $Y_{2t} = 0$ presque sûrement. Sachant que X_{2t} est une \mathbb{H} – représentation minimale de Y_{2t} , on en déduit que $X_{2t} = 0$. Donc X_t se réduit à sa composante purement non déterministe. ■

Remarque 18 *En résumé, si Y_t admet des représentations finies, sa densité spectrale est rationnelle en $e^{i\lambda}$ et sa fonction de saut n'a qu'un nombre fini de discontinuités.*

2.2.4 Représentation état-espace d'un processus ARMA

Dans ce paragraphe, on va montrer qu'il est toujours possible de représenter les modèles ARMA sous la forme de Kalman-Bucy ou état-espace. En effet, supposons que $r = \max(p, q + 1)$, alors un processus ARMA(p, q) centré X_t s'écrit tel que :

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_r X_{t-r} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_{r-1} \varepsilon_{t-r+1}$$

ou bien

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_r B^r) X_t = (1 + \theta_1 B + \dots + \theta_{r-1} B^{r-1}) \varepsilon_t \quad (2.20)$$

Considérons l'équation d'état suivante :

$$Y_{t+1} = F Y_t + V \quad (2.21)$$

où Y_{t+1} est un vecteur dont la $j^{\text{ème}}$ composante est notée $Y_{j,t+1}$, F la matrice de transition définie telle que

$$F = \begin{pmatrix} \varphi_1 & \varphi_2 & \dots & \varphi_{r-1} & \varphi_r \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

et V le vecteur innovant défini tel que :

$$V = \begin{pmatrix} \varepsilon_{t+1} \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

L'équation d'observation s'écrit telle que :

$$X_t = (1 \quad \theta_1 \quad \theta_2 \quad \dots \quad \theta_{r-1}) Y_t \quad (2.22)$$

On va vérifier que les relations 2.21 et 2.22 décrivent le même processus que la relation 2.20. En effet, il est évident que la seconde ligne de l'équation 2.21 nous donne $Y_{2,t+1} = Y_{1t}$, la troisième ligne nous donne $Y_{3,t+1} = Y_{2t} = Y_{1,t-1}$ et en général on a :

$$Y_{j,t+1} = B^{j-1} Y_{1,t+1}$$

De l'équation 2.21 on déduit que :

$$Y_{1,t+1} = (\varphi_1 + \varphi_2 B + \varphi_3 B^2 + \dots + \varphi_r B^{r-1}) Y_{1t} + \varepsilon_{t+1}$$

ou bien

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_r B^r) Y_{1,t+1} = \varepsilon_{t+1} \quad (2.23)$$

Par ailleurs, l'équation 2.22 s'écrit telle que :

$$X_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_{r-1} B^{r-1}) Y_{1t} \quad (2.24)$$

Si on multiplie la relation 2.24 par le polynome $(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_r B^r)$ et en utilisant l'égalité 2.23, on retrouve l'équation 2.20 :

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_r B^r) X_t = (1 + \theta_1 B + \dots + \theta_{r-1} B^{r-1}) \varepsilon_t$$

La forme de représentation état-espace est très utile dans les problèmes de prévision, et de plus elle nous permet d'utiliser toutes les propriétés des modèles markoviens.

2.3 Dualité des représentations et principe du retournement du temps

2.3.1 Représentation duale d'un processus aléatoire

Supposons que l'on dispose d'un processus X_t identifié comme un $ARMA(p, q)$ qui peut être exprimé tel que :

$$P(B)X_t = Q(B)\varepsilon_t \quad (2.25)$$

Box et Jenkins ([6]) ont constaté que s'ils remplaçaient l'opérateur B par son dual F dans l'expression 2.25, ils obtenaient le modèle retourné ou modèle dual possédant les mêmes propriétés que le modèle originel et dans lequel la valeur X_t du processus est exprimée en fonction de ses valeurs futures. Cette heureuse constatation a permis à ses auteurs de définir la notion de "backforecast" qui leurs permettait de procéder à la prévision des valeurs du processus survenues avant le début des observations.

Il a été établi dans ([2], [6], [7], [9], [22], [30]) qu'un processus $ARMA(p, q)$ est stationnaire si les racines de l'équation $Q(B) = 0$ sont à l'extérieur du cercle unité et il est inversible si les racines de l'équation $P(B) = 0$ sont aussi à l'extérieur du cercle unité. Si la condition d'inversibilité est satisfaite, alors la relation 2.25 peut s'écrire sous la forme :

$$\prod_{i=1}^p (1 - G_i B) X_t = \prod_{j=1}^q (1 - H_j B) \varepsilon_t \quad (2.26)$$

où $\{G_i^{-1}, i = 1, 2, \dots, p\}$ et $\{H_j^{-1}, j = 1, 2, \dots, q\}$ sont les racines respectivement des polynômes $P(B)$ et $Q(B)$ et de plus :

$$\begin{aligned} G_i &< 1 & i = 1, 2, \dots, p \\ H_j &< 1 & j = 1, 2, \dots, q \end{aligned}$$

Donc, un processus $ARMA(p, q)$ stationnaire admet une structure de covariance unique mais la réciproque n'est en général pas vraie. On obtient l'unicité réciproque en introduisant les conditions d'inversibilité.

Il est clair que la structure de covariance du modèle 2.25 ne change pas si on y remplace l'opérateur B par son dual F . C'est ainsi qu'on obtient la représentation "backward" d'un processus $ARMA(p, q)$ qui s'écrit telle que :

$$P(F) X_t = Q(F) e_t$$

où e_t est aussi un processus d'innovation tel que $E(e_t^2) = E(\varepsilon_t^2)$.

Nous verrons plus loin la nécessité de déterminer des valeurs du processus antérieures à la série chronologique observée, dans le but d'exprimer la fonction de vraisemblance du processus. Supposons que l'on veuille déterminer la valeur $X_{-\ell}, \ell \geq 0$ du processus aléatoire. La structure de probabilité du modèle est déterminée indifféremment par le processus "forward" ou le processus "backward". Ainsi, grâce à la stationnarité du processus, la variable $X_{-\ell}$ a la même relation de probabilité avec la suite $\{X_1, X_2, \dots, X_N\}$ que la variable $X_{N+\ell+1}$ avec la suite $\{X_N, X_{N-1}, \dots, X_1\}$.

Pour essayer de comprendre et d'analyser cette procédure et les concepts qui s'y rattachent, nous allons essayer de formaliser le problème en termes d'espaces gaussiens ([9], [11]). En général, l'étude des processus aléatoires et l'analyse des séries chronologiques ont pour but pratique essentiellement les prévisions du comportement futur du processus ou des réalisations futures des séries chronologiques. L'instant ou l'origine à partir duquel on procède aux prévisions (dans une direction spécifiée du temps) est considéré comme l'instant présent du phénomène étudié. Avant cet instant c'est le passé et après cet instant se trouve le futur.

Considérant $(X_t, t = 1, 2, \dots, N)$ une réalisation d'un processus aléatoire centré et stationnaire et soit \mathcal{H} l'espace gaussien engendré par X_t . Si on considère la valeur X_N comme origine de prévision, on notera $\overrightarrow{\mathcal{H}}_N^-$ le sous-espace de Hilbert représentant le passé avant l'instant $t = N$ et $\overrightarrow{\mathcal{H}}_N^+$ le sous-espace de Hilbert représentant le futur après l'instant $t = N$. $\overrightarrow{\mathcal{H}}_N^-$ est alors l'espace

engendré par l'ensemble des valeurs $(X_t, t \leq N)$ et $\overrightarrow{\mathcal{H}}_N^+$ est l'espace engendré par l'ensemble des valeurs $(X_t, t \geq N + 1)$.

Nous allons maintenant parcourir l'espace temps dans le sens opposé, i.e. nous supposons que la série chronologique a été réalisée dans cet ordre $\{X_N, X_{N-1}, \dots, X_1\}$.

Si on considère la valeur X_1 comme origine de prévision, on notera $\overleftarrow{\mathcal{H}}_1^-$ le sous-espace de Hilbert représentant le passé avant l'instant $t = 1$ (selon le sens défini) et $\overleftarrow{\mathcal{H}}_1^+$ le sous-espace de Hilbert représentant le futur après l'instant $t = 1$ (toujours dans le sens défini). Alors, en considérant $E^{\overrightarrow{\mathcal{H}}_N^-}$ le projecteur du futur sur le passé, $E^{\overleftarrow{\mathcal{H}}_1^-}$ sera par dualité le projecteur du passé sur le futur. Cette façon de concevoir la représentation d'une série chronologique dans le sens naturel de développement du temps et dans le sens contraire est la base de la construction de l'algorithme de l'aller-retour.

2.3.2 Ordre d'un processus aléatoire

Le problème de la représentation du processus dual a été abordé par Box et Jenkins d'un point de vue analytique, en ce sens que la similitude entre le modèle "forward" et le modèle "backward" a été justifiée par le fait que les polynômes caractéristiques des deux modèles admettent les mêmes racines. En termes d'espaces gaussiens, on peut voir aisément que les produits scalaires (i.e., les fonctions de covariance) définis sur les espaces de Hilbert engendrés sont identiques du fait de leur symétrie. Il reste à comparer l'ordre des deux modèles. Il a été établi ([16], [35]) que l'ordre, noté O , d'une représentation ou d'un modèle est défini tel que :

$$O(X_t) = \dim E^{\overrightarrow{\mathcal{H}}_N^-} \left(\overrightarrow{\mathcal{H}}_N^+ \right)$$

Et par suite, si on note \overleftarrow{X}_t le processus retourné ou dual du processus X_t , on a :

$$O(\overleftarrow{X}_t) = \dim E^{\overleftarrow{\mathcal{H}}_1^-} \left(\overleftarrow{\mathcal{H}}_1^+ \right)$$

Par ailleurs, par rapport au segment d'observation $\{X_1, X_2, \dots, X_N\}$ du processus, les espaces de Hilbert $\overrightarrow{\mathcal{H}}_N^+$ et $\overleftarrow{\mathcal{H}}_1^+$ sont symétriques à cause de la stationnarité du processus, de même que les espaces de Hilbert $\overrightarrow{\mathcal{H}}_N^-$ et $\overleftarrow{\mathcal{H}}_1^-$. Avant de poursuivre, nous rappelons un résultat important sur les espaces de Hilbert.

Proposition 19 Soient H_1 et H_2 deux espaces de Hilbert, alors :

$$\dim E^{H_1}(H_2) = \dim E^{H_2}(H_1) \quad (2.27)$$

Cette relation est valable même en dimension infinie.

Démonstration : Définissons H_3 par :

$$H_3 = E^{H_1}(H_2) \oplus H_2$$

Il est clair qu'on a :

$$H_3 \subset H_1 \perp H_2 \ominus E^{H_1}(H_2)$$

Sachant que $H_3 \perp E^{H_1}(H_2)$, il vient que $H_3 \perp H_2$, d'où $E^{H_2}(H_1) = E^{H_2}(E^{H_1}(H_2))$, par conséquent :

$$\dim E^{H_1}(H_2) \leq \dim E^{H_2}(H_1)$$

En faisant jouer à H_1 et H_2 un rôle symétrique on déduit la relation 2.27. ■

Alors, d'après les résultats établis dans ([18], [35]), on obtient la proposition suivante :

Proposition 20 Le processus X_t et son retourné ou processus dual ont le même ordre, i.e. :

$$O(X_t) = \dim E^{\vec{\mathcal{H}}_N^-}(\vec{\mathcal{H}}_N^+) = \dim E^{\overleftarrow{\mathcal{H}}_1^-}(\overleftarrow{\mathcal{H}}_1^+) = O(\overleftarrow{X}_t)$$

Démonstration : En posant $H_1 = E^{\vec{\mathcal{H}}_N^-}(\vec{\mathcal{H}}_N^+)$ et $E^{\overleftarrow{\mathcal{H}}_1^-}(\overleftarrow{\mathcal{H}}_1^+) = H_2$ et en appliquant la proposition 19 on déduit la proposition 20. ■

On rappelle sous forme de remarque, un résultat dû à Ruckebush ([35]).

Remarque 21 Si H est la matrice de Hankel infinie associé au processus X_t , alors :

$$O(X_t) = \text{rang}(H)$$

Chapitre 3

Approximation de la vraisemblance d'un processus gaussien stationnaire

3.0.3 Introduction

Soit $X^N = (X_1, X_2, \dots, X_N)'$ l'observation d'un processus gaussien stationnaire et centré, de densité spectrale $f(\lambda, \theta)$, $\theta \in \Theta$. Lorsque cela ne sera pas nécessaire, nous ne mentionnerons pas θ . Nous supposons le modèle identifié et Θ un sous-ensemble de \mathbb{R}^k . Les problèmes statistiques qui nous intéressent sont généralement ceux de l'estimation du paramètre θ et ceux des tests sur des parties Θ_0 de Θ . Les techniques d'estimation et de tests sont fondées sur des méthodes de vraisemblance.

Nous allons nous intéresser en premier lieu à une méthode assez intuitive quoique peu générale de calcul et d'approximation de la vraisemblance au moyen du périodogramme ([6], [15], [22], [38]). Cette approximation a été introduite par Whittle et dans laquelle la forme quadratique dans la vraisemblance est approchée par $\frac{N}{4\pi} \int \frac{I_N(\lambda)}{f(\lambda, \theta)} d\lambda$, où I_N est le périodogramme. Cette approximation, valable sous des hypothèses assez générales, permet par la même occasion de montrer que, sous ces mêmes hypothèses ([6], [9]), la somme finie des carrés des innovations $\sum_{j=1}^N \varepsilon_j^2$ est une bonne approximation de la forme quadratique dans la vraisemblance de la série chronologique. Par

ailleurs, Box et Jenkins ont proposé l'expression de la fonction de vraisemblance à l'aide des espérances conditionnelles des innovations ε_j , $j \in \mathbb{Z}$, du processus qu'ils ont appelé fonction de vraisemblance conditionnelle. Elle est proportionnelle à la somme des carrés des innovations de l'échantillon $\sum_{j=1}^N \varepsilon_j^2$.

Les deux approximations sont donc essentiellement de même nature, l'une permettant de se guider sur les usages habituels de régression, l'autre nous conduit vers la statistique asymptotique classique ([14]), puisque $\frac{N}{4\pi} \int \frac{I_N(\lambda)}{f(\lambda, \theta)} d\lambda$ est à peu près le logarithme d'une famille exponentielle de paramètre $\frac{1}{f(\lambda, \theta)}$ et de statistique I_N .

3.1 Les différents types d'approximations

Soit $X^N = (X_1, X_2, \dots, X_N)'$ l'observation d'un processus gaussien stationnaire et centré $X = (X_t)_{t \in \mathbb{Z}}$. Il est supposé posséder une densité spectrale f .

Soit $\Gamma_N = E({}^t X^N X^N)$ la matrice de covariance de X^N . La ln-vraisemblance L_N vaut alors (si Γ_N est inversible) :

$$L_N = -\frac{1}{2} \left(\ln(\det \Gamma_N) + N \ln(2\pi) + {}^t X^N \Gamma_N^{-1} X^N \right)$$

Sauf dans le cas où le processus est autorégressif, Γ^{-1} est peu manipulable, et il est pratiquement impossible d'obtenir des résultats concrets d'estimation ou de test en utilisant L_N . Whittle a été le premier à introduire sans justification mathématique rigoureuse une approximation L_N^W de L_N ([38]), donnée par la relation suivante :

$$L_N^W = -\frac{N}{2} \left(C + \frac{1}{2\pi} \int_{-\pi}^{+\pi} \ln f(\lambda) d\lambda + \frac{1}{2\pi} \int_{-\pi}^{+\pi} \frac{I_N(\lambda)}{f(\lambda)} d\lambda \right)$$

où

$$I_N(\lambda) = \frac{1}{N} \left| \sum_{j=1}^N X_j \exp(ij\lambda) \right|^2$$

Le deuxième type d'approximation consiste à considérer tout processus aléatoire comme limite de processus autorégressif ([6], [9], [22]). Ce type d'approximation a été largement popularisé par Box et Jenkins.

Dans la littérature il n'est pas fait mention d'une justification mathématique ou de comparaison de ces procédures d'approximation des vraisemblances. Alors, nous allons essayer de préciser quelques justifications. Pour simplifier et sans perdre en généralité dans ce que l'on avance, on va restreindre le problème au cas suivant :

$$I. \quad a. \ln f \in L^1(d\lambda) \quad b. \frac{1}{f} \in L^1(d\lambda) \quad (3.1)$$

En effet, la première hypothèse équivaut à ce que le processus soit complètement non déterministe ([19], [34]) et la seconde signifie que le processus est linéairement indépendant ([34]). Sous la condition 3.1, et de manière équivalente on peut écrire toujours sous p_f :

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \quad p.s. \text{ ou dans } L^2$$

ε_t étant un bruit blanc stationnaire dit processus d'innovation.

Si on pose :

$$h(\exp(i\lambda)) = \sum_{j=0}^{\infty} c_j \exp(ij\lambda)$$

h est une fonction analytique dans le disque unité de module de carré intégrable ($h \in H^2$) et $f = |h|^2$. De plus, si on suppose :

$$II. \quad f > 0 \text{ et } h^{-1}(\exp(i\lambda)) = \sum_{j=0}^{\infty} d_j \exp(ij\lambda) \quad (3.2)$$

$$\text{où} \quad \ln d_0 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \ln f(\lambda) d\lambda$$

Alors, sous les conditions 3.1 et 3.2 h^{-1} a les mêmes propriétés que h . Plus précisément, on supposera que le processus bruit blanc est normé, et que l'on a :

$$\varepsilon_t = \sum_{j=0}^{\infty} d_j X_{t-j}$$

L'espace de Hilbert \mathcal{H} engendré par les $(X_t)_{t \in \mathbb{Z}}$ est canoniquement isomorphe à $L^2(f(\lambda), d\lambda)$, l'isométrie canonique étant définie par $X_t \rightarrow \exp(it\lambda)$. L'image de l'opérateur de translation $X_t \rightarrow X_{t+1}$ est la multiplication par $\exp(i\lambda)$. Nous utiliserons sans plus de commentaire cette isométrie pour faire des calculs. On notera la fonction d'autocovariances du processus par $R(n)$ et on a :

$$f(\lambda) = \sum_{n=-\infty}^{+\infty} R(n) \exp(in\lambda)$$

avec $R(n) = R(-n)$

On notera $E^{\mathcal{H}_t}$ l'opérateur de projection de $\mathcal{H} \rightarrow \mathcal{H}_t$ où \mathcal{H}_t est l'espace de Hilbert engendré par $\{X_j, j \leq t\}$.

Lemme 22 *Pour tout processus satisfaisant les conditions 3.1 et 3.2, on a :*

$$\tau X^N \Gamma_N^{-1} X^N = \sum_{t=-\infty}^N \{E^{\mathcal{H}_t}(\varepsilon_t)\}^2$$

Démonstration : En effet, si on pose :

$$E^{\mathcal{H}_t}(\varepsilon_t) = F_t X^N \quad \text{et} \quad X^N = \sum_{t=-\infty}^N M_t \varepsilon_t$$

où M_t est une matrice définie à partir de la représentation de X_t , et F_t est une matrice.

Par ailleurs, si on pose :

$$E(X^N \varepsilon_t) = M_t \quad \text{et} \quad E(\varepsilon_t^\tau X^N) = F_t \Gamma_N$$

On a :

$$\Gamma_N = \sum_{t=-\infty}^N \Gamma_N \tau F_t F_t \Gamma_N$$

car $\Gamma_N = E(X^N \tau X^N)$ implique $\Gamma_N = \sum_{t=-\infty}^N M \tau M_t$, car $E(\varepsilon_t \varepsilon_u) = 0$ si $t \neq u$.

Et, en multipliant des deux côtés par $\Gamma_N^{-1} X^N$ et ${}^\tau X^N \Gamma_N^{-1}$ on obtient le résultat.

Alors, pour $t > 0$, posons $\varepsilon_t = \eta_t + \alpha_t$ (η et α dépendant de N), avec :

$$\begin{aligned}\eta_t &= \sum_{j=0}^{t-1} d_j X_{t-j} \quad \text{si } t \leq N \\ \alpha_t &= \sum_{j=t}^{\infty} d_j X_{t-j} \quad \text{si } t \leq N \\ \eta_t &= \sum_{j=t-N}^{t-1} d_j X_{t-j} \quad \text{si } t > N\end{aligned}$$

On a alors :

$$\sum_{t=1}^{\infty} \eta_t^2 = \sum_{t=1}^{\infty} \left(\sum_{j=0}^{\infty} d_j X_{t-j}^N \right)^2 = \|d * \tilde{X}^N\|_{L^2}^2$$

où $\tilde{X}^N = (X_t^N)_{t \in \mathbb{Z}}$ tel que :

$$X_t^N = \begin{cases} X_t & \text{si } 1 \leq t \leq N \\ 0 & \text{si } t \notin [1, N] \end{cases}$$

et $d = (d_j)_{j \in \mathbb{N}}$. D'où, en appliquant l'inégalité de Bessel-Parseval, on a (voir [15]) :

$$\sum_{t=1}^{\infty} \eta_t^2 = \frac{N}{2\pi} \int_{-\pi}^{+\pi} \frac{I_N(\lambda)}{f(\lambda)} d\lambda$$

$\sum_{t=1}^{\infty} \eta_t^2$ conduit donc à approximation L_N^W de ${}^\tau X^N \Gamma_N^{-1} X^N$. ■

La forme de l'approximation liée aux innovations est par ailleurs $L^{BJ} = \sum_{t=1}^N \varepsilon_t^2$. Nous verrons un peu plus loin que :

$$\sum_{t=N+1}^{\infty} \eta_t^2 = o_{L_f^2}(1) \quad \text{et} \quad \sum_{t=1}^N \{\varepsilon_t - E^{\mathcal{H}_t}(\varepsilon_t)\}^2 = o_{L_f^2}(1)$$

Il est nécessaire d'étudier les formes quadratiques ${}^\tau X^N \Gamma_N^{-1} X^N$ non seulement sous une loi de probabilité f mais sous d'autres lois probabilités g . Dans le cadre des tests d'hypothèses sur le modèle il faudra étudier des expressions

du type ${}^{\tau} X^N (\Gamma_N^{-1}(f) - \Gamma_N^{-1}(g)) X^N$, où $\Gamma_N(f)$ est la matrice d'autocovariances associée à f et $\Gamma_N(g)$ celle associée à g . La variable aléatoire ε_t n'a évidemment le sens d'une innovation que sous la loi f , sinon elle représente seulement $\sum_{j=0}^{\infty} d_j(\cdot)$ où $d_j(\cdot)$ est soit $d_j(f)$ soit $d_j(g)$. L_N^W a donc une expression analytique simple sous f ou sous g , mais il n'en est pas de même pour L_N^{BJ} . Cependant, sur la ln-vraisemblance, les deux approximations se prêtent très mal aux calculs asymptotiques. Nous serons donc amenés à introduire un troisième type d'approximation.

3.2 Méthode d'approximation de Toeplitz et Szegö

Dans ce paragraphe nous proposons une autre méthode d'approximation de la vraisemblance que l'on va appeler "approximation de Toeplitz-Szegö". L'essentiel des résultats que l'on propose sont basés sur les travaux Davies ([12]), de Szegö ([20]) et de Dzaparidge ([14]). Nous rappelons d'abord que dans la suite $L^p(-\pi, \pi)$ sera noté $L^p(d\lambda)$, $1 < p < \infty$. Si $h \in L^1(d\lambda)$, on notera $T_N(h)$ sa transformée de Toeplitz ([20]), i.e. la matrice de taille (N, N) de terme général :

$$h_{m-n} = \int_{-\pi}^{+\pi} \exp\{i\lambda(m-n)\} h(\lambda) d\lambda \quad 1 < m, n < N$$

où h_k désigne le $k^{\text{ème}}$ coefficient de Fourier de h . $T_N(h)$ est une matrice hermitienne et on a en particulier :

$$\Gamma_N = E_f(X^N {}^{\tau} X^N) = T_N(h)$$

La matrice de covariance est donc la transformée de Toeplitz de la densité spectrale ([15], [20]).

Dans la suite, on s'intéresse au cas des fonctions h de la classe \mathcal{C} suivante :

1. h et $h^{-1} \in L^{\infty}$
 2. $\|h\|^2 = \sum_{k=-\infty}^{+\infty} kh_k^2 < \infty$
- (3.3)

La norme ci-dessus induit des propriétés intéressantes du type algèbre de Banach. Les densités spectrales seront choisies dans la classe \mathcal{C} . De toute manière pour avoir des théorèmes de type Szegö, des propriétés de ce genre

sont indispensables.

Sur l'espace de Banach des fonctions h telles que $\|h\| < \infty$, on peut mettre une norme équivalente ainsi définie :

$$\|h\|^2 = \sum_{n=1}^{\infty} |\omega_h(n)|^2$$

où

$$|\omega_h(n)|^2 = \sup_{|a| < \frac{1}{n}} \int |h(\lambda + a) - h(\lambda)|^2 d\lambda$$

On suppose qu'on a deux probabilités gaussiennes F et G définies sur \mathbb{R}^Z et possédant des densités spectrales f et g dans la classe \mathcal{C} . La condition 3.3 implique en particulier que pour tout N , $T_N f$ et $T_N g$ sont régulières. On peut alors écrire :

$$L_N^S = \frac{1}{2} \{ \Sigma_N + {}^\tau X^N ((T_N f)^{-1} - (T_N g)^{-1}) X^N \}$$

où

$$\Sigma_N = \ln \left(\frac{\det T_N f}{\det T_N g} \right)$$

Si on considère la transformation $Y_N = T_N^{-\frac{1}{2}}(f) X^N$, alors on a $E_F(Y {}^\tau Y) = I_N$ (identité), et en posant

$$A_N = I - T_N^{\frac{1}{2}} f T_N^{-1} g T_N^{\frac{1}{2}} f$$

où on note

$$T_N^{\frac{1}{2}} f = (T_N f)^{\frac{1}{2}}$$

On obtient alors :

$$L_N^S = \frac{1}{2} (\Sigma_N + {}^\tau Y^N A_N Y^N)$$

Par ailleurs, L_N^S peut être écrite telle que :

$$L_N^S = \frac{1}{2} \{ \Sigma_N + {}^\tau X^N (T_N f^{-1} - T_N g^{-1}) X^N \} = \frac{1}{2} (\Sigma_N + {}^\tau Y^N C_N Y^N)$$

où

$$C_N = T_N^{\frac{1}{2}} f (T_N f^{-1} - T_N g^{-1}) T_N^{\frac{1}{2}} f$$

On peut montrer encore que :

$$L_N^W = \frac{1}{2} \left\{ \Sigma_N + N \int I_N(\lambda) \left(\frac{1}{f(\lambda)} - \frac{1}{g(\lambda)} \right) d\lambda \right\}$$

où

$$I_N(\lambda) = \frac{1}{N} \left| \sum_{k=1}^N X_k \exp(ik\lambda) \right|^2$$

Si R_N est une suite de matrices symétriques, nous noterons $\mu(R_N)$ la probabilité $\frac{1}{N} \sum_{i=1}^N \delta_{\mu_i^N}$, où $(\mu_i^N)_{i=1, \dots, N}$ est l'ensemble des valeurs propres de R_N et δ_a la masse unité en a .

Dans la suite nous allons généraliser deux types de résultats. Le premier est le théorème de Szegö ([20]) :

Théorème 23 (de Szegö) *Si $h \in L^\infty$, alors $\mu(T_N h) \xrightarrow{\text{Lois}} \mathcal{L}h(U)$, où U est une variable aléatoire uniforme et $\mathcal{L}h(U)$ désigne la loi de $h(U)$.*

Le théorème de Szegö s'applique donc à $B_N = T_N \psi$, où $\psi = 1 - \frac{f}{g}$. Nous allons montrer que la conclusion subsiste pour A_N et C_N .

Théorème 24 *Si $f, g \in \mathcal{C}$, alors :*

$$\begin{aligned} \mu(A_N) &\xrightarrow{\text{Loi}} \mathcal{L}h(U) \\ \text{et } \mu(C_N) &\xrightarrow{\text{Loi}} \mathcal{L}h(U) \end{aligned}$$

Corollaire 25 *Si $f, g \in \mathcal{C}$, $\phi_N(t) = E \exp(tL_N)$, alors :*

$$\phi_N^{\frac{1}{N}}(t) \xrightarrow{\quad\quad\quad} \exp L(t)$$

où

$$L(t) = \frac{1}{4\pi} \int_{-\pi}^{+\pi} \left\{ t \ln \frac{f(\lambda)}{g(\lambda)} - (1-t) \frac{f(\lambda)}{g(\lambda)} \right\} d\lambda$$

Un autre type de résultat concerne l'approximation de la vraisemblance.

Proposition 26 *Si $f, g \in \mathcal{C}$, alors :*

$$E_F (L_N - L_N^S)^2 = o_{L^2}(1)$$

La proposition subsiste si on y remplace L_N ou L_N^S par L_N^W .

De plus, cette proposition implique que :

$$E_F \exp(\gamma |L_N - L_N^S|) < \infty \quad \text{pour } \gamma \text{ assez petit}$$

Nous verrons plus loin d'autres versions de cette proposition.
Un autre résultat concerne l'approximation L_N^{BJ} de la vraisemblance.

Proposition 27 Si $f \in \mathcal{C}$ où \mathcal{C} est la classe des fonctions f définies plus haut, alors:

$$|L_N^{BJ}(X^N) - L_N(X^N)| = o_{L^p}(1)$$

Démonstration : On peut montrer que la proposition 27 se déduit de la proposition 26. En effet, soit :

$$\gamma_j = \sum_{k=j}^{\infty} |d_k|$$

De plus, on suppose que :

$$\sum_j j\gamma_j^2 < \infty$$

Alors, si $f \in \mathcal{C}$, on a :

$$E \sum_{t=N+1}^{\infty} \left(\sum_{j=t-N+1}^{t-1} d_j X_{t-j} \right)^2 < \sum_{j=0}^{\infty} j\gamma_j^2$$

Et

$$\begin{aligned} E \left\{ \varepsilon_t^2 - (E^{\mathcal{H}_t}(\varepsilon_t))^2 \right\} &< 2 \sum_{t=1}^N E(\alpha_t^2) < 2 \sum_{t=1}^{\infty} Var(\alpha_t^2) \\ &< 2 \sum_{t=1}^{\infty} \sum_{j=t+1}^{\infty} d_j \gamma_j^2 = o_{L^2}(1) \end{aligned}$$

d'où le résultat.

■

On va maintenant démontrer le théorème 24.

Démonstration : Soit $h \neq 0$ et $x \in \mathbb{R}^N$, alors :

$${}^{\tau} x T_N(h) x = \int \left| \sum_{j=1}^N x_j \exp(i\lambda j) \right|^2 h(\lambda) d\lambda$$

Si μ_j^N est la suite des valeurs propres de $T_N(h)$, alors $\mu_j^N \in [ess \inf(h), ess \sup(h)]$
 $\forall j = 1, \dots, N$ et $\forall N \in \mathbb{N}$.

Remarquons aussi que la démonstration du théorème de Szegö ([15], [20]) consiste justement à montrer que :

$$\int_{\mathbb{R}} \ln(1 - tx) d\mu(T_N h)(x)$$

et converge vers

$$\int_{\mathbb{R}} \ln(1 - tx) h(x) dx$$

A partir de là on déduit la convergence des moments et celle de $\mu(T_N h)$. ■

Par ailleurs

$$E \left\{ \exp \left(\frac{1}{2} t^r Y^N T_N(\psi) Y^N \right) \right\} = \frac{1}{2} \int_{\mathbb{R}} \ln(1 - tx) d\mu(T_N \psi)$$

On en déduit par le théorème de Szegö :

Lemme 28 Si $1 - \frac{f}{g}$ est bornée, alors :

$$\frac{1}{N} \ln E_F(\exp t L_N^S) = \ln \frac{\psi_N^S(t)}{N}$$

converge vers $L(t)$, pour $t \in [\text{ess inf}(\psi), \text{ess sup}(\psi)]$ si $\text{ess inf}(\psi) < 0$ (resp. $\text{ess sup}(\psi) > 0$), mais ces bornes sont $-\infty$ (resp. $+\infty$) si $\text{ess inf}(\psi) > 0$ (resp. $\text{ess sup}(\psi) < 0$).

Lemme 29 Considérons maintenant les fonctions $\phi_1, \dots, \phi_j, \dots, \phi_\ell$ de classe \mathcal{C} , de coefficients de Fourier notés $\phi_{k,j}$ et soit :

$$D_N(\phi_1, \dots, \phi_\ell) = T_N(\phi_1, \dots, \phi_\ell) - (T_N \phi_1, \dots, T_N \phi_\ell)$$

Alors :

$$|T_N D_N| < 4 \sum_{j=1}^{\ell} \|(\phi_1, \dots, \phi_j)\| \|(\phi_{j+1}, \dots, \phi_\ell)\|$$

Démonstration : En effet, soit $A_N \cap \mathbb{Z}$, où $A_N =]-\infty, 0] \cap [N+1, +\infty[$.
Alors

$$\begin{aligned} (D_N(\phi_1, \dots, \phi_\ell))_{\alpha, \beta} &= \sum_{j=1}^{\ell} \sum_{\substack{i \neq j \\ 1 \leq i \leq \ell}} \sum_{\lambda_j \in A_N} \sum_{\lambda_i \in \mathbb{Z}} T_N \phi_{1, \alpha - \lambda_1} T_N \phi_{2, \lambda_1 - \lambda_2} \dots \\ &\quad \dots T_N \phi_{j, \lambda_{j-1} - \lambda_j} \dots T_N \phi_{\ell, \lambda_\ell - \beta} \end{aligned}$$

d'où

$$\begin{aligned}
|Tr(D_N)| &= \left| Tr \left(\sum_{j=1}^{\ell} \sum_{\alpha=1}^N \sum_{\lambda_j \in A_N} T_N(\phi_1, \dots, \phi_j)_{\alpha-\lambda_j} T_N(\phi_{j+1}, \dots, \phi_{\ell})_{j-\alpha} \right) \right| \\
&< \sum_{j=1}^{\ell} \sum_{\alpha=1}^N \left\{ \sum_{\lambda_j \in A_N} T_N(\phi_1, \dots, \phi_j)_{\alpha-\lambda_j}^2 \right\}^{\frac{1}{2}} \left\{ \sum_{\lambda_j \in A_N} T_N(\phi_{j+1}, \dots, \phi_{\ell})_{\alpha-\lambda_j}^2 \right\}^{\frac{1}{2}}
\end{aligned}$$

Or

$$\sum_{\alpha=1}^N \sum_{\lambda_j \in A_N} \phi_{j', \alpha-\lambda_j}^2 < 2 \sum_{\alpha \in \mathbb{Z}} \alpha \phi_{j', \alpha}^2$$

Donc

$$|Tr(D_N)| < 4 \sum_{j=1}^{\ell} \|\phi_1, \dots, \phi_j\| \|\phi_{j+1}, \dots, \phi_{\ell}\|$$

D'où le lemme 29. ■

Pour m strictement positive constante d'isomorphisme entre la norme $\|\cdot\|$ et la norme $\|\|\cdot\|\|$, on a donc :

$$|Tr D_N| < 4m \sum_{j=1}^{\ell} \|\|\phi_1, \dots, \phi_j\|\| \|\|\phi_{j+1}, \dots, \phi_{\ell}\|\|$$

Supposons maintenant que pour tout j , on ait $\|\phi_j\|_{\infty} < \rho$, $\|\phi_j\| < A$. On a, en utilisant la définition de $\|\|\phi_j\|\|$:

$$|Tr D_N| < \sum_{j=1}^{\ell} 4mA^2 j(\ell-j)\rho^{\ell-2} < \sum_{j=1}^{\ell} 4mA^2 \ell^3 \rho^{\ell-2}$$

Soient $U_1(z), \dots, U_{\ell}(z)$ des séries entières de rayon de convergence > 1 . Posons :

$$U_j(z) = \sum_{n=0}^{\infty} u_{n,j}(z) \quad \text{et} \quad |U_j|(z) = \sum_{n=0}^{\infty} |u_{n,j}|(z)$$

Alors, $U\phi$ et UM sont définies de manière évidente si ϕ est une fonction telle que $\|\phi\|_{\infty} < 1$ et si M est une matrice telle que $\|M\| < 1$. Remarquons que $\|h\|_{\infty} < 1$ implique $\|T_N h\| < 1$. Posons enfin $\bar{U} = |U''| + 3|U''| + 2|U|$.

Lemme 30 *Sous les hypothèses du lemme 29, on a :*

$$|Tr E_N| < A^2 m \rho^{-4} \prod_{j=1}^{\ell} \overline{U}_j(\rho)$$

où

$$E_N(U_1, \dots, U_\ell, \phi_1, \dots, \phi_\ell) = \prod_{j=1}^{\ell} T_N(U_j \phi_j) - \prod_{j=1}^{\ell} U_j(T_N \phi_j)$$

Démonstration : Il est clair que l'on a :

$$|Tr(E_N)| < \sum_{\substack{n_1, \dots, n_\ell \\ 0 < n_j < \infty}} \left(\prod_{j=1}^{\ell} |u_{j, n_j}| \right) Tr \left\{ \prod_{j=1}^{\ell} T_N(\phi_j^{n_j}) - \prod_{j=1}^{\ell} (T_N \phi_j)^{n_j} \right\}$$

Par ailleurs,

$$\begin{aligned} Tr \left\{ \prod_{j=1}^{\ell} T_N(\phi_j^{n_j}) - \prod_{j=1}^{\ell} (T_N \phi_j)^{n_j} \right\} &< \left| Tr \left(\prod_{j=1}^{\ell} T_N(\phi_j^{n_j}) - T_N \prod_{j=1}^{\ell} \phi_j^{n_j} \right) \right| \\ &+ \left| Tr \left\{ T_N \left(\prod_{j=1}^{\ell} \phi_j^{n_j} \right) - \prod_{j=1}^{\ell} (T_N \phi_j)^{n_j} \right\} \right| \end{aligned}$$

avec

$$\begin{aligned} \left| Tr \left\{ \prod_{j=1}^{\ell} T_N(\phi_j^{n_j}) - T_N \prod_{j=1}^{\ell} \phi_j^{n_j} \right\} \right| &< 4 \sum_{j=1}^{\ell} \|\phi_1^{n_1} - \phi_j^{n_j}\| \|\phi_{j+1}^{n_{j+1}} \phi_\ell^{n_\ell}\| \\ &< 4mA^2 \rho^{-2} \rho^{\sum_{j=1}^{\ell} n_j} (n_1 + \dots + n_\ell)^3 \\ &< 4mA^2 \rho^{-2} \prod_{j=1}^{\ell} n_j^3 \rho^{n_j} \end{aligned}$$

et

$$\left| Tr \left\{ T_N \left(\prod_{j=1}^{\ell} \phi_j^{n_j} \right) - \prod_{j=1}^{\ell} (T_N \phi_j)^{n_j} \right\} \right| < 4mA^2 \rho^2 \prod_{j=1}^{\ell} n_j^3 \rho^{n_j}$$

et d'après le lemme 29, on a :

$$\sum_{n_j=0}^{\infty} |u_{j, n_j}| n_j^3 \rho^{n_j} < \rho^{-2} \overline{U}_j(\rho)$$

D'où le lemme 30. ■

Reprenons la démonstration du théorème 24.

Démonstration : On sait que :

$$|\ln \phi_N(t) - \ln \phi_N^S(t)| = \frac{1}{2} \left| \sum_{i=1}^N \{ \ln(1 - \nu_i^N t) - \ln(1 - \mu_i^N t) \} \right|$$

où μ_i^N est la suite des valeurs propres de A_N et ν_i^N celles de B_N . Les fonctions $\ln \phi_N(t)$ et $\ln \phi_N^S(t)$ sont toutes deux définies pour tout $t \in [0.1]$ et $\|\mu_i^N\|_\infty < 1$. Alors, pour tout $t \in [0.1]$, on a :

$$\begin{aligned} |\ln \phi_N(t) - \ln \phi_N^S(t)| &= \left| \sum_{k=0}^{\infty} \frac{t^k}{k!} \text{Tr} \left\{ (A_N)^k - (I - B_N)^k \right\} \right| \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \sum_{\ell=1}^k k_\ell \text{Tr} \left\{ A_N^\ell - \overline{B}_N^\ell \right\} \end{aligned}$$

où $\overline{B}_N = I - B_N$.

Posons : $U_1(z) = (1 - z)^{\frac{1}{2}}$, $U_2(z) = (1 - z)^{-1}$, $\psi_1 = 1 - \frac{f}{M_1}$ et $\psi_2 = 1 - \frac{g}{M_1}$ telles que $\max(\|f\|_\infty, \|g\|_\infty) < \infty$. De plus,

$$\begin{aligned} (T_N f)^{\frac{1}{2}} &= M_1^{\frac{1}{2}} U_1(T_N \psi_1) \\ (T_N g)^{-1} &= M_1^{-1} U_2(T_N \psi_2) \end{aligned}$$

et

$$\begin{aligned} A_N &= U_1(T_N \psi_1) U_2(T_N \psi_2) U_1(T_N \psi_1) \\ B_N &= T_N \{ U_1(\psi_1) U_2(\psi_2) U_1(\psi_1) \} \end{aligned}$$

Alors

$$\begin{aligned} |\text{Tr}(A_N^\ell - B_N^\ell)| &= \left| \text{Tr} \{ T_N (U_1 \psi_1 U_2 \psi_2 U_1 \psi_1) \}^\ell - U_1^\ell(T_N \psi_1) U_2^\ell(T_N \psi_2) U_1^\ell(T_N \psi_1) \right| \\ &< \left| \text{Tr} \left\{ (T_N (U_1 \psi_1 U_2 \psi_2 U_1 \psi_1))^\ell - T_N \left| (U_1 \psi_1 U_2 \psi_2 U_1 \psi_1)^\ell \right| \right\} \right| \\ &\quad + \left| \text{Tr} \{ T_N (U_1^\ell \psi_1 U_2^\ell \psi_2 U_1^\ell \psi_1) - U_1^\ell(T_N \psi_1) U_2^\ell(T_N \psi_2) U_1^\ell(T_N \psi_1) \} \right| \\ &= a_\ell + b_\ell \end{aligned}$$

D'après l'isomorphisme vu plus haut, on a :

$$a_\ell < 4m_1 \ell^3 \|U_1 \psi_1 U_2 \psi_2 U_1 \psi_1\|^2 \|U_1 \psi_1 U_2 \psi_2 U_1 \psi_1\|_\infty^{\ell-2}$$

avec

$$\begin{aligned} \|U_1\psi_1U_2\psi_2U_1\psi_1\|_\infty &< \left\| \frac{f}{g} \right\|_\infty \\ \|U_1\psi_1U_2\psi_2U_1\psi_1\| &< 2m \max(\|f\|_\infty, \|g^{-1}\|_\infty) \times \max(\|f\|, \|g\|) \\ \|g^{-1}\| &< 2m \|g^{-1}\|_\infty \|g\| \end{aligned}$$

Et d'après le lemme 30, on déduit que :

$$b_\ell < 4m_1 \left\| \overline{U}_1^\ell(\psi_1) \overline{U}_2^\ell(\psi_2) \overline{U}_1^\ell(\psi_1) \right\| \times \|\psi_1\psi_2\psi_1\|_\infty$$

Cependant, on remarque que l'on peut modifier légèrement le résultat du lemme 30 car les coefficients de U_1 et U_2 sont positifs à l'exception du premier coefficient de U_1 . On peut donc remplacer dans le lemme la majoration

$$\sum |u_{n,1}| |u_{m,2}| |u_{\ell,1}|$$

par

$$\sum (u_{n,1}u_{m,2}u_{\ell-1,1}\rho^{\ell+m+n} + 2|u_{n,1}u_{m,2}\rho^{\ell+m+n}|)$$

où

$$\rho = \max\left(\frac{\|f\|_\infty}{M_1}, \frac{\|g\|_\infty}{M_1}\right) < s$$

Un calcul immédiat nous donne :

$$\begin{aligned} \left| \overline{U}_i^k(\rho) \right| &< 16 \left| \overline{U}_i^{k-3}(\rho) \right| \overline{U}_i(\rho) \quad i = 1, 2 \\ \text{et} \quad U_1(\rho)U_2(\rho)U_1(\rho) &= 1 \end{aligned}$$

Finalement, il existe une constante $K = 24m\overline{U}_1^2(\rho)\overline{U}_2^2(\rho)\left\| \frac{f}{g} \right\|_\infty^{-4}$ indépendante de k telle que l'on ait :

$$|Tr(A_N^k - B_N^k)| < K$$

On a donc d'après le lemme 28 :

$$\frac{1}{N} \ln \phi_N(t) \xrightarrow{\quad\quad\quad} L(t) \quad 0 < t < 1$$

Puisque

$$\frac{1}{N} \left(|Tr(A_N^k)|^{\frac{1}{k}} - |Tr(B_N^k)|^{\frac{1}{k}} \right) < \frac{1}{N} |Tr(A_N^k - B_N^k)|^{\frac{1}{k}}$$

Il est aisé de conclure que $\frac{1}{N} \lim Tr(A_N^k) = \frac{1}{N} \lim Tr(B_N^k)$, donc $K \lim_N \mu(A_N) = \lim_N \mu(B_N)$ puisque les distributions $\mu(A_N)$ sont contenues dans un compact fixe de \mathbb{R} . Ainsi se termine la démonstration du théorème 24. ■

Chapitre 4

Détermination de la vraisemblance inconditionnelle d'un processus aléatoire par la méthode de l'aller-retour

Au chapitre précédent on a exprimé la vraisemblance L_N à l'aide des espérances conditionnelles des innovations ε_j , $j \in \mathbb{Z}$, du processus. La forme quadratique de L_N a pour approximation naturelle $\sum_{j=1}^N \varepsilon_j^2$ ([6]). La méthode de Box et Jenkins est une application particulière de cette approximation dans le cas d'un spectre rationnel. Les ε_j ne sont en général pas calculables puisqu'ils dépendent de l'observation avant l'instant $t = 0$. Cependant, une procédure itérative permet dans ce cas de calculer la vraisemblance. Comme il semble que la convergence de cet algorithme n'a pas été établie, nous avons consacré ce chapitre à l'étude du problème.

4.1 Position du problème

Supposons que l'étape de l'identification se soit soldée par le choix d'un modèle adéquat pour la série chronologique étudiée. Pour les étapes suivantes de l'étude, i.e., l'estimation des paramètres et les tests d'hypothèses, il est nécessaire d'exprimer la fonction de vraisemblance du processus avec autant

de précision que possible. Car, il est bien établi qu'une bonne approximation de la fonction de vraisemblance produira de meilleurs résultats pour l'estimation et les tests des paramètres du modèle. Par ailleurs, pour exprimer la vraisemblance d'une série chronologique centrée et identifiée par un modèle *ARMA*, nous avons besoin de valeurs initiales. Pour cela, nous allons utiliser la procédure proposée par Box et Jenkins ([6]) basée sur le principe de la prévision du passé et appelée "backforecast" ([7], [10]). Mais, le problème de la convergence de cette procédure se pose. Pour répondre à cette question, on se propose de multiplier le nombre de cycles dans la procédure de Box et Jenkins. Et ainsi, on va générer un mouvement de va et vient entre le futur et le passé de la série chronologique en utilisant les deux formes duales de la représentation d'un processus. A chaque passage de cet algorithme les valeurs passées et futures de la série chronologique sont déterminées par prévision. A partir d'un certain nombre K de cycles il apparait que les valeurs que l'on détermine à chaque passage ne changent pratiquement plus, et ainsi on soupçonne l'existence d'un point fixe.

4.2 Approximation de la vraisemblance exacte d'un processus *MA*

4.2.1 La vraisemblance exacte d'un processus *MA*(q)

Supposons qu'on associe à la série chronologique centrée X_1, X_2, \dots, X_N , un modèle *MA*(q) stationnaire défini tel que :

$$X_t = Q(B) \varepsilon_t \quad (4.1)$$

L'expression 4.1 peut être écrite sous la forme :

$$\varepsilon_t = X_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (4.2)$$

où ε_t est un processus d'innovations de distribution $N(0, \sigma_\varepsilon^2)$.

En notant $\Theta = (\theta_1 \ \theta_2 \ \dots \ \theta_q)'$, la fonction ln-vraisemblance de ce modèle peut s'exprimer telle que :

$$\ln(L(\Theta, \sigma_\varepsilon^2)) = l(\Theta, \sigma_\varepsilon^2) = C - N \ln \sigma_\varepsilon - \frac{1}{2\sigma_\varepsilon^2} S(\Theta/X_1, \dots, X_N)$$

$$\text{où } S(\Theta/X_1, \dots, X_N) = \sum_{t=-\infty}^N \varepsilon_t^2(\Theta/X_1, \dots, X_N) \quad (4.3)$$

Il est à noter que le logarithme de la vraisemblance conditionnelle $l = \ln L$ fait intervenir les observations uniquement à travers la somme des carrés $S(\Theta/X_1, \dots, X_N)$. Donc, pour déterminer la fonction de vraisemblance inconditionnelle, il suffit de calculer la somme des carrés inconditionnelle. Il a été établi qu'en fait $S(\Theta/X_1, \dots, X_N)$ peut être approchée par une somme finie, car à partir d'un certain indice $t = 1 - q$ les ε_t sont négligeables.

On remarque d'après la relation 4.2, que la somme de carrés inconditionnelle 4.3 ne peut être calculée que si l'on se donne les vecteurs des valeurs initiales X^0 et ε^0 définis tels que :

$$X^0 = (X_0 \quad X_{-1} \quad \dots \quad X_{1-p})' \quad \text{et} \quad \varepsilon^0 = (\varepsilon_0 \quad \varepsilon_{-1} \quad \dots \quad \varepsilon_{1-q})'$$

Le problème donc est la détermination des valeurs initiales X^0 et ε^0 .

4.2.2 Détermination des valeurs initiales par l'algorithme de l'aller-retour

Considérons N observations X_1, X_2, \dots, X_N d'un processus aléatoire $MA(q)$ centré et stationnaire. Ces observations vérifient alors la relation 4.1.

La méthode de l'aller-retour proposée par Box-Jenkins ([6]), permet de déterminer les q valeurs initiales du processus aléatoire, qui serviront à écrire explicitement l'expression de la vraisemblance inconditionnelle de l'échantillon aléatoire. L'idée de l'aller-retour est basée sur le principe du retournement du temps et de la prévision du passé ([6], [35]). En effet, il a été établi dans ([6], [35]) qu'il est possible de représenter un processus aléatoire soit dans le sens croissant du temps, i.e. sur l'espace de Hilbert $\overrightarrow{\mathcal{H}}_N$ engendré par $\{X_1, X_2, \dots, X_N\}$, ou dans le sens contraire, i.e. sur l'espace de Hilbert $\overleftarrow{\mathcal{H}}_N$ engendré par $\{X_N, X_{N-1}, \dots, X_1\}$. Cependant, la structure de covariance du phénomène ne change pas.

La représentation duale de 4.1 s'obtient en y remplaçant B par F ([2], [6], [30]) telle que :

$$X_t = Q(F) e_t \quad \text{où} \quad e_t \sim N(0, \sigma_e^2) \quad \forall t \quad (4.4)$$

Dans l'espace $\overrightarrow{\mathcal{H}}_N$ les valeurs $(X_{N+\ell}, \ell > 0)$ du processus sont estimées par prévision d'origine N . Et dans l'espace $\overleftarrow{\mathcal{H}}_N$ les valeurs X_0, X_{-1}, \dots seront estimées par prévision d'origine 1 ([2], [4], [6], [22]).

La fonction de prévision d'origine N pour le processus forward est définie telle que :

$$\widehat{X}_{N+\ell} = Q(B) \widehat{\varepsilon}_{N+\ell} \quad (4.5)$$

avec

$$\widehat{X}_{N+l} = \begin{cases} \widehat{X}_{N+l} & \text{si } l > 0 \\ X_{N+l} & \text{si } l \leq 0 \end{cases} \quad \text{et} \quad \widehat{\varepsilon}_{N+l} = \begin{cases} 0 & \text{si } l > 0 \\ \varepsilon_{N+l} & \text{si } l \leq 0 \end{cases}$$

Il est évident que $\widehat{X}_{N+l} = 0$ dès que $l \geq q + 1$.

Nous allons adopter les notations suivantes :

- le vecteur des prévisions d'origine N des q valeurs futures du processus X_t obtenues au $k^{\text{ème}}$ passage de l'algorithme d'aller-retour est noté par :

$$\widehat{X}_{(N,q)}^{(k)} = \left(\widehat{X}_{N+1}^{(k)} \quad \widehat{X}_{N+2}^{(k)} \quad \dots \quad \widehat{X}_{N+q}^{(k)} \right)'$$

- le vecteur des q dernières innovations du processus forward obtenu au $k^{\text{ème}}$ passage de l'algorithme est noté tel que :

$$\vec{\eta}^{(k)} = \left(\varepsilon_{N-q+1}^{(k)} \quad \varepsilon_{N-q+2}^{(k)} \quad \dots \quad \varepsilon_N^{(k)} \right)'$$

- la matrice triangulaire Θ d'ordre q est définie telle que :

$$\Theta = \begin{pmatrix} \theta_q & \theta_{q-1} & \dots & \theta_1 \\ 0 & \theta_q & \dots & \theta_2 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \theta_q \end{pmatrix}$$

De la relation 4.5 et pour $l = 1, 2, \dots, q$, on obtient au $k^{\text{ème}}$ passage de l'algorithme l'équation :

$$\widehat{X}_{(N,q)}^{(k)} + \Theta \vec{\eta}^{(k)} = 0 \quad (4.6)$$

Considérons les notations suivantes :

- le vecteur des q dernières observations du processus forward est noté tel que :

$$\vec{X}_{(q)} = \left(X_{N-q+1} \quad X_{N-q+2} \quad \dots \quad X_N \right)'$$

- la matrice diagonale G_B d'ordre q sera notée telle que :

$$G_B = \text{diag} \left(G_{N-q}(B) \quad G_{N-q+1}(B) \quad \dots \quad G_{N-1}(B) \right)$$

où

$$G_{N-\ell}(B) = \sum_{j=0}^{N-\ell} a_j B^j \quad \ell = 1, 2, \dots, q$$

- la matrice symétrique A d'ordre q sera notée telle que :

$$A = \begin{pmatrix} a_{N-q+1} & a_{N-q+2} & \dots & a_N \\ a_{N-q+2} & a_{N-q+3} & \dots & a_{N-1} \\ \dots & \dots & \dots & \dots \\ a_N & a_{N+1} & \dots & a_{N+q-1} \end{pmatrix}$$

où

$$a_m = \sum_{j=1}^q \theta_j a_{m-j} \text{ avec } a_0 = 1 \text{ et } a_1 = \theta_1$$

- le vecteur des prévisions des q valeurs du processus X_t avant l'instant $t = 1$, au $k^{\text{ème}}$ passage de l'algorithme est noté :

$$\widehat{X}_{(1,q)}^{(k)} = \left(\widehat{X}_0^{(k)} \quad \widehat{X}_{-1}^{(k)} \quad \dots \quad \widehat{X}_{1-q}^{(k)} \right)'$$

Au $k^{\text{ème}}$ passage de l'algorithme, le calcul des innovations à l'aide du processus forward est obtenu à l'aide de l'équation :

$$\overrightarrow{\eta}^{(k)} = G_B \overrightarrow{X}_{(q)} + A \widehat{X}_{(1,q)}^{(k-1)} \quad (4.7)$$

La fonction de prévision d'origine 1 et d'horizon ℓ pour le processus backward, est définie telle que ([2], [6]) :

$$\widehat{X}_{1-\ell} = Q(F) \widehat{e}_{1-\ell} \quad (4.8)$$

où

$$\widehat{X}_{1-\ell} = \begin{cases} \widehat{X}_{1-\ell} & \text{si } \ell > 0 \\ X_{1-\ell} & \text{si } \ell \leq 0 \end{cases} \text{ et } \widehat{e}_{1-\ell} = \begin{cases} 0 & \text{si } \ell > 0 \\ e_{1-\ell} & \text{si } \ell \leq 0 \end{cases}$$

Il est évident que $\widehat{X}_{1-\ell} = 0$ dès que $\ell \geq q + 1$.

Considérons les notations suivantes :

- le vecteur des q premières innovations pour le processus backward sera noté tel que :

$$\overleftarrow{e}^{(k)} = \left(e_q^{(k)} \quad e_{q-1}^{(k)} \quad \dots \quad e_1^{(k)} \right)'$$

Au $k^{\text{ème}}$ passage de l'algorithme d'aller-retour et en utilisant la relation 4.8 pour $\ell = 1, 2, \dots, q$, on obtient un système d'équations qu'on résume tel que :

$$\widehat{X}_{(1,q)}^{(k)} + \Theta \overleftarrow{e}^{(k)} = 0 \quad (4.9)$$

En notant :

- G_F la matrice diagonale d'ordre q définie telle que :

$$G_F = \text{diag} (G_{N-q}(F) \quad G_{N-q+1}(F) \quad \dots \quad G_{N-1}(F))$$

où

$$G_{N-\ell}(F) = \sum_{j=0}^{N-\ell} a_j F^j \quad \ell = 1, 2, \dots, q$$

- le vecteur des q dernières observations pour le processus backward sera noté tel que :

$$\overleftarrow{X}_{(q)} = (X_q \quad X_{q-1} \quad \dots \quad X_1)'$$

Alors, au $k^{\text{ème}}$ passage de l'algorithme d'aller-retour, le calcul des innovations à l'aide du processus backward est obtenu à l'aide de l'équation suivante :

$$\overleftarrow{e}^{(k)} = G_F \overleftarrow{X}_{(q)} + A \widehat{X}_{(N,q)}^{(k)} \quad (4.10)$$

Proposition 31 Soit $\{ \widehat{X}_{(1,q)}^{(k)} \}_{k \geq 1}$ la suite des vecteurs des valeurs initiales obtenues aux différents passages de l'algorithme d'aller-retour. Alors, cet algorithme converge si la suite $(\widehat{X}_{(1,q)}^{(k)})_{k \geq 1}$ admet un point fixe.

Démonstration : Des relations 4.6, 4.7, 4.9 et 4.10, on déduit le système d'équations :

$$\begin{cases} \widehat{X}_{(N,q)}^{(k)} + \Theta G_B \overrightarrow{X}_{(q)} + \Theta A \widehat{X}_{(1,q)}^{(k-1)} = 0 \\ \widehat{X}_{(1,q)}^{(k)} + \Theta G_F \overleftarrow{X}_{(q)} + \Theta A \widehat{X}_{(N,q)}^{(k)} = 0 \end{cases}$$

D'où, on obtient :

$$\widehat{X}_{(1,q)}^{(k)} = \Theta A \Theta G_B \overrightarrow{X}_{(q)} - \Theta G_F \overleftarrow{X}_{(q)} + (\Theta A)^2 \widehat{X}_{(1,q)}^{(k-1)} \quad (4.11)$$

Si la matrice ΘA est contractante, alors le point fixe est obtenu en passant à la limite dans 4.11. ■

Proposition 32 La matrice ΘA est une matrice contractante, i.e. $\|\Theta A\| < 1$.

Démonstration : On sait que :

$$a_N = \theta_1 a_{N-1} + \theta_2 a_{N-2} \quad (4.12)$$

Alors, il est clair que :

$$\Theta A = \begin{pmatrix} a_{N+1} & a_{N+2} \\ \theta_2 a_N & \theta_2 a_{N+1} \end{pmatrix}$$

D'où, on a :

$$\|\Theta A\| = |\theta (a_{N+1}^2 - a_N a_{N+2})|$$

De plus, on remarque que les racines du polynôme caractéristique associé à l'équation de récurrence 4.12 sont les mêmes que celles de l'équation définie telle que :

$$r^2 - \theta_1 r - \theta_2 = 0 \quad (4.13)$$

Notons les deux racines de l'équation 4.13 par ρ_1 et ρ_2 .

On rappelle que le polynôme caractéristique d'un processus en moyennes mobiles d'ordre 2 est $Q(B) = 1 - \theta_1 B - \theta_2 B^2$. Il admet toutes ses racines à l'extérieur du cercle unité à cause de la propriété de causalité du processus ([2], [4], [6], [9], [22]). De plus, il est évident que le polynôme $Q(B)$ est le polynôme dual du polynôme 4.12. Par conséquent, les racines de l'équation 4.13 sont à l'intérieur du cercle unité.

La solution générale de l'équation de récurrence 4.12 est telle que :

$$\begin{aligned} a_N &= \lambda \rho_1^N + \mu \rho_2^N \quad \text{si } \rho_1 \neq \rho_2 \\ a_N &= \delta \rho^N \quad \text{si } \rho_1 = \rho_2 = \rho \end{aligned}$$

Par conséquent :

$$\|\Theta A\| = |\theta (a_{N+1}^2 - a_N a_{N+2})| = \left| \theta_2 \lambda \mu (\rho_1 \rho_2)^N (\rho_1 - \rho_2)^2 \right|$$

Sachant que les constantes λ and μ sont données par les équations $\lambda + \mu = 1$ et $\lambda \rho_1 + \mu \rho_2 = \theta_1$, alors :

$$\begin{aligned} \lambda &= \frac{\rho_1}{\sqrt{\Delta}} \quad \text{and} \quad \mu = 1 - \lambda = -\frac{\rho_2}{\sqrt{\Delta}} \\ \text{où } \Delta &= \theta_1^2 + 4\theta_2 \end{aligned}$$

Donc,

$$\|\Theta A\| = \left| \theta_2 (\rho_1 \rho_2)^{N+1} \right|$$

D'où

$$\|\Theta A\| = |\theta_2| \text{Max}^{2N+2} (|\rho_1|, |\rho_2|) < 1$$

Finalement, en notant le vecteur limite par $\widehat{X}_{(1,q)}^{(\infty)} = \left(\widehat{X}_0^{(\infty)} \quad \widehat{X}_{-1}^{(\infty)} \quad \dots \quad \widehat{X}_{1-q}^{(\infty)} \right)'$, le passage à la limite dans l'expression 4.11 nous donne :

$$\widehat{X}_{(1,q)}^{(\infty)} = [I_q - (\Theta A)^2]^{-1} \left\{ \Theta A \Theta G_B \overrightarrow{X}_{(q)} - \Theta G_F \overleftarrow{X}_{(q)} \right\}$$

où I_q est la matrice unité d'ordre q ■

Il reste à établir que le vecteur $\widehat{X}_{(1,q)}^{(\infty)}$ coïncide avec le vecteur X^0 des valeurs du processus avant l'instant $t = 1$, au sens de la topologie associée au produit scalaire dans $\overrightarrow{\mathcal{H}}_N$.

Proposition 33 *Soit $X = (X_1, X_2, \dots, X_N)'$ un système de générateurs de l'espace de Hilbert $\overrightarrow{\mathcal{H}}_N$. Alors*

$$E \left\{ \widehat{X}_{(1,q)}^{(\infty)} X' \right\} = E \left\{ X^0 X' \right\} = \Gamma$$

Démonstration : Considérons le cas $q = 2$ et rappelons les notations adoptées précédemment :

$$\Theta = \begin{pmatrix} \theta_2 & \theta_1 \\ 0 & \theta_2 \end{pmatrix} \quad A = \begin{pmatrix} a_{N-1} & a_N \\ a_N & a_{N+1} \end{pmatrix}$$

$$G_B \overrightarrow{X}_{(2)} = \begin{pmatrix} \sum_{j=0}^{N-2} a_j B^j X_{N-1} \\ \sum_{j=0}^{N-1} a_j B^j X_N \end{pmatrix} \quad G_F \overleftarrow{X}_{(2)} = \begin{pmatrix} \sum_{j=0}^{N-2} a_j F^j X_2 \\ \sum_{j=0}^{N-1} a_j F^j X_1 \end{pmatrix}$$

Calculons :

$$E \left(\widehat{X}_{(1,2)}^{(\infty)} X' \right) = [I_2 - (\Theta A)^2]^{-1} E \left(\Theta A \Theta G_B \overrightarrow{X}_{(2)} X' - \Theta G_F \overleftarrow{X}_{(2)} X' \right)$$

On a :

$$E \left(G_B \overrightarrow{X}_{(2)} X' \right) = \begin{pmatrix} \sum_{j=0}^{N-2} a_j \gamma_{|N-j-2|} & \sum_{j=0}^{N-2} a_j \gamma_{|N-j-3|} & \dots & \sum_{j=0}^{N-2} a_j \gamma_{j+1} \\ \sum_{j=0}^{N-1} a_j \gamma_{|N-j-1|} & \sum_{j=0}^{N-1} a_j \gamma_{|N-j-2|} & \dots & \sum_{j=0}^{N-1} a_j \gamma_j \end{pmatrix}$$

Par ailleurs, la fonction d'autocovariances γ_m d'un $MA(2)$ vérifie les conditions suivantes :

$$(C) \left\{ \begin{array}{l} \frac{1}{\sigma_\varepsilon^2} \sum_{j=0}^{N-2} a_j \gamma_{|N-r-j|} = \begin{cases} \theta_1 a_{N-1} + \theta_2^2 a_{N-2} & \text{si } r = 2 \\ \theta_2 a_{N-1} & \text{si } r = 3 \\ 1 & \text{si } r = N-1 \\ -\theta_1 & \text{si } r = N \end{cases} \\ \frac{1}{\sigma_\varepsilon^2} \sum_{j=0}^{N-2} a_j \gamma_{|N-r-j|} = 0 & \text{si } 4 \leq r \leq N-2 \\ \frac{1}{\sigma_\varepsilon^2} \sum_{j=0}^{N-1} a_j \gamma_{|N-r-j|} = \begin{cases} \theta_1 a_N + \theta_2^2 a_{N-1} & \text{si } r = 1 \\ \theta_2 a_N & \text{si } r = 2 \\ 0 & \text{si } r = N \\ -\theta_1 & \text{si } r = N-1 \end{cases} \\ \frac{1}{\sigma_\varepsilon^2} \sum_{j=0}^{N-1} a_j \gamma_{|N-r-j|} = 0 & \text{si } 3 \leq r \leq N-2 \end{array} \right.$$

Donc

$$E \left(G_B \vec{X}_{(2)} X' \right) = \sigma_\varepsilon^2 \begin{pmatrix} \theta_2^2 a_{N-2} + \theta_1 a_{N-1} & \theta_2 a_{N-1} & \dots & 1 & -\theta_1 \\ \theta_2^2 a_{N-1} + \theta_1 a_N & \theta_2 a_N & \dots & 0 & 1 \end{pmatrix} \quad (4.14)$$

En procédant de la même manière que précédemment, $E \left(G_F \overleftarrow{X}_{(2)} X' \right)$ peut être exprimée telle que :

$$E \left(G_F \overleftarrow{X}_{(2)} X' \right) = \begin{pmatrix} \sum_{j=0}^{N-2} a_j \gamma_{j+1} & \sum_{j=0}^{N-2} a_j \gamma_j & \dots & \sum_{j=0}^{N-2} a_j \gamma_{|N-j-2|} \\ \sum_{j=0}^{N-1} a_j \gamma_j & \sum_{j=0}^{N-1} a_j \gamma_{|j-1|} & \dots & \sum_{j=0}^{N-1} a_j \gamma_{|N-j-1|} \end{pmatrix}$$

D'après les conditions (C), l'expression ci-dessus s'écrit alors telle que :

$$E \left(G_F \overleftarrow{X}_{(2)} X' \right) = \sigma_\varepsilon^2 \begin{pmatrix} -\theta_1 & 1 & \dots & \theta_2 a_{N-1} & \theta_2^2 a_{N-2} + \theta_1 a_{N-1} \\ 1 & 0 & \dots & \theta_2 a_N & \theta_2^2 a_{N-1} + \theta_1 a_N \end{pmatrix} \quad (4.15)$$

Notons :

$$\Phi = \begin{pmatrix} 1 & -\theta_1 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad = \begin{pmatrix} -\theta_1 & 1 \\ 1 & 0 \end{pmatrix}$$

Il est aisé de remarquer que :

$$\begin{aligned}
-\Theta &= \frac{1}{\sigma_\varepsilon^2} \Gamma = \begin{pmatrix} -\theta_1 + \theta_1\theta_2 & -\theta_2 \\ -\theta_2 & 0 \end{pmatrix} \\
-A\Gamma &= \sigma_\varepsilon^2 \begin{pmatrix} \theta_2^2 a_{N-2} + \theta_1 a_{N-1} & \theta_2 a_{N-1} \\ \theta_2^2 a_{N-1} + \theta_1 a_N & \theta_2 a_N \end{pmatrix} \\
\Theta\Phi &= \begin{pmatrix} \theta_2 & \theta_1 - \theta_1\theta_2 \\ 0 & \theta_2 \end{pmatrix} \\
A\Theta\Phi &= \begin{pmatrix} \theta_2 a_{N-2} & \theta_1 a_{N-1} + \theta_2^2 a_{N-2} \\ \theta_2 a_N & \theta_1 a_N + \theta_2^2 a_{N-1} \end{pmatrix}
\end{aligned}$$

Ainsi, les expressions 4.14 et 4.15 peuvent être exprimées plus simplement sous la forme :

$$\begin{aligned}
E \left(G_B \overrightarrow{X}_{(2)} X' \right) &= \sigma_\varepsilon^2 \begin{pmatrix} A\Theta\Phi & : & 0 & : & \Phi \end{pmatrix} \\
E \left(G_F \overleftarrow{X}_{(2)} X' \right) &= \sigma_\varepsilon^2 \begin{pmatrix} & : & 0 & : & A\Theta\Phi \end{pmatrix}
\end{aligned}$$

Finalement, on obtient :

$$\begin{aligned}
E \left(\widehat{X}_{(1,2)}^{(\infty)} X' \right) &= \sigma_\varepsilon^2 [I_2 - (\Theta A)^2]^{-1} \begin{pmatrix} (\Theta A)^2 \Theta & -\Theta & : & 0 & : & 0 \end{pmatrix} \\
&= \sigma_\varepsilon^2 [I_2 - (\Theta A)^2]^{-1} [(\Theta A)^2 - I_2] \Theta = -\sigma_\varepsilon^2 \Theta = \Gamma
\end{aligned}$$

Ainsi, on montre bien que le vecteur-limite $\widehat{X}_{(1,q)}^{(\infty)}$ coïncide avec le vecteur des valeurs initiales X^0 , relativement à la norme associée avec le produit scalaire défini sur l'espace de Hilbert $L^2(f)$, (f est la loi de probabilité du processus).
■

4.3 Approximation de la vraisemblance exacte d'un processus *ARMA*

4.3.1 La vraisemblance exacte d'un processus *ARMA*(p, q)

Soit X_1, X_2, \dots, X_N une série chronologique identifiée par un processus *ARMA*(p, q) et donc elle vérifie la relation

$$P(B) X_t = Q(B) \varepsilon_t \quad (4.16)$$

où $P(B)$ et $Q(B)$ sont deux polynômes en B de degrés respectifs p et q .
 Supposons que $E(X_t) = 0$, alors la relation 4.16 peut s'écrire en extension sous la forme :

$$\varepsilon_t = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \dots - \varphi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (4.17)$$

Le calcul des innovations ε_t à l'aide de l'équation itérative 4.17, ne peut se faire que si l'on se donne $p + q$ valeurs initiales.

Supposons que le processus des innovations ε_t soit gaussien. Alors, la ln-vraisemblance de la série chronologique peut être exprimée telle que :

$$l(\phi; \vartheta; \sigma_\varepsilon^2) = \ln \{L(\phi; \vartheta; \sigma_\varepsilon^2)\} = c(\phi; \vartheta) - \frac{N}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} S(\phi, \vartheta) \quad (4.18)$$

où $c(\phi; \vartheta)$ est une fonction de ϕ et de ϑ , et $S(\phi, \vartheta)$ est la somme des carrés inconditionnelle définie telle que :

$$S(\phi, \vartheta) = \sum_{t=-\infty}^N \varepsilon_t^2(\phi, \vartheta, X_N, X_{N-1}, \dots) \quad (4.19)$$

Comme il a été établi dans ([6], [9]), pour N assez grand la relation 4.18 est dominée par le terme $\frac{S(\phi, \vartheta)}{2\sigma_\varepsilon^2}$. Cela veut dire que les contours de la somme des carrés dans l'espace des paramètres $(\phi; \vartheta)$ sont quasiment les mêmes contours que ceux de la ln-vraisemblance. Il vient en particulier que les estimateurs des moindres carrés des paramètres sont très approximativement les mêmes que ceux obtenus par la méthode du maximum de vraisemblance.

Il a aussi été établi dans ([2], [6], [22], [37]) que le modèle 4.16 peut être exprimé sous la forme d'un processus en moyennes mobiles d'ordre infini. Cependant, pour les processus qui ont un intérêt pratique les poids ψ_j décroissent et s'annulent assez rapidement. Pour cela, on peut approcher la somme des carrés 4.19 avec une certaine précision par la somme des carrés associée à un processus en moyennes mobiles fini d'ordre q .

En pratique, les valeurs des innovations passées ε_t ($t < 1$) peuvent être calculées récursivement à l'aide de l'équation 4.17 en partant de l'instant $t = 1 - q$ avant lequel les innovations ε_t sont considérés comme négligeables. Et donc, la somme des carrés 4.19 peut être exprimée telle que :

$$S(\phi, \vartheta) = \sum_{t=1-q}^N \varepsilon_t^2(\phi, \vartheta, X, \widehat{X}^0, \varepsilon^0) \quad (4.20)$$

où on adopte les notations suivantes :

$$\begin{aligned} X &= (X_1 \ X_2 \ \dots \ X_N)' \\ \phi &= (\varphi_1 \ \varphi_2 \ \dots \ \varphi_p)' \\ \vartheta &= (\theta_1 \ \theta_2 \ \dots \ \theta_q)' \\ \widehat{X}^0 &= (\widehat{X}_0 \ \widehat{X}_{-1} \ \dots \ \widehat{X}_{1-p})' \\ \varepsilon^0 &= (\varepsilon_0 \ \varepsilon_{-1} \ \dots \ \varepsilon_{1-q})' \end{aligned}$$

Les estimateurs des paramètres ϕ and ϑ sont obtenus en maximisant la vraisemblance 4.18 ou ce qui revient au même, en minimisant la somme des carrés 4.20 en utilisant les méthodes mathématiques traditionnelles. Ces dernières sont itératives et requièrent la donnée de valeurs initiales. Pour cela, nous préconisons la méthode proposée par Box and Jenkins pour la détermination de ces valeurs en utilisant le principe de ce qu'on appelle la "back-forecast" ou bien l'algorithme de l'"aller-retour" ([10], [11]). Dans la suite de ce paragraphe notre principale préoccupation sera l'étude de la convergence de l'algorithme de l'"aller-retour" pour les modèles mixtes en considérant le cas le plus simple.

4.3.2 Détermination des valeurs initiales par l'algorithme de l'aller-retour

La forme générale d'un processus $ARMA(1, 1)$ peut être exprimée telle que :

$$X_t - \alpha X_{t-1} = \varepsilon_t - \beta \varepsilon_{t-1}$$

Alors, les innovations ε_t peuvent être calculées récursivement telles que :

$$\varepsilon_t = X_t - \alpha X_{t-1} + \beta \varepsilon_{t-1} \quad t = 1, 2, \dots, N \quad (4.21)$$

Soit \widehat{X}_0 la valeur initiale et soit \widehat{X}_{N+1} la prévision de la valeur future X_{N+1} . En procédant de la même manière que dans le paragraphe précédent, on obtient au k^{th} passage de l'algorithme de l'"aller-retour" les équations suivantes :

$$\widehat{X}_{N+1}^{(k)} = (\alpha - \beta) \sum_{j=0}^{N-1} \beta^j X_{N-j} + \beta^N \left(\alpha \widehat{X}_0^{(k-1)} - \beta \varepsilon_0^{(k-1)} \right) \quad (4.22)$$

quand on utilise la représentation "forward" Et

$$\widehat{X}_0^{(k)} = (\alpha - \beta) \sum_{j=0}^{N-1} \beta^j X_{j+1} + \beta^N \left(\alpha \widehat{X}_{N+1}^{(k)} - \beta e_{N+1}^{(k)} \right) \quad (4.23)$$

quand on utilise la représentation "backward".

De plus, il a été établi dans ([6], [9], [22]) que les prévisions à l'horizon ℓ en utilisant la représentation "forward" ou la représentation "backward", sont obtenues telles que :

$$\begin{aligned} \widehat{X}_{N+\ell} &= \alpha^{\ell-1} \widehat{X}_{N+1} \\ \text{et } \widehat{X}_{-\ell} &= \alpha^\ell \widehat{X}_0 \end{aligned}$$

Alors, au k^{th} passage de l'algorithme de l'"aller-retour" et à l'horizon ℓ , les innovations $\varepsilon_0^{(k)}$ et $e_{N+1}^{(k)}$ sont déterminées telles que :

$$\varepsilon_0^{(k,\ell)} = \frac{(1 - \alpha^2) \left(1 - (\alpha\beta)^\ell \right)}{1 - \alpha\beta} \widehat{X}_0^{(k)} \quad (4.24)$$

$$e_{N+1}^{(k,\ell)} = \frac{(1 - \alpha^2) \left(1 - (\alpha\beta)^\ell \right)}{1 - \alpha\beta} \widehat{X}_{N+1}^{(k)} \quad (4.25)$$

Faire des prévisions à horizon infini à chaque passage de l'algorithme, consiste à passer à la limite sur ℓ dans les équations 4.24 et 4.25. De plus, on sait que $\alpha\beta < 1$. Alors, on obtient

$$\varepsilon_0^{(k,\infty)} = \frac{(1 - \alpha^2)}{1 - \alpha\beta} \widehat{X}_0^{(k)} \quad (4.26)$$

$$e_{N+1}^{(k,\infty)} = \frac{(1 - \alpha^2)}{1 - \alpha\beta} \widehat{X}_{N+1}^{(k)} \quad (4.27)$$

Des équations 4.22, 4.23, 4.26 et 4.27, et en reprenant les notations utilisées plus haut, on déduit :

$$\begin{aligned} \widehat{X}_{N+1}^{(k)} &= (\alpha - \beta) G_{N-1}(B) X_N + \frac{\beta^N (\alpha - \beta)}{1 - \alpha\beta} \widehat{X}_0^{(k)} \\ \widehat{X}_0^{(k+1)} &= (\alpha - \beta) G_{N-1}(F) X_1 + \frac{\beta^N (\alpha - \beta)}{1 - \alpha\beta} \widehat{X}_{N+1}^{(k)} \end{aligned}$$

où

$$G_{N-\ell}(B) = \sum_{j=0}^{N-\ell} a_j B^j \quad \ell = 1, 2, \dots, q$$

and

$$G_{N-\ell}(F) = \sum_{j=0}^{N-\ell} a_j F^j \quad \ell = 1, 2, \dots, q$$

En notant

$$\Lambda = \frac{(\alpha - \beta)}{1 - \alpha\beta} \beta^N$$

On obtient finalement :

$$\widehat{X}_0^{(k+1)} = \Lambda(\alpha - \beta) G_{N-1}(B) X_N + (\alpha - \beta) G_{N-1}(F) X_1 + \Lambda^2 \widehat{X}_0^{(k)} \quad (4.28)$$

Proposition 34 *L'équation de récurrence 4.28 admet un point fixe si $|\Lambda| < 1$.*

Démonstration : Rappelons qu'un processus *ARMA* (1, 1) est stationnaire si $|\alpha| < 1$ et il est inversible si $|\beta| < 1$ voir ([6], [9], [10], [22]). De plus, il est clair que $\left| \frac{(\alpha - \beta)}{1 - \alpha\beta} \right| < M$. Par conséquent, $|\Lambda| = \left| \frac{(\alpha - \beta)}{1 - \alpha\beta} \beta^N \right| < M |\beta^N| < 1$. ■

Finalement, le point fixe ou la valeur limite $\widehat{X}_0^{(\infty)}$ est obtenu tel que :

$$\widehat{X}_0^{(\infty)} = \frac{\Lambda(\alpha - \beta) G_{N-1}(B) X_N + (\alpha - \beta) G_{N-1}(F) X_1}{1 - \Lambda^2}$$

Proposition 35 *Soit (X_1, X_2, \dots, X_N) un système de générateurs de l'espace de Hilbert \mathcal{H}_N . Si le vecteur $\widehat{X}_0^{(\infty)}$ est la valeur recherchée, alors la condition suivante est satisfaite :*

$$E \left(X_0^{(\infty)} X_k \right) = \begin{cases} \gamma_1 & \text{si } k = 1 \\ \alpha^{k-1} \gamma_1 & \text{si } k \geq 2 \end{cases}$$

Démonstration : Calculons

$$E \left(X_0^{(\infty)} X_1 \right) = \frac{(\alpha - \beta)}{1 - \Lambda^2} [\Lambda E(G_{N-1}(B) X_N X_1) + E(G_{N-1}(F) X_1 X_1)]$$

D'une part, on calcule :

$$\begin{aligned} E(G_{N-1}(B) X_N X_1) &= \sum_{j=0}^{N-1} \beta^j E(X_{N-j} X_1) = \sum_{j=0}^{N-1} \beta^j \gamma_{N-j-1} \\ &= \beta^{N-1} \gamma_0 + \frac{\alpha^{N-1} - \beta^{N-1}}{\alpha - \beta} \gamma_1 \end{aligned}$$

Et d'autre part, on a :

$$\begin{aligned} E(G_{N-1}(F) X_1 X_1) &= \sum_{j=0}^{N-1} \beta^j E(X_{j+1} X_1) = \sum_{j=0}^{N-1} \beta^j \gamma_j \\ &= \gamma_0 + \beta \gamma_1 + \frac{\alpha \beta^2}{1 - \alpha \beta} \gamma_1 - \frac{\alpha^{N-1} \Lambda}{\alpha - \beta} \gamma_1 \end{aligned}$$

Par conséquent,

$$\Lambda E(G_{N-1}(B) X_N X_1) + E(G_{N-1}(F) X_1 X_1) = \frac{1 - \alpha \beta}{1 - \alpha^2} (1 - \Lambda^2) \sigma_\epsilon^2$$

Il a été établi dans ([2], [6], [9]) que :

$$\begin{cases} \gamma_0 = \frac{1 + \beta^2 - 2\alpha\beta}{1 - \alpha^2} \sigma_\epsilon^2 \\ \gamma_1 = \frac{(1 - \alpha\beta)(\alpha - \beta)}{1 - \alpha^2} \sigma_\epsilon^2 \\ \gamma_k = \alpha \gamma_{k-1} \quad k \geq 2 \end{cases}$$

D'où

$$E(X_0^{(\infty)} X_1) = \frac{(\alpha - \beta)(1 - \alpha\beta)}{1 - \alpha^2} \sigma_\epsilon^2 = \gamma_1$$

Considérons maintenant le cas $k = 2$.

Alors

$$E(X_0^{(\infty)} X_2) = \frac{\alpha - \beta}{1 - \Lambda^2} [\Lambda E(G_{N-1}(B) X_N X_2) + E(G_{N-1}(F) X_1 X_2)]$$

Ou bien

$$\frac{1 - \Lambda^2}{\alpha - \beta} E(X_0^{(\infty)} X_2) = \Lambda E(G_{N-1}(B) X_N X_2) + E(G_{N-1}(F) X_1 X_2)$$

En premier lieu, on a :

$$\begin{aligned}
E(G_{N-1}(B) X_N X_2) &= \sum_{j=0}^{N-1} \beta^j E(X_{N-j} X_2) = \sum_{j=0}^{N-1} \beta^j \gamma_{|N-j-2|} \\
&= \beta^{N-1} \gamma_1 + \beta^{N-2} \gamma_0 + \beta^{N-3} \gamma_1 + \sum_{j=0}^{N-4} \beta^j \gamma_{|N-j-2|} \\
&= (\beta^N - \alpha \beta^{N-1} + \beta^{N-2} - \alpha^{N-2}) \frac{\gamma_1}{\beta - \alpha} + \beta^{N-2} \gamma_0
\end{aligned}$$

Et ensuite, on obtient :

$$\begin{aligned}
E(G_{N-1}(F) X_1 X_2) &= \sum_{j=0}^{N-1} \beta^j \gamma_{|j-1|} = \gamma_1 + \beta \gamma_0 + \beta^2 \gamma_1 + \sum_{j=3}^{N-1} \beta^j \gamma_{|j-1|} \\
&= \gamma_1 + \beta \gamma_0 + \frac{\beta^2 \gamma_1}{1 - \alpha \beta} - \frac{\Lambda \alpha^{N-2}}{\alpha - \beta} \gamma_1
\end{aligned}$$

Par conséquent,

$$\begin{aligned}
\frac{1 - \Lambda^2}{\alpha - \beta} E(X_0^{(\infty)} X_2) &= (\beta^N - \alpha \beta^{N-1} + \beta^{N-2}) \frac{\Lambda}{\beta - \alpha} \gamma_1 + \\
&\quad + \Lambda \beta^{N-2} \gamma_0 + \gamma_1 + \beta \gamma_0 + \frac{\beta}{1 - \alpha \beta} \gamma_1 \\
&= -\frac{\Lambda^2 (1 - \alpha \beta)}{\beta^2 (\alpha - \beta)^2} \{ (1 + \beta^2 - \alpha \beta) \gamma_1 + (\beta - \alpha) \gamma_0 \} + \\
&\quad + \gamma_1 + \beta \gamma_0 + \frac{\beta^2 \gamma_1}{1 - \alpha \beta} \\
&= \frac{\alpha (1 - \alpha \beta)}{1 - \alpha^2} (1 - \Lambda^2) \sigma_\epsilon^2
\end{aligned}$$

Donc

$$E(X_0^{(\infty)} X_2) = \alpha \gamma_1$$

Pour terminer montrons, qu'en général, on a :

$$E(X_0^{(\infty)} X_h) = \alpha^{h-1} \gamma_1 \quad \forall h \geq 2$$

Un premier calcul nous donne :

$$\begin{aligned}
E(G_{N-1}(B) X_N X_h) &= \sum_{j=0}^{N-1} \beta^j \gamma_{N-j-h} = \sum_{j=0}^{N-h-2} \beta^j \gamma_{N-j-h} + \sum_{j=N-h+2}^{N-1} \beta^j \gamma_{j+h-N} \\
&\quad + \beta^{N-h-1} \gamma_1 + \beta^{N-h} \gamma_0 + \beta^{N-h+1} \gamma_1 \\
&= \frac{\alpha \beta^N [1 - (\alpha \beta)^{h-2}]}{(1 - \alpha \beta) \beta^{h-2}} \gamma_1 + \frac{\alpha}{\beta - \alpha} (\beta^{N-h-1} - \alpha^{N-h-1}) \gamma_1 \\
&\quad + \beta^{N-h-1} \gamma_1 + \beta^{N-h} \gamma_0 + \beta^{N-h+1} \gamma_1 \\
&= \frac{\alpha \gamma_1 \Lambda}{(\alpha - \beta) \beta^{h-2}} - \frac{\alpha^{h-1}}{\alpha - \beta} \Lambda \gamma_1 - \frac{\alpha(1 - \alpha \beta) \gamma_1 \Lambda}{(\alpha - \beta)^2 \beta^{h+1}} + \frac{\alpha^{N-h}}{\alpha - \beta} \gamma_1 + \\
&\quad + \frac{1 - \alpha \beta}{(\alpha - \beta) \beta^{h+1}} \Lambda \gamma_1 + \frac{1 - \alpha \beta}{(\alpha - \beta) \beta^h} \Lambda \gamma_0 + \frac{1 - \alpha \beta}{(\alpha - \beta) \beta^{h-1}} \Lambda \gamma_1 \\
&= \frac{\Lambda(1 - \alpha \beta)}{(\alpha - \beta) \beta^{h+1}} \left\{ \frac{\alpha \beta^3}{1 - \alpha \beta} \gamma_1 - \frac{\alpha}{\alpha - \beta} \gamma_1 + \beta^2 \gamma_1 + \gamma_1 + \beta \gamma_0 \right\} + \\
&\quad + \frac{\alpha^{N-h}}{\alpha - \beta} \gamma_1 - \frac{\alpha^{h-1}}{\alpha - \beta} \Lambda \gamma_1
\end{aligned}$$

Il est évident que :

$$\frac{\alpha \beta^3}{1 - \alpha \beta} \gamma_1 - \frac{\alpha}{\alpha - \beta} \gamma_1 + \beta^2 \gamma_1 + \gamma_1 + \beta \gamma_0 = 0 \quad (4.29)$$

D'où

$$E(G_{N-1}(B) X_N X_h) = \frac{\alpha^{N-h}}{\alpha - \beta} \gamma_1 - \frac{\alpha^{h-1}}{\alpha - \beta} \gamma_1 \Lambda \quad (4.30)$$

De plus

$$\begin{aligned}
E(G_{N-1}(F) X_1 X_h) &= \sum_{j=0}^{N-1} \beta^j \gamma_{j-h-1} = \sum_{j=0}^{h-3} \beta^j \gamma_{h-j-1} + \sum_{j=h+1}^{N-1} \beta^j \gamma_{j-h+1} \\
&\quad + \beta^{h-2} \gamma_1 + \beta^{h-1} \gamma_0 + \beta^h \gamma_1 \\
&= \frac{\alpha^{h-1}}{\alpha - \beta} \gamma_1 - \frac{\alpha \beta^{h-2}}{\alpha - \beta} \gamma_1 + \frac{\alpha \beta^{h+1}}{\alpha - \beta} \gamma_1 - \frac{\alpha^{N-h}}{\alpha - \beta} \Lambda \gamma_1 + \\
&\quad + \beta^{h-2} \gamma_1 + \beta^{h-1} \gamma_0 + \beta^h \gamma_1 \\
&= \beta^{h-2} \left(\frac{\alpha \beta^3}{1 - \alpha \beta} \gamma_1 - \frac{\alpha}{\alpha - \beta} \gamma_1 + \gamma_1 + \beta \gamma_0 + \beta^2 \gamma_1 \right) + \\
&\quad + \frac{\alpha^{h-1}}{\alpha - \beta} \gamma_1 - \frac{\alpha^{N-h}}{\alpha - \beta} \Lambda \gamma_1
\end{aligned}$$

D'après l'équation 4.29, on a :

$$E(G_{N-1}(F) X_1 X_h) = \frac{\alpha^{h-1}}{\alpha - \beta} \gamma_1 - \frac{\alpha^{N-h}}{\alpha - \beta} \Lambda \gamma_1 \quad (4.31)$$

Finalement, en combinant les équations 4.30 et 4.31, on obtient

$$\begin{aligned}
E\left(X_0^{(\infty)} X_h\right) &= \frac{\alpha - \beta}{1 - \Lambda^2} [\Lambda E(G_{N-1}(B) X_N X_h) + E(G_{N-1}(F) X_1 X_h)] \\
&= \frac{\alpha - \beta}{1 - \Lambda^2} \left[\frac{\alpha^{h-1}}{\alpha - \beta} \gamma_1 - \frac{\alpha^{h-1}}{\alpha - \beta} \Lambda^2 \gamma_1 \right] = \alpha^{h-1} \gamma_1
\end{aligned}$$

■

Ainsi, on vient d'établir que la valeur limite $X_0^{(\infty)}$ coïncide avec la vraie valeur \widehat{X}_0 du processus antérieure à l'instant $t = 1$ relativement à la norme associée au produit scalaire défini sur l'espace de Hilbert \mathcal{H} . De plus, il est clair qu'on peut arriver au même résultat pour un processus *ARMA* d'ordre quelconque. Néanmoins, les problèmes d'écriture et de formulation doivent être surmontés.

Chapitre 5

Rapport des vraisemblances de deux processus gaussiens

Dans cette dernière partie on va procéder à l'extension des résultats du chapitre trois aux problèmes de tests. On calcule le test de Neyman-Pearson en situation exponentielle. Il est dans ce cas équivalent aux tests classiques d'ajustement type porte-manteau. Nous avons repris quelques travaux concernant le cas continu que l'on a essayé d'innover avec l'introduction de l'approximation de Toeplitz-Szegö. Notre but est de situer le cadre naturel de développements ultérieurs, i.e, appliquer une certaine forme du théorème de Szegö ([28]) au cas continu pour obtenir des approximations des équations de Wiener-Hopf faciles à manier. C'est à ce travail que nous allons nous attacher de manière à approfondir dans le cas continu les résultats de Dzaparidze ([14], [15]) et Pissarenko ([32], [33]).

5.1 Equations de Wiener- Hopf

Ce paragraphe rassemble un certain nombre de développements essentiellement de synthèse concernant les équations de Wiener-Hopf. Le lien de ces dernières avec la théorie des opérateurs de Toeplitz et leurs applications aux problèmes de statistique des processus gaussiens ([33]) sont mis en évidence. On s'intéressera plus particulièrement au calcul et à l'approximation de la vraisemblance de deux processus gaussiens à temps continu associés à des mesures gaussiennes absolument continues l'une par rapport à l'autre. Nous allons nous limiter au cas des processus à temps continu ([14], [15], [31]).

Soit $\delta(\zeta)$ la masse de Dirac en ζ . Pour une fonction f localement sommable par rapport à la mesure de Lebesgue $d\zeta$ nous identifierons f et la distribution de Schwartz associée. Soit C une distribution de Schwartz du type :

$$C(\zeta) = \delta(\zeta) - C_a(\zeta)$$

Soit $T^{+, \eta}$ l'opérateur de troncage défini tel que :

$$T^{+, \eta}(f) = 0 \quad \text{sur} \quad [0, \eta]^c$$

L'équation de Wiener-Hopf de 2ème membre une distribution g définie telle que :

$$g(\zeta) = g_s \delta(\zeta) + g_a(\zeta)$$

s'écrit, pour $\eta > 0$ fixé, telle que :

$$T^{+, \eta}(C * f) = g$$

L'inconnue f est du même type que g . On peut écrire aussi cette équation en séparant les deux parties des mesures locales f et g sous la forme :

$$\left\{ \begin{array}{l} f_s = g_s \quad \text{trivialement vérifiée pour } g \text{ localement intégrable} \\ f_a(\zeta) - \int_0^\eta C_a(\zeta - \eta) f_a(\eta) d\eta = g_a(\zeta) - g_s C_a(\zeta) \quad 0 \leq \zeta \leq \eta \end{array} \right.$$

Remarque 36 La deuxième équation se réduit à la forme dite fonction de l'équation de Wiener-Hopf lorsque $g_s = 0$, soit :

$$f(\zeta) - \int_0^\eta C(\zeta - \eta) f(\eta) d\eta = g(\zeta)$$

Cette équation a été largement étudiée par les analystes et les probabilistes.

5.2 Le rapport de vraisemblances de deux processus gaussiens

5.2.1 Condition d'existence

Pour simplifier, nous supposerons que les processus étudiés sont centrés et stationnaires sur \mathbb{R} . L'intervalle d'observation sera noté $[0, T]$. Pour beaucoup

de résultats que nous allons rappeler dans ce cadre ([4], [14], [15], [33]), on en fait une démonstration dans un cadre tout à fait général de processus gaussiens sur $[0, T]$.

Soit C une fonction symétrique définie de $[0, T] \times [0, T] \rightarrow \mathbb{R}$ telle que :

$$C(s, t) = C(t, s)$$

Soient P_0, P_1 deux mesures gaussiennes stationnaires centrées sur \mathbb{R}^T et soient $\xi(t, \cdot)$ les applications coordonnées de $\mathbb{R}^T \rightarrow \mathbb{R}$. Soit $H_0 = H_1$ la fermeture dans $L^2(\mathbb{R}^T, \mathcal{B}, P_0 = P_1)$ de l'espace des combinaisons linéaires finies du type :

$$\sum C(s, t) \xi(s, \omega) \xi(t, \omega)$$

Soient $\Phi_0(\cdot, \omega), \Phi_1(\cdot, \omega)$ les mesures aléatoires des représentations spectrales de P_0 et P_1 (i.e. $\xi(t, \omega) = \int_{\mathbb{R}} e^{i\lambda t} \Phi(d\lambda, \omega)$). Tout élément de $H_0 = H_1$ est du type :

$$h(\omega, \omega') = \int \int_{\mathbb{R} \times \mathbb{R}} \varphi(\lambda, \mu) \Phi_0(d\lambda, \omega) \Phi_1(d\mu, \omega')$$

L'application $h \rightarrow \varphi$ est une isométrie canonique de $H_0 = H_1$ dans $L^2(F_0 = F_1)$ où F_0 et F_1 sont les mesures spectrales associées à Φ_0 et Φ_1 .

Rappelons que les mesures P_0 et P_1 sont équivalentes si et seulement si l'identité est un isomorphisme de L^2 . On note :

$$h^s(\omega) = \sum C(s, t) \xi(s, \omega) \xi(t, \omega)$$

Si $P_0 \sim P_1$ alors l'application $h^s \rightarrow h$ est un isomorphisme de $H_s(P_0)$ sur $H_0 = H_1$, où $H_s(P_0)$ est l'espace engendré par les combinaisons du type h^s .

Théorème 37 *Pour que P_0 et P_1 soient équivalentes, il faut et il suffit que l'on ait :*

$$B_0(s, t) - B_1(s, t) = \iint \xi(s, \omega) \xi(t, \omega') \psi(\omega, \omega') dP_0(\omega) dP_1(\omega') \quad (5.1)$$

où B_0 et B_1 sont les fonctions de corrélations associées à P_0 et P_1 et $\psi \in H_0 = H_1$.

On a alors :

$$\frac{dP_1}{dP_0}(\omega) = a e^{-\frac{1}{2}\psi(\omega, \omega)} \quad \text{où } a = \text{constante}$$

Démonstration : Remarquons d'abord que si la condition 5.1 est réalisée, alors $P_0 \sim P_1$. En effet, en notant $C.\xi$ pour $\sum C(s_i) \xi(s, \omega)$, on a alors

$$\begin{aligned} \left(\|C\xi\|_{L^2(P_0)} - \|C\xi\|_{L^2(P_1)} \right)^2 &= \left(\sum \sum C(s) C(t) b(s, t) \right)^2 \\ &= \left(\int \int C\xi(\omega) C\xi(\omega') \psi(\omega, \omega') dP_0(\omega) dP_1(\omega') \right)^2 \\ &= \|C\xi\|_{L^2(P_0)}^2 \|C\xi\|_{L^2(P_1)}^2 \|\psi\|_{L^2(P_0 \times P_1)} \end{aligned}$$

Supposons maintenant que $P_0 \sim P_1$. On a $dP_1 = \phi dP_0$ et donc

$$b(s, t) = \int \int \xi(s, \omega) \xi(t, \omega') \psi_1(\omega, \omega') dP_0(\omega) dP_0(\omega')$$

où $\psi_1 = \psi\phi \in L^2(P_0 \times P_0)$.

La symétrie de $b(s, t)$ montre que l'on peut choisir $\psi_1 \in H_0 \otimes H_0$ (en prenant $\frac{1}{2}(\psi_1(x, y) + \psi_1(y, x))$ éventuellement).

Pour démontrer le théorème, on utilise la démonstration suivante de type martingale introduite par différents auteurs russes dont Rosanov [33].

Soit $\Delta_n \subset [0, T]$, Δ_n fini et la suite $(\Delta_n)_n$ croissante. Supposons que $\bigvee_n H_{\Delta_n}$

soit dense dans $H(P_0)$ où $\bigvee_n H_{\Delta_n}$ est le sous-espace de $H(P_0)$ engendré par

$\{\xi(t, \omega), t \in \Delta_n\}$. Il est clair que si $P_0 \sim P_1$ alors les restrictions P_0^n et P_1^n de P_0 et P_1 à la σ -algèbre engendrée par $\{\xi(t, \omega), t \in \Delta_n\}$, sont telles que $P_0^n \sim P_1^n$. Supposons que :

$$\frac{dP_0^n}{dP_1^n} = p_n$$

Alors p_n est une martingale. Nous allons travailler sur la sous-martingale $\ln p_n$ définie telle que :

$$\ln p_n = -\frac{1}{2} \xi^{\Delta_n} (B_{1, \Delta_n}^{-1} - B_{0, \Delta_n}^{-1}) \xi^{\Delta_n} + \ln \det (B_{1, \Delta_n} B_{0, \Delta_n}^{-1})$$

où ξ^{Δ_n} est le vecteur $\{\xi(t, \omega), t \in \Delta_n\}$ et B_{i, Δ_n} est la matrice de covariance sous P_i de ξ^{Δ_n} .

Considérons $Y^{\Delta_n} = B_{0, \Delta_n}^{-\frac{1}{2}} \xi^{\Delta_n}$, alors $Cov_{P_0}(Y^{\Delta_n}) = I_{\Delta_n}$, où I_{Δ_n} est l'identité. D'où

$$\ln p_n = -\frac{1}{2} Y^{\Delta_n} \left(B_{0, \Delta_n}^{\frac{1}{2}} B_{1, \Delta_n}^{-1} B_{0, \Delta_n}^{\frac{1}{2}} - I_{\Delta_n} \right) Y^{\Delta_n}$$

Désignons par σ_k^n les valeurs propres de la matrice $B_{0,\Delta_n}^{\frac{1}{2}} B_{1,\Delta_n}^{-\frac{1}{2}}$ et par $Z^{\Delta_n} = (Z_1, \dots, Z_{\Delta_n})$ ses vecteurs propres. Alors

$$\ln p_n = \frac{1}{2} \left(\sum_{k=1}^{\Delta_n} \left(\frac{1 - \sigma_k^2}{\sigma_k^2} \right) z_k^2 + \sum_{k=1}^{\Delta_n} \ln \sigma_k^2 \right)$$

Remarquons que la loi de Z^{Δ_n} sous P_1 est donnée par la densité :

$$\left(\frac{1}{2\pi} \right)^{\frac{\Delta_n}{2}} \exp \left(-\frac{1}{2} \sum_{k=1}^{\Delta_n} z_k^2 \right) \exp \left(\frac{1}{2} \sum_{k=1}^{\Delta_n} \left(\frac{1 - \sigma_k^2}{\sigma_k^2} \right) z_k^2 \prod_{k=1}^{\Delta_n} \sigma_k \right)$$

Donc, les variables Z_k sont indépendantes et $E_{P_1}(Z_k^2) = \sigma_k^2$.

Par ailleurs, d'après ([14], [15]) et ([33]) on a :

$$E_{P_1}(\ln p_n) = \frac{1}{2} \sum_{k=1}^{\Delta_n} \{ (1 - \sigma_k^2) + \ln \sigma_k^2 \}$$

et

$$Var_{P_0}(\ln p_n) = \frac{1}{2} \sum_{k=1}^{\Delta_n} \left(\frac{1 - \sigma_k^2}{\sigma_k^2} \right)$$

Comme $P_0 \sim P_1$, alors il existe des constantes $0 < m < M < \infty$ telles que $m < \sigma_k^n < M$ pour tout n et tout k . D'où

$$Var_{P_1}(\ln p_n) \simeq \sum_{k=1}^{\Delta_n} (1 - \sigma_k^2) \simeq E_{P_1}(\ln p_n)$$

Finalement, si $\sup_n E_{P_1}(\ln p_n) < \infty$, alors $\ln p_n$ converge *p.s* d'après le théorème des martingales et le théorème de Radon-Nicodyn. ■

Une autre démonstration consiste à noter A un ensemble tel que par exemple $P_0(A) = 0$, $P_1(A) \neq 0$ et une suite d'évènements A_n telle que $A_n \downarrow A$ pour $\frac{P_0 + P_1}{2}$, voir ([4], [14]).

Soit \mathcal{F}_n la σ -algèbre $\{A_n, A_n^C, \emptyset, \Omega\}$. Il est admis que

$$E_{P_0}^{\mathcal{F}_n}(p_n) = \frac{P_1(A_n)}{P_0(A_n)} 1_{A_n} + \frac{1 - P_1(A_n)}{1 - P_0(A_n)} 1_{A_n^C}$$

On a d'une part, en désignant par $K(P, Q)$ l'information de Kullback de deux probabilités P et Q :

$$E_{P_1} \ln E_{P_0}^{\mathcal{F}_n}(p_n) \rightarrow \infty$$

Car

$$\begin{aligned} E_{P_1} \ln E_{P_0}^{\mathcal{F}_n} (p_n) &= \ln \frac{P_1(A_n)}{P_0(A_n)} P_1(A_n) + \ln \frac{1 - P_1(A_n)}{1 - P_0(A_n)} (1 - P_1(A_n)) \\ &= K(P_{1/\mathcal{F}_n}, P_{0/\mathcal{F}_n}) \rightarrow \infty \end{aligned}$$

D'autre part, pour toute σ -algèbre \mathcal{B}_n telle que $\mathcal{F}_n \subset \mathcal{B}_n$, on a :

$$\sup_n E_{P_1} (\ln p_n) = K(P_{1/\mathcal{B}_n}, P_{0/\mathcal{B}_n}) < \infty$$

Et

$$K(P_{1/\mathcal{F}_n}, P_{0/\mathcal{F}_n}) \leq K(P_{1/\mathcal{B}_n}, P_{0/\mathcal{B}_n})$$

Donc, si $\sup_n E_{P_1} (\ln p_n) < \infty$, on a $P_0 \sim P_1$. Si $\sup_{k,n} (\sigma_{k,n}^2) = \infty$ ou $\inf_{k,n} (\sigma_{k,n}^2) = 0$, il y a orthogonalité car alors il existerait des $Z_{k,n}$ tels que :

$$E_{P_0} (Z_{k,n}^2) \rightarrow 0 \quad \text{et} \quad E_{P_1} (Z_{k,n}^2) \rightarrow 1$$

Il reste alors à étudier le cas où $\sup_n E_{P_1} (\ln p_n) = \infty$. On sait qu'il existe m et M tels que $0 < m < \sigma_{k,n}^2 < M$ pour tout k et tout n . Soit A_n l'évènement défini tel que :

$$A_n = \left\{ \ln p_n - E_{P_1} (\ln p_n) > \frac{1}{2} E_{P_1} (\ln p_n) \right\}$$

On a alors, d'après l'inégalité de Tchébichev

$$P_0(A_n) \leq \frac{\text{Var}_{P_0}(\ln p_n)}{\frac{1}{4} (E_{P_1}(\ln p_n))^2} \sim \frac{C}{E_{P_1}(\ln p_n)} \rightarrow 0$$

De plus

$$P_1(A_n) = 1 - P_1(\neg A_n) \geq 1 - \frac{\text{Var}_{P_1}(\ln p_n)}{\frac{1}{4} (E_{P_1}(\ln p_n))^2} \gtrsim 1 - \frac{C'}{E_{P_1}(\ln p_n)} \rightarrow 1$$

Nous concluons ce paragraphe par l'énoncé suivant :

Théorème 38 $P_0 \sim P_1$ si et seulement si on a :

$$\sup_n K(P_1^n, P_0^n) < \infty \quad \text{ou bien} \quad \sup_n \text{Var}_{P_0} \ln \left(\frac{dP_1^n}{dP_0^n} \right) < \infty$$

Ces deux conditions sont équivalentes.

5.3 L'équation de Wiener-Hopf du rapport de vraisemblances

Le logarithme du rapport de vraisemblance de deux processus gaussiens est défini tel que :

$$\ln p_n = -\frac{1}{2} \ln \det (B_{1,\Delta_n} B_{0,\Delta_n}^{-1}) + {}^t X^{\Delta_n} (B_{1,\Delta_n}^{-1} - B_{0,\Delta_n}^{-1}) X^{\Delta_n} \quad (5.2)$$

Soit $d(s, t)$, $s, t \in \Delta_n$ le terme général de la matrice $B_{1,\Delta_n}^{-1} - B_{0,\Delta_n}^{-1}$. On pose :

$$\varphi_n (\lambda, \mu) = \sum_{s,t} d(s, t) \exp i (\lambda s - \mu t)$$

En remplaçant $X(t, \cdot)$ par sa représentation spectrale dans 5.2, on a :

$$\ln p_n = -\frac{1}{2} \ln \det (B_{1,\Delta_n} B_{0,\Delta_n}^{-1}) + \iint \varphi_n (\lambda, \mu) \Phi_0 (d\lambda) \Phi_0 (d\mu)$$

De plus

$$E_{P_0} (\ln p_n) = \frac{1}{2} \int \varphi_n (\lambda, \mu) dF_0(\lambda) - \frac{1}{2} \ln \det (B_{1,\Delta_n} B_{0,\Delta_n}^{-1})$$

et

$$Var_{P_0} (\ln p_n) = \frac{1}{2} \iint \{\varphi_n (\lambda, \mu)\}^2 dF_0(\lambda) dF_0(\mu)$$

Donc, $\sup_n (Var_{P_0} (\ln p_n)) < \infty$ équivaut à $\sup_n (\|\varphi_n\|_{L^2_{(F_0, F_0)}}) < \infty$ ou bien à $\sup_n (\|\varphi_n\|_{L^2_{(F_0, F_1)}}) < \infty$ puisque $F_0 \sim F_1$.

Mais

$$\begin{aligned} & \iint \exp \{-i (\lambda s - \mu t)\} \varphi_n (\lambda, \mu) dF_0(\lambda) dF_1(\mu) \\ &= \iint \sum_{s', t'} d(s', t') \exp \{i (\lambda s' - \mu t') - i (\lambda s - \mu t)\} dF_0(\lambda) dF_0(\mu) \\ &= \sum_{s', t'} B_{0, s' - s} \left(B_{1, (s', t')}^{-1} - B_{0, (s', t')}^{-1} \right) B_{0, t' - t} = b(s, t) \end{aligned}$$

Donc, on peut dire que φ_n est solution de :

$$\iint \exp \{-i (\lambda s - \mu t)\} \varphi_n (\lambda, \mu) dF_0(\lambda) dF_1(\mu) = b(s, t) \quad s, t \in \Delta_n \quad (5.3)$$

L'équation 5.3 est une équation de type Wiener-Hopf pour l'ensemble Δ_n , ou équation de Wiener-Hopf de type fini.

Soit Δ un sous-ensemble de $[0, T]$ telle que la suite $\Delta_n \uparrow \Delta$. Nous allons passer à la limite sur φ_n . A cet effet, considérons l'hyperplan affine :

$$L_\Delta = \left\{ \begin{array}{l} \text{fermeture des combinaisons linéaires des fonctions} \\ \psi(\lambda, \mu) = \frac{1}{b(s,t)} \exp\{i(\lambda s - \mu t)\} + c(s', t') \exp\{i(\lambda s' - \mu t')\} \\ \text{où } (s', t') \text{ est un point tel que } b(s', t') = 0 \\ \text{et } (s, t) \text{ est un point tel que } b(s, t) \neq 0 \end{array} \right\} \quad (5.4)$$

Considérons

$$\iint \exp\{i(\lambda s - \mu t)\} \varphi(\lambda, \mu) F_0(d\lambda) F_1(d\mu) = b(s, t) \quad (s, t) \in \Delta^2$$

Soit φ une solution de $\varphi_0 = \frac{\varphi}{\|\varphi\|}$. Supposons que $0 \notin L_\Delta$, alors φ est la projection de 0 sur L_Δ dans $L^2(F_0 \times F_1)$. En effet, $\langle \psi - \varphi_0, \varphi_0 \rangle = 0$ pour $\varphi \in L_\Delta^2$, i.e.

$$\begin{aligned} & \frac{1}{b(s,t)} \iint \exp\{i(\lambda s - \mu t)\} \varphi_0(\lambda, \mu) F_0(d\lambda) F_1(d\mu) + \dots \\ & \dots c(s', t') \iint \exp\{i(\lambda s' - \mu t')\} \varphi_0(\lambda, \mu) F_0(d\lambda) F_1(d\mu) = 0 \end{aligned}$$

Or, $c(s', t')$ est choisi arbitrairement dans 5.4, donc la deuxième intégrale dans l'expression précédente est nulle et par conséquent φ_0 est bien une solution non nulle.

Réciproquement, s'il y a une solution, alors $0 \notin L_\Delta^2$ car ceci implique que $\|\varphi_0\|^2 = 0$ ou bien $\|\varphi_0\|^2 = 1$.

Considérons alors la suite φ_n associée à la suite Δ_n . On a alors :

$$L_{\Delta_1}^2 \subset L_{\Delta_2}^2 \subset \dots \subset L_{\Delta_n}^2 \subset \dots \subset L_\Delta^2$$

Ces injections sont des isométries canoniques. Par ailleurs, φ_n est une martingale bornée de L^2 (projections sur $L_{\Delta_n}^2$).

Comme $\sup_n \left\{ \|\varphi_n\|_{L^2(F_0 \times F_1)}^2 \right\} = \sup_n \{Var_{P_0}(\ln p_n)\} < \infty$, φ_n converge vers φ . On en déduit immédiatement que :

$$\eta_n = \iint \varphi_n(\lambda, \mu) \Phi_0(\omega, d\lambda) \Phi_1(\omega', d\lambda)$$

converge dans L^2 vers :

$$\eta = \iint \varphi(\lambda, \mu) \Phi_0(\omega, d\lambda) \Phi_1(\omega, d\lambda)$$

Si on pose :

$$\begin{aligned}
\ln P(\omega) &= \lim_n E_{P_0}(\ln p_n) + \lim_n (\ln p_n - E_{P_0}(\ln p_n)) \\
&= \lim_n E_{P_0}(\ln p_n) - \frac{1}{2} \lim_n \eta_n \\
&= \ln \left(\exp \left(-\frac{1}{2} \|\varphi_n\|^2 \right) - \frac{1}{2} \iint \varphi(\lambda, \mu) \Phi_0(\omega, d\lambda) \Phi_1(\omega, d\lambda) \right)
\end{aligned}$$

d'où la solution de l'équation de Wiener-Hopf conduisant à la vraisemblance $\frac{dP_0}{dP_1}$.

5.4 Solution dans le cas rationnel de l'équation de vraisemblance

Soit φ une fonction vérifiant l'équation de Wiener-Hopf telle que :

$$\int_{\mathbb{R}^2} \exp \{i(\lambda s - \mu t)\} \varphi(\lambda, \mu) F_0(d\lambda) F_1(d\mu) = b(s - t)$$

Pour déterminer le rapport de vraisemblance des lois P_0 et P_1 , il faut donc résoudre l'équation ci-dessus. Des calculs explicites ont été faits dans le cas où F_0 et F_1 sont des fractions rationnelles ([6], [32], [33]).

En effet, si $F_0(d\lambda) = f_0(\lambda) d\lambda$ et $F_1(d\mu) = f_1(\mu) d\mu$, on peut toujours ramener le problème à la résolution de deux équations unidimensionnelles en posant :

$$K_s(\mu) = \int_{\mathbb{R}} \exp(i\lambda s) \varphi(\lambda, \mu) f_0(\lambda) d\lambda$$

et

$$\int_{\mathbb{R}} \exp(i\mu t) K_s(\mu) f_1(\mu) d\mu = b(s - t)$$

Il suffit donc de savoir résoudre une équation du type :

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(i\lambda t) \varphi(\lambda) \left[\frac{Q(i\lambda)}{P(i\lambda)} \right]^2 d\lambda = x(t) \quad 0 \leq t \leq T$$

où P, Q, x sont donnés et P est un polynome de degré r , Q un polynome de degré s et $x \in L^2$.

Si x admet des dérivées d'ordre $r - s$ dans L^2 , on se se replace dans une situation classique en théorie des distributions.

Dans la suite nous allons résumer les travaux de Rosanov ([33]).

Considérons X_t , $t \in [0, T]$ un processus gaussien stationnaire défini sur un espace de probabilité (Ω, \mathcal{A}, P) . Soit $\mathcal{B}(\Omega)$ la sous σ -algèbre des boréliens de Ω sur laquelle on définit deux mesures de probabilité gaussiennes P_1 et P_2 telles que :

$$\begin{aligned} E_{P_1}(X_t) &= E_{P_2}(X_t) = 0 \\ E_{P_1}(X_t X_s) &= R_1(s, t) \\ E_{P_2}(X_t X_s) &= R_2(s, t) \end{aligned}$$

La stationnarité du processus X_t nous assure que $R_1(s, t) = R_1(s - t)$ et $R_2(s, t) = R_2(s - t)$ et on peut les exprimer sous la forme :

$$\begin{aligned} R_1(s - t) &= \int \exp\{i\lambda(s - t)\} F_1(d\lambda) \\ R_2(s - t) &= \int \exp\{i\lambda(s - t)\} F_2(d\lambda) \end{aligned}$$

où $F_1(d\lambda)$ et $F_2(d\lambda)$ sont les mesures spectrales associées à P_1 et P_2 .

De plus, on suppose que P_1 et P_2 sont équivalentes pour utiliser des résultats établis sur les mesures gaussiennes équivalentes. Considérons l'ensemble $H_N(P_1)$ des combinaisons linéaires des variables X_t tel que ses éléments Y_N s'écrivent comme suit :

$$Y_N = \sum_{k=1}^N c_k X_{t_k} \quad t_k \in \Delta \subset [0, T], \text{ card}(\Delta) = N$$

Alors

$$E_{P_1} \left(\sum_{k=1}^N c_k X_{t_k} \right)^2 \approx E_{P_2} \left(\sum_{k=1}^N c_k X_{t_k} \right)^2 \quad (5.5)$$

ou bien en utilisant les mesures spectrales :

$$\int \left| \sum c_k \exp(i\lambda t_k) \right|^2 F_1(d\lambda) \approx \int \left| \sum c_k \exp(i\lambda t_k) \right|^2 F_2(d\lambda) \quad (5.6)$$

ou ce qui revient au même

$$F_1(d\lambda) \approx F_2(d\lambda) \quad (5.7)$$

L'équivalence 5.7 est vérifiée dès que l'on a :

$$0 < \lim_{\lambda \rightarrow \infty} \frac{F_2(d\lambda)}{F_1(d\lambda)} \leq \overline{\lim}_{\lambda \rightarrow \infty} \frac{F_2(d\lambda)}{F_1(d\lambda)} < \infty$$

Considérons les fonctions $U(\lambda, \mu) \in L^2(F_1, F_2)$ satisfaisant à la condition suivante :

$$\iint |U(\lambda, \mu)|^2 F_1(d\lambda) F_2(d\mu) < \infty$$

Alors, nous allons reprendre le théorème d'existence dans le cas rationnel.

Proposition 39 *Les probabilités P_1 et P_2 sont équivalentes si et seulement si la différence des fonctions de corrélation $b(s, t) = R_1(s - t) - R_2(s - t)$ s'écrit sous la forme :*

$$b(s, t) = \iint \exp\{-i(\lambda s - \mu t)\} U(\lambda, \mu) F_1(d\lambda) F_2(d\mu) \quad 0 \leq s, t \leq T \quad (5.8)$$

De plus, la densité $p(\omega) = \frac{P_2(d\omega)}{P_1(d\omega)}$ est donnée par la relation :

$$p(\omega) = D \exp\left\{-\frac{1}{2} \iint U(\lambda, \mu) \Phi(d\lambda) \overline{\Phi(d\mu)}\right\} \quad (5.9)$$

où $U(\lambda, \mu)$ est la solution de l'équation 5.8.

Démonstration : Soit $\Delta_1 \subset \Delta_2 \subset \dots$, une suite monotone de sous-ensembles finis de $[0, T]$ telle que $\bigcup_{n=1}^{\infty} \Delta_n$ soit partout dense dans l'intervalle $[0, T]$. Soit \mathcal{B}_n la σ -algèbre engendrée par les variables $X_t, t \in \Delta_n$. D'après la condition 5.6 les distributions P_1 et P_2 sont équivalentes sur \mathcal{B}_n , donc $E^{\mathcal{B}_n} \{p(\omega)\} = \frac{P_2(d\omega)}{P_1(d\omega)}$ est leur densité. ■

Lemme 40 *P_1 et P_2 sont équivalentes si et seulement si :*

$$\sup_n D_1 \ln \{p(\omega)\} < \infty \quad (5.10)$$

où D_1 est le symbole du moment d'ordre deux centré par rapport à P_1 .

Démonstration : On peut représenter $\ln \{p(\omega)\}$ sous la forme :

$$\ln \{p(\omega)\} = -\frac{1}{2} \sum_k \left[\ln \sigma_k^2 + \frac{1 - \sigma_k^2}{\sigma_k^2} \eta_k^2(\omega) \right] \quad (5.11)$$

En effet, la vraisemblance $p_n(\omega)$ s'écrit sous telle que :

$$p_n(\omega) = \left(\frac{\det R_1}{\det R_2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} {}^\tau X (R_2^{-1} - R_1^{-1}) X \right\}$$

où $X = \{X_t\}$, $t \in \Delta_n$.

Il est toujours possible de trouver une base dans le sous-espace engendré par les X_t , $t \in \Delta_n$, dans laquelle R_1^{-1} et R_2^{-1} se ramènent à deux formes diagonales simultanément. Alors, considérons la transformation :

$$X \longrightarrow \eta = O' R_1^{-\frac{1}{2}} X \quad \text{où } O'O = I \quad (5.12)$$

En appliquant cette transformation à la forme quadratique $Q = {}^\tau X (R_2^{-1} - R_1^{-1}) X$ on obtient \tilde{Q} telle que :

$$\tilde{Q} = {}^\tau \eta \left(O' R_1^{-\frac{1}{2}} R_2^{-1} R_1^{\frac{1}{2}} O - I \right) \eta$$

Le vecteur η est tel que :

$$\begin{aligned} E_{P_1}(\eta_t \eta_s) &= \delta_{st} \\ E_{P_2}(\eta_t \eta_s) &= \sigma_{st}^2 \delta_{st} \end{aligned}$$

où δ_{st} est le symbole de Kronecker et σ_{st}^2 sont les valeurs propres de la matrice définie positive $B = R_1^{\frac{1}{2}} R_2 R_1^{-\frac{1}{2}}$.

Alors,

$$O'B^{-1}O = \text{diag} \left(\sigma_1^{-2} \quad \sigma_2^{-2} \quad \dots \quad \sigma_n^{-2} \right)$$

et

$$\det (R_1^{-1} R_2) = \det B = \prod_k \sigma_k^2$$

Finalement

$$\tilde{Q} = \sum_k \left(\frac{1}{\sigma_k^2} - 1 \right) \eta_k^2$$

D'après la transformation 5.12 on a :

$$E_{P_1}(\eta_k) = E_{P_2}(\eta_k) = 0 \quad E_{P_1}(\eta_k^2) = 1 \quad E_{P_2}(\eta_k^2) = \sigma_k^2 \quad k = 1, 2, \dots, n$$

Alors il vient que

$$E_{P_2} \{ \ln p_n(\omega) \} = -\frac{1}{2} \sum_k \{ \ln \sigma_k^2 + 1 - \sigma_k^2 \}$$

et

$$D_1 \ln \{p_n(\omega)\} = \sum_k \frac{(1 - \sigma_k^2)^2}{\sigma_k^4}$$

En effet, rappelons que la matrice B est telle que $I - B$ est de Hilbert-Schmidt, i.e. si $(\mu_k)_{k=1, \dots, n}$ est le spectre de $I - B$ alors :

$$\sum_k \mu_k < \infty \quad \text{et} \quad \mu_k = 1 - \sigma_k^2$$

Par conséquent

$$\begin{aligned} E_{P_2} \{\ln p_n(\omega)\} &= -\frac{1}{2} \sum_k \{\ln \sigma_k^2 + 1 - \sigma_k^2\} \\ &\approx -\frac{1}{2} \sum_k \{\ln(1 - \mu_k) + \mu_k\} \approx \frac{1}{4} \sum_k \mu_k^2 \\ &\approx \frac{1}{4} \sum_k (1 - \sigma_k^2)^2 \end{aligned}$$

Finalement

$$D_1 \ln \{p_n(\omega)\} \approx \sum_k (1 - \sigma_k^2)^2 \approx E_{P_2} \{\ln p_n(\omega)\}$$

D'où l'on déduit que la condition 5.10 est équivalente à :

$$E_{P_2} \{\ln p_n(\omega)\} < \infty$$

La démonstration du lemme 40 est ainsi terminée. ■

Reprenons la démonstration de la proposition 39, et exprimons la vraisemblance en fonction de X_t directement tel que :

$$\ln p_n(\omega) = -\frac{1}{2} \left\{ \ln \frac{\det R_2}{\det R_1} + \sum_{s, t \in \Delta_n} C(s, t) X_s X_t \right\}$$

Puisque P_1 et P_2 sont équivalentes, alors le sous-espace engendré par les quantités $\eta_n = \sum_{s, t \in \Delta_n} C(s, t) X_s X_t$ est isométrique à l'espace H^2 des éléments symétriques de la forme :

$$h_n = \eta_n - E_1 \eta_n = \sum_{s, t \in \Delta_n} C(s, t) \{X_s X_t - R_1(s, t)\}$$

Alors on peut écrire

$$\ln p_n(\omega) = E_1 \{ \ln p_n(\omega) \} - \frac{1}{2} \sum_{s,t \in \Delta_n} C(s,t) \{ X_s X_t - R_1(s,t) \}$$

où $\{C(s,t)\} = \{R_2^{-1}(s,t)\} - \{R_1^{-1}(s,t)\}$.

Introduisons la mesure spectrale stochastique $\Psi(d\lambda, d\mu)$ telle que :

$$\Psi(\Delta_1 \times \Delta_2) = \Phi(\Delta_1) \overline{\Phi(\Delta_2)} - F(\Delta_1 \cap \Delta_2)$$

Il est facile de voir que tout élément h_n de H^2 peut s'écrire sous la forme :

$$h_n = \int \int U_n(\lambda, \mu) \Psi(d\lambda, d\mu)$$

où $U_n(\lambda, \mu) = \sum_{s,t \in \Delta_n} C(s,t) \exp\{i(\lambda s - \mu t)\} \in L^2_{\Delta_n}(F_1 \times F_2)$ qui est partout dense dans $L^2(F_1 \times F_2)$.

Il est équivalent de dire que $\lim_{n \rightarrow \infty} U_n(\lambda, \mu) = U(\lambda, \mu) \in L^2_{[0,T]}(F_1 \times F_2)$, et donc

$$\ln p_n(\omega) = \lim_{n \rightarrow \infty} E_1 \{ \ln p_n(\omega) \} - \frac{1}{2} \int \int U(\lambda, \mu) \Psi(d\lambda, d\mu)$$

Pour terminer, vérifions que chacune des fonctions $U_n(\lambda, \mu)$ satisfait à l'équation 5.8. Alors, d'après la définition de $U_n(\lambda, \mu)$ on a :

$$\begin{aligned} & \iint \exp\{-i(\lambda s - \mu t)\} \sum_{s',t' \in \Delta_n} C(s',t') \exp\{i(\lambda s' - \mu t')\} F_1(d\lambda) F_2(d\mu) \\ &= \sum_{s',t' \in \Delta_n} C(s',t') \int \exp\{-i\lambda(s - s')\} F_1(d\lambda) \int \exp\{i\mu(t - t')\} F_2(d\mu) \\ &= \sum_{s',t' \in \Delta_n} R_1(s,s') C(s',t') R_2(t,t') = \{R_1(R_2^{-1} - R_1^{-1})R_2\}(s,t) \\ &= (R_1 - R_2)(s,t) = b(s,t) \end{aligned}$$

Revenons à l'équation 5.8 pour remarquer que c'est une équation intégrale du type Wiener-Hopf de la fonction $U(\lambda, \mu)$. La solution de ce type d'équations, considérée comme élément de $L^2_{[0,T]}$ est souvent unique. On sait que la condition d'existence d'une telle solution est qu'il existe une constante C telle que :

$$\left| \sum_{s,t} C(s,t) b(s,t) \right|^2 \leq C \int \left| \sum_{s,t} C(s,t) \exp\{i(\lambda s - \mu t)\} \right|^2 F_1(d\lambda) F_2(d\mu)$$

Enfin, nous énonçons un résultat ([33]) utile pour la recherche de la solution de l'équation 5.8 dans le cas où $F_1(d\lambda)$ est absolument continue de densité $f_1(\lambda)$ ayant un bon comportement asymptotique :

Proposition 41 *Une condition nécessaire et suffisante d'existence d'une solution pour l'équation 5.8 est que $b(s, t)$ ait les dérivées partielles $\frac{\partial^{2n-1}b(s,t)}{\partial s^{n-1}\partial t^n}$ et $\frac{\partial^{2n-1}b(s,t)}{\partial s^n\partial t^{n-1}}$ absolument continues et :*

$$\int_0^T \int_0^T \left| \frac{\partial^{2n}b(s,t)}{\partial s^n\partial t^n} \right|^2 dsdt < \infty$$

5.5 Le lemme de Neyman-Pearson

5.5.1 Cas discret

Soient F et G deux mesures gaussiennes centrées et stationnaires sur $\mathbb{R}^{\mathbb{Z}}$. On désigne par X^N l'observation (X_1, X_2, \dots, X_N) d'un processus aléatoire. On suppose que pour tout N assez grand, F et G admettent des densités F_N et G_N par rapport à la mesure de Lebesgue sur \mathbb{R}^N . Pour qu'il en soit ainsi, il faut et il suffit que F et G aient des densités spectrales f et g telles que $\ln f, \ln g \in L_T^1$ (s'il n'en est pas ainsi, le problème statistique étudié est trivial).

On pose $L_N(t) = \frac{1}{N} \log \phi_N(t)$ avec $\phi_N(t) = E_F \exp \left\{ t \ln \frac{dG_N}{dF_N} \right\}$, $t \in [0, 1]$. On désigne par $(T_N f)_N$ et $(T_N g)_N$ les suites des matrices de Toeplitz associées à f et g . Alors, on a :

$$\ln \frac{dG_N}{dF_N} = \frac{1}{2} \log \{ \det (T_N^{-1} g T_N f) \} - \frac{1}{2} \tau X^N (T_N^{-1} f - T_N^{-1} g) X^N$$

Par ailleurs, comme la covariance de $T_N^{-\frac{1}{2}} f X^N$ est la matrice unité I_N , alors en désignant par $(\mu_i^N)_{i=1, \dots, N}$ les valeurs propres de $A_N = I_N - T_N^{-\frac{1}{2}} f T_N^{-1} g T_N^{\frac{1}{2}} f$, on a :

$$NL_N(t) = \frac{t}{2} \log \{ \det (T_N^{-1} g T_N f) \} - \frac{1}{2} \sum_{i=1}^N \ln (1 - \mu_i^N t)$$

A N fixé, pour $|t| < \frac{1}{\sup_{1 \leq i \leq N} |\mu_i^N|}$ on a :

$$\begin{aligned} \sum_{i=1}^N \ln(1 - \mu_i^N t) &= - \sum_{p=1}^{\infty} \frac{t^p}{p} \text{Tr}(A_N^p) \\ &= - \sum_{\ell=1}^{\infty} (-1)^\ell \frac{1}{\ell!} \text{Tr} (T_N f T_N^{-1} g)^\ell \left(\frac{t}{1-t} \right)^\ell + N \ln(1-t) \end{aligned}$$

Soit $z \in \mathbb{C}$. Comme $T_N f$ et $T_N g$ sont régulières, on a dans un voisinage ouvert de 0, $T_N(zf + g)$ régulière et donc d'après ([5]) :

$$\begin{aligned} \frac{d}{dz} \ln \{ \det (T_N(zf + g)) \} &= \text{Tr} \{ T_N f T_N^{-1} (zf + g) \} \\ \frac{d}{dz^\ell} \ln \{ \det (T_N(zf + g)) \} \Big|_{z=0} &= (-1)^{\ell-1} (\ell-1)! \text{Tr} (T_N f T_N^{-1} g)^\ell \end{aligned}$$

D'où

$$\sum_{i=1}^N \ln(1 - \mu_i^N t) = - \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \ln \{ \det (T_N(zf + g)) \} \Big|_{z=0}^{(\ell)} \left(\frac{t}{1-t} \right)^\ell + N \ln(1-t)$$

De l'analyticit  au voisinage de 0 de la fonction $z \longrightarrow \ln \{ \det (T_N(zf + g)) \}$, on d duit

$$\sum_{i=1}^N \ln(1 - \mu_i^N t) = \ln \left\{ \det T_N \left(\frac{t}{1-t} f + g \right) \right\}$$

D'où

$$\begin{aligned} NL_N(t) &= \frac{t}{2} \ln \{ \det (T_N^{-1} g T_N f) \} - \\ &\quad - \frac{1}{2} \left[\ln \left\{ \det T_N \left(\frac{t}{1-t} f + g \right) \right\} - \ln (\det T_N g) + N \ln(1-t) \right] \end{aligned}$$

Finalement

$$NL_N(t) = \frac{1}{2} [t \ln (\det T_N f) + (1-t) \ln (\det T_N g) - \ln \{ \det T_N (tf + (1-t)g) \}]$$

Cette formule est valable au voisinage de 0.

Mais la fonction $z \longrightarrow \ln \det (T_N(zf + (1-z)g))$ est analytique dans un \mathbb{C} -voisinage de tout point de $]0, 1[$, car $\ln(tf + (1-t)g) \in L^1$ par convexit  et donc $T_N(tf + (1-t)g)$ est r guli re pour $t \in [0, 1]$.

Par suite, par prolongement analytique la formule est valable pour $t \in]0, 1[$ et par continuité pour $t \in [0, 1]$.

D'après le théorème élémentaire de Szegö, si $\ln f \in L^1$, $\frac{1}{N} \ln \det(T_N f) \longrightarrow \frac{1}{2\pi} \int_{[0, T]} \ln f(\theta) d\theta$ quand $N \longrightarrow +\infty$, et donc

Proposition 42 *Si $\ln f, \ln g \in L^1_{[0, T]}$ et $t \in [0, 1]$, alors*

$$L_N(t) \longrightarrow L(t) \quad N \longrightarrow \infty$$

avec
$$L(t) = \frac{1}{4\pi} \int_{-\pi}^{+\pi} [t \ln f(\theta) + (1-t) \ln g(\theta) - \ln(tf(\theta) + (1-t)g(\theta))] d\theta$$

A partir de ce résultat, on déduit comme dans ([12]), la formule de Chernoff et les propriétés asymptotiques du lemme de Neyman-Pearson.

5.5.2 Cas continu

Dans le paragraphe précédent, nous avons calculé la transformée de Laplace de la vraisemblance. On pourra utiliser cette formule pour T assez grand pour déterminer une borne de Chernoff pour $P \left\{ \ln \frac{dF}{dG}(0, T) > \lambda \right\}$. Le calcul exact de la loi de $\ln \frac{dF}{dG}$ est compliqué puisque ψ est solution de l'équation :

$$\int_0^T \int_0^T \xi(s, \omega) \xi(t, \omega') \psi(\omega, \omega') dP(\omega) dP_1(\omega') = B(s, t) - B_1(s, t)$$

Cependant, pour T suffisamment grand et ξ assez régulier, il semble raisonnable de discrétiser le problème et d'appliquer la méthode précédente au processus $\left(\xi_{\frac{kT}{N}} \right)_{k=1, 2, \dots, N}$ ce qui correspond d'ailleurs aux procédures usuelles de calcul numérique.

Bibliographie

- [1] Akaiké H. Markovian Representation of Stochastic Process by Canonical Variables, SIAM Journal on Control (1976)
- [2] Anderson O.D. Time Series Analysis : The Box-Jenkins Approach. Butterworths, London (1975).
- [3] Aoki M. State Space Modelling of Time Series, Springer-Verlag, Berlin (1987)
- [4] Azencot R. et Dacunha-Castelle D. Séries d'observations irrégulières, Modélisation et prévision, Masson, Paris (1984)
- [5] Baxter H. A norm inequality for a "finite section" Wiener-Hopf equation. Illinois J. Math. 7, pp. 97-103 (1963)
- [6] Box G.E.P. and Jenkins G.M. Time Series Analysis, Forecasting and Control. Holden Day, San Francisco (1976).
- [7] Breidt F.J. and Davis R.A. Time-reversibility, identifiability and independence of innovations for stationary time series. J. Time Series Analysis, 13, 377-390 (1991)
- [8] Brockwell P.J. and Davis R.A. Applications of innovation representations in time series analysis, Probability and Statistics, Essays in Honor of Franklin A. Graybill, J.N. Srivastava 5th ed.) Elsevier, Amsterdam, 61-84 (1988a)
- [9] Brockwell P.J. and Davis R.A. Time Series : "Theory and Methods". Springer-Verlag, New York, 2nd Ed. (1996).
- [10] Broersen P.M.T. and de Waele S. "Empirical Time Series Analysis and Maximum Likelihood Estimation". Proc. 2nd IEEE Benelux Signal Processing (SPS 2000), Hilvarenbeek, The Netherlands, (2000).
- [11] Dakhmouche M. Determination of Initial Values for the Inconditional Likelihood Function of a Moving Average Process. Far East Journal of Theoretical Statistics, 17(1) (2005), pp. 31-42.

- [12] Davies R.B. Asymptotic inference in stationary time series. *Adv. App. Prob.* 5, pp. 469-497 (1973)
- [13] Davis M.H.A. and Vinter R.B. *Stochastic Modelling and Control*, Chapman and Hall, London (1985)
- [14] Dzaparidze K. Estimateurs asymptotiquement efficaces pour les paramètres du spectre d'une série chronologique gaussienne. Publication de l'Univ. de Tbilissi (1977)
- [15] Dzaparidze K. *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*, Springer-Verlag, Tokyo (1986)
- [16] Faure P. "Réalizations markoviennes de processus stationnaires", Thèse de Doctorat d'état, Rapport Laboria n°13 (1973)
- [17] Feller W. *An Introduction to Probability Theory and Its Applications...*, V 2, 2nd Ed., John Wiley & Sons, New York (1971)
- [18] Gantmacher. "Théorie des matrices", T 1, Dunod (1966)
- [19] Grenander U. and Rosenblatt M. *Statistical Analysis of Time Series*, John Wiley, New York (1957)
- [20] Grenander U. et Szegö G. *Toeplitz forms and their applications*. Univ. of Calif. Press, Berkeley (1968)
- [21] Guyon X. et Prum B. Estimation et tests pour des champs gaussiens indexés par \mathbb{Z}^p , Preprint, Orsay (1978)
- [22] Hamilton J.D. *Time Series Analysis*. Princeton University Press, Princeton (1994).
- [23] Hannan E.J. *Multiple Time Series*, John Wiley, New York (1970)
- [24] Hannan E.J. and Deistler M. *The Statistical Theory of Linear Systems*, John Wiley, New York (1988)
- [25] Hajek J. On a property of normal distributions for arbitrary stochastic processes, *Czechoslovak Math. J.*, 8, pp. 82-87 (1958)
- [26] Ibragimov I., Rosanov Y.A. *Processus aléatoires gaussiens*, Ed. Mir, Moscou (1974)
- [27] Kalman R.E. and Bucy. "New results in linear filtering and prediction theory", *Trans-ASME, J. Basic Engrg, Series D*, V. 83, pp. 95-107 (1961)
- [28] Le Cam L. *Notes on asymptotic methods in statistical decision theory*, Centre de Recherches Mathématiques, Université de Montréal (1974)

- [29] Lehmann E.L. Testing Statistical Hypothesis, John Wiley & Sons, Inc., New York (1959)
- [30] Makridakis S. "A Survey of Time Series". Int. Stat. Rev., Longman, V. 44, 29-70 (1976)
- [31] Neveu.J "Processus aléatoires gaussiens", Presse de l'université de Montréal (1966)
- [32] Pisarenko V.F. On the computation of the likelihood ratio for gaussian processes with a rational spectrum, Theory Prob. Applications, 10, pp. 299-303 (1965)
- [33] Rosanov Y.A. On gaussian distribution densities and Wiener-Hopf integral equations, Theory Prob. Applications, 11, pp. 152-160 (1966)
- [34] Rosanov Y.A. Processus aléatoires, Ed. Mir, Moscou (1975)
- [35] G. Ruckebusch. Représentation markovienne des processus gaussiens stationnaires. Thèse de Doctorat, Paris VI, Paris (1975).
- [36] Stoica P. and Moses R. "Introduction to Spectral Analysis", Upper Sadd River, New Jersey, Printice Hall (1997)
- [37] Yule G.U. Why do we sometimes get nonsense-correlations between time series? A study on sampling and the nature of time series. Journal of the Royal Statistical Society, 89, pp. 1-64 (1926)
- [38] Whittle P. "Estimation and information in stationary time series", Ark. Math., 2, 423-434, (1953).
- [39] Wiener N. Generalized harmonic analysis, Acta Math.,55, p. 117 (1930)
- [40] Wilkinson J.H. The Algebraic Eigenvalue Problem, Clarendon Press, Oxford (1965)

ملخص

في هذه المساهمة تمحور عملنا حول دراسة و تحليل للسلاسل الزمنية و العمليات التصادفية للزمن أحادية البعد لأن هناك العديد من الجوانب غير مدروسة في هذا الميدان. ولذلك فهو من الضروري تلخيص و تقديم النتائج المحصل عليها لمستعملي هذه التقنيات.

تطرقنا خلال هذا العمل الى تقديم مختلف نماذج تمثيل العمليات التصادفية للزمن و خاصة النموذج التمثيلي "كلمان-بوسي" و نموذج "بوكس-جنكينس"، و هذا الأخير مشهور و سهل الاستعمال. كذلك أستعملنا التحليل الطيفي للسلاسل الزمنية و هو مبني على تحويل "فوريي" و يسمى عادة الشكل الدوري. بعدها تطرقنا لدراسة المترajحات أو معقوليات العمليات التصادفية للزمن، و نعلم أنه حتى اذا كان المجال الزمني للدراسة متقطع، مشكلة حساب هذه المترajحات معقد و لهذا الغرض أستعملنا بعض التقريبات للمترajحات و خاصة تقريب "ويتل"، "بوكس-جنكينس" وكذلك تقريب "توبليز-زيقو". فتقريب "بوكس-جنكينس" مثلا مبني على مجموع المربعات التي لا يمكن حسابها إلا بعد تحديد قيم للمعلومة التصادفية التي تحققت قبل الحصول على السلسلة الزمنية، و هذه القيم نحصل عليها باستعمال خوارزم الذهاب و الرجوع.

أخيرا و بعد تلخيص أعمال "رزنوف و أبرجموف" حول حساب المتراجحات،
عندما يكون الزمن مستمرا باستعمال معادلات "فينار-هوبف"، حاولنا تطبيق هذه
التقاريبات في اطار اختبار الفرضيات.