



HAL
open science

Estimation et test dans les modèles paramétriques de processus stationnaires

Tuan Pham Dinh

► **To cite this version:**

Tuan Pham Dinh. Estimation et test dans les modèles paramétriques de processus stationnaires. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1975. tel-00286247

HAL Id: tel-00286247

<https://theses.hal.science/tel-00286247>

Submitted on 9 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée à

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE
INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

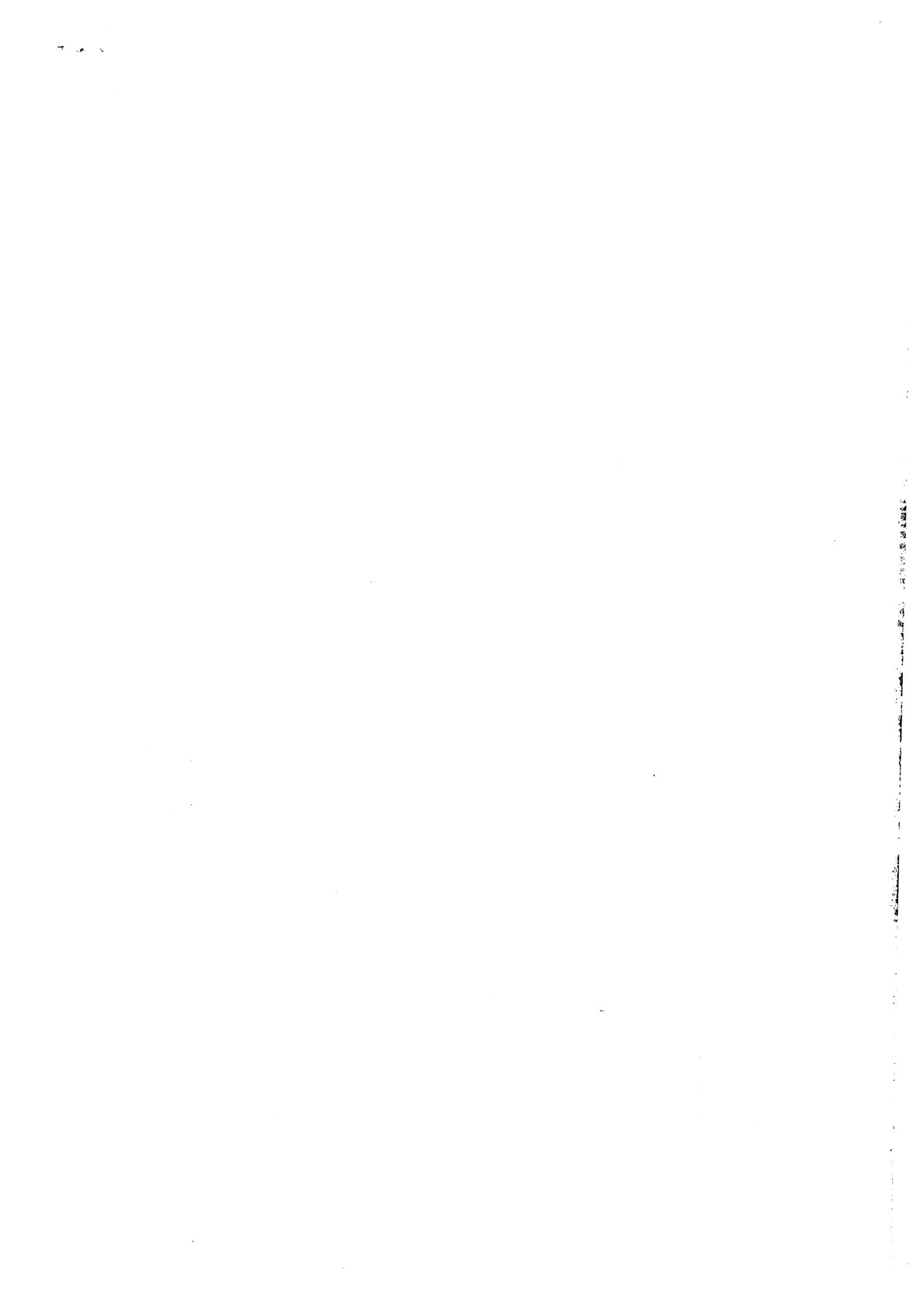
POUR OBTENIR LE ~~GRADUÉ~~ DE
DOCTEUR ES SCIENCES MATHÉMATIQUES

PHAM - DINH - Tuan

**ESTIMATION ET TEST DANS
LES MODELES PARAMETRIQUES
DE PROCESSUS STATIONNAIRES**

Soutenue le 27 janvier 1975 devant la Commission d'Examen

Monsieur	J.R. BARRA	Président
Messieurs	F. BRODEAU Ph. COURREGÉ M. METIVIER M.B. PRIESTLEY B. VAN CUTSEM	Examineurs



Je tiens à exprimer ma profonde reconnaissance à :

Monsieur le Professeur J.R. BARRA qui, par ses conseils et son encouragement, m'a beaucoup aidé dans la réalisation de ce travail et qui a bien voulu me faire l'honneur de présider ce jury.

Messieurs les Professeurs M. METIVIER, de l'Université de Rennes et M.B. PRIESTLEY de l'Université de Manchester pour l'intérêt qu'ils ont porté à mon travail et pour leur acceptation de prendre part au jury.

Messieurs B. VAN CUTSEM, Professeur, F. BRODEAU, Maître de Conférences et Ph. COURREGÉ, Maître de recherches au C.N.R.S. qui ont bien voulu accepter de prendre part au jury.

Je remercie Madame TREVISAN qui a assuré la dactylographie du manuscrit et au personnel du service de Reproduction du Laboratoire qui en a réalisé le tirage avec beaucoup de soins.

PHAN-DINH-Tuan

M. MICHEL SOUTIF
M. GABRIEL CAU

Présidents M. Louis NEEL
Vice-Présidents MM. Lucien BONNETAIN
Jean BENOIT

MEMBRES DU CORPS ENSEIGNANT DE L'USMG

PROFESSEURS TITULAIRES

MM. ANGLES D'AURIAC	Mécanique des fluides
ARNAUD Georges	Clinique des maladies infectieuses
ARNAUD Paul	Chimie
AUBERT Guy	Physique
AYANT Yves	Physique approfondie
Mme BARBIER Marie-Jeanne	Electrochimie
MM. BARBIER Jean-Claude	Physique expérimentale
BARBIER Reynold	Géologie appliquée
BARJON Robert	Physique nucléaire
BARNOUD Fernand	Biosynthèse de la cellulose
BARRA Jean-René	Statistiques
BARRIE Joseph	Clinique chirurgicale
BEAUDOING André	Pédiatrie
BERNARD Alain	Mathématiques Pures
Mme BERTRANDIAS Fraçoise	Mathématiques Pures
MM. BEZES Henri	Chirurgie générale
BLAMBERT Maurice	Mathématiques Pures
BOLLIET Louis	Informatique (IUT B)
BONNET Georges	Electrotechnique
BONNET Jean-Louis	Clinique ophtalmologique
BONNET-EYMARD Joseph	Pathologie médicale
BOUCHERLE André	Chimie et Toxicologie
BOUCHEZ Robert	Physique nucléaire
BOUSSARD Jean-Claude	Mathématiques Appliquées
BRAVARD Yves	Géographie
CABANEL Guy	Clinique rhumatologique et hydrologie
CALAS François	Anatomie
CARRAZ Gilbert	Biologie animale et pharmacodynamie
CAU Gabriel	Médecine légale et Toxicologie
CAUQUIS Georges	Chimie organique
CHABAUTY Claude	Mathématiques Pures
CHARACHON Robert	Oto- Rhino- Laryngologie
CHATEAU Robert	Thérapeutique
CHIBON Pierre	Biologie animale
COEUR André	Pharmacie chimique et chimie analytique
CONTAMIN Robert	Clinique gynécologique
COUDERC Pierre	Anatomie pathologique
CRAYA Antoine	Mécanique
Mme DEBELMAS Anne-Marie	Matière médicale
MM. DEBELMAS Jacques	Géologie générale
DEGRANGE Charles	Zoologie
DEPORTES Charles	Chimie minérale
DESRE Pierre	Métallurgie
DESSAUX Georges	Physiologie animale
DODU Jacques	Mécanique appliquée
DOLIQUE Jean-Michel	Physique des plasmas
DREYFUS Bernard	Thermodynamique
DUCROS Pierre	Cristallographie
DUGOIS Pierre	Clinique de Dermatologie et Syphilligraphie
FAU René	Clinique neuro-psychiatrique

MM. GAGNAIRE Didier	Chimie physique
GALLISSOT François	Mathématiques Pures
GALVANI Octave	Mathématiques Pures
GASTINEL Noël	Analyse numérique
GAVEND Michel	Pharmacologie
GEINDRE Michel	Electroradiologie
GERBER Robert	Mathématiques Pures
GERMAIN Jean-Pierre	Mécanique
GIRAUD Pierre	Géologie
KAHANE André	Physique générale
KLEIN Joseph	Mathématiques Pures
KOSZUL Jean-Louis	Mathématiques Pures
KRAVTCHENKO Julien	Mécanique
KUNTZMANN Jean	Mathématiques Appliquées
LACAZE Albert	Thermodynamique
LACHARME Jean	Biologie végétale
LAJZEROWICZ Joseph	Physique
LATREILLE René	Chirurgie générale
LATURAZE Jean	Biochimie pharmaceutique
LAURENT Pierre	Mathématiques Appliquées
LEDRU Jean	Clinique médicale B
LLIBOUTRY Louis	Géophysique
LONGQUEUE Jean-Pierre	Physique nucléaire
LOUP Jean	Géographie
Mlle LUTZ Elisabeth	Mathématiques Pures
MALGRANGE Bernard	Mathématiques Pures
MALINAS Yves	Clinique obstétricale
MARTIN-NOEL Pierre	Séméiologie médicale
MAZARE Yves	Clinique médicale A
MICHEL Robert	Minéralogie et pétrographie
MOURIQUAND Claude	Histologie
MOUSSA André	Chimie nucléaire
NEEL Louis	Physique du solide
OZENDA Paul	Botanique
PAYAN Jean-Jacques	Mathématiques Pures
PEBAY-PEYROULA Jean-Claude	Physique
RASSAT André	Chimie systématique
RENARD Michel	Thermodynamique
REULOS René	Physique industrielle
RINALDI Renaud	Physique
ROGET Jean	Clinique de pédiatrie et de puériculture
DE ROUGEMONT Jacques	Neuro-chirurgie
SEIGNEURIN Raymond	Microbiologie et Hygiène
SENGEL Philippe	Zoologie
SOUTIF Michel	Physique générale
TANCHE Maurice	Physiologie
TRAYNARD Philippe	Chimie générale
VAILLANT François	ZOologie
VALENTIN Jacques	Physique nucléaire
VAUQUOIS Bernard	Calcul électronique
Mme VERAIN Alice	Pharmacie galénique
MM. VERAIN André	Physique
VEYRET Paul	Géographie
VIGNAIS Pierre	Biochimie médicale
YOCCOZ Jean	Physique nucléaire théorique

PROFESSEURS ASSOCIES

MM. ASCARELLI Gianni	Physique
CHEEKE John	Thermodynamique
GILLESPIE John	I. S. N.
ROCKAFELLAR Ralph	Mathématiques Appliquées
WOHLFARTH Erich	Physique du solide

MM.	COHEN-ADDAD Jean Pierre	Spectrométrie physique
	COLOMB Maurice	Biochimie médicale
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie
	CYROT Michel	Physique du solide
	DELOBEL Claude	M.I.A.G.
	DUSSAUD René	Mathématique (OUS)
Mme	ETEPRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine légale
	FONTAINE Jean Marc	Mathématiques Pures
	GENSAC Pierre	Botanique
	GIDON Maurice	Géologie
	GRIFFITHS Michaël	Mathématiques Appliquées
	GROS Yves	Physique (stag.)
	GROULADE Joseph	Biochimie médicale
	GUITTON Jacques	Chimie
	IVANES Marcel	Electricité
	JALBERT Pierre	Histologie
	KRAKOWIAK Sacha	Mathématique Appliquées
Mme	LAJZEROWICZ Jeannine	Physique
MM.	LEROY Philippe	Mathématiques
	LOISEAUX Jean Marie	Physique nucléaire
	MACHE Régis	Physiologie végétale
	MAGNIN Robert	Hygiène et Médecine préventive
	MARECHAL Jean	Mécanique
	MARTIN-BOUYER Michel	Chimie (CUS)
	MICHOULIER Jean	Physique (I.U.T. "A")
Mme	MINIER Colette	Physique
MM.	MICOUD Max	Maladie infectieuses
	NEGRE Robert	Mécanique
	PARAMELLE Bernard	Pneumologie
	PECCOUD François	Analyse (IUT B)
	PEFFEN René	Métallurgie
	PELMONT Jean	Physiologie animale
	PERRET Jean	Neurologie
	PHELIP Xavier	Rhumatologie
	RACHAIL Michel	Médecine interne
	RACINET Claude	Gynécologie et obstétrique
	RAYNAUD Hervé	MI.A.G.
	RENAUD Maurice	Chimie
Mme	RENAUDET Jacqueline	Bactériologie
M.	RICHARD Lucien	Botanique
Mme	RINAUDO Marguerite	Chimie macromoléculaire
MM.	ROMIER Guy	Mathématiques (IUT B)
	SHOM Jean Claude	Chimie Générale
	STIEGLITZ Paul	Anesthésiologie
	STOEBNER Pierre	Anatomie pathologique
	VROUSOS Constantin	Radiologie

MAITRES DE CONFERENCES ASSOCIES

MM.	CRABBEE Pierre	C.E.R.M.O.
	CABOT	Mathématiques Appliquées
	CURRIE Jan	Mathématiques Appliquées

CHARGES DE FONCTIONS DE MAITRES DE CONFERENCES

MM.	BARGE Michel	Neuro-chirurgie
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculai:
	CORDONNIER Daniel	Néphrologie
	DENIS Bernard	Cardiologie
	KOLODIE Lucien	Hématologie
	RAMBAUD Pierre	Pédiatrie
	ROCHAT Jacques	Hygiène et hydrologie

"MEMBRES DU CORPS ENSEIGNANT DE L'I.N.P.G."PROFESSEURS TITULAIRES

MM. BENOIT Jean	Radioélectricité
BESSON Jean	Electrochimie
BONNETAIN Lucien	Chimie Minérale
BONNIER Etienne	Electrochimie, Electrometallurgie
BRISSONNEAU Pierre	Physique du solide
BUYLE-BODIN Maurice	Electronique
COUMES André	Radioélectricité
FELICI Noël	Electrostatique
PAUTHENET René	Physique du solide
PERRET René	Servomécanismes
SANTON Lucien	Mécanique
SILBER Robert	Mécanique des fluides

PROFESSEUR ASSOCIE

M. BOUDOURIS Georges	Radioélectricité
----------------------	------------------

PROFESSEURS SANS CHAIRE

MM. BLIMAN Samuel	Electronique
BLOCH Daniel	Physique du solide et cristallographie
COHEN Joseph	Electrotechnique
DURAND François	Metallurgie
MOREAU René	Mécanique
POLOUJADOFF Michel	Electrotechnique
VEILLON Gérard	Informatique fondamentale et appliquée
ZADWORNY François	Electronique

MAITRES DE CONFERENCES

MM. BOUVARD Maurice	Génie mécanique
CHARTIER Germain	Electronique
FOULARD Claude	Automatique
GUYOT Pierre	Chimie minérale
JOUBERT Jean Claude	Physique du solide
LACOUME Jean Louis	Géophysique
LANCIA Roland	Physique atomique
LESPINARD Georges	Mécanique
MORET Roger	Electrotechnique nucléaire
ROBERT François	Analyse numérique
SABONNADIÈRE Jean Claude	Informatique fondamentale et appliquée
Mme SAUCIER Gabrièle	Informatique fondamentale et appliquée

MAITRE DE CONFERENCES ASSOCIE

M. LANDAU Ioan Doré	Automatique
---------------------	-------------

CHARGE DE FONCTIONS DE MAITRES DE CONFERENCES

M. ANCEAU François	Mathématiques appliquées
--------------------	--------------------------

TABLE DES MATIERES

INTRODUCTION -----	1
CHAPITRE 1 - METHODE GENERALE D'ESTIMATION DANS LES MODELES DE PROCESSUS STATIONNAIRE CENTRE EN TEMPS DISCRET	
1. Différents modèles de processus en temps discret -----	8
2. Le "modèle fini" général -----	12
3. Fonction de vraisemblance et ses approximations asympto- tiques -----	15
4. Propriétés asymptotiques de la fonction de vraisemblance	21
5. Construction de l'estimateur -----	24
ANNEXE : Démonstrations de lemmes et théorèmes -----	31
CHAPITRE 2 - SPECIFICATION DE LA METHODE GENERALE DU CHAPITRE PRECEDENT	
1. Estimation et test d'adéquation du modèle de processus auto-régressif -----	46
2. Estimation et test dans le modèle de processus en moyen- ne mobile -----	61
3. Estimation dans le modèle de processus mixte -----	75
CHAPITRE 3 - EXTENSION DE LA METHODE DU CHAPITRE PRECEDENT AU CAS DU MODELE DE REGRESSION A RESIDUS STATIONNAIRES	
1. Le modèle -----	83
2. Les approximations de la fonction de vraisemblance -----	86
3. Propriétés asymptotiques de la fonction de vraisemblance	91
4. Construction de l'estimateur -----	100
5. Sur l'estimation des coefficients de régression -----	102
CHAPITRE 4 - ESTIMATION DES PARAMETRES D'UN PROCESSUS GAUSSIEN EN TEMPS CONTINU STATIONNAIRE CENTRE DE DENSITE SPECTRALE RATIONNELLE	
1. Définition du problème et notations -----	108
2. Calcul de la fonction de vraisemblance -----	109
3. Fonction de vraisemblance asymptotique -----	114
4. Propriétés asymptotiques de la fonction L_T -----	117
5. Construction de l'estimateur -----	120
ANNEXE : Démonstrations des lemmes et théorèmes -----	127

CHAPITRE 5 - EXTENSION DE LA METHODE D'ESTIMATION DU CHAPITRE PRECEDENT

1. Estimation dans le modèle d'évolution stochastique linéaire 151

2. Estimation dans le modèle de régression à résidus stationnaires ----- 158

Bibliographie ----- 177

I N T R O D U C T I O N

Dans ce travail, nous étudions d'une manière systématique les méthodes d'estimation et test dans les modèles courants de séries chronologiques, qui s'appuient sur la méthode du maximum de vraisemblance et nous généralisons ces méthodes aux problèmes d'estimation correspondant au cas des processus gaussiens en temps continu.

Le premier chapitre est consacré au problème d'estimation dans le modèle général de processus :

$$I.1. \quad X(t) = \sum_{j=0}^{\infty} A_{\theta}(j) \varepsilon(t-j)$$

où $\varepsilon(t)$, $t \in \mathbb{Z}$, sont des vecteurs aléatoires indépendants, centrés, de même loi ayant une matrice de covariance G et où les matrices $A_{\theta}(j)$, $j=0,1,\dots$, dépendent d'un paramètre $\theta \in \Theta \subset \mathbb{R}^k$. Le paramètre spécifiant le modèle est $v = (\theta, G)$.

Whittle [29] a résolu le problème d'estimation de v par maximisation de l'approximation suivante de la fonction log-vraisemblance :

$$L_N(v) = -\frac{N}{2} \text{Log det}(2\pi G) - \frac{N}{4\pi} \int_{\pi}^{\pi} \text{Tr}\{f_v^{-1}(\lambda) I^N(\lambda)\} d\lambda$$

où $f_v(\cdot)$ est la densité spectrale du processus et où $I^N(\cdot)$ est le périodogramme.

Box et Jenkins [6] dans les cas particuliers des processus (scalaire) autorégressif, en moyenne mobile et mixte a estimé les paramètres par minimisation de la somme des carrés des résidus (c'est-à-dire la somme des carrés de $\varepsilon(t, \theta)$, ces derniers étant calculés en fonction de $X(s)$, $0 < s < t$ et de θ à partir de la définition du modèle). Sa méthode

généralisée au cas multivariable par Wilson [27] équivaut à maximiser l'approximation suivante de la fonction log-vraisemblance :

$$L_N(v) = -\frac{N}{2} \text{Log det } (2\pi G) - \frac{N}{2} \sum_{t=1}^N \varepsilon(t, \theta) G^{-1} \varepsilon(t, \theta)$$

où $\varepsilon(t, \theta)$ est défini comme ci-dessus.

Il a été démontré (Walker [30], Pierce [23] ...) que les fonctions L_N et L_N possèdent les propriétés analogues à celle d'une fonction log-vraisemblance classique. Un point important est néanmoins ignoré dans ces travaux est de savoir dans quelle mesure L_N (resp. L_N) est une approximation de la fonction log-vraisemblance exacte \mathcal{L}_N (le processus étant supposé gaussien). Pour répondre à cette question, nous montrons que la différence $\delta_N(v) = \mathcal{L}_N(v) - L_N(v)$, ainsi que ses dérivées premières et secondes, calculées au point v^* (v^* étant la vraie valeur de v) sont bornées en probabilité quand $N \rightarrow \infty$. Ceci montre que l'estimateur de Box et Jenkins, s'il est consistant est asymptotiquement efficace et asymptotiquement équivalent à l'estimateur consistant de maximum de vraisemblance si ce dernier existe. D'autre part, nous montrons que les bornes supérieures des valeurs absolues de $\delta_N'(v) = L_N'(v) - L_N'(v)$ et de ses dérivées premières et secondes sur un voisinage de v^* sont bornées en probabilité quand $N \rightarrow \infty$. Cela entraîne que les méthodes d'estimation de Box et Jenkin et de Whittle sont équivalentes à l'ordre N^{-1} près.

La maximisation de L_N (resp L_N) peut s'effectuer à l'aide de l'algorithme de Newton-Raphson. Mieux, nous montrons que si l'estimateur initial est convenablement choisi, alors la première itération est déjà un estimateur asymptotiquement efficace. dans le chapitre II, l'application de cet algorithme au cas des processus en moyenne mobile et mixte en utilisant la fonction L_N permet de retrouver les procédures d'estimation d'Hannan [14]. Ce résultat est déjà mis en évidence par Akaike [2]; toutefois, notre travail dégage une deuxième procédure équivalente : celle qui utilise la fonction L_N .

Le problème de test d'adéquation du modèle auto-régressif et en moyenne mobile est aussi considéré. Notre méthode est une adaptation de la méthode générale de Neyman [21]. On construira la statistique de test à partir des dérivées de la fonction log-vraisemblance calculées pour les vraies valeurs des paramètres. Comme ces paramètres sont inconnus, on les remplacera par leur estimateur $N^{1/2}$ consistants (i.e. la différence entre l'estimateur et la vraie valeur du paramètre, multipliée par $N^{1/2}$ est bornée en probabilité). Le test de Neyman est équivalent au test de maximum de vraisemblance (Moran [19]) mais possède l'avantage d'être plus facile à calculer. Dans le cas de l'adéquation du modèle auto-régressif, nous avons ainsi construit deux tests :

Le premier test est basé sur les matrices de covariances empiriques partielles. Si le modèle est auto-régressif d'ordre q , ces matrices, d'ordre plus grand que q , sont asymptotiquement indépendantes et de loi gaussienne centrée. Ce résultat généralise un résultat bien connu dans le cas du processus auto-régressif univariable : les coefficients de corrélation empiriques partiels d'ordre plus grand que q sont asymptotiquement indépendantes, de loi normale centrée (Anderson [3]).

Le deuxième test peut être identifié avec le test de Quenouille généralisé obtenu par Hannan [14]. Toutefois, notre travail établit la liaison entre ce test et le test général de Neyman et par là dégage une méthode de calcul pratique plus simple.

Enfin, dans le chapitre III, nous étendons les résultats du premier chapitre au modèle de régression :

$$X(t) = \sum_{j=0}^{\infty} B(j) Z(t-j) + Y(t)$$

où $\{Y(t), t \in \mathbb{Z}\}$ est un processus stationnaire centré du type (1.1). Notre travail généralise celui de Pierce [23] en étudiant un vectoriel plus générale. Nous avons aussi ajouté les résultats sur la comparaison de la fonction log-vraisemblance exacte avec ses approximations L_N et \tilde{L}_N .

La normalité asymptotique du vecteur des dérivées de L_N (resp L_N) est aussi redémontrée d'une manière **rigoureuse**.

Dans la deuxième partie nous considérons le problème d'estimation dans les processus en temps continu. L'analogie du processus mixte en temps continu est le processus gaussien stationnaire centré de densité spectrale rationnelle, soit :

$$1.2. \quad f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{p(i\lambda)}{q(i\lambda)} \right|^2$$

où p et q sont deux polynômes de coefficients du plus haut degré unité et de degrés respectifs n et m ($n < m$).

Dans le chapitre IV, nous étudions donc l'estimation des paramètres de la densité spectrale d'un processus gaussien stationnaire centré de densité spectrale (1.2.). Nous commençons par calculer la dérivée de Radon-Nykodym de la probabilité image d'un tel processus, par rapport à une mesure dominante. Il existe des résultats généraux sur l'équivalence et le calcul des dérivées entre les mesures gaussiennes (voir Neveu 20b), mais ce ne sont pas des résultats constructifs. C'est grâce à un théorème de Girsanov [13] que nous pouvons déterminer une expression explicite de la fonction log-vraisemblance exacte du modèle, à partir de laquelle nous dérivons une approximation asymptotique notée L_N . Cette dernière est définie comme suit :

Le processus observé $\{X(t), t \in [0, T]\}$ peut être représenté à partir d'un mouvement brownien $\{\xi(t), t > 0\}$ par :

$$X(t) = p\left(\frac{d}{dt}\right) Y(t)$$

$$Y^{(m-1)}(t) - Y^{(m-1)}(0) - \int_0^t q\left(\frac{d}{ds}\right) Y(s) ds = \sigma \xi(t)$$

où $q(u) = q(u) - u^m$. D'autre part on définit un processus :

$\{\phi(t), t \in [0, T]\}$ tel que :

$$\xi(t) = W(t) - \int_0^t \phi(s) ds \quad \text{où} \quad W(t) = \frac{1}{\sigma} [X^{(m-n-1)}(t) - X^{(m-n-1)}(0)]$$

La fonction L_N est alors

$$L_N = \int_0^T \phi(t) dW(t) - \frac{1}{2} \int_0^T |\phi(t)|^2 dt$$

Cette fonction peut s'écrire formellement :

$$L_N = -\frac{1}{2} \int_0^T \left(\frac{d\xi}{dt} \right)^2 dt - \int_0^T \left(\frac{dW}{dt} \right)^2 dt$$

ce qui fait apparaître sa ressemblance avec la fonction log-vraisemblance de Box et Jenkins dans le cas discret.

Nous établissons ensuite la validité de l'approximation L_N . Nous montrons que la différence de L_N et la fonction log-vraisemblance exacte ainsi que ses dérivées premières et secondes calculées pour les vraies valeurs du paramètre, sont bornées en probabilité quand $T \rightarrow \infty$. Nous montrons parallèlement l'existence d'une version régulière de la fonction de vraisemblance qui pour chaque réalisation du processus, est une fonction deux fois continûment dérivable par rapport aux paramètres. Ce problème d'existence d'une version régulière de la fonction de vraisemblance, triviale dans le cas discret demande une démonstration assez compliquée dans le cas continu.

Finalement, nous étudions le calcul numérique des estimateurs. L'algorithme utilisé est encore du type de Newton-Raphson, lequel fournit dès la première itération un estimateur asymptotiquement efficace. Avec quelques modifications, on obtient un algorithme très simple permettant l'estimation des coefficients des polynômes p et q de la densité spectrale du processus.

Dzharidze dans [10] et [11] a aussi considéré le problème d'estimation des paramètres de la densité spectrale d'un processus gaussien stationnaire centré de densité spectrale rationnelle. Il utilise une approximation de la fonction log-vraisemblance semblable à celle de Whittle dans le cas discret :

$$\frac{T}{4\pi} \int_{-\infty}^{\infty} \text{Log}[f_0(\lambda)/f(\lambda)] d\lambda + \int_{-\infty}^{\infty} [f_0^{-1}(\lambda) - f^{-1}(\lambda)] I^T(\lambda) d\lambda$$

où $f(\cdot)$ et $f_0(\cdot)$ sont respectivement les densités spectrales correspondantes à une valeur quelconque et à une valeur donnée du paramètre et où $I^T(\cdot)$ est le périodogramme.

L'inconvénient de cette fonction est qu'il faut donner une interprétation spéciale de la deuxième intégrale qui diverge. D'autre part la méthode d'estimation de Dzhaparidze fait intervenir la résolution d'un système non linéaire. Remarquons aussi que d'après notre théorème 3.3. du chapitre IV, la fonction log-vraisemblance de Dzhaparidze et notre fonction L_N sont asymptotiquement équivalentes : la différence entre elles et ses dérivées jusqu'à l'ordre deux par rapport au paramètre admettent des bornes supérieures sur un voisinage de la vraie valeur du paramètre, bornées en probabilité quand $T \rightarrow \infty$.

Dans le dernier chapitre, nous considérons deux extensions du modèle étudié au chapitre précédent. Le premier est celui du processus gaussien vectoriel stationnaire centré de densité spectrale :

$$f(\lambda) = \frac{1}{2\pi} (A + i\lambda I)^{-1} G (A + i\lambda I)^{-1}$$

Ce modèle correspond à l'état stationnaire d'un système dynamique stochastique régi par l'équation :

$$dX(t) = AX(t)dt + G^{1/2} d\xi(t)$$

où $\xi(t)$ est un mouvement brownien vectoriel standard.

Notre travail rejoint celui de Le Breton [17] mais dans un contexte différent.

La deuxième extension est le modèle de régression (scalaire) :

$$X(t) = \sum_{j=1}^k \mu_j z_j(t) + Y(t)$$

où $\{Y(t), t \in \mathbb{R}\}$ est un processus gaussien stationnaire centré de densité spectrale rationnelle. La démarche est la même que celle du chapitre IV. Nous déterminons une approximation de la fonction log-vraisemblance dont nous établissons la validité. Tous les résultats du chapitre IV peuvent être alors généralisés sans difficulté.

C H A P I T R E 1

METHODE GENERALE D'ESTIMATION DANS LES MODELES DE PROCESSUS STATIONNAIRES CENTRES EN TEMPS DISCRET

Trois modèles de processus stationnaires centrés en temps discret sont couramment employés dans la littérature : le modèle de processus autorégressif, celui de processus en moyenne mobile et celui de processus mixte. Ces modèles peuvent être groupés sous le titre "modèles finis" car ils sont décrits à l'aide d'un nombre fini de paramètres. En fait, la loi de probabilité du processus dépend aussi d'une loi de probabilité sur \mathbb{R}^p , p étant la dimension du processus (vectoriel), mais on verra dans la suite que la spécification de cette loi est indifférente en ce qui concerne le problème d'estimation des paramètres du modèle.

La méthode générale d'estimation dans les modèles finis consiste à maximiser une fonction qui est une approximation de la fonction de vraisemblance dans le cas où le processus est gaussien. Elle sera étudiée dans ce chapitre et son application dans les cas particuliers sera l'objet du chapitre suivant.

§ 1. DIFFERENTS MODELES DE PROCESSUS EN TEMPS DISCRET

Nous rappelons ici brièvement les définitions et principales propriétés des différents "modèles finis". Pour les détails, le lecteur pourra consulter [6] ou [14].

Tous les processus considérés dans la suite sont vectoriels de p composantes. D'autre part $\xi(t)$, $t \in \mathbb{Z}$ désignera toujours une suite de vecteurs aléatoires indépendants centrés de même loi, ayant une matrice de covariance notée G .

Définition 1.1.

"Un processus stationnaire centré $\{X(t), t \in \mathbb{Z}\}$ sera dit en moyenne mobile d'ordre q s'il peut être représenté sous la forme :

$$(1.1.) \quad X(t) = \sum_{j=0}^q B(j) \varepsilon(t-j)$$

où $B(j)$, $j = 0, \dots, q$ sont des matrices d'ordre p avec $B(0) = I_p$ matrice unité".

Théorème 1.1.

"Avec les notations précédentes si $\det \left[\sum_{j=0}^q B(j) z^j \right]$ n'a pas de

zéro de module inférieur ou égal à l'unité, alors les $\varepsilon(t)$ peuvent être calculés en fonction de $X(s)$, $s < t$ comme suit :

$$(1.2.) \quad \varepsilon(t) = \sum_{j=0}^{\infty} A(j) X(t-j)$$

où les matrices $A(j)$, $j=0,1,\dots$, sont les coefficients du développement :

$$\left[\sum_{j=0}^q B(j) z^j \right]^{-1} = \sum_{j=0}^{\infty} A(j) z^j \quad |z| < 1 "$$

Définition 1.2.

"Un processus stationnaire centré $\{X(t), t \in \mathbb{Z}\}$ sera dit auto-régressif d'ordre q s'il vérifie :

$$(1.3.) \quad \sum_{j=0}^q A(j) X(t-j) = \varepsilon(t)$$

où $\varepsilon(t)$ est indépendant de $X(s)$, $s < t$ et où $A(j)$ sont des matrices d'ordre p avec $A(0) = I_p$

Théorème 1.2.

"Une condition nécessaire et suffisante pour qu'il existe un processus stationnaire $\{X(t), t \in \mathbb{Z}\}$ vérifiant les conditions de la définition 1.2 est que $\det \left[\sum_{j=0}^q A(j) z^j \right]$ n'ait pas de zéro de module inférieur ou égal à l'unité. Dans ce cas le processus $\{X(t), t \in \mathbb{Z}\}$ est donné par :

(1.4.) $X(t) = \sum_{j=0}^{\infty} B(j) \varepsilon(t-j)$

où les $B(j)$, $j = 0, 1, \dots$, sont les coefficients du développement :

$$\left[\sum_{j=0}^q A(j) z^j \right]^{-1} = \sum_{j=0}^{\infty} B(j) z^j \quad |z| < 1 \quad "$$

Théorème 1.3.

"La densité spectrale du processus auto-régressif (1.3.) est :

$$(1.5.) \quad f(\lambda) = \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1} \frac{G}{2\pi} \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1}$$

et sa fonction de covariance vérifie les équations (de Yule-Walker) :

$$(1.6.) \quad \sum_{j=0}^q A(j) R(j-k) = 0 \quad k = 1, 2, \dots$$

$$\sum_{j=0}^q A(j) R(j) + G = 0 \quad "$$

Définition 1.3.

"Le processus stationnaire centré $\{X(t), t \in \mathbb{Z}\}$ est dit mixte d'ordre (q, r) s'il vérifie :

$$(1.7.) \quad \sum_{j=0}^q A(j) X(t-j) = \sum_{j=0}^r B(j) \varepsilon(t-j)$$

où $\varepsilon(t)$ est indépendant de $X(s)$, $s < t$ et où $A(j)$, $j = 0, 1, \dots, q$, $B(j)$, $j = 0, \dots, r$ sont des matrices d'ordre p avec $A(0) = B(0) = I_p$ "

Théorème 1.4.

"Il existe un processus stationnaire $\{X(t), t \in \mathbb{Z}\}$ vérifiant les conditions de la définition 1.3 si $\det \begin{bmatrix} q \\ \sum_{j=0} A(j) z^j \end{bmatrix}$ n'a pas de zéro de module supérieur ou égal à l'unité. Dans ce cas, ce processus est donné par :

$$(1.8.) \quad X(t) = \sum_{j=0}^{\infty} C(j) \begin{bmatrix} r \\ \sum_{k=0} B(k) \varepsilon(t-j-k) \end{bmatrix}$$

où les $C(j)$ sont les coefficients du développement :

$$\begin{bmatrix} q \\ \sum_{j=0} A(j) z^j \end{bmatrix}^{-1} = \sum_{j=0}^{\infty} C(j) z^j \quad |z| < 1$$

Si en plus $\det \begin{bmatrix} r \\ \sum_{j=0} B(j) z^j \end{bmatrix}$ n'a pas de zéro de module inférieur

ou égal à l'unité, on peut exprimer $\varepsilon(t)$ en fonction de $X(s)$, $s < t$ comme suit :

$$(1.9.) \quad \varepsilon(t) = \sum_{j=0}^{\infty} D(j) \begin{bmatrix} q \\ \sum_{k=0} A(k) X(t-k) \end{bmatrix}$$

où les $D(j)$, $j = 0, 1, \dots$, sont les coefficients du développement :

$$\left[\sum_{j=0}^{\infty} B(j, z^j) \right]^{-1} = \sum_{j=0}^{\infty} D(j) z^j \quad |z| < 1 "$$

Théorème 1.5.

"La densité spectrale du processus mixte (1.7.) est :

$$(1.10.) \quad f(\lambda) = \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1} \left[\sum_{k=0}^r B(k) e^{ik\lambda} \right] \frac{G}{2} \left[\sum_{k=0}^r B(k) e^{ik\lambda} \right]^* \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^*^{-1}$$

et sa fonction de covariance vérifie le système d'équations :

$$(1.11.) \quad \sum_{j=0}^q \sum_{k=0}^q A(j) R(j-k-u)^t A(k) = \sum_{j=0}^{r-u} B(j) G^t B(j+u) \quad u = 0, 1, \dots, r$$

$$\sum_{j=0}^k A(j) R(j-k) = 0 \quad k > r \quad "$$

§ 2. LE "MODELE FINI" GENERAL

Les modèles précédents sont des cas particuliers du modèle général :

$$(2.1.) \quad X(t) = \sum_{j=0}^{\infty} A_{\theta}(j) \cdot \varepsilon(t-j) \quad A_{\theta}(0) = I_p$$

où les $\varepsilon(t)$, $t \in \mathbb{Z}$ sont des vecteurs aléatoires indépendants centrés, de même loi ayant une matrice de covariance G et où les matrices $A_{\theta}(j)$, $j = 1, 2, \dots$, dépendent d'un paramètre vectoriel $\theta \in \mathbb{W} \subset \mathbb{R}^k$.

Pour que la série (2.1.) converge, il faut et il suffit que :

$$\sum_{j=0}^{\infty} \text{Tr} \left[G A_{\theta}(j)^t A_{\theta}(j) \right] < + \infty.$$

En fait, nous faisons des hypothèses beaucoup plus forte sur les coefficients $A_\theta(j)$. Ce sont des hypothèses de travail pour le problème d'estimation du paramètre θ .

Nous noterons dans toute la suite :

$$(2.2.) \quad h_\theta(z) = \sum_{j=0}^{\infty} A_\theta(j) z^j \quad |z| < 1$$

$$(2.3.) \quad f_\nu(\lambda) = h_\theta(e^{i\lambda}) \frac{G}{2\pi} h_\theta^*(e^{i\lambda})$$

où ν désigne le couple (θ, G) . La vraie valeur de ν (resp. θ , resp. G) sera notée ν^* (resp. θ^* , resp. G^*).

Les deux hypothèses suivantes sont supposées être vérifiées dans toute la suite :

H1. "La fonction $(\lambda, \theta) \mapsto h_\theta^{-1}(e^{i\lambda})$ est définie continue dans $[-\pi, \pi] \times V(\theta)$ où $V(\theta)$ est un voisinage de θ^ , et admet des dérivées partielles jusqu'à l'ordre 2 par rapport à θ , continues dans le même domaine".*

H2. "Les coefficients de Fourier $C_\theta(j)$, $j \in \mathbb{Z}$ de la fonction $h_\theta^{-1}(e^{i\lambda})$ sont nuls pour $j < 0$ et sont deux fois dérivables par rapport à θ et vérifient :

$$(2.4.) \quad \sum_j j ||C_{\theta^*}(j)|| < +\infty$$

$$(2.5.) \quad \sum_j j ||(\frac{\partial}{\partial \theta_h} C_\theta(j))_{\theta=\theta^*}|| < +\infty$$

$$(2.6.) \quad \sum_j j ||(\frac{\partial^2}{\partial \theta_h \partial \theta_l} C_\theta(j))_{\theta=\theta^*}|| < +\infty \quad "$$

Nous notifierons, lorsque l'hypothèse suivante, assez forte, sera faite :

$H2'$: C'est H2 avec en plus : il existe un voisinage $V(\theta)$ de θ tel que :

$$(2.7.) \quad \sum_j \sup_{\theta \in V(\theta^*)} \left\| \frac{\partial^2}{\partial \theta_h \partial \theta_l} C_\theta(j) \right\| < + \infty \quad "$$

Remarques :

1.- La série $\sum_j C_\theta(j) z^j$ est d'après H2 absolument convergente

pour $|z| < 1$. L'égalité $h_\theta^{-1}(e^{i\lambda}) h_\theta(e^{i\lambda}) = \mathbb{I}_p$ donne alors :

$$(2.8.) \quad \sum_{j=0}^k C_\theta(j) A_\theta(k-j) = 0 \quad k = 1, 2, \dots$$

et $C_\theta(0) = \mathbb{I}_p$. Il s'en suit que $\left| \sum C_\theta(j) z^j \right| h_\theta(z) = \mathbb{I}_p$. Les $C_\theta(j)$ sont donc les coefficients de la fonction h_θ^{-1} , analytique dans $|z| < 1$.

2.- L'hypothèse H2 implique :

$$(2.9.) \quad \sum_j \sup_{\theta \in V(\theta^*)} \left\| \frac{\partial}{\partial \theta_h} C_\theta(j) \right\| < + \infty$$

$$(2.10.) \quad \sum_j \sup_{\theta \in V(\theta^*)} \|C_\theta(j)\| < + \infty$$

En effet, d'après le théorème des accroissements finis :

$$\frac{\partial}{\partial \theta_h} C_\theta(j) = \left(\frac{\partial}{\partial \theta_h} C_\theta(j) \right)_{\theta=\theta^*} + \sum_{l=1}^k (\theta_l - \theta_l^*) \left(\frac{\partial^2}{\partial \theta_h \partial \theta_l} C_\theta(j) \right)_{\theta=\theta^*}$$

où θ^* est un point sur le segment joignant θ et θ^* . En tenant compte de (2.5.) et (2.7.), on obtient (2.9.). De même façon, on démontre (2.10.).

Les hypothèses H1 et H2 sont raisonnables. Elles sont satisfaites si $h_\theta(z)$ est une fraction rationnelle en z dont les coefficients sont des

fonctions deux fois continument dérivables de θ et dont les pôles et zéros sont de module strictement supérieur à 1 pour $\theta = \theta^*$. On vérifie également que dans ce cas H_2' est aussi satisfaite.

§ 3. FONCTION DE VRAISEMBLANCE ET SES APPROXIMATIONS ASYMPTOTIQUES

Nous considérons le cas où le processus (2.1.) du § 2 est gaussien. La densité de la loi de probabilité P_v^X du vecteur aléatoire X , de composantes $X_j(t)$, $j = 1, \dots, p$, $t = 1, \dots, N$ est :

$$\{\det[2\pi\Gamma_N(v)]\}^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{x} \Gamma_N^{-1}(v) \mathbf{x}\right\}$$

où $\Gamma_N(v)$ est la matrice de covariance de X .

Le logarithme de la fonction de vraisemblance est alors :

$$\mathcal{L}_N(v) = -\frac{1}{2} \text{Log det } [2\pi\Gamma_N(v)] - \frac{1}{2} \mathbf{x} \Gamma_N^{-1}(v) \mathbf{x}$$

Cette fonction est trop compliquée pour être utilisable. Nous sommes amenés à la remplacer par des approximations asymptotiques ($N \rightarrow \infty$) plus facile à manipuler. Une approximation valable serait une fonction $L_N(v)$ possédant toutes les propriétés habituelles d'une fonction log-vraisemblance (voir Cramer [7]) qui est telle que la différence $\delta_N(v) = \mathcal{L}_N(v) - L_N(v)$ soit deux fois dérivables par rapport à θ , sa valeur et celle de ses dérivées premières et secondes calculées en v , étant bornées en probabilité quand $N \rightarrow \infty$.

Rappelons que : "les vecteurs aléatoires U_N , $N = 1, 2, \dots$, sont dits bornés en probabilité quand $N \rightarrow \infty$ si quelque soit $N > N_0$, pour tout $\epsilon > 0$ il existe $A(\epsilon)$ tel que $:P\{||U_N|| > A(\epsilon)\} < \epsilon$ ". Ainsi, si pour un $\alpha > 0$, $E||U_N||^\alpha < K$, $\forall N > N_0$ alors U_N , $N = 1, 2, \dots$, sont bornés en probabilité quand $N \rightarrow \infty$.

Nous allons définir deux approximations de \mathcal{L}_N proposées par Box et Jenkins [6] et par Whittle [29] et nous montrons que la condition ci-dessus est vérifiée.

Pour faciliter la lecture les démonstrations des lemmes et théorèmes de ce paragraphe et des paragraphes suivants sont groupées dans une annexe.

Théorème 3.1.

"Avec les notations introduites :

$$t_X \Gamma_N^{-1}(\nu) X = \sum_{t=-N}^N t_X F(t) G^{-1} F(t) X$$

où $F(t)$ est la matrice définie par :

$$F(t) X = E_\nu (\epsilon(t) | X)$$

($E_\nu (\epsilon(t) | X)$ étant calculée suivant le modèle 2.1. avec $\nu = (\theta, G)$ comme paramètre)"

Remarque :

Notre résultat étend un résultat de Box et Jenkins [6] (p. 272) concernant le processus en moyenne mobile.

Maintenant, nous avons, compte-tenu de (2.8.)

$$(3.2.) \quad \epsilon(t) = \sum_{j=0}^{\infty} C_\theta(j) X(t-j)$$

Une approximation de $E_\nu (\epsilon(t) | X)$ serait $\sum_{j=0}^{t-1} C_\theta(j) X(t-j)$. Cela

revient à remplacer l'espérance conditionnelle $E_\nu (X(t) | X)$, $t < 0$ par son espérance qui est nulle.

Lemme 3.1.

"Pour tout $s < 0$, $j = 1, \dots, p$ l'application $\nu \rightarrow E (X_j(s) | X)$ à valeurs dans l'espace vectoriel de dimension $p.N$ des variables aléatoires combinaisons linéaires de $X_n(t)$, $h = 1, \dots, p$, $t = 1, \dots, N$ est deux fois dérivable et on a pour tout N :

$$\{E_{v^*} \left\| \left(\frac{\partial}{\partial v_j} E_v(X(s) | X) \right)_{v=v^*} \right\|^2\}^{1/2} < K_j$$

$$\{E_{v^*} \left\| \left(\frac{\partial}{\partial v_j \partial v_h} E_v(X(s) | X) \right)_{v=v^*} \right\|^2\}^{1/2} < K_{jh} \quad "$$

Lemme 3.2.

"Soient $U(t)$, $t \in \mathbb{Z}$, des vecteurs aléatoires vérifiant $E \|U(t)\|^2 < A^2$, $t \in \mathbb{Z}$ et $A(j)$, $j = 0, 1, \dots$ des matrices telles que :

$$\sum_j \|A(j)\| < K$$

où $\|A\| = \text{Sup} \{ \|Ax\|; \|x\| < 1 \}$ est la norme d'opérateur de A . Posons :

$$\begin{aligned} \delta_N(t) &= \sum_{j=t}^{\infty} A(j) U(t-j) && 1 < t < N \\ &= \sum_{j=t-N}^{t-1} A(j) U(t-j) && t > N \end{aligned}$$

alors on a

$$\sum_{t=1}^N \{E \|\delta_N(t)\|^2\}^{1/2} < A.K \quad N$$

$$\sum_{t=N+1}^{\infty} \{E \|\delta_N(t)\|^2\}^{1/2} < A.K \quad N "$$

L'application des lemmes 3.1. et 3.2. donne immédiatement :

Lemme 3.3.

"Soient :

$$(3.3.) \quad \delta(t, \theta) = \sum_{j=t}^{\infty} C_{\theta}(j) \varepsilon(t-j)$$

alors l'application $v \rightarrow E_v(\delta(t, \theta) | X)$ est deux fois dérivables par rapport à v et on a pour tout N :

$$\sum_{t=1}^N \{E_{v^*} \| |E_{v^*}(\delta(t, \theta) | X) | \|^2\}^{1/2} < K$$

$$\sum_{t=1}^N \{E_{v^*} \| |(\frac{\partial}{\partial v_j} E_v(\delta(t, \theta) | X))_{v=v^*} | \|^2\}^{1/2} < K_j$$

$$\sum_{t=1}^N \{E_{v^*} \| |(\frac{\partial^2}{\partial v_j \partial v_h} E_v(\delta(t, \theta) | X))_{v=v^*} | \|^2\}^{1/2} < K_{jh}$$

et des majorations analogues avec $\delta(t, \theta)$ à la place de $E_v(\delta(t, \theta) | X)$ "

Le lemme 3.3 permet de justifier l'approximation de $E_v(\varepsilon(t) | X)$

par $\sum_{j=0}^{t-1} C_\theta(j) X(t-j)$ car $E_v(\delta(t, \theta) | X)$ est précisément la différence entre

les deux termes précédents. Maintenant :

Lemme 3.4.

"On a :

$$\text{Log}[\det \Gamma_N(v)] = N \text{Log} \det G + \sum_{n=1}^N \text{Log} \det [I + G^{-1} H_n(v)]$$

où $H_n(v)$ sont des matrices positives, deux fois dérivables par rapport à v et vérifient :

$$\sum_n \| |H_n(v) | \| < + \infty$$

$$\sum_n \| |(\frac{\partial}{\partial v_j} H_n(v))_{v=v^*} | \| < + \infty$$

$$\sum_n \| |(\frac{\partial^2}{\partial v_j \partial v_h} H_n(v))_{v=v^*} | \| < + \infty "$$

Nous obtenons alors :

Théorème 3.2.*"Soit :*

$$(3.4.) \quad L_N(v) = -\frac{N}{2} \text{Log det } (2\pi G) - \frac{1}{2} \sum_{t=1}^N \varepsilon(t, \theta) G^{-1} \varepsilon(t, \theta)$$

où :

$$(3.5.) \quad \varepsilon(t, \theta) = \sum_{j=0}^{t-1} C_\theta(j) X(t-j)$$

alors le logarithme de la densité de P_v^X est $\mathcal{L}_N(v) = L_N(v) + \delta_N(v)$ où $\delta_N(v)$ est deux dérivables par rapport à v , sa valeur et celles de ses dérivées calculées en v^* étant bornées en probabilité quand $N \rightarrow \infty$ "

L'approximation (3.4.) de \mathcal{L}_N est l'extension au cas multivariable de celle de Box et Jenkins. Une deuxième approximation, due à Whittle dif-
fère de la précédente en remplaçant la somme finie dans (3.4.) par la somme infinie :

$$\frac{1}{2} \sum_{t=1}^{\infty} \varepsilon(t, \theta) G^{-1} \varepsilon(t, \theta)$$

où $\varepsilon(t, \theta)$, pour $t > N$ est défini par :

$$(3.6.) \quad \varepsilon(t, \theta) = \sum_{j=t-N}^{t-1} C_\theta(j) X(t-j) \quad t > N$$

En posant $X_N(t) = X(t)$ si $1 < t < N$ et $X_N(t) = 0$ sinon, on voit que $\varepsilon(t, \theta)$ est la convolution de la série $C_\theta(j)$ et $X_N(t)$. Sa série de

Fourier $\tilde{\varepsilon}(\lambda, \theta) = \sum_{t=1}^{\infty} e^{-i\lambda t} \varepsilon(t, \theta)$ est alors $h_\theta(e^{i\lambda}) \sum_{t=1}^N X(t) e^{-i\lambda t}$. D'où :

$$\sum_{t=1}^{\infty} \varepsilon(t, \theta) G^{-1} \varepsilon(t, \theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\varepsilon(t, \theta)] G^{-1} \varepsilon(t, \theta) d\lambda$$

$$(3.7.) \quad \sum_{t=1}^{\infty} \epsilon(t, \theta) G^{-1} \epsilon(t, \theta) = \frac{N}{2\pi} \int_{-\pi}^{\pi} \text{Tr}[f^{-1}(\lambda) I^N(\lambda)] d\lambda$$

où

$$(3.8.) \quad I^N(\lambda) = \frac{1}{2\pi N} \left[\begin{array}{c} N \\ \sum_{t=1}^N e^{i\lambda t} x(t) \end{array} \right] \left[\begin{array}{c} N \\ \sum_{t=1}^N e^{i\lambda t} x(t) \end{array} \right]^*$$

est le périodogramme".

Théorème 3.3.

"Soit :

$$\tilde{L}_N(v) = -\frac{N}{2} \text{Log det}(2\pi G) - \frac{N}{4\pi} \int_{-\pi}^{\pi} \text{Tr}[f_v^{-1}(\lambda) I^N(\lambda)] d\lambda$$

alors la différence $\delta_N(v) = L_N(v) - \tilde{L}_N(v)$ est deux fois dérivable par rapport à v , sa valeur et celles de ses dérivées calculées en v^* , étant bornées en probabilité quand $N \rightarrow \infty$.

Si l'hypothèse H2' est vérifiée, alors il existe un voisinage $V(v^*)$ de v^* tel que les variables aléatoires :

$$\sup_{v \in V(v^*)} |\delta_N(v)|, \quad \sup_{v \in V(v^*)} \left| \frac{\partial}{\partial v_j} \delta_N(v) \right|, \quad \sup_{v \in V(v^*)} \left| \frac{\partial^2}{\partial v_j \partial v_l} \delta_N(v) \right|$$

soient bornées en probabilité quand $N \rightarrow \infty$ "

L'intérêt de la dernière assertion du théorème est d'assurer l'équivalence complète entre L_N et \tilde{L}_N . Ces fonctions ont mêmes propriétés asymptotiques et les estimateurs obtenus par maximisation de l'une ou l'autre ne diffèrent que par un terme $N^{-1/2} \epsilon_N$ où $\epsilon_N \rightarrow 0$ en probabilité quand $N \rightarrow \infty$.

§ 4. PROPRIÉTÉS ASYMPTOTIQUES DE LA FONCTION DE VRAISEMBLANCE

La fonction L_N (resp. \tilde{L}_N) possèdent toutes les propriétés habituelles d'une fonction log-vraisemblance relative à des observations indépendantes. Ces propriétés ont déjà obtenues dans les travaux de Box et Jenkins [6], Pierce [23], Walker [30] ... Ces auteurs ont considéré le cas univariable. La généralisation au cas multi-variable n'est pas difficile. Toutefois, les résultats suivants ont leur importance et sont utiles pour la suite. Remarquons qu'aucune hypothèse n'est faite sur la loi de $\varepsilon(t)$ sinon qu'elle possède une matrice de covariance G .

Nous commençons par remplacer $\varepsilon(t, \theta)$ défini par (3.5.) par :

$$(4.1.) \quad \bar{\varepsilon}(t, \theta) = \sum_{j=0}^{\infty} C_{\theta}(j) X(t-j)$$

Une démonstration analogue (et plus simple) à celle du théorème 3.2 montre que l'erreur :

$$\delta_N(v) = \frac{1}{2} \sum_{t=1}^N \bar{\varepsilon}(t, \theta) G^{-1} \bar{\varepsilon}(t, \theta) - \frac{1}{2} \sum_{t=1}^N \varepsilon(t, \theta) G^{-1} \varepsilon(t, \theta)$$

et ses dérivées premières et secondes (qui existent), calculées en v^* sont bornées en probabilité quand $N \rightarrow \infty$.

Ecrivons maintenant :

$$(4.2.) \quad L_N(v) = -\frac{N}{2} \text{Log det } (2\pi G) - \frac{N}{2} \text{Tr } G^{-1} \cdot \bar{G}(\theta) + \delta_N(v)$$

où :

$$(4.3.) \quad \bar{G}(\theta) = \frac{1}{N} \sum_{t=1}^N \bar{\varepsilon}(t, \theta) \bar{\varepsilon}(t, \theta)$$

Nous voyons que le comportement asymptotique de L_N et de ses dérivées découle de celui de $\bar{G}(\theta)$ et ses dérivées. A l'aide du théorème ergodique (Doob [9]), on a :

Lemme 4.1.

"Quand $N \rightarrow \infty$, \bar{G} et ses dérivées premières par rapport à θ_j et ses dérivées secondes par rapport à θ_j, θ_ℓ calculées au point θ^* , convergent presque sûrement respectivement vers G^* , vers 0 et vers :

$$\int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \theta_j} h_{\theta}^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} f_{\nu}(\lambda) \left(\frac{\partial}{\partial \theta_\ell} h_{\theta}^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} d\lambda +$$

$$\int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \theta_\ell} h_{\theta}^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} f_{\nu}(\lambda) \left(\frac{\partial}{\partial \theta_j} h_{\theta}^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} d\lambda \quad "$$

On en déduit :

Théorème 4.1.

"Quand $N \rightarrow \infty$, le vecteur des dérivées premières et la matrice des dérivées secondes de $N^{-1} L_N$ (resp. $N^{-1} \tilde{L}_N$, resp. $N^{-1} \mathcal{L}_N$) calculés en θ^* convergent en probabilité respectivement vers 0 et vers :

$$-K = - \begin{pmatrix} S & 0 \\ 0 & H \end{pmatrix}$$

où : S et H sont les matrices de terme général :

$$(4.4.) \quad S_{j\ell} = \int_{-\pi}^{\pi} \text{Tr} \left\{ G^{-1} \left(\frac{\partial}{\partial \theta_j} h_{\theta}^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} f_{\nu}(\lambda) \left(\frac{\partial}{\partial \theta_\ell} h_{\theta}^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} \right\} d\lambda$$

$$(4.5.) \quad H_{(j,\ell)(j',\ell')} = (G^{-1})_{jj'} + (G^{-1})_{\ell\ell'} + (G^{-1})_{j\ell'} + (G^{-1})_{j'\ell} \quad \begin{matrix} j < \ell \text{ et} \\ j' < \ell' \end{matrix}$$

$$H_{(j,j)(j',\ell')} = (G^{-1})_{jj'} + (G^{-1})_{j\ell'} \quad j' < \ell'$$

$$H_{(j,j)(\ell,\ell)} = \frac{1}{2} (G^{-1})_{j\ell}^2 \quad "$$

Nous allons étendre le résultat précédent au cas où les dérivées de $N^{-1}L_N$ sont calculées en un estimateur consistant de v .

Lemme 4.2.

"Soit la variable aléatoire :

$$Z_v^N = \int_{-\pi}^{\pi} \text{Tr} [p_v(\lambda) I^N(\lambda)] d\lambda$$

où $(v, \lambda) \rightarrow p_v(\lambda)$ est une fonction matricielle continue sur $V(v^*) \times [-\pi, \pi]$, $V(v^*)$ étant un voisinage de v^* . Alors :

$$\forall \varepsilon > 0 \quad \exists \eta > 0 : \quad \text{Sup}_{v: ||v-v^*|| < \eta} |Z_v^N - Z_{v^*}^N| < \varepsilon M_N$$

où M_N tend en probabilité vers une constante quand $N \rightarrow \infty$.

En particulier si v_N est un estimateur consistant de v , alors

$$Z_{v_N}^N - Z^N \rightarrow 0 \text{ en probabilité quand } N \rightarrow \infty$$

Ce lemme et le théorème 3.3 donnent immédiatement :

Théorème 4.2.

"Les conclusions du théorème 4.1 relatives à \tilde{L}_N restent vraies quand v est remplacé par son estimateur consistant v_N . Si l'hypothèse H_2' est vérifiée, ce résultat s'applique aussi pour la fonction L_N "

*Théorème 4.3.

"Le vecteur des dérivées de $N^{-1/2} L_N$ (resp. $N^{-1/2} \tilde{L}_N$, resp. $N^{-1/2} \mathcal{L}_N$) par rapport à θ , calculé en v^* converge en loi selon N vers la loi gaussienne centrée de matrice de covariance S .

Si les $\varepsilon(t)$ possèdent des moments d'ordre 4, le vecteur des dérivées de $N^{-1/2} L_N$ (resp. $N^{-1/2} \tilde{L}_N$, resp. $N^{-1/2} \mathcal{L}_N$) calculé en v^* convergent en loi selon N vers la loi gaussienne centrée de matrice de covariance.

$$\begin{pmatrix} S & 0 \\ 0 & \tilde{H} \end{pmatrix}$$

où $\tilde{H} = H$ si les cumulants d'ordre 4 de $\varepsilon(t)$ sont nuls."

§ 5. CONSTRUCTION DE L'ESTIMATEUR

Une méthode d'estimation de v consiste à maximiser L_N (resp. \tilde{L}_N) ou plus exactement de résoudre le système

$$(5.1.) \quad \frac{\partial}{\partial v_j} L_N(v) = 0$$

$$(5.2.) \quad (\text{resp. } \frac{\partial}{\partial v_j} \tilde{L}_N(v) = 0)$$

Nous appelons estimateur de maximum de vraisemblance un estimateur obtenu de cette façon quoique L_N (resp. \tilde{L}_N) n'est une approximation de la fonction log-vraisemblance que dans le cas gaussien.

Pour fixer les idées nous travaillons sur la fonction \tilde{L}_N . On obtiendra les mêmes résultats relatifs à la fonction L_N si l'hypothèse H2' est vérifiée. Nous supposons d'autre part que la matrice S du théorème 4.1. est non-singulière (la matrice H étant non singulière car G l'est).

La fonction L_N possède toutes les propriétés d'une fonction log-vraisemblance ordinaire. On peut s'attendre à ce que le résultat classique sur la consistance de l'estimateur de maximum de vraisemblance reste valable, i.e. :

"Il existe pour chaque N un estimateur v_N , fonction seulement de $X(1), \dots, X(N)$ tel que quand $N \rightarrow \infty$, v_N converge en probabilité vers v^ et est solution de (5.2.) avec probabilité tendant vers 1".*

Ce résultat peut être démontré en utilisant un raisonnement analogue à celui d'Aitchinson et Silvey [1], et en appuyant sur le résultat suivant qui est une conséquence du lemme 4.2 :

$$\varepsilon > 0, \quad \eta > 0 : \quad \sup_{v: ||v - v^*|| < \eta} \left| \left(\frac{\partial^2 L_N}{\partial \theta_j \partial v_\ell} \right) (v) - \left(\frac{\partial^2 L_N}{\partial v_j \partial v_\ell} \right) (v) \right| < \varepsilon M_{j\ell}^N$$

où les $M_{j\ell}^N$ tendent en probabilité vers des constantes quand $N \rightarrow \infty$.

Nous ne nous intéressons pas à la construction d'un estimateur consistant de maximum de vraisemblance de v . Par contre nous allons définir un algorithme qui fournit un estimateur approchant ce dernier et asymptotiquement efficace.

L'idée est de calculer la solution de (5.2.) à l'aide de l'algorithme de Raphon-Newton. Si $v^{(n)}$ est le n-ième estimateur de v , le (n+1)ième estimateur sera :

$$(5.3.) \quad v^{(n+1)} = v^{(n)} - K_N^{-1}(v^{(n)}) k_N(v^{(n)})$$

où $k_N(v)$ et $K_N(v)$ sont respectivement le vecteur des dérivées premières et la matrice des dérivées secondes de $-N^{-1} L_N$.

Comme $K_N(v^{(n)})$ converge en probabilité quand $N \rightarrow \infty$ vers la matrice K (théorème 4.1, $v^{(n)}$ étant supposé consistant), une généralisation de l'algorithme (5.3.) consiste à remplacer $K_N(v^{(n)})$ par un estimateur consistant K_N de K .

Si la suite $\{v^{(n)}\}$ converge, sa limite sera évidemment l'estimateur de maximum de vraisemblance. La convergence de cette suite dépend du choix de l'estimateur initial $v^{(0)}$. Nous allons montrer que si $v^{(0)}$ est convenablement choisi, la première itération est déjà un bon estimateur. Pour cela nous introduisons d'après Neyman [21], la notion de $N^{1/2}$ consistance :

"L'estimateur φ_N de $\varphi(v)$ est dit $N^{1/2}$ consistant si $N^{-1/2}[\varphi_N - \varphi(v^*)]$ est borné en probabilité quand $N \rightarrow \infty$ "

D'autre part, pour simplifier les énoncés des résultats, on note $u_N \sim v_N$ si $(u_N - v_N) \rightarrow 0$ en probabilité quand $N \rightarrow \infty$ ($\{u_N\}$ et $\{v_N\}$ étant deux suites de vecteurs aléatoires).

Théorème 5.1.

"Soient v_N un estimateur $N^{1/2}$ consistant de v et K_N un estimateur consistant de K alors l'estimateur :

$$\tilde{v}_N = v_N - K_N^{-1} k_N(v_N)$$

vérifie $\sqrt{N}(v_N - v) \sim -\sqrt{N} K^{-1} k_N(v^*)$

Si les cumulants d'ordre 4 de $\varepsilon(t)$ sont nuls, on a en plus $\sqrt{N} \tilde{v}_N \sim \sqrt{N} \hat{v}_N$ où \hat{v}_N est l'estimateur consistant de maximum de vraisemblance de v ."

Nous remarquons que la matrice K du théorème 4.1. à la forme :

$$K = \begin{pmatrix} S & 0 \\ 0 & H \end{pmatrix}$$

ce qui suggère le remplacement de l'algorithme (5.3.) par :

$$(5.4.) \quad \theta^{(n+1)} = \theta^{(n)} - S_N^{-1} s_N(\theta^{(n)}, G^{(n)})$$

$$(5.5.) \quad G^{(n+1)} = G^{(n)} - H_N^{-1} h_N(\theta^{(n)}, G^{(n)})$$

où S_N et H_N sont respectivement un estimateur consistant de S et de H et où $s_N(\theta, G)$, $h_N(\theta, G)$ sont respectivement le vecteur des dérivées de $-NL_N$ par rapport à θ et par rapport à G (Dans (5.5.), G est supposée écrite sous forme vectorielle).

Maintenant, pour θ fixé, un calcul élémentaire montre que \tilde{L}_N est maximisé quand :

$$(5.6.) \quad G = \hat{G}(\theta) = \frac{1}{N} \sum_{t=1}^{\infty} \varepsilon(t, \theta)^t \varepsilon(t, \theta)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} h_{\theta}^{-1}(e^{i\lambda}) I^N(\lambda) h^{*-1}(e^{i\lambda}) d\lambda$$

Cela suggère le remplacement de (5.4.) par :

$$(5.7.) \quad G^{(n+1)} = \hat{G}(\theta^{(n)})$$

Le théorème suivant justifie l'algorithme (5.4.) et (5.7.).

Théorème 5.2.

"Soient θ_N un estimateur $N^{1/2}$ consistant de θ et G_N , S_N respectivement un estimateur consistant de G et de S (S étant définie par (4.4.)). Alors l'estimateur :

$$G_N = \hat{G}(\theta_N)$$

est consistant et l'estimateur :

$$\tilde{\theta}_N = \theta_N - S_N^{-1} s_N(\theta_N, G_N)$$

où $s_N(\theta, G)$ est le vecteur des dérivées de $-N^{-1} \tilde{L}_N$ par rapport à θ , vérifie :

$$\sqrt{N} (\tilde{\theta}_N - \theta^*) \sim - \sqrt{N} S^{-1} s_N(\theta^*, G^*)$$

D'autre part si $(\hat{\theta}_N, \hat{G}_N)$ est l'estimateur consistant de maximum de vraisemblance de (θ, G) alors $\sqrt{N} G_N \sim \sqrt{N} \hat{G}_N$ et $\sqrt{N} \theta_N \sim \sqrt{N} \hat{\theta}_N$ "

Nous terminons ce paragraphe par deux remarques :

Remarque 1 :

L'algorithme connu sous le nom "régression non linéaire" (voir Box et Jenkins [6], Wilson [27]) est en fait un cas particulier de l'algorithme général (5.4.) et (5.7.). Cet algorithme est défini de la façon suivante :

Pour G fixé, la maximisation de L_N est équivalente à la minimisation de :

$$(5.8.) \quad Q_N(\theta) = \frac{1}{2} \sum_{t=1}^N t \epsilon(t, \theta) G^{-1} \epsilon(t, \theta)$$

En linéarisant $\epsilon(t, \cdot)$ autour de $\theta^{(0)}$, un estimateur initial de θ :

$$\epsilon(t, \theta) \simeq \epsilon(t, \theta^{(0)}) + \sum_j \left(\frac{\partial}{\partial \theta_j} \epsilon(t, \theta) \right)_{\theta = \theta^{(0)}} (\theta_j - \theta_j^{(0)})$$

on obtient une forme quadratique en θ , soit :

$$(5.9.) \quad Q_N(\theta) \simeq Q_N(\theta^{(0)}) + {}^t [\theta - \theta^{(0)}] A_N(\theta^{(0)}, G) [\theta - \theta^{(0)}] + a_N(\theta^{(0)}, G) \cdot [\theta - \theta^{(0)}]$$

où $A_N(\theta, G)$ est la matrice de termes généraux :

$$\sum_{t=1}^N {}^t \left[\frac{\partial}{\partial \theta_j} \epsilon(t, \theta) \right] G^{-1} \frac{\partial}{\partial \theta_l} \epsilon(t, \theta)$$

et où $a_N(\theta, G)$ est le vecteur de composantes :

$$\sum_{t=1}^N {}^t \left[\frac{\partial}{\partial \theta_j} \epsilon(t, \theta) \right] G^{-1} \epsilon(t, \theta)$$

Comme G est inconnue, on la remplace par son estimateur initial $G^{(0)} = \hat{G}(\theta^{(0)})$, la minimisation du deuxième membre de (5.9.) fournit un deuxième estimateur $\theta^{(1)}$ de θ . En itérant la procédure on obtient une suite $\{\theta^{(n)}, G^{(n)}\}$ dont la limite, si elle existe est l'estimateur de maximum de vraisemblance de (θ, G) .

Maintenant un calcul élémentaire montre que $\theta^{(1)}$ est défini par

$$\theta^{(1)} - \theta^{(0)} = - [A_N(\theta^{(0)}, G^{(0)})]^{-1} \cdot \alpha_{N(\theta^{(0)}, G^{(0)})}$$

Où $\alpha_N(\theta, G)$ n'est autre que $N s_N(\theta, G)$ et $N^{-1} A_N(\theta^{(0)}, G^{(0)})$ converge en probabilité vers S . En effet, un raisonnement analogue à la démonstration du théorème 3.3. montre que :

$$A_{N,j\ell}(\theta, G) = \sum_{t=1}^{\infty} t \left[\frac{\partial}{\partial \theta_j} \varepsilon(t, \theta) \right] G^{-1} \frac{\partial}{\partial \theta_\ell} \varepsilon(t, \theta) + \delta_{N,j\ell}(\nu)$$

où $\text{Sup} \{ |\delta_{N,j\ell}(\nu)| ; \nu \in V(\nu^*) \}$ est borné en probabilité quand $N \rightarrow \infty$. L'application du lemme 4.2 montre alors :

$$N^{-1} A_{N,j\ell}(\theta^{(0)}, G^{(0)}) \sim N^{-1} \sum_{t=1}^{\infty} t \left(\frac{\partial}{\partial \theta_j} \varepsilon(t, \theta) \right)_{\theta=\theta^*} G^{-1} \left(\frac{\partial}{\partial \theta_\ell} \varepsilon(t, \theta) \right)_{\theta=\theta^*}$$

$$N^{-1} \sum_{t=1}^N t \left(\frac{\partial}{\partial \theta_j} \varepsilon(t, \theta) \right)_{\theta=\theta^*} G^{-1} \left(\frac{\partial}{\partial \theta_\ell} \varepsilon(t, \theta) \right)_{\theta=\theta^*}$$

La dernière expression converge presque sûrement quand $N \rightarrow \infty$ vers S_j d'après le théorème ergodique.

Remarque 2 :

La valeur maximum de L_N pour θ fixé est d'après ce qui précède $-(N/2) \text{Log} [\det \hat{G}(\theta)] + C^{te}$. On peut penser à estimer θ par maximisation de $-\text{Log} \det \hat{G}(\theta)$. Pour ce faire, appliquons l'algorithme de Raphon-Newton, et calculons les dérivées premières et secondes de $\text{Log}[\det G(\theta)]$ qui sont respectivement :

$$\text{Tr} \hat{G}^{-1}(\theta) \frac{\partial}{\partial \theta_j} \hat{G}(\theta) = -N \left(\frac{\partial L_N}{\partial \theta_j} \right) (\theta, \hat{G}(\theta))$$

et :

$$\begin{aligned} & \text{Tr } \hat{G}^{-1}(\theta) \frac{\partial^2}{\partial \theta_j \partial \theta_\ell} G(\theta) + \text{Tr } G^{-1}(\theta) \frac{\partial}{\partial \theta_j} G(\theta) G^{-1}(\theta) \frac{\partial}{\partial \theta_\ell} G(\theta) \\ &= -N \left(\frac{\partial^2 L_N}{\partial \theta_j \partial \theta_\ell} \right) (\theta, \hat{G}(\theta)) + \varepsilon_N(\theta) \end{aligned}$$

Si θ_N est un estimateur consistant de θ , alors, d'après les lemmes 4.1 et 4.2., on a $\hat{G}(\theta_N) \sim G$, $(\partial \hat{G} / \partial \theta_j)(\theta_N) \sim 0$, d'où $\varepsilon_N(\theta_N) \sim 0$. Nous sommes donc amenés à l'algorithme (5.4.) et

$$(5.7') \quad G^{(n)} = \hat{G}(\theta^{(n)})$$

qui est essentiellement le même que (5.5.) et (5.7.).

ANNEXE : DEMONSTRATIONS DES LEMMES ET THEOREMES

Démonstration du théorème 3.1.

Ce théorème découle directement du résultat suivant :

"Soient X_1, \dots, X_n et $\varepsilon_{ij}, j = 1, \dots, n_i, i \in \mathbb{N}$ des variables aléatoires de carrés intégrables centrées définies sur un espace probabilisé (Ω, \mathcal{G}, P) telles que :

(i) X_1, \dots, X_n appartiennent au sous-espace vectoriel fermé engendré par $\varepsilon_{ij}, j = 1, \dots, n_i, i$.

(ii) ε_{ij} est non corrélé avec $\varepsilon_{i',j'}$ pour tout $i' \neq i$ et tout j' .

On suppose en plus que la matrice Λ de covariance du vecteur aléatoire $X = {}^t(X_1 \dots X_n)$ et les matrices G_i de covariance des vecteurs aléatoires $\varepsilon_i = (\varepsilon_{i1} \dots \varepsilon_{in_i})$ sont toutes non singulières. Alors

$$\Lambda^{-1} = \sum_{i \in \mathbb{N}} {}^t F_i G_i^{-1} F_i$$

où F_i est la matrice d'ordre (n_i, n) telle que $F_i X$ soit le vecteur des projections de $\varepsilon_{i1}, \dots, \varepsilon_{in_i}$ sur le sous-espace des combinaisons linéaires de X_1, \dots, X_n "

En effet, on peut écrire d'après (i) et (ii) :

$$X = \sum_i A_i \varepsilon_i$$

où les matrices A_i sont uniquement déterminées par :

$$E(X {}^t \varepsilon_i) = A_i G_i$$

D'autre part, d'après la définition de F_i :

$$E(\varepsilon_i^\dagger X) = E[(F_i X)^\dagger X] = F_i \Lambda$$

Par suite $A_i = \Lambda^\dagger F_i G^{-1}$ et :

$$\Lambda = E(X^\dagger X) = E[(\sum_i \Lambda^\dagger F_i G^{-1} \varepsilon_i)^\dagger X] = \Lambda (\sum_i F_i G^{-1} F_i) \Lambda$$

Comme Λ est inversible, on a le résultat.

Démonstration du lemme 3.1.

Nous noterons $L^2(f_V)$ l'espace de Hilbert (Rozanov [24], p. 30) des fonctions vectorielles sur $[-\pi, \pi]$ telles que :

$$\int \varphi^*(\lambda) f_V(\lambda) \varphi(\lambda) d\lambda < +\infty$$

muni du produit scalaire :

$$\varphi, \psi \in L^2(f_V) \quad \langle \varphi, \psi \rangle = \int_{-\pi}^{\pi} \varphi^*(\lambda) f_V(\lambda) \psi(\lambda) d\lambda$$

Nous savons que l'espace $L^2(f_V)$ est isomorphe au sous-espace fermé des variables aléatoires de carrés intégrables (selon P_V^X) engendré par $X_j(t)$, $j = 1, \dots, p, t$, par l'isomorphisme :

$$e_j e^{i\lambda t} \leftrightarrow X_j(t)$$

où e_j est le j -ième vecteur de la base canonique de \mathbb{R}^p .

Nous noterons $L_N^2(f_V)$ le sous-espace de $L^2(f_V)$ engendré par $e_j e^{i\lambda t}$, $j = 1, \dots, p, t = 1, \dots, N$ et π_V^N le projecteur orthogonal de $L^2(f_V)$ sur $L_N^2(f_V)$. L'élément $\pi_V^N(e_j e^{i\lambda s})$ correspond alors par l'isomorphisme précédent à $E_V(X_j(s) | X)$ et est donné par :

$$\int_{-\pi}^{\pi} e^{-i\lambda t} f_{\nu}(\lambda) \pi_{\nu}^N(e_j e^{i\lambda s}) d\lambda = \int_{-\pi}^{\pi} e^{-i\lambda t} f_{\nu}(\lambda) e_j e^{i\lambda s} d\lambda \quad t=1, \dots, N$$

Nous allons montrer que si l'application $\nu \rightarrow k_{\nu} \in L^2(f_{\nu})$ est continûment dérivable, alors il est de même pour l'application $\nu \rightarrow \pi_{\nu}^N k_{\nu}$ et que la dérivée de cette dernière est :

$$\frac{\partial}{\partial \theta_j} (\pi_{\nu}^N k_{\nu}) = \pi_{\nu}^N \left[\frac{\partial}{\partial \theta_j} k_{\nu} + (f_{\nu}^{-1} \frac{\partial}{\partial \theta_j} f_{\nu}) (k_{\nu} - \pi_{\nu}^N k_{\nu}) \right]$$

Comme l'application $\nu \rightarrow f_{\nu} \in C([- \pi, \pi])$ où $C([- \pi, \pi])$ est l'espace des fonctions continues sur $[- \pi, \pi]$ muni de la norme du max, est deux fois dérivable (à cause H1), on voit alors que l'application $\nu \rightarrow \pi_{\nu}^N k_{\nu}$ est deux fois dérivable par rapport à ν si l'application $\nu \rightarrow k_{\nu}$ l'est.

En prenant $k_{\nu} = e_j e^{i\lambda s}$, $s < 0$ et en remarquant que la norme de $\pi_{\nu}^N (k_{\nu})$ dans $L^2(f_{\nu})$ est inférieure à celle de k_{ν} , on obtient le lemme.

Il nous reste à démontrer l'assertion que nous venons d'énoncer. Soit donc $\delta \nu$ un accroissement de ν et δk_{ν} , $\delta(\pi_{\nu}^N k_{\nu})$ et δf_{ν} les accroissements correspondants de k_{ν} de $\pi_{\nu}^N k_{\nu}$ et f_{ν} . On a :

$$\begin{aligned} \int_{-\pi}^{\pi} e^{-i\lambda t} f_{\nu}(\lambda) \delta(\pi_{\nu}^N k_{\nu})(\lambda) d\lambda = \\ \int_{-\pi}^{\pi} e^{-i\lambda t} \{ f_{\nu}(\lambda) \delta k_{\nu}(\lambda) + \delta f_{\nu}(\lambda) [k_{\nu+\delta \nu}(\lambda) - \pi_{\nu+\delta \nu}^N k_{\nu+\delta \nu}(\lambda)] \} d\lambda \end{aligned}$$

Par conséquent :

$$\delta \pi_{\nu}^N k_{\nu} = \pi_{\nu}^N [\delta k_{\nu} + f_{\nu}^{-1} \delta f_{\nu} (k_{\nu+\delta \nu} - \pi_{\nu+\delta \nu}^N k_{\nu+\delta \nu})]$$

On peut montrer que la métrique de $L^2(f_{\nu+\delta \nu})$ est équivalente de celle de $L^2(f_{\nu})$ si $\delta \nu$ est suffisamment petit.

Plus exactement, il existe $\varepsilon > 0$ tel que :

$$\|\delta v\| < \varepsilon \quad C \|\varphi\|_{L^2(f_v)} < \|\varphi\|_{L^2(f_{v+\delta v})} < C' \|\varphi\|_{L^2(f_v)}$$

En effet, à cause de H1 : $\lambda > 0$:

$$\|\delta v\| < \varepsilon \quad \left\| f_{v+\delta v}^{1/2}(\lambda) f_v^{-1/2}(\lambda) \right\| < C' \quad \lambda \in [-\pi, \pi]$$

où $A^{1/2}$ désigne la racine carrée hermitienne positive de A et où A

$\|A\| = \text{Sup}\{\|Ax\|, \|x\| < 1\}$ est la norme d'opérateur de A.

Par conséquent :

$$\|\varphi\|_{L^2(f_{v+\delta v})}^2 = \int_{-\pi}^{\pi} \left\| f_{v+\delta v}^{1/2}(\lambda) \varphi(\lambda) \right\|^2 d\lambda < C'^2 \|\varphi\|_{L^2(f_v)}^2$$

De même façon :

$$\|\delta v\| < \varepsilon \quad \left\| f_v^{1/2}(\lambda) f_{v+\delta v}^{-1/2}(\lambda) \right\| < 1/C \quad \lambda \in [-\pi, \pi]$$

ce qui entraîne le résultat.

D'après ce qui précède, la norme de $k_{v+\delta v}^{-\pi N} k_{v+\delta v}$ dans $L^2(f_v)$ reste bornée quand $\delta v \rightarrow 0$. Par suite $\delta(\pi_v^N k_v)$ tend vers 0 avec δv et l'application $v \rightarrow \pi_v^N k_v$ est continue. En prenant $\delta v_i = h$ si $i = j$ et $\delta v_i = 0$ si $i \neq j$ et calculant la limite $h^{-1} \delta(\pi_v^N k_v)$ on obtient le résultat cherché.

Démonstration du lemme 3.2.

On a :

$$E \left\| \delta_N(t) \right\| < A \sum_{j=t}^{\infty} \left\| A(j) \right\| \quad 1 < t < N$$

$$E \left\| \delta_N(t+N) \right\| < A \sum_{j=t}^{\infty} \left\| A(j) \right\| \quad 1 < t$$

comme :

$$\sum_{t=1}^{\infty} \sum_{j=t}^{\infty} \left\| A(j) \right\| \quad \sum_{j=1}^{\infty} j \left\| A(j) \right\| < K$$

on déduit le lemme.

Démonstration du lemme 3.4.

Ecrivons :

$$X(n+1) = U(n+1) + \sum_{j=1}^n A(j) X(j)$$

où $U(n+1)$ est non corrélé avec $X(1), \dots, X(n)$. Si $G_{n+1}(v)$ est la matrice de covariance de $U(n+1)$ et T la matrice correspondant à la transformation :

$$(X(1), \dots, X(n+1)) \rightarrow (X(1), \dots, X(n), U(n+1))$$

alors la matrice de covariance Γ_{n+1} de $(X(1) \dots X(n+1))$ est :

$$\Gamma_{n+1} = T \begin{pmatrix} \Gamma_n & 0 \\ 0 & G_{n+1} \end{pmatrix} {}^t T$$

Comme T est triangulaire de termes diagonaux égaux à 1, on a :

$$\det[\Gamma_{n+1}(v)] = \det \Gamma_n(v) \det G_{n+1}(v)$$

Or d'après (3.2) :

$$\begin{aligned} X(n+1) &= \varepsilon(n+1) - \sum_{j=1}^{\infty} C_{\theta}(j) X(n+1-j) \\ &= \varepsilon(n+1) - \delta(n+1, \theta) - \sum_{j=1}^n C_{\theta}(j) X(n+1-j) \end{aligned}$$

où $\delta(n+1, \theta)$ est donné par (3.3). Par identification on a :

$$U(n+1) = \varepsilon(n+1) - [\delta(n+1, \theta) - E_{\nu}(\delta(n+1, \theta) | X(1), \dots, X(n))]$$

Le crochet du second membre est non corrélé avec $\varepsilon(n+1)$.

Si $H_{n+1}(\nu)$ est sa matrice de covariance, on a : $G_{n+1}(\nu) = G + H_{n+1}(\nu)$. D'où :

$$\text{Log} [\det \Gamma_N(\nu)] = N \text{Log} \det G + \sum_{n=1}^N \text{Log} \det [I + G^{-1} H_n(\nu)]$$

Maintenant le terme général de la matrice $H_n(\nu)$ est :

$$(H_n(\nu))_{j\ell} = \int_{-\pi}^{\pi} (\varphi_j^{(n)}(\lambda)) f_{\nu}(\lambda) \varphi_{\ell}^{(n)}(\lambda) d\lambda = \langle \varphi_j^{(n)}, \varphi_{\ell}^{(n)} \rangle_{L^2(f_{\nu})}$$

où $\varphi_j^{(n)}$ est l'élément de $L^2(f_{\nu})$ correspondant à $\delta_j(n, \theta) -$

$E(\delta_j(n, \theta) | X(1), \dots, X(n-1))$ (voir la démonstration du lemme 3.1).

En utilisant le fait que l'application $\nu \rightarrow f_{\nu} \in C([-\pi, \pi])$ est continûment dérivable, on montre que :

$$\begin{aligned} \frac{\partial}{\partial \nu_{\alpha}} (H_n(\nu))_{j\ell} &= \langle \varphi_j^{(n)}, \frac{\partial}{\partial \nu_{\alpha}} \varphi_{\ell}^{(n)} \rangle + \langle \frac{\partial}{\partial \nu_{\alpha}} \varphi_j^{(n)}, \varphi_{\ell}^{(n)} \rangle \\ &+ K \langle \varphi_j^{(n)}, (f_{\nu}^{-1} \frac{\partial}{\partial \nu_{\alpha}} f_{\nu}) \varphi_{\ell}^{(n)} \rangle \end{aligned}$$

D'où la majoration :

$$\begin{aligned} \left| \frac{\partial}{\partial \nu_{\alpha}} (H_n(\nu))_{j\ell} \right| &< \| \varphi_j^{(n)} \| \| \frac{\partial}{\partial \nu_{\alpha}} \varphi_{\ell}^{(n)} \| + \| \frac{\partial}{\partial \nu_{\alpha}} \varphi_j^{(n)} \| \| \varphi_{\ell}^{(n)} \| \\ &+ K \| \varphi_j^{(n)} \| \| \varphi_{\ell}^{(n)} \| \end{aligned}$$

D'après le lemme 3.3, les séries :

$$\sum_n \|\varphi_j^{(n)}\|, \quad \sum_n \left\| \frac{\partial}{\partial v_\alpha} \varphi_j^{(n)} \right\|$$

sont absolument sommables quand $v = v^*$. La série

$$\sum_n \left(\frac{\partial}{\partial v_\alpha} H_n(v) \right)_{j\ell}$$

est donc absolument sommable. Les autres résultats du lemme se démontrent d'une manière analogue.

Démonstration du théorème 3.2.

$$E_v(\varepsilon(t) | X) = \varepsilon(t, \theta) + E_v(\delta(t, \theta) | X)$$

En tenant compte du lemme 3.4, on obtient :

$$\begin{aligned} \delta_N(v) &= -\frac{1}{2} \sum_{n=1}^N \text{Log det} [I + G^{-1} H_n(v)] \\ &\quad - \frac{1}{2} \sum_{t=1}^N E_v^\dagger(\delta(t, \theta) | X) G^{-1} E(\varepsilon(t) | X) \\ &\quad - \frac{1}{2} \sum_{t=1}^N \varepsilon(t) G^{-1} E_v(\delta(t, \theta) | X) \end{aligned}$$

Comme $E_{v^*} [\|E_{v^*}(\varepsilon(t) | X)\|^2]$ et $E_{v^*} [\| \varepsilon(t, \theta) \|^2]$ sont bornés pour tout t et

$$0 < \text{Log}[\text{det}(I + G^{-1} H_n)] = \sum_{i=1}^p \text{Log}(1 + \lambda_i) < \sum_{i=1}^p \lambda_i = \text{Tr}(G^{-1} H_n)$$

($\lambda_1, \dots, \lambda_p$ étant les valeurs propres de $G^{-1} H_n$), les lemmes 3.2 et 3.4 montrent que :

$$E_v \| \delta_N(v) \| < K \quad \forall N$$

Par suite $\delta_N(v^*)$ est borné en probabilité quand $N \rightarrow \infty$. Les autres résultats du théorème se démontrent d'une manière analogue.

Démonstration du théorème 3.3.

On a d'après (3.7)

$$\delta_N(v) = \frac{1}{2} \sum_{t=N+1}^{\infty} {}^t \varepsilon(t, \theta) G^{-1} \varepsilon(t, \theta)$$

où $\varepsilon(t, \theta)$, $t > N$ est donné par (3.6). A l'aide du lemme 3.2 on montre facilement que δ_N et ses dérivées premières et secondes au point v sont bornées en norme L^1 pour tout N . D'où la première partie du théorème.

Supposons que $H2'$ soit vérifiée, on a :

$$\sup_{\theta \in V(\theta^*)} \|\varepsilon(t, \theta)\| < \sum_{j=t-N}^{t-1} \left[\sup_{\theta \in V(\theta^*)} \|C(j)\| \right] \|X(t-j)\|$$

Notons $S(t)$ le crochet du second membre. Il existe alors un voisinage $V(v^*)$

de v^* tel que $|\delta_N(v)| < K \sum_{t=N}^{\infty} S^2(t)$, $\forall v \in V(v^*)$.

Or d'après le lemme 3.2. :

$$\sum_{t=N}^{\infty} \{E_{v^*} [S(t)]^2\}^{1/2} < C \quad \forall N$$

par suite $\{E_{v^*} [S(t)]^2\}^{1/2} < C$ et :

$$\sum_{t=N}^{\infty} E_{v^*} [S(t)]^2 < C^2$$

D'où :

$$E_{v^*} \left\{ \sup_{v \in V(v^*)} |\delta_N(v)| \right\} < C^2 K \quad \forall N$$

De la même façon, on obtient :

$$\sup_{v \in V(v^*)} \left| \frac{\partial}{\partial \theta_j} \delta_N(v) \right| < K \cdot \sum S_j(t) S(t)$$

où :

$$S_j(t) = \sup_{\theta \in V(\theta^*)} \left\| \frac{\partial}{\partial \theta_j} C_{\theta}(l) \right\| \|X(t-l)\|$$

et des majorations analogues. On en déduit alors le théorème.

Démonstration du lemme 4.1.

Nous avons :

$$\frac{\partial \bar{G}}{\partial \theta_j}(\theta^*) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t)^\dagger V_j(t) + V_j(t)^\dagger \varepsilon(t)$$

$$\begin{aligned} \frac{\partial \bar{G}}{\partial \theta_j \partial \theta_\ell}(\theta^*) &= \frac{1}{N} \sum_{t=1}^N V_j(t)^\dagger V_\ell(t) + V_\ell(t)^\dagger V_j(t) + \varepsilon(t)^\dagger V_{j\ell}(t) \\ &\quad + V_{j\ell}(t)^\dagger \varepsilon(t) \end{aligned}$$

où $V_j(t)$, $V_{j\ell}(t)$ sont respectivement les dérivées de $\bar{\varepsilon}(t, \theta)$ par rapport à θ_j et à θ_j, θ_ℓ calculées en θ^* . Le lemme découle alors du théorème ergodique en remarquant que :

$$E [V_j(t)^\dagger V_\ell(t)] = \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \theta_j} h_\theta^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} f_\nu(\lambda) \left(\frac{\partial}{\partial \theta_\ell} h_\theta^{-1}(e^{i\lambda}) \right)_{\theta=\theta^*} d\lambda$$

Démonstration du théorème 4.1.

Nous avons :

$$N^{-1} \left(\frac{\partial L_N}{\partial \theta_j} \right) (\nu^*) = -\frac{1}{2} \text{Tr} [G^{-1} \left(\frac{\partial \bar{G}}{\partial \theta_j} \right) (\theta^*)] + N^{-1} o(N)$$

$$N^{-1} \left(\frac{\partial L_N}{\partial G_{j\ell}} \right) (\nu^*) = \frac{1}{2} \text{Tr} \{ G^{-1} \left(\frac{\partial G}{\partial G_{j\ell}} \right) G^{-1} [\bar{G}(\theta^*) - G] \} + N^{-1} o(N)$$

$$N^{-1} \left(\frac{\partial^2 L_N}{\partial \theta_j \partial \theta_\ell} \right) (\nu^*) = -\frac{1}{2} \text{Tr} [G^{-1} \left(\frac{\partial \bar{G}}{\partial \theta_j \partial \theta_\ell} \right) (\theta^*)] + N^{-1} o(N)$$

$$N^{-1} \left(\frac{\partial^2 L_N}{\partial \theta_j \partial G_{rs}} \right) (\nu^*) = \frac{1}{2} \text{Tr} [G^{-1} \left(\frac{\partial G}{\partial G_{rs}} \right) G^{-1} \frac{\partial \bar{G}}{\partial \theta_j} (\theta^*)] + N^{-1} o(N)$$

$$N^{-1} \left(\frac{\partial^2 L_N}{\partial G_{j\ell} \partial G_{j'\ell'}} \right) (v) = \frac{1}{2} \text{Tr} \left\{ \left[\frac{\partial}{\partial G_{j'\ell'}} (G^{-1} \frac{\partial G}{\partial G_{j\ell}} G^{-1}) \right]_{G=G^*} [\bar{G}(\theta) - G^*] \right\} \\ - \frac{1}{2} \text{Tr} \left(G^{-1} \frac{\partial G}{\partial G_{j\ell}} G^{-1} \frac{\partial G}{\partial G_{j'\ell'}} \right) + N^{-1} O(N)$$

où $O(N)$ est une variable aléatoire bornée en probabilité quand $N \rightarrow \infty$.

Le résultat du théorème découle alors du lemme 4.1 en remarquant que :

$$\frac{\partial G}{\partial G_{j\ell}} = E_{j\ell} + E_{\ell j} \quad \text{si } j < \ell \\ = E_{jj} \quad \text{si } j = \ell$$

où $E_{j\ell}$ est la matrice dont tous les éléments sont nuls sauf celui sur la j -ième colonne et la ℓ -ième ligne qui vaut 1.

Démonstration du lemme 4.2

On peut sans restreindre la généralité supposer $V(v^*)$ compact. La fonction $(v, \lambda) \rightarrow p_v(\lambda)$ est alors uniformément continue. Pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que :

$$||v-v'|| < \eta, \quad |\lambda-\lambda'| < \eta \quad : \quad ||p_v(\lambda) - p_{v^*}(\lambda')|| < \varepsilon$$

Ecrivons :

$$I^N(\lambda) = (2\pi N)^{-1} \tilde{X}_N(\lambda) \tilde{X}_N^*(\lambda) \quad ; \quad \tilde{X}_N(\lambda) = \sum_{t=1}^N e^{i\lambda t} X(t)$$

on obtient :

$$||v-v^*|| < \eta \quad |Z_v^N - Z_{v^*}^N| = \frac{1}{2\pi N} \int_{-\pi}^{\pi} ||p_v(\lambda) - p_{v^*}(\lambda) \tilde{X}_N(\lambda)||^2 d\lambda \\ < \frac{\varepsilon}{2\pi N} \int_{-\pi}^{\pi} \tilde{X}_N(\lambda) \tilde{X}_N^*(\lambda) d\lambda \\ = \varepsilon \left[\frac{1}{N} \sum_{t=1}^N X(t) X(t) \right]$$

Le crochet du second membre converge presque sûrement suivant N vers la constante $M = E(\|X(0)\|^2)$.

Maintenant si v_N est un estimateur consistant de v , on a : ($\alpha > 0$).

$$\begin{aligned} P\{|Z_{v_N}^N - Z_{v^*}^N| < \varepsilon(M+\alpha)\} &> P\{M_N < M+\alpha, \|v_N - v^*\| < \varepsilon\} \\ &> 1 - P\{M_N > M-\alpha\} - P\{\|v_N - v\| > \varepsilon\} \end{aligned}$$

Le second membre tend vers 1 quand $N \rightarrow \infty$. D'où le lemme.

Démonstration du théorème 4.3.

La démonstration suivante est due à Pierce [23].

Ecrivons les dérivées de $N^{-1/2} L_N(v)$ sous la forme :

$$N^{-1/2} \left(\frac{\partial L_N}{\partial v_j} \right) (v) = N^{-1/2} \sum_{t=1}^N {}^t \varepsilon(t) G^{-1} v_j(t) + N^{-1/2} O(N)$$

où $O(N)$ est borné en probabilité quand $N \rightarrow \infty$ et où :

$$v_j(t) = \sum_{k=1}^{\infty} \left(\frac{\partial}{\partial \theta_j} C_{\theta}(k) \right)_{\theta=\theta^*} X(t-k)$$

Toute combinaison de ces dérivées est donc de la forme

$$N^{-1/2} \sum_{t=1}^N W(t) + N^{-1/2} O(N)$$

avec :

$$W(t) = \sum_j C_j {}^t \varepsilon(t) G^{-1} v_j(t)$$

Si \mathcal{a}_t désigne la sous-tribu engendrée par $X(s)$, $s < t$, alors $E(W(t) | \mathcal{a}_{t-1}) = 0$. Comme les $W(s)$, $s < t-1$ sont \mathcal{a}_{t-1} mesurables on a :

$$E [W(t) | W(t-1), W(t-2), \dots] = 0$$

Le processus $W(t)$, $t = 1, 2, \dots$, est d'autre part stationnaire ergodique avec :

$$\begin{aligned} \sigma^2[W(t)] &= \sum_{j, \ell} C_j C_\ell E [\varepsilon(t) G^{-1} V_j(t)^t V_\ell(t) G^{-1} \varepsilon(t)] \\ &= \sum_{j, \ell} C_j C_\ell E \{ \text{Tr} [G^{-1} \varepsilon(t)^t \varepsilon(t) G^{-1} V_j(t)^t V_\ell(t)] \} \\ &= \sum_{j, \ell} C_j C_\ell \text{Tr} \{ E [G^{-1} \varepsilon(t)^t \varepsilon(t) G^{-1}] E [V_j(t)^t V_\ell(t)] \} \\ &= \sum_{j, \ell} C_j C_\ell S_j = \sigma^2 < \infty \end{aligned}$$

où S_j est donnée dans le théorème 3.1.

D'après un théorème de Billingsley [5] (p. 206) : $N^{-1/2} \sum_{t=1}^N W(t)$

converge en loi selon N vers la loi normale centrée de variance σ^2 . La première partie du théorème est donc démontrée.

Maintenant, la dérivée de $N^{-1/2} L_N$ par rapport à $G_{j\ell}$, $j < \ell$ calculée en v s'écrit :

$$N^{-1/2} \frac{1}{2} \text{Tr} \left\{ (G^{-1} \frac{\partial G}{\partial G_{j\ell}} G^{-1}) \left[\sum_{t=1}^N (\varepsilon(t)^t \varepsilon(t) - G) \right] \right\} + N^{-1/2} O(N)$$

Posons $\xi(t) = G^{-1} \varepsilon(t)$ la première expression devient :

$$N^{-1/2} \sum_{t=1}^N [\xi_j(t) \xi_\ell(t) + \xi_\ell(t) \xi_j(t)] / 2 - (G^{-1}) \quad \text{si } j \neq \ell$$

$$N^{-1/2} \sum_{t=1}^N [\xi_j(t)^2 - G_{jj}^{-1}] / 2 \quad \text{si } j = \ell$$

qui est donc de la forme $N^{-1/2} \sum r_{j\ell}(t)$. Les $r_{j\ell}(t)$ sont indépendantes de

\mathcal{A}_{t-1} , centrées et admettent des variances finies si les moments d'ordre 4 de $\varepsilon(t)$ existent. Dans ce cas toute combinaison linéaire de $r_{j\ell}(t)$ et $W(t)$ vérifie les conditions du théorème de Billingsley. D'où la convergence en loi selon N du vecteur des dérivées de $N^{-1/2} L_N$, calculé en v vers une loi gaussienne centrée.

Maintenant la covariance entre $r_{j\ell}(t)$ et $W(t)$ est clairement nulle et celle entre $r_{j\ell}(t)$ et $r_{j'\ell'}(t)$, d'après un calcul explicite, est égale à $H_{j\ell j'\ell'}$, si les cumulants d'ordre 4 de $\varepsilon(t)$ sont nuls.

Démonstration du théorème 5.1.

D'après le théorème des accroissements finis :

$k_{Nj}(v_N) = k_{Nj}(v^*) + \sum_{\ell} \left(\frac{\partial}{\partial v_{\ell}} k_{Nj}(v) \right)_{v=v_N^*} (v_{N\ell} - v_{\ell}^*)$ où v_N est un point sur le segment joignant v_N et v et par conséquent est un estimateur consistant de v . Le théorème 4.2 permet d'écrire :

$k_N(v_N) = k_N(v) + K_N(v_N - v) + \delta_N(v_N - v)$ où δ_N est une matrice aléatoire tendant vers 0 en probabilité quand $N \rightarrow \infty$

Par suite :

$$\sqrt{N}(v_N - v) = -\sqrt{N} K_N^{-1} k_N(v) + (K_N \delta_N) \cdot \sqrt{N}(v_N - v)$$

Le dernier terme du second membre tend vers 0 en probabilité selon N . En effet, $\sqrt{N}(v_N - v^*)$ est borné en probabilité, pour tout $\varepsilon > 0$, il existe $A(\varepsilon)$ tel que :

$$P \{ \sqrt{N} \|v_N - v^*\| > A \} < \varepsilon/2$$

Soit maintenant $\eta > 0$, il existe $N_0(\varepsilon, \eta)$ tel que :

$$N > N_0 : P \{ \|K_N \delta_N\| > \eta/A \} < \varepsilon/2$$

D'où :

$$\begin{aligned}
 P \{ | (K_N \delta_N) \cdot \sqrt{N}(v_N - v^*) | < \eta \} \\
 &> P \{ | |K_N \delta_N| | < \eta/A, \sqrt{N} | |v_N - v^*| | < A \} \\
 &> 1 - \epsilon/2 - \epsilon/2 = 1 - \epsilon
 \end{aligned}$$

On en déduit : $\sqrt{N}(\tilde{v}_N - v^*) \sim -\sqrt{N} K_N^{-1} k_N(v^*) \sim -\sqrt{N} K^{-1} k_N(v^*)$

Remplaçons maintenant dans les calculs précédents v_N par \hat{v}_N où \hat{v}_N est l'estimateur consistant de maximum de vraisemblance de v , on obtient :

$$0 = k_N(\hat{v}_N) = k_N(v^*) + K_N(v_N - v^*) + N(v_N - v)$$

où $\delta_N^* \rightarrow 0$ en probabilité quand $N \rightarrow \infty$.

Par suite :

$$\sqrt{N}(\hat{v}_N - v^*) + \sqrt{N} K_N^{-1} k_N(v^*) = (K_N \delta_N^*) \cdot \sqrt{N}(\hat{v}_N - v^*)$$

comme $K_N \delta_N^* \sim 0$ et $\sqrt{N} K_N^{-1} k_N(v^*)$ est convergent en loi, d'après un théorème de Billingsley [5b] (p. 62) on a $\sqrt{N}(\hat{v}_N - v) \sim -\sqrt{N} K_N^{-1} k_N(v^*) \sim \sqrt{N}(\tilde{v}_N - v)$.

D'où le théorème.

Démonstration du théorème 5.2.

L'application du lemme (4.2.), compte-tenu de (5.6.) montre que :

$$\hat{G}(\theta_N) - \hat{G}(\theta^*) \rightarrow 0 \text{ en probabilité } (N \rightarrow \infty).$$

D'autre part, le théorème 3.3 et le lemme 4.1 montrent que :

$$\hat{G}(\theta^*) \rightarrow G \text{ en probabilité } (N \rightarrow \infty).$$

D'où la consistance de $\tilde{G}_N = \hat{G}(\theta_N)$.

Maintenant, nous avons :

$$\sqrt{N} [s_{N,j}(\theta_N, G_N) - s_{N,j}(\theta_N, G^*)] = \frac{\sqrt{N}}{2} \text{Tr}[(G^{*-1} - G_N) \left(\frac{\partial \hat{G}}{\partial \theta_j} \right) (\theta_N)]$$

Le second membre tend vers 0 en probabilité selon N. En effet,

$$\left(\frac{\partial \hat{G}}{\partial \theta_j}\right)(\theta_N) = \left(\frac{\partial \hat{G}}{\partial \theta_j}\right)(\theta^*) + \sum_{\ell} (\theta_{N\ell} - \theta_{\ell}^*) \left(\frac{\partial^2 \hat{G}}{\partial \theta_j \partial \theta_{\ell}}\right)(\theta_N')$$

où θ_N' est un point sur le segment joignant θ_N et θ .

Or d'après le lemme (4.2.), compte-tenu de (5.6.)

$$\frac{\partial^2 \hat{G}}{\partial \theta_j \partial \theta_{\ell}}(\theta_N') - \frac{\partial^2 \hat{G}}{\partial \theta_j \partial \theta_{\ell}}(\theta^*) \rightarrow 0 \text{ en probabilité } (N \rightarrow \infty)$$

Par suite $(\partial^2 \hat{G} / \partial \theta_j \partial \theta_{\ell})(\theta_N')$ converge en probabilité, vers une limite quand $N \rightarrow \infty$ (lemme 4.1.).

Comme θ_N est $N^{1/2}$ consistant et G_N est consistant, le terme :

$$\sqrt{N} \operatorname{Tr}\{(G^{*-1} - G_N^{-1}) \sum_{\ell} (\theta_{N\ell} - \theta_{\ell}^*) \left(\frac{\partial^2 \hat{G}}{\partial \theta_j \partial \theta_{\ell}}\right)(\theta_N')\}$$

tend vers 0 en probabilité selon N.

D'autre part, un calcul explicite montre que la variance de $(\partial \bar{G} / \partial \theta_j)(\theta^*)$ où \bar{G} est donné par (4.1.) tend vers 0 comme N^{-1} . On en déduit que $\sqrt{N}(\partial \bar{G} / \partial \theta_j)(\theta^*)$ est borné en probabilité et :

$$\sqrt{N} \operatorname{Tr} G^{-1} - G_N^{-1} \left(\frac{\partial \bar{G}}{\partial \theta_j}\right)(\theta^*) \rightarrow 0 \text{ en probabilité } (N \rightarrow \infty).$$

La démonstration des résultats comme $\sqrt{N}(\tilde{\theta}_N - \theta^*) \sim -\sqrt{N} S^{-1} s_N(\theta^*)$ et $\sqrt{N} \tilde{\theta}_N \sim \sqrt{N} \hat{\theta}_N$ est maintenant essentiellement la même que celle du théorème 5.1.

Finalement :

$$\sqrt{N} \hat{G}(\theta_N) - \hat{G}(\hat{\theta}_N) = \sqrt{N} \sum_j (\theta_{Nj} - \hat{\theta}_{Nj}) \left(\frac{\partial \hat{G}}{\partial \theta_j}\right)(\theta_N')$$

où θ_N' est un point sur le segment joignant θ_N et θ_N . Comme θ_N et $\hat{\theta}_N$ sont $N^{1/2}$ -consistants les considérations précédentes montrent que le second membre tend vers 0 en probabilité selon N. D'où le théorème.

CHAPITRE II

SPECIFICATION DE LA METHODE GENERALE

DU CHAPITRE PRECEDENT

Nous allons appliquer la méthode générale d'estimation décrite dans le chapitre précédent au cas des modèles de processus auto-régressif, en moyenne mobile et mixte, en l'adaptant à chaque cas particulier de façon à utiliser au mieux la structure particulière du modèle. Nous retrouvons ainsi les méthodes d'estimation d'Hannan [14] et de Box et Jenkins [6] dans le cas des processus en moyenne mobile et mixte. D'autre part, on obtient des résultats intéressants dans le cas du processus auto-régressif, conduisant à des tests simples de l'ordre d'auto-régression du modèle.

§ 1. ESTIMATION ET TEST D'ADEQUATION DU MODELE DE PROCESSUS AUTO-REGRESSIF

Considérons le modèle de processus auto-régressif d'ordre q défini au chapitre I, § 1.

$$(1.1.) \quad \sum_{j=0}^q A(j) X(t-j) = \varepsilon(t) \quad A(0) = I_p$$

On suppose que les matrices $A(j)$ $j = 1, \dots, q$ qui constituent le paramètre θ du modèle vérifient :

$$\det[1 + \sum A(j) z^j] \neq 0 \quad z : |z| < 1$$

On est alors dans un cas particulier du modèle général du chapitre 1, § 2; clairement les hypothèses H1-H2' sont vérifiées.

1.1. Fonction de vraisemblance

En remarquant que la densité spectrale du processus (1.1.) est :

$$f(\lambda) = \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1} \frac{G}{2\pi} \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^* -1$$

la deuxième forme du logarithme de la fonction de vraisemblance approchée est ((3.9.), chap. I) :

$$\tilde{L}_N = -\frac{N}{2} \text{Log det}(2\pi G) - \frac{N}{2} \sum_{j,\ell} \int_{-\pi}^{\pi} \text{Tr} \left[{}^t A(j) G^{-1} A(\ell) I^N(\lambda) \right] e^{i(j-\ell)\lambda} d\lambda$$

où $I^N(\lambda)$ est le périodogramme défini par (3.8.), chap. I. Ce dernier est aussi la transformée de Fourier de la fonction de covariance empirique, soit :

$$I^N(\lambda) = \frac{1}{2\pi} \sum_{|u| < N} e^{-iu\lambda} R^N(u)$$

où $R^N(u)$ est définie par :

$$\begin{aligned} (1.2.) \quad R^N(u) &= \frac{1}{N} \sum_{t=1}^{N-u} X(t) {}^t X(t+u) && 0 < u < N \\ &= \frac{1}{N} \sum_{t=|u|+1}^N X(t) {}^t X(t+u) && -N < u < 0 \\ &= 0 && |u| > N \end{aligned}$$

Par suite :

$$(1.3.) \quad \tilde{L}_N = \frac{N}{2} \text{Log det}(2\pi G) - \frac{N}{2} \sum_{j=0}^q \sum_{\ell=0}^q \text{Tr} [{}^t A(\ell) G^{-1} A(j) R^N(j-\ell)]$$

Cette fonction est une forme quadratique en $A_{\alpha\beta}(j)$.

Nous noterons $D^N(j)$ la matrice de terme général $\partial L_N / \partial A_{\alpha\beta}(j)$.

Un calcul élémentaire montre que :

$$(1.4.) \quad D^N(j) = -N \sum_{\ell=0}^q G^{-1} A(\ell) R^N(\ell-j)$$

et par suite les dérivées secondes de L_N sont :

$$\frac{\partial^2 L_N}{\partial A_{\alpha\beta}(j) \partial A_{\gamma\delta}(\ell)} = N (G^{-1})_{\alpha\gamma} R_{\delta\beta}^N(\ell-j)$$

Pour simplifier les écritures nous noterons $\mathbf{a} = {}^t(A(1) \dots A(q))$ et $\mathbf{a}_{\cdot\alpha}$ le α -ième vecteur colonne de \mathbf{a} . La matrice des dérivées secondes de L_N par rapport aux différentes composantes de $\mathbf{a}_{\cdot\alpha}$ et $\mathbf{a}_{\cdot\beta}$ est alors $(G^{-1})_{\alpha\beta} \Gamma^N$ où :

$$\Gamma^N = \begin{pmatrix} R^N(0) & R^N(-1) & \dots & R^N(1-q) \\ R^N(1) & R^N(0) & \dots & R^N(2-q) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ R^N(q-1) & R^N(q-2) & \dots & R^N(0) \end{pmatrix}$$

Le théorème suivant qui découle du théorème ergodique est une version plus fine du théorème 4.1. du chapitre I (la convergence en probabilité est remplacée par la convergence presque sûre).

Théorème 1.1.

"Quand $N \rightarrow \infty$, $D^N(j)$, $j = 1, \dots, q$ convergent presque sûrement vers 0 et Γ^N converge presque sûrement vers :

$$\Gamma = \begin{pmatrix} R(0) & R(-1) & \dots & R(1-q) \\ R(1) & R(0) & \dots & R(2-q) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ R(q-1) & R(q-2) & \dots & R(0) \end{pmatrix}$$

où R est la fonction de covariance du processus (1.1.)"

La matrice $G^{-1} \otimes \Gamma$ qui correspond à la matrice S du théorème 4.1 du chapitre I est non singulière d'après le :

Théorème 1.2.

"La matrice Γ est non singulière".

Démonstration

Si Γ est singulière, il existe des vecteurs $\lambda(1), \dots, \lambda(q)$ non tous nuls tels que :

$$\sum_{j=1}^q \lambda(j) X(j) = 0 \quad \text{p.s}$$

Soit k le plus grand j tel que $\lambda(j) \neq 0$, alors $\lambda(k) X(k)$ est une combinaison linéaire de $X(k-j)$, $j > 1$. Or d'après la relation d'auto-régression :

$$\varepsilon(k) = X(k) + \sum_{j=1}^q A(j) X(k-j)$$

par suite $\lambda(k)\varepsilon(k)$ est aussi une combinaison linéaire de $X(k-j)$, $j > 1$, ce qui contredit le fait que $\varepsilon(k)$ est non corrélé avec ces derniers. D'où le théorème.

D'autre part, comme un processus auto-régressif d'ordre q est aussi auto-régressif d'ordre $q+s$ (avec $A(j) = 0$, $j = q+1, \dots, q+s$), le théorème 4.2 du chapitre I donne :

Théorème 1.3.

"Les matrices $N^{-1/2} D^N(j)$, $j = 1, \dots, q+s$ définies par (1.4.) et (1.2.) convergent en loi quand $N \rightarrow \infty$ vers des matrices aléatoires gaussiennes centrées dont la structure de covariance est donnée par :

$$\lim_{N \rightarrow \infty} N^{-1} E[D_{\alpha}^N(j) D_{\beta}^N(l)] = (G^{-1})_{\alpha\beta} R(-l)$$

où $D_{\alpha}^N(j)$ désigne la α -ième ligne de $D^N(j)$ et où R désigne la fonction de covariance du processus".

1.2. Estimation des paramètres

La maximisation de L_N revient à résoudre le système d'équation $D^N(j) = 0$ $j = 1, \dots, q$. Les estimateurs de maximum de vraisemblance (relatifs à L_N) $A^N(1), \dots, A^N(q)$ de $A(1), \dots, A(q)$ sont donc solutions des équations (de Yule-Walker) :

$$(1.5.) \quad R^N(-j) + \sum_{\ell=1}^N A^N(\ell) R^N(\ell-j) = 0 \quad j = 1, \dots, q$$

compte-tenu de (1.4.), on a :

$$(1.6.) \quad \sum_{\ell=1}^N [A^N(\ell) - A(\ell)] R^N(\ell-j) = N^{-1} G D^N(j) \quad j = 1, \dots, q$$

Les théorèmes 1.1. et 1.2. nous donnent alors :

Théorème 1.4.

"Les estimateurs $A^N(1), \dots, A^N(q)$ sont consistants presque-sûrement".

Nous notons $a^N = {}^t(A^N(1) \dots A^N(q))$. La loi asymptotique de a^N est donné par le théorème suivant qui est une spécification du théorème 4.2 du chapitre I et qui peut être aussi démontré directement à partir des théorèmes 1.1 - 1.4.

Théorème 1.5.

"Quand $N \rightarrow \infty$, $\sqrt{N}(a^N - a^*)$, a^* étant la vraie valeur de a , converge en loi vers une matrice aléatoire gaussienne centrée telle que la matrice de covariance entre son α -ième et son β -ième vecteur colonne est $G_{\alpha\beta} \Gamma^{-1}$ "

1.3. Test d'adéquation du modèle auto-régressif

On considère le problème de test de l'hypothèse H_0 : les observations proviennent d'un processus auto-régressif d'ordre q contre l'alternative que ce processus est auto-régressif d'ordre $q+s$. En choisissant s suffisamment grand, la classe des alternatives est suffisamment riche pour approcher convenablement la structure réelle du processus.

Notre procédure de test est une adaptation de la procédure générale de Neyman [21] et Moran [19]. Elle consiste à transformer la suite $N^{-1/2} D^N(j)$, $j = 1, 2, \dots, q+s$ en une suite de matrices aléatoires

$\eta^N(j)$, $j = 1, 2, \dots, q+s$ asymptotiquement indépendantes. Pour cela, nous calculons les résidus de la régression de $N^{-1/2} D^N(j)$ sur $N^{-1/2} D^N(\ell)$, $\ell = 1, \dots, j-1$, en utilisant les covariance limites entre les différentes variables $D_{\alpha\beta}^N(\ell)$. La matrice $\eta^N(q+j)$, ($j > 1$) (ou plus exactement une approximation de celle-ci) sert alors à tester l'hypothèse que le modèle est auto-régressif d'ordre $j-1$ contre l'alternative qu'il est d'ordre $q+1$.

D'après le théorème 1.3, la structure de covariance limite des $D^N(j)$ est directement liée à la fonction de covariance R du processus $X(t)$. Cela suggère le calcul des coefficients de régression de $D^N(j)$ sur $D^N(\ell)$, $1 < \ell < j$ à partir des coefficients d'auto-régression du processus $X(t)$.

Définition 1.1.

"Soit $\{X(t), t \in \mathbb{Z}\}$ un processus stationnaire du second ordre de fonction de covariance R . On appelle coefficients d'auto-régression (resp. d'auto-régression conjuguée) d'ordre m , les matrices $A(m, j)$ (resp. $\bar{A}(m, j)$), $j=1, \dots, m$, telles que :

$$\varepsilon(t, m) = X(t) + \sum_{j=1}^m A(m, j) X(t-j)$$

$$\text{(resp. } \bar{\varepsilon}(t, m) = X(t) + \sum_{j=1}^m \bar{A}(m, j) X(t+j)$$

soit non corrélé avec $X(t-j)$ (resp. $X(t+j)$), $j=1, \dots, m$.

D'autre part, on appelle matrice de covariance résiduelle d'auto-régression (resp. d'auto-régression conjuguée) d'ordre m , la matrice de covariance $G(m)$ (resp. $\bar{G}(m)$) du vecteur $\varepsilon(t, m)$ (resp. $\bar{\varepsilon}(t, m)$) et matrice de covariance partielle d'ordre m la matrice $\Delta(m) = E[\varepsilon(t, m) \varepsilon^T(t-m-1, m)]$

Les matrices $A(m, j)$, $\bar{A}(m, j)$, $j = 1, \dots, m$, $G(m)$, $\bar{G}(m)$ et $\Delta(m)$ précédentes ne dépendent en fait que de la fonction R et peuvent être calculées à l'aide du :

Théorème 1.6.

"Les matrices $A(j,m)$, $j = 1, \dots, m$ et $\bar{A}(m,j)$ vérifient respectivement les systèmes d'équations :

$$(1.7.) \quad R(-k) + \sum_{j=1}^m A(m,j) R(j-k) = 0 \quad k = 1, \dots, m$$

$$(1.8.) \quad R(k) + \sum_{j=1}^m \bar{A}(m,j) R(k-j) = 0 \quad k = 1, \dots, m$$

et les matrices $G(m)$, $\bar{G}(m)$, $A(m)$ sont données par :

$$(1.9.) \quad G(m) = R(0) + \sum_{j=1}^m A(m,j) R(j)$$

$$(1.10.) \quad \bar{G}(m) = R(0) + \sum_{j=1}^m \bar{A}(m,j) R(-j)$$

$$(1.11.) \quad \Delta(m) = R(-m-1) + \sum_{j=1}^m A(m,j) R(j-m-1)$$

$$= R(-m-1) + \sum_{j=1}^m R(j-m-1) \bar{A}(m,j) "$$

Démonstration :

Le système (1.7.) (resp. 1.8.) s'obtient en écrivant que $\varepsilon(t,m)$ (resp. $\bar{\varepsilon}(t,m)$) de la définition 1.1 est non corrélé avec $X(t-k)$ (resp. $X(t+k)$) $k = 1, \dots, m$. La formule (1.9.) (resp. (1.10.)) découle du fait que $G(m)$ $G(m) = E[\varepsilon(t,m) \bar{X}(t)]$ (resp. $\bar{G}(m) = E[\bar{\varepsilon}(t,m) X(t)]$) et la formule (1.11.) découle de $\Delta(m) = E[\varepsilon(t,m) \bar{X}(t-m-1)] = E[X(t+m+1) \bar{\varepsilon}(t,m)]$

Remarques :

1.- Les coefficients et matrices de covariances résiduelles d'auto-régression conjuguée relatifs à la fonction de covariance R sont les coefficients et matrices de covariances résiduelles d'auto-régression relatifs à la fonction de covariance tR , (${}^tR(u) = R(-u)$).

2.- Si R est la fonction de covariance d'un processus auto-régressif d'ordre q , de paramètres $A(1), \dots, A(q)$ et G , alors les coefficients et la matrice de covariance résiduelle d'auto-régression d'ordre m ($m > q$) relatifs à R sont : $A(m, j) = A(j)$ si $j < q$ et $A(m, j) = 0$ si $q < j < m$ et $G(m) = G$.

Or tR est aussi la fonction de covariance d'un autre processus auto-régressif d'ordre q car la densité spectrale associée à tR est aussi l'inverse d'un polynôme. Par conséquent, si $m > q$ on a $\bar{A}(m, j) = \bar{A}(q, j)$, $j = 1, \dots, q$, $\bar{A}(m, j) = 0$, $q < j < m$ et $\bar{G}(m) = \bar{G}(q)$.

3.- Les coefficients d'auto-régression et d'autres peuvent être calculés récursivement à l'aide d'un algorithme proposé par Durbin (voir [6]) généralisé par Whittle [29]. Cet algorithme est basé sur les relations de récurrence :

$$(1.12.) \quad A(m+1, m+1) \bar{G}(m) = -\Delta(m)$$

$$(1.13.) \quad A(m+1, j) = A(m, j) + A(m+1, m+1) \bar{A}(m, m+1-j) \quad j = 1, \dots, m$$

$$(1.14.) \quad \bar{A}(m+1, m+1) G(m) = -{}^t\Delta(m)$$

$$(1.15.) \quad \bar{A}(m+1, j) = \bar{A}(m, j) + \bar{A}(m+1, m+1) A(m, m+1-j) \quad j = 1, \dots, m$$

Quant à $G(m)$ et $\bar{G}(m)$ elles peuvent être calculées soit à l'aide de (1.9.) et (1.10.), soit à l'aide des relations de récurrences suivantes qui sont une conséquence simple de (1.12.) - (1.15.) :

$$(1.16.) \quad G(m+1) = G(m) + A(m+1, m+1) {}^t\Delta(m+1)$$

$$(1.17.) \quad \bar{G}(m+1) = \bar{G}(m) + \bar{A}(m+1, m+1) \Delta(m+1).$$

Si R est la fonction de covariance d'un processus auto-régressif d'ordre q , alors $\Delta(q) = 0$. On retrouve les conclusions de la remarque 2, à savoir $\bar{A}(m, j) = \bar{A}(q, j)$ $j = 1, \dots, q$, $\bar{A}(m, j) = 0$, $q < j < m$ et $\bar{G}(m) = \bar{G}(q)$.

Revenons maintenant à notre problème. Nous noterons dans toute la suite $A(m,j)$ (resp. $\bar{A}(m,j)$) $j = 1, \dots, m$ et $G(m)$ (resp. $\bar{G}(m)$) les coefficients et la matrice de covariance résiduelle d'auto-régression (resp. d'auto-régression conjuguée) d'ordre m et $\Delta(m)$ la matrice de covariance partielle d'ordre m , du processus auto-régressif (1.1.). Alors :

Théorème 1.7.

"Quand $N \rightarrow \infty$, les matrices aléatoires :

$$\eta^N(m) = N^{-1/2} [D^N(m) + \sum_{j=1}^{m-1} D^N(m-j) {}^t \bar{A}(m-1, j)] \quad m = 1, \dots, q+s$$

convergent en loi vers des matrices aléatoires gaussiennes centrées indépendantes $\eta(1), \dots, \eta(q+s)$ telles que $E [{}^t \eta_{\alpha.}(m) \eta_{\beta.}(m)] = (G^{-1})_{\alpha\beta} \bar{G}(m-1)$ ($\eta_{\alpha.}(m)$ désigne la α -ième ligne de $\eta(m)$)"

Démonstration

D'après le théorème 1.3., la matrice de covariance limite entre ${}^t(N^{-1/2} D_{\alpha.}^N(j))$ et ${}^t(N^{-1/2} D_{\beta.}^N(\ell))$ est $(G^{-1})_{\alpha\beta}$ fois celle entre $X(-j)$ et $X(-\ell)$. Par suite la matrice de covariance entre ${}^t \eta_{\alpha.}(m)$ et ${}^t \eta_{\beta.}(\ell)$ est $(G^{-1})_{\alpha\beta}$ fois celle entre :

$$\bar{\varepsilon}(-m, m-1) = X(-m) + \sum_{j=1}^{m-1} \bar{A}(m-1, j) X(-m+j)$$

et :

$$\bar{\varepsilon}(-\ell, \ell-1) = X(-\ell) + \sum_{j=1}^{\ell-1} \bar{A}(\ell-1, j) X(-\ell+j)$$

Or $\varepsilon(-m, m-1)$ est par définition non corrélé avec $X(-m+j)$, $j = 1, \dots, m-1$, donc avec $\bar{\varepsilon}(-\ell, \ell-1)$, pour tout $\ell < m$. On en déduit que $\eta(m)$ est indépendante de $\eta(\ell)$, $\ell < m$.

D'autre part, la matrice de covariance de $\bar{\varepsilon}(-m, m-1)$ est $\bar{G}(m-1)$, d'où celle entre ${}^t \eta_{\alpha.}(m)$ et ${}^t \eta_{\beta.}(m)$ est $(G^{-1})_{\alpha\beta} \bar{G}(m-1)$

Les matrices $\eta^N(m)$ ne sont pas des statistiques car elles dépendent des paramètres inconnus. Définissons des estimateurs consistants $A^N(m,j)$, $\bar{A}^N(m,j)$, $j = 1, \dots, m$, $G^N(m)$, $\bar{G}^N(m)$, $\Delta^N(m)$ par la résolution de (1.7.) et (1.8.) et les formules (1.9.)-(1.11.) où $R(u)$ est remplacée par $R^N(u)$. On pourra alors dans l'expression de $\eta^N(m)$, $m = q+1, \dots, q+s$ remplacer $A(m-1,j)$, $j = 1, \dots, m-1$:

(i) soit par $A^N(m-1,j)$; dans ce cas le résultat ne dépend plus de $A(j)$, $j = 1, \dots, q$

(ii) soit par $\bar{A}^N(q,j)$ si $1 < j < q$ et par 0 si $q < j < m-1$; dans ce cas $A(j)$, $j = 1, \dots, q$ seront remplacés par $\bar{A}(q,j)$.

Le théorème suivant montre que ces opérations ne changent pas la loi asymptotique de $\eta^N(m)$, $m = q+1, \dots, q+s$.

Théorème 1.8.

"On a

$$D^N(m) + \sum_{j=1}^{m-1} D^N(m-j) {}^t A^N(m-j) = -NG^{-1} \Delta^N(m-1) \quad m > q+s$$

Quand $N \rightarrow \infty$, les statistiques $\sqrt{N} \Delta^N(m-1)$, $m = q+1, \dots, q+s$, ainsi que les statistiques $\sqrt{N} H^N(m)$, $m = q+1, \dots, q+1$

$$H^N(m) = \sum_{j=0}^q \sum_{k=0}^q A^N(q,k) R^N(-m+j-k) {}^t \bar{A}^N(q,k)$$

$(A^N(q,0) = \bar{A}^N(q,0) = I_p)$ convergent en loi vers des matrices aléatoires indépendantes de même loi gaussienne centrée telle que la covariance entre les termes indexés par (α, β) et (γ, δ) est $G_{\alpha\gamma}(\bar{G}(q))_{\beta\delta}$, et en plus $\sqrt{N} [H^N(m) - \Delta^N(m-1)] \rightarrow 0$ en probabilité".

Démonstration

On a, en posant $A^N(m-1,0) = A(0) = I_p$

$$\begin{aligned} \sum_{j=0}^m D^N(m-j)^t \bar{A}^N(m-1,j) &= -NG^{-1} \sum_{j=0}^{m-1} \left[\sum_{\ell=0}^q A(\ell) R^N(\ell-m+j) \right]^t \bar{A}^N(m-1,j) \\ &= -NG^{-1} \sum_{\ell=0}^q A(\ell) \cdot \left[\sum_{j=0}^{m-1} R^N(\ell-m+j)^t \bar{A}^N(m-1,j) \right] \end{aligned}$$

D'après (1.8.) et (1.11.) le crochet du second membre est nul si $\ell \geq 1$ ($m > q$) et est égal à $\Delta^N(m-1)$ si $\ell = 0$. D'où la première partie du théorème.

Comme $\bar{A}^N(m-1,j)$, $j = 1, \dots, m-1$ sont des estimateurs consistants de $\bar{A}(m-1,j)$, $j = 1, \dots, m-1$, la différence de $\sqrt{N} \Delta^N(m-1)$ et $-G\eta^N(m)$ où $\eta^N(m)$ est définie dans le théorème 1.7, tend en probabilité vers 0 quand $N \rightarrow \infty$. On en déduit la deuxième partie du théorème (en remarquant que $\bar{G}(m) = \bar{G}(q)$, $m > q$).

Maintenant, pour $m > q$, $\bar{A}(m-1,j) = 0$, $q < j < m-1$ et $\bar{A}^N(m-1,j)$, $j = 1, \dots, q$ peuvent être estimés d'une manière consistante par $A^N(q,j)$, $j = 1, \dots, q$. Par conséquent, la différence entre $\eta^N(m)$ et :

$$\hat{\eta}^N(m) = N^{-1/2} \sum_{j=0}^q D^N(m-j)^t \bar{A}^N(q,j)$$

converge en probabilité vers 0 quand $N \rightarrow \infty$.

Or :

$$\begin{aligned} \hat{\eta}^N(m) &= -\sqrt{N} \cdot G^{-1} \sum_{j=0}^q \left[\sum_{k=0}^q A(k) R^N(k+j-m) \right]^t \bar{A}^N(q,j) \\ &= -\sqrt{N} G^{-1} \sum_{k=0}^q A(k) \left[\sum_{j=0}^q R^N(k+j-m)^t \bar{A}^N(q,j) \right] \end{aligned}$$

Comme $\sqrt{N} [A^N(q,k) - A(k)]$, $k = 1, \dots, q$ sont convergentes en loi (théorème 1.4.) et que le crochet du second membre converge en probabilité vers 0 (d'après (1.8.)), on voit que $\hat{\eta}^N(m)$ ne diffère de :

$$\sqrt{N} G^{-1} \sum_{k=0}^q \sum_{j=0}^q A^N(q,k) R^N(k+j-m) \mathbf{t} A^N(q,j) \quad (m > q)$$

que par un terme tendant en probabilité vers 0. D'où le théorème.

Nous sommes maintenant en mesure de définir des tests du modèle auto-régressif d'ordre q :

Test basé sur les covariances partielles

Ce test est basé sur la statistique :

$$\sum_{m=q+1}^{q+s} \text{Tr} \{ \Delta^N(m-1) [\bar{G}^N(m-1)]^{-1} \mathbf{t} \Delta^N(m-1) [G^N(m-1)]^{-1} \} = \sum_{m=q+1}^{q+s} \text{Tr} [A^N(m,m) \bar{A}^N(m,m)]$$

D'après le théorème 1.8., cette statistique, multipliée par N converge en loi quand $N \rightarrow \infty$ vers la loi du χ^2 à s degrés de liberté.

Ce test a été proposé dans le cas univariable ($p=1$) par Box et Jenkins [6] pour déterminer l'ordre d'auto-régression. Dans ce cas $G(m) = \bar{G}(m)$ et par suite les statistiques $\sqrt{N} A^N(m,m)$, $m = q+1, \dots, q+s$ sont asymptotiquement indépendantes, de loi normale centrée réduite. La statistique de

test est alors basée sur $N \sum_{m=q+1}^{q+s} [A(m,m)]^2$ qui suit approximativement la loi du χ^2 à s degrés de liberté.

Test de Quenouille généralisé

Ce test est basé sur la statistique :

$$Q^N = \sum_{m=q+1}^{q+s} \text{Tr} \{ H^N(m) [G^N(q)^{-1} t_{H^N(m)} G^N(q)]^{-1} \}$$

D'après le théorème 1.8. NQ^N converge en loi quand $N \rightarrow \infty$ vers la loi du χ^2 à sp degrés de liberté. Dans le cas univariable ($p=1$) cette statistique devient :

$$\sum_{m=q+1}^{q+s} [H^N(m)/G^N(q)]^2$$

où

$$H^N(m) = \sum_{j=0}^q \sum_{k=0}^q R^{N(m-j-k)} A^N(j) A^N(k)$$

On reconnaît alors la statistique de Quenouille (voir Anderson [1], p. 218). Notre test est donc une généralisation au cas multivariable du test de Quenouille. On peut montrer aussi que notre test est le même que celui d'Hannan [14] p. 400. Néanmoins sa statistique de test, définie à partir de la factorisation d'une densité spectrale, semble moins naturelle et est difficile à calculer numériquement.

Test basé sur les résidus d'auto-régression :

La valeur maximum de la fonction de vraisemblance sous l'hypothèse que le processus est auto-régressif d'ordre q (resp. $q+s$) est :

$$-\frac{N}{2} \text{Log}[\det 2\pi G^N(q)] - \frac{N}{2} p \quad (\text{resp. } -\frac{N}{2} \text{Log}[\det 2\pi G^N(q+s)] - \frac{N}{2} p)$$

La méthode de maximisation de vraisemblance conduit donc à la statistique de test : $N\{\text{Log}[\det G^N(q)] - \text{Log}[\det G^N(q+s)]\}$.

On peut montrer que cette statistique est asymptotiquement équivalente aux précédentes (i.e. leurs différences tendent vers 0 en probabilité).

En effet, d'après (1.18.) et (1.16.) :

$$G^N(m) - G^N(m+1) = A^N(m+1, m+1) \bar{A}^N(m+1, m+1) G^N(m)$$

ce qui montre l'équivalence asymptotique de :

$$N \sum_{m=q+1}^{q+s} \text{Tr}[A^N(m,m) \bar{A}^N(m,m)]$$

et $N \text{Tr}\{[G^N(q) - G^N(q+s)] [G^N(q+s)]^{-1}\}$. Cette dernière statistique est asymptotiquement équivalente à $N\{\text{Log}[\det G^N(q+s)] - \text{Log}[\det G(q)]\}$ d'après l'inégalité :

$$\text{Tr } F - \frac{1}{2} \text{Tr}(F^2) < \text{Log } \det (I+F) < \text{Tr } F$$

quelque soit la matrice définie positive F (cette inégalité vient du fait que $x - x^2/2 < \text{Log}(1+x) < x$; $x > 0$)

Remarques sur les tests précédents

Les trois tests précédents sont asymptotiquement équivalents sans l'hypothèse à tester H_0 . Ce résultat ne permet pas néanmoins de comparer ces tests au point de vue puissance. Ce sont tous des tests consistants pour définir leurs puissances asymptotiques, il faudrait considérer une suite d'alternatives H_N convergeant vers H_0 ; par exemple H_N est l'hypothèse $A(q+j) = \xi^N(j)$, $j = 1, \dots, s$ où $\sqrt{N} \xi^N(j)$ converge vers une limite finie. On peut alors se demander si les puissances de chacun des tests précédents ont une limite et ensuite les comparer.

Le test basé sur les covariances partielles est particulièrement adapté pour tester les hypothèses multiples correspondant aux différents ordres d'auto-régression. Les statistiques $\sqrt{N} \Delta^N(m-1)$ sont asymptotiquement indépendantes et chaque $\sqrt{N} \Delta^N(m-1)$ est la statistique naturelle pour tester l'hypothèse que le modèle est d'ordre $m-1$ contre celle que le modèle est d'ordre m .

La statistique de Quenouille a l'avantage de ne faire intervenir que les coefficients d'auto-régression et d'auto-régression conjuguée jusqu'à l'ordre q mais de ce fait, elle est liée à l'hypothèse que le modèle est d'ordre q et semble moins adaptée pour le problème du choix de l'ordre d'auto-régression. D'autre part, avec l'algorithme de Durbin généralisé, le calcul de la suite $A^N(m,m)$ et $\bar{A}(m,m)$ n'est pas plus long que celle de la suite $H^N(m)$.

§ 2. ESTIMATION ET TEST DANS LE MODELE DE PROCESSUS EN MOYENNE MOBILE

Considérons le modèle de processus centré en moyenne mobile d'ordre q défini au chapitre I, § 1 :

$$X(t) = \sum_{j=0}^q B(j) \varepsilon(t-j) \quad B(0) = I_p$$

On suppose que les matrices $B(1), \dots, B(q)$ qui constituent le paramètre θ du modèle vérifient :

$$\det[1 + \sum B(j) z^j] \neq 0 \quad z : |z| < 1$$

de façon que les hypothèses $H1, H2, H2'$ du chapitre I, § 2 soient vérifiées. Nous noterons conformément aux notations de ce paragraphe :

$$h(z) = \sum_{j=0}^q B(j) z^j$$

$$h^{-1}(z) = \sum_{j=0}^{\infty} C(j) z^j$$

et :

$$f(\lambda) = \frac{1}{2\pi} h(e^{i\lambda}) G^{-1} h(e^{i\lambda})$$

(Les indices étant omis pour alléger les écritures).

Les deux formes approchées du logarithme de la fonction de vraisemblance s'écrivent :

$$(2.1) \quad L_N = -\frac{N}{2} \text{Log}(\det 2\pi G) - \frac{1}{2} \sum_{t=1}^N \varepsilon_N(t, \theta) G^{-1} \varepsilon_N(t, \theta)$$

$$(2.2) \quad \tilde{L}_N = -\frac{N}{2} \text{Log}(\det 2\pi G) - \frac{N}{2} \int_{-\pi}^{\pi} \text{Tr} h^{-1}(e^{i\lambda}) G^{-1} h^{-1}(e^{i\lambda}) I^N(\lambda) d\lambda$$

où :

$$(2.3) \quad \varepsilon_N(t, \theta) = \sum_{j=0}^{t-1} C(j) X(t-j) \quad 1 < t < N$$

et où $I^N(\lambda)$ est le périodogramme défini par (3.8), chapitre I.

Nous allons distinguer le cas univariable et le cas multivariable de façon à mettre en évidence les difficultés qu'on rencontre en généralisant les calculs relatifs au premier cas.

2.1. Estimation dans le cas univariable ($p=1$)

Lemme 2.1

"Les $\epsilon_N(t, \theta)$, $t = 1, \dots, N$ et leurs dérivées peuvent être calculées à l'aide des équations de récurrence suivantes :

$$(2.4) \quad \epsilon_N(t, \theta) = - \sum_{j=1}^q B(j) \epsilon_N(t-j, \theta) + X(t) \quad t = 1, \dots, N$$

$$(2.5) \quad \frac{\partial}{\partial B(\ell)} \epsilon_N(t, \theta) = - \sum_{j=1}^q B(j) \frac{\partial}{\partial B(\ell)} \epsilon_N(t-j, \theta) - \epsilon_N(t-\ell, \theta)$$

de conditions initiales $\epsilon_N(t, \theta) = \frac{\partial}{\partial B(\ell)} \epsilon_N(t, \theta) = 0 \quad t < 0$ "

Démonstration

D'après leur définition, les coefficients $C(j)$, $j = 0, 1, \dots$, vérifient :

$$\sum_{j=0}^{\min(q, u)} B(j) C(u-j) = \begin{cases} 0 & \text{si } u > 0 \\ 1 & \text{si } u = 0 \end{cases}$$

D'où :

$$\sum_{j=0}^q B(j) \epsilon_N(t-j, \theta) = \sum_{j=0}^q B(j) \sum_{k=0}^{t-j-1} C(k) X(t-j-k)$$

$$= \sum_{u=0}^{t-1} \sum_{j=0}^{\min(q, u)} A(j) C(u-j) X(t-u) = X(t)$$

En dérivant (2.4) par rapport à $B(\ell)$, on obtient (2.5).

D'après le lemme si $u_N(t, \theta)$ est solution de :

$$(2.6) \quad u_N(t, \theta) = - \sum_{j=1}^q B(j) u_N(t-j, \theta) - \varepsilon_N(t, \theta), \quad t = 1, \dots, N$$

de conditions initiales $u_N(t, \theta) = 0$, $t < 0$, alors :

$$(2.7) \quad \frac{\partial}{\partial B(\ell)} \varepsilon_N(t) = u_N(t-\ell, \theta)$$

Les dérivées premières de L_N s'écrivent donc :

$$(2.8) \quad \frac{\partial L_N}{\partial B(\ell)} = \frac{1}{G} \sum_{t=1}^N u_N(t-\ell, \theta) \varepsilon_N(t, \theta)$$

D'après (2.6), on a aussi :

$$(2.9) \quad \frac{\partial L_N}{\partial B(\ell)} = - \frac{1}{G} \sum_{j=0}^q \sum_{t=1}^N u_N(t-\ell, \theta) u_N(t-j, \theta) B(j)$$

D'autre part, d'après la remarque 1 à la fin du paragraphe 5 du chapitre I, les variables aléatoires

$$(2.10) \quad \frac{1}{G} \sum_{t=1}^N \frac{\partial}{\partial B(j)} \varepsilon(t, \theta) \frac{\partial}{\partial B(\ell)} \varepsilon(t, \theta) = \frac{1}{G} \sum_{t=1}^N u_N(t-j, \theta) u_N(t-\ell, \theta)$$

et $-\partial^2 L_N / \partial B(j) \partial B(\ell)$ convergent en probabilité vers les mêmes limites si θ est remplacé par un estimateur consistant θ_N . Les limites en question sont $S_{j\ell}$ données par (4.4) du chapitre I. Dans le cas présent nous avons :

$$S_{j\ell} = \frac{1}{G} \int_{-\pi}^{\pi} \frac{\partial}{\partial B(j)} h^{-1}(e^{i\lambda}) f(\lambda) \frac{\partial}{\partial B(\ell)} h^{-1}(e^{i\lambda}) d\lambda$$

Un calcul simple montre que :

$$(2.11) \quad S_{j\ell} = \frac{1}{G} \int_{-\pi}^{\pi} |h(e^{i\lambda})|^{-2} e^{i(j-\ell)\lambda} d\lambda$$

Les relations (2.9) et (2.10) suggèrent l'introduction de la "fonction de covariance empirique" $\Lambda_N(s, \theta)$ de la série $u_N(t, \theta)$:

$$(2.12) \quad \Lambda_N(s, \theta) = \frac{1}{N} \sum_{t=1}^N u_N(t, \theta) u_N(t+s, \theta) = \Lambda_N(-s, \theta) \quad s > 0.$$

On peut montrer que la différence entre $N\Lambda_N(j-l, \theta)$ et

$$\sum_{t=1}^N u_N(t-j, \theta) u_N(t-l, \theta)$$

est bornée en probabilité quand $N \rightarrow \infty$ si θ est remplacé par un estimateur consistant θ_N . Cela vient du fait que $u_N(N-s, \theta_N)$ ainsi que $u_N(s', \theta_N)$ sont bornés en probabilité (s, s' étant fixés).

Les dérivées premières de $-N^{-1/2} L_N$ calculées en θ_N sont donc approximativement :

$$-N^{-1/2} \frac{1}{G} [\Lambda_N(j, \theta_N) + \sum_{\ell=1}^q \Lambda_N(j-\ell, \theta_N) B(\ell)]$$

et les dérivées secondes de $-N^{-1} G L_N$ calculées en θ_N sont approximativement $\Lambda_N(j-l, \theta_N)/G$.

Nous noterons $U_N(\theta)$, $u_N(\theta)$ respectivement la matrice de terme général $\Lambda_N(j-l)$ et le vecteur de composantes $\Lambda_N(j, \theta)$. Le vecteur $s_N(\theta)$ des dérivées de $-N^{-1} L_N^{-1}$ s'écrit alors :

$$(2.13) \quad s_N(\theta) = -\frac{1}{G} [u_N(\theta) + U_N(\theta)\theta] + \delta_N(\theta)$$

où $\theta = {}^t(B(1), \dots, B(q))$ et où $N\delta_N(\theta_N)$ est bornée en probabilité si θ_N est un estimateur consistant de θ .

L'application du théorème 5.2 du chapitre I montre immédiatement :

Théorème 2.1.

"Avec les notations introduites, si θ_N est un estimateur $N^{1/2}$ consistant de θ , alors l'estimateur

$$\hat{\theta}_N = 2\theta_N - U_N^{-1}(\theta_N) u_N(\theta_N)$$

est tel que $\sqrt{N}(\hat{\theta}_N - \theta)$ converge en loi vers la loi gaussienne centrée de matrice de covariance S . De plus si θ_N est l'estimateur consistant de maximum de vraisemblance de θ , alors $\sqrt{N}(\theta_N - \theta) \xrightarrow{P} 0$ en probabilité selon N ".

Considérons maintenant l'estimation basée sur la fonction \tilde{L}_N . Cette dernière peut s'écrire :

$$\tilde{L}_N = -\frac{N}{2} \text{Log det}(2\pi G) - \frac{1}{2G} \sum_{t=1}^{\infty} [\varepsilon(t, \theta)]^2$$

où

$$(2.14) \quad \varepsilon(t, \theta) = \sum_{j=t-N}^{t-1} C(j) X(t-j) \quad t > N$$

On voit facilement que $\varepsilon(t, \theta)$, $t = 1, 2, \dots$ vérifient encore (2.4) à condition de poser $X(t) = 0$ $t > N$. Les calculs sont donc les mêmes que précédemment excepté le fait que dans la formule (2.12) la sommation est étendue sur toutes les valeurs de t allant de 1 à l'infini.

On peut d'autre part, effectuer les calculs dans le "domaine spectral", c'est-à-dire en partant de (2.2) :

$$(2.15) \quad \frac{\partial \tilde{L}_N}{\partial B(\lambda)} = \frac{N}{G} \int_{-\pi}^{\pi} |h(e^{i\lambda})|^{-2} h^{-1}(e^{i\lambda}) I^N(\lambda) e^{i\lambda} d\lambda$$

$$(2.16) \quad \frac{\partial^2 \tilde{L}_N}{\partial B(j) \partial B(\lambda)} = -\frac{N}{G} \int_{-\pi}^{\pi} |h(e^{i\lambda})|^{-4} I^N(\lambda) e^{i(\lambda-j)\lambda} d\lambda$$

Par identification, l'analogue de $\Lambda(s, \theta)$ est :

$$(2.17) \quad \tilde{\Lambda}_N(s, \theta) = \int_{-\pi}^{\pi} |h(e^{i\lambda})|^{-4} I^N(\lambda) e^{-is\lambda}$$

On retrouve alors la méthode d'Hannan [14], à ceci près que le second membre de (2.17) est remplacé par l'expression approchée :

$$\frac{2\pi}{N} \sum_{k=1}^N |h e^{is \cdot (2\pi k/N)}|^{-4} I^N(2\pi k/N) e^{-is(2\pi k/N)}$$

2.2. Estimation du cas multivariable

Dans le cas multivariable, les $\varepsilon_N(t)$ peuvent être encore calculés à l'aide de l'équation de récurrence (2.5). Leurs dérivées par rapport à $B_{\alpha\beta}(\ell)$ sont solutions de :

$$\frac{\partial \varepsilon_N(t, \theta)}{\partial B_{\alpha\beta}(\ell)} = - \sum_{j=1}^q B(j) \frac{\partial \varepsilon_N(t, \theta)}{\partial B_{\alpha\beta}(\ell)} - E_{\alpha\beta} \cdot \varepsilon_N(t-\ell, \theta)$$

où $E_{\alpha\beta}$ est la matrice dont tous les termes généraux sont nuls sauf celui sur la α -ième ligne et β -ième colonne qui vaut 1.

La difficulté du cas multivariable est que nous ne pouvons plus calculer les dérivées de $\varepsilon_N(t, \theta)$ par rapport à $B_{\alpha\beta}(\ell)$ à l'aide de la résolution d'une équation de récurrence. On doit en effet résoudre p^2 équations :

$$u_N^{(\alpha\beta)}(t, \theta) = - \sum_{j=1}^q B(j) u_N^{(\alpha\beta)}(t-j, \theta) - E_{\alpha\beta} \varepsilon_N(t, \theta)$$

de façon à avoir les dérivées de $\varepsilon_N(t, \theta)$:

$$\frac{\partial \varepsilon_N(t, \theta)}{\partial B_{\alpha\beta}(\ell)} = u_N^{(\alpha\beta)}(t-\ell, \theta)$$

On peut bien sûr généraliser le théorème 2.1, mais les calculs sont beaucoup plus lourds. La matrice U_N est d'ordre $p^2 q$.

Par contre, une généralisation des calculs dans le "domaine spectral", relatifs à la fonction L_N conduit à la résolution d'un système d'ordre pq seulement. Cela vient du fait que la matrice S est produit tensoriel de G avec une matrice d'ordre pq . En effet, d'après (4.8) du chapitre I :

$$\begin{aligned} S_{(\alpha, \beta, j)(\gamma, \delta, \ell)} &= \int_{-\pi}^{\pi} \text{Tr} \left[G^{-1} \frac{\partial}{\partial B_{\alpha\beta}(j)} h^{-1}(e^{i\lambda}) f(\lambda) \frac{\partial}{\partial B_{\gamma\delta}(\ell)} h^{-1}(e^{i\lambda}) \right] d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr} \left[G^{-1} h^{-1}(e^{i\lambda}) E_{\alpha\beta} G E_{\delta\gamma} h^{-1}(e^{i\lambda}) \right] e^{i(j-\ell)\lambda} d\lambda \end{aligned}$$

En posant :

$$(2.18) \quad \Lambda(s) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} f^{-1}(\lambda) e^{-is\lambda} d\lambda$$

On a :

$$(2.19) \quad S_{(\alpha, \beta, j)(\delta, \gamma, \ell)} = \Lambda_{\gamma\alpha}(\ell-j) \cdot G_{\beta\delta} = \Lambda_{\alpha\gamma}(j-\ell) G_{\beta\delta}$$

Maintenant, notons $D_{\alpha\beta}^N(j, \theta, G)$ la dérivée de \tilde{L}_N par rapport à $B_{\alpha\beta}(j)$, on a :

$$D_{\alpha\beta}^N(j; \theta, G) = \frac{N}{2\pi} \int_{-\pi}^{\pi} \text{Tr} h^{-1}(e^{i\lambda}) E_{\beta\alpha} f^{-1}(\lambda) I^N(\lambda) e^{-ij\lambda} d\lambda$$

$$(2.20) \quad D^N(j; \theta, G) = \frac{N}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\lambda) I^N(\lambda) h^{-1}(e^{i\lambda}) e^{-ij\lambda} d\lambda$$

D'où en posant :

$$(2.21) \quad \Lambda^N(s; \theta, G) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} f^{-1}(\lambda) I^N(\lambda) f^{-1}(\lambda) e^{-is\lambda} d\lambda$$

$$(2.22) \quad D^N(j, \theta, G) = N \sum_{\ell=0}^q \Gamma^N(j; \theta, G) B(\ell) G$$

Les formules (2.21) et (2.22) généralisent (2.17) et (2.9) compte-tenu de (2.12).

Lemme 2.1.

"Soient (θ_N, G_N) un estimateur consistant de (θ, G) , alors $\Lambda^N(s; \theta_N, G_N)$ converge en probabilité vers Γ quand $N \rightarrow \infty$ "

Démonstration

Nous savons que la mesure $\Gamma^N(\lambda) d\lambda$ converge faiblement presque sûrement selon N vers $f(\lambda)d\lambda$. En effet, pour tout $\alpha \in \mathbb{C}^P$, la mesure positive $\alpha \Gamma^N(\lambda) \alpha d\lambda$, ayant pour coefficients de Fourier $\alpha R^N(u) \alpha$, convergeant presque sûrement selon N vers $\alpha R(u) \alpha$, converge donc faiblement presque sûrement selon N vers la mesure $\alpha f(\lambda) \alpha d\lambda$ (théorème de Lévy). Or pour toute matrice Q , on a :

$$(\alpha + \beta)^* Q(\alpha + \beta) - (\alpha - \beta)^* Q(\alpha - \beta) - i[(\alpha + i\beta)^* Q(\alpha + i\beta) - (\alpha - i\beta)^* Q(\alpha - i\beta)] = 4\alpha^* Q \beta$$

En prenant pour α, β les vecteurs de la base canonique de \mathbb{R}^P , on a le résultat. Le lemme découle alors du lemme 4.2 du chapitre I.

Théorème 2.2.

"Soient $\theta_N = (B_N(1), \dots, B_N(q))$ un estimateur $N^{1/2}$ consistant de θ et G_N un estimateur consistant de G , et $\{B_N^i(j), j = 1, \dots, q\}$ la solution de :

$$\Lambda^N(\ell; \theta_N, G_N) + \sum_{j=1}^q \Lambda^N(\ell - j; \theta_N, G_N) B_N^i(j) = 0 \quad \ell = 1, \dots, q$$

alors les estimateurs :

$$\tilde{B}_N(j) = 2B_N(j) - B_N^i(j) \quad j = 1, \dots, q$$

sont tels que $\sqrt{N}[\tilde{B}_N(j) - B(j)]$, $j = 1, \dots, q$ convergent en loi selon N vers la loi gaussienne centrée de structure de covariance :

$$\lim_{N \rightarrow \infty} N \text{Cov}\{[\tilde{B}_N(j)]_{\alpha\beta}, [\tilde{B}_N(\ell)]_{\delta\gamma}\} = (\Gamma^{-1})_{(\alpha, j)(\gamma, \ell)} G_{\beta\delta}$$

où Γ est la matrice de terme général $\Gamma_{(\alpha, j)(\gamma, \ell)} = \Lambda_{\alpha\gamma}(\ell - j)$

D'autre part si θ_N est l'estimateur consistant de maximum de vraisemblance de θ , alors $\sqrt{N}(\theta_N - \theta) \rightarrow \infty$ en probabilité selon N''

Démonstration

L'application du théorème 5.2 du chapitre I, compte-tenu du lemme précédent montre que l'estimateur $(B_N(j) - B_N''(j), j = 1, \dots, q)$ où les $B_N''(j)$ sont solutions de

$$\sum_{\gamma, \delta, \ell} \Lambda_{\alpha\gamma}^N(j-\ell, \theta_N, G_N) (G_N)_{\beta\delta} B_N''(\ell)_{\gamma\delta} = - N^{-1} D_{\alpha\beta}^N(j, \theta_N, G_N)$$

possède les propriétés du théorème.

Or le système précédent s'écrit sous forme matricielle :

$$\begin{aligned} \sum_{\ell=1}^q \Lambda^N(j-\ell; \theta_N) B_N'' G_N &= - N D^N(j, \theta_N, G_N) \\ &= \sum_{j=0}^q \Lambda^N(j-\ell; \theta_N, G_N) B_N(\ell) G_N \end{aligned}$$

Par suite $B_N'(j)$ définie dans le théorème est égale à $B_N''(j) + B_N(j)$. D'où $\tilde{B}_N(j) = B_N(j) - B_N''(j)$ et le théorème est démontré.

Remarque :

L'intégrale :

$$\Lambda^N(s, \theta, G) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} f^{-1}(\lambda) I^N(\lambda) f^{-1}(\lambda) e^{-is\lambda} d\lambda$$

n'est pas commode à calculer. La méthode d'Hannan consiste à la remplacer par :

$$\frac{1}{2\pi N} \sum_{k=0}^{N-1} f^{-1}(2\pi k/N) I^N(2\pi k/N) f^{-1}(2\pi k/N) e^{-is(2\pi k/N)}$$

Une autre méthode consiste à passer au 'domaine temporel' par la transformée de Fourier :

$$(2.23) \quad \Lambda^N(s; \theta, G) = \frac{1}{N} \sum_{t=-\infty}^{+\infty} v_N(t; \theta, G)^{\dagger} v_N(t; \theta, G)$$

où $v_N(t; \theta, G)$ sont les coefficients de Fourier de

$$2\pi f(\lambda)^{-1} \sum_{t=1}^N x(t) e^{i\lambda t} \text{ et par conséquent vérifient :}$$

$$(2.24) \quad v_N(t; \theta, G) = - \sum_{j=1}^q {}^{\dagger} B(j) v_N(t+j, \theta, G) + G^{-1} \epsilon_N(t, \theta)$$

où les $\epsilon_N(t, \theta)$ sont donnés par (2.4).

La relation de récurrence (2.24) ne permet pas de calculer $v_N(t, \theta, G)$ car on n'a pas les valeurs initiales. Toutefois, comme $v_N(t, \theta, G)$ tend exponentiellement vers 0 quant $t \rightarrow \infty$, on peut les supposer nulles à partir d'un rang $N+M$. Un calcul détaillé montrera que l'erreur commise sur $\Gamma^N(s, \theta, G)$ est négligeable quand $N \rightarrow \infty$.

Nous terminons ce paragraphe en indiquant un choix possible de l'estimateur initial $B_N(j)$, $j = 1, \dots, q$. Soit :

$$(2.25) \quad f_N(\lambda) = \sum_{|u| < q} R^N(u) e^{-i\lambda u}$$

Nous savons (Hannan [14], p. 66) que si $\det[f_N(\lambda)] \neq 0$, il existe une factorisation :

$$(2.26) \quad f_N(\lambda) = \frac{1}{2\pi} \left[\sum_{j=0}^q B_N(j) e^{ij\lambda} \right] G_N \left[\sum_{j=0}^q B_N(j) e^{ij\lambda} \right]; B(0) = I_p$$

vérifiant :

$$(2.27) \quad \det[\sum B_N(j) z^j] \neq 0 \quad \forall z : |z| < 1.$$

Les $B_N(j)$, $j = 1, \dots, q$, G_N peuvent être déterminées à partir de $R^N(u)$; $u = 0, 1, \dots, q$ comme l'unique solution de :

$$(2.28) \quad R^N(u) = \sum_{j=0}^{q-u} B_N(j) G_N^t B_N(j+u) \quad u = 0, 1, \dots, q$$

vérifiant (2.27).

On voit alors que $B_N(j)$, $j = 1, \dots, q$, G_N sont des fonctions continûment dérivables de $R^N(u)$, $u = 0, 1, \dots, q$ dans le domaine où elles sont définies. Comme $\sqrt{N}[R^N(u) - R(u)]$, $u = 0, 1, \dots, q$ sont asymptotiquement gaussiennes (Hannan [14] p. 228), d'après un théorème de Cramer, il en est de même pour $\sqrt{N}[B_N(j) - B(j)]$, $j = 1, \dots, q$, $\sqrt{N}(G_N - G)$. D'où le $N^{1/2}$ - consistence de ces estimateurs.

La détermination de $B_N(j)$, $j = 1, \dots, q$, G_N n'est pas commode, on peut l'éviter en remarquant que les estimateurs $B_N^t(j)$, $j = 1, \dots, q$ du théorème 2.2. sont aussi $N^{1/2}$ - consistants et leur calcul ne fait intervenir que :

$$\frac{1}{4\pi^2} \int_{-\pi}^{\pi} f_N^{-1}(\lambda) I^N(\lambda) f_N^{-1}(\lambda) e^{-is\lambda} d\lambda = \Gamma^N(s; \theta_N, G_N)$$

Hannan [14] a montré que l'on peut remplacer l'intégrale précédente par une sommation sur les valeurs $2\pi k/N$ de λ sans changer les propriétés asymptotiques des estimateurs obtenus. Pour estimer G , on peut utiliser soit :

$$(G_N^t)^{-1} = \Lambda^N(0; \theta_N, G_N) + \sum_{j=1}^q \Lambda^N(-j; \theta_N, G_N) B_N^t(j)$$

soit :

$$G_N = \frac{1}{N} \sum_{t=1}^N \epsilon_N(t, \theta_N)^t \epsilon(t, \theta_N)$$

§ 2.3. TEST D'HYPOTHESES DANS LE MODELE DE PROCESSUS EN MOYENNE MOBILE

Nous considérons le problème de test de l'hypothèse H_0 que le processus est en moyenne d'ordre q contre l'alternative qu'il est d'ordre $q+s$. Pour cela nous rappelons ce résultat fondamental qui est la spécification du théorème 4.3., compte-tenu de (2.19).

Théorème 2.3.

"Sous l'hypothèse H_0 , les matrices $N^{-1/2} D^N(j; \theta, G)$, $j=1, \dots, q+s$ définies par (2.22) et (2.21) convergent en loi quand $N \rightarrow \infty$, vers des matrices aléatoires gaussiennes centrées dont la structure de covariance est donnée par :

$$\lim_{N \rightarrow \infty} N^{-1} E [D_{\cdot\alpha}^N(j; \theta, G) {}^t(D_{\cdot\beta}^N(l; \theta, G))] = \Lambda(j-l) \cdot G_{\alpha\beta}$$

où $D_{\cdot\alpha}^N(j; \theta, G)$ désigne la α -ième colonne de $D^N(j; \theta, G)$ et où Λ est donné par (2.18)"

Le théorème 2.3 ressemble beaucoup au théorème 1.2; les tests de ce paragraphe peuvent être construits de la même manière que ceux du paragraphe 1.3. Il faut toutefois faire attention au fait que $D^N(j; \theta, G)$ dépend des paramètres θ, G qui doivent être estimés.

Soit donc θ_N un estimateur $N^{1/2}$ -consistant de θ et G_N un estimateur consistant de G . La démonstration du théorème 5.2, chapitre I, montre que :

$$(2.29) \quad \lim_{N \rightarrow \infty} N^{-1/2} [D^N(j, \theta_N, G_N) - D^N(j, \theta_N, G^*)] = 0 \text{ (en probabilité)}$$

D'autre part la démonstration du théorème 5.1 montre que l'on peut écrire :

$$\begin{aligned}
 -N^{-1} D_{\alpha\beta}^N(j, \theta_N, G) &= -N^{-1} D^N(j, \theta^*, G^*) + \sum_{\gamma\delta\ell} S_{(\alpha,\beta,j)(\gamma,\delta,\ell)} [B_N^{(\ell)} - B^{*(\ell)}] \gamma\delta \\
 &\quad + \varepsilon_{(\alpha,\beta,j)(\gamma,\delta,\ell)}^N [B_N^{(\ell)} - B^{*(\ell)}] \gamma\delta
 \end{aligned}$$

où S est donnée par (2.19) et $\varepsilon^N \rightarrow 0$ en probabilité quand $N \rightarrow \infty$

Comme $\theta_N = (B^N(\ell), \ell = 1, \dots, q)$ est $N^{1/2}$ consistant :

$$(2.30) \quad N^{-1/2} D^N(j, \theta_N, G^*) - D^N(j, \theta^*, G^*) = -\sqrt{N} \sum_{\ell=1}^q \Lambda(j-\ell) [B^N(\ell) - B^{*(\ell)}] G^* + o(N)$$

où $o(N) \rightarrow 0$ en probabilité quand $N \rightarrow \infty$.

Répetons maintenant la démonstration du théorème 1.7 où Λ joue le rôle de R et ${}^t D^N(j, \theta^*, G)$ joue le rôle de $D^N(j)$. Remarquons que d'après (2.18) : Λ est la fonction de covariance d'un processus auto-régressif d'ordre q (où les coefficients sont ${}^t B(1), \dots, {}^t B(q)$ et où la matrice de covariance des résidus est \bar{G}^{-1}). Par suite si $\bar{B}(q,1), \dots, \bar{B}(q,q)$ et $\bar{G}(q)$ sont les coefficients et la matrice de covariance résiduelle d'auto-régression conjuguée relatifs à Λ , alors : "les matrices aléatoires :

$$\eta^N(m) = N^{-1/2} [D^N(m; \theta^*, G) + \sum_{j=1}^q \bar{B}(q,j) D^N(m-j, \theta^*, G)] \quad m = q+1, \dots, q+s$$

convergent en loi quand $N \rightarrow \infty$ vers des matrices aléatoires indépendantes de même loi gaussienne centrée dont la matrice de covariance entre les α -ième et β -ième vecteurs lignes est $\bar{G}(q) G_{\alpha\beta}$ ".

Soit $m \in \{q+1, \dots, q+s\}$ et $\bar{B}^N(j)$ un estimateur consistant de $\bar{B}(q,j)$ si $1 \leq j \leq q$ et de 0 si $q < j < m$. Il est clair que la matrice :

$$\hat{\eta}^N(m) = N^{-1/2} [D^N(m; \theta^*, G) + \sum_{j=1}^{m-1} \bar{B}^N(j) D^N(m-j, \theta^*, G)]$$

ne diffère de $\eta^N(m)$ que par un terme tendant vers 0 en probabilité quand $N \rightarrow \infty$. Maintenant, compte-tenu de (2.29) et (2.30) la différence entre :

$$(2.31) \quad N^{+1/2} [D^N(m, \theta_N, G_N) + \sum_{j=1}^{m-1} \bar{B}^N(j) D^N(m-j; \theta_N, G_N)]$$

et $\hat{\eta}^N(m)$ est :

$$\sqrt{N} \sum_{j=0}^m \sum_{\ell=1}^q \bar{B}^N(j) \Lambda(m-j-\ell) [B^N(\ell) - B^*(\ell)] G^* + O(N)$$

où $B^N(0) = Ip$ et où $O(N)$ tend vers 0 en probabilité quand $N \rightarrow \infty$.

Or le premier terme de l'expression précédente s'écrit :

$$\sum_{\ell=1}^q \left[\sum_{j=0}^m \bar{B}^N(j) \Lambda(m-\ell-j) \right] \sqrt{N} [B^N(\ell) - B^*(\ell)] G^*$$

Le crochet tend vers 0 en probabilité quand $N \rightarrow \infty$ d'après (1.11), par suite la différence entre $\hat{\eta}^N(m)$ et (2.31) tend vers 0 en probabilité selon N .

Comme estimateur $\bar{B}^N(j)$, on peut prendre soit $\bar{B}^N(j) = B^N(q, j)$ $j < q$, $\bar{B}^N(j) = 0$, $j > q$, soit $\bar{B}^N(j) = B^N(m-1, j)$, où les $B^N(m, j)$ sont les coefficients d'auto-régression conjuguée relatifs à Λ . On obtient donc le théorème suivant qui est l'analogie du théorème 1.7.

Théorème 2.4.

"Quand $N \rightarrow \infty$, les statistiques $\sqrt{N} H^N(m)$, $m = q+1, \dots, q+s$ où

$$H^N(m) = \sum_{j=0}^q \sum_{\ell=0}^q \bar{B}^N(q, j) \Lambda^N(m-j-\ell, \theta_N, G_N) B^N(\ell) \quad (\bar{B}^N(q, 0) = B^N(0) = Ip)$$

convergent en loi vers des matrices aléatoires indépendantes de même loi gaussienne centrée telle que la covariance entre les termes indexés par (α, β) et (γ, δ) est $(G(q))_{\alpha\gamma} (G^{-1})_{\beta\delta}$

On a le même résultat pour les statistiques $\sqrt{N} \Delta^N(m)$, $m = q+1, \dots, q+s$, où $\Delta^N(m)$ est la matrice de covariance partielle d'ordre m relative à $\Gamma^N(\cdot, \theta_N, G_N)$ est en plus $\sqrt{N} [H^N(m) - \Delta^N(m-1)]$ $q < m < q+s$ tend vers 0 en probabilité selon N .

§ 3. ESTIMATION DANS LE MODELE DE PROCESSUS MIXTE

Considérons le modèle de processus mixte d'ordre (q,r) défini au chapitre 1, § 1 :

$$(3.1) \quad \sum_{j=0}^q A(j) X(t-j) = \sum_{j=0}^r B(j) \varepsilon(t-j) \quad A(0) = B(0) = I_p$$

Nous ne restreignons pas la généralité en supposant que les matrices $A(j)$ (resp. $B(j)$) dépendent d'un paramètre vectoriel $a \in R^\alpha$ (resp; $b \in R^\beta$) de façon que :

$$\phi(z) = \sum_{j=0}^q A(j) z^j = \sum_{k=0}^{\alpha} a_k \phi_k(z) \quad a_0 = 1$$

$$\psi(z) = \sum_{j=0}^r B(j) z^j = \sum_{k=0}^{\beta} b_k \psi_k(z) \quad b_0 = 1$$

Nous notons $\theta = (a,b), v = (\theta, G)$ et nous supposons que pour la vraie valeur de θ , $\det [\phi(z)]$ et $\det [\psi(z)]$ ne s'annulent pas dans le disque fermé $|z| < 1$ du plan complexe. Ainsi les hypothèses H1-H2' du chapitre 1, §2 sont vérifiées.

Introduisons l'opérateur symbolique : $Z : ZX(t) = X(t-1)$.

De même manière que dans le paragraphe précédent, on a :

Lemme 3.1.

"Les $\varepsilon_N(t, \theta), t \in \mathbb{Z}$ intervenant dans la fonction L_N (ou \tilde{L}_N) vérifient l'équation de récurrence :

$$(3.2) \quad \varepsilon_N(t, \theta) = - \sum_{j=1}^q B(j) \varepsilon_N(t-j) + \phi(Z) X_N(t)$$

où $X_N(t) = X(t), 1 < t < N$ et $X_N(t) = 0$ sinon. Les dérivées de

$\varepsilon_N(t, \theta)$ par rapport à a_j , notées $U_j^N(t, \theta)$ et par rapport à b_j , notées $V_j^N(t, \theta)$ vérifient respectivement les équations de récurrence :

$$(3.3) \quad u_j^N(t, \theta) = \sum_{\ell=1}^q B(\ell) \dot{u}_j^N(t-\ell, b) + \phi_j(Z) X_N(t)$$

$$(3.4) \quad v_j^N(t, b) = \sum_{\ell=1}^a B(\ell) v_j^N(t-\ell, \theta) - \psi_j(Z) \varepsilon_N(t)$$

D'après (3.3) on voit que $u_j^N(t, \theta)$ ne dépend en fait que de b .

On la notera $u_j^N(t, b)$ quand on veut expliciter cette indépendance.

Les dérivées de $-N^{-1}L_N$ sont données par :

$$(3.5) \quad -\frac{1}{N} \left(\frac{\partial L_N}{\partial a_j} \right) (v) = \ell_j^N(v) = \frac{1}{N} \sum_{t=1}^N t u_j^N(t, \theta) G^{-1} \varepsilon_N(t, \theta)$$

$$(3.6) \quad -\frac{1}{N} \left(\frac{\partial L_N}{\partial b_j} \right) (v) = k_j^N(v) = \frac{1}{N} \sum_{t=1}^N t v_j^N(t, \theta) G^{-1} \varepsilon_N(t, \theta)$$

D'autre part si $v_N = (\theta_N, G_N)$ est un estimateur consistant de v .
D'après la remarque 1 à la fin du paragraphe 5, chapitre I, on a :

$$(3.7) \quad -\frac{1}{N} \left(\frac{\partial^2 L_N}{\partial a_j \partial a_k} \right) (v_N) = \frac{1}{N} \sum_{t=1}^N t u_j^N(t, b_N) G_N^{-1} u_k^N(t, b_N) = U_{jk}^N(v_N)$$

$$(3.8) \quad -\frac{1}{N} \left(\frac{\partial^2 L_N}{\partial a_j \partial b_k} \right) (v_N) \sim \frac{1}{N} \sum_{t=1}^N t u_j^N(t, b_N) G_N^{-1} v_k^N(t, \theta_N) = \Omega_{jk}^N(v_N)$$

$$(3.9) \quad -\frac{1}{N} \left(\frac{\partial^2 L_N}{\partial b_j \partial b_k} \right) (v_N) \sim \frac{1}{N} \sum_{t=1}^N t v_j^N(t, \theta_N) G_N^{-1} v_k^N(t, \theta_N) = V_{jk}^N(v_N)$$

Le signe $x_N \sim y_N$ signifie que la différence $x_N - y_N$ converge en probabilité vers 0 quand $N \rightarrow \infty$.

D'après le théorème 4.2 du chapitre I, les matrices aléatoires $U^N(v_N)$, $\Omega^N(v_N)$, $V^N(v_N)$ convergent en probabilité quand $N \rightarrow \infty$ vers des matrices notées U , Ω , V . Ces matrices peuvent être calculées à l'aide de la formule 4.8 du chapitre I.

L'application du théorème 5.2 est immédiate. Elle montre que :
 "Si $\theta_N = (a_N, b_N)$ est $N^{1/2}$ -consistant alors l'estimateur :

$$(3.10) \quad \tilde{\theta}_N = \begin{pmatrix} \tilde{a}_N \\ \tilde{b}_N \end{pmatrix} = \begin{pmatrix} a_N \\ b_N \end{pmatrix} - \begin{pmatrix} U^N(v_N) & \Omega^N(v_N) \\ t_\Omega^N(v_N) & V^N(v_N) \end{pmatrix} \cdot \begin{pmatrix} \ell^N(v_N) \\ k^N(v_N) \end{pmatrix}$$

est tel que $\sqrt{N}(\tilde{\theta}_N - \theta^*)$ converge en loi vers la loi gaussienne centrée de matrice de covariance :

$$\begin{pmatrix} U & \Omega \\ t_\Omega & V \end{pmatrix}^{-1}$$

et en plus si $\hat{\theta}_N$ est l'estimateur de maximum de vraisemblance de θ , alors

$$\sqrt{N} \tilde{\theta}_N \sim \sqrt{N} \hat{\theta}_N "$$

Nous avons remarqué que $\frac{\partial}{\partial a_j} \varepsilon_N(t, \theta) = u_j^N(t, b)$ est indépendante de a . Cela entraîne :

$$\varepsilon_N(t, \theta) = \sum_{j=0}^{\alpha} a_j u_j^N(t, b)$$

($u_0^N(t, b)$ étant défini de la même façon que $u_j^N(t, b)$). Par suite $\ell^N(\theta, G)$ est une forme linéaire en a :

$$(3.11) \quad \ell^N(\theta, G) = u^N(b, G) \cdot a + u_0^N(b, G)$$

où $u^N(b, G)$ est écrit à la place de $u^N(v)$ car ce dernier est indépendant de a et où $u_0^N(b, G)$ est le vecteur de composantes :

$$(3.12) \quad u_{0j}^N(b, G) = \frac{1}{N} \sum_{t=1}^N t u_0^N(t, b) G^{-1} u_j^N(t, b)$$

On voit alors que si b, G sont donnés, la valeur de a qui maximise L_N peut être calculée directement, elle est :

$$(3.13) \quad \hat{a}(b, G) = - U^N(b, G)^{-1} U_0^N(b, G)$$

Nous avons donc intérêt à séparer l'estimation de b à celle de a , cette séparation est précisée par :

Théorème 3.1.

"Avec les notations introduites, si $\theta_N = (a_N, b_N)$ est un estimateur $N^{1/2}$ -consistant de θ , G_N , un estimateur consistant de G , alors les estimateurs \tilde{a}_N, \tilde{b}_N définis par :

$$a'_N = \hat{a}(b_N, G_N)$$

$$v'_N = ((a'_N, b_N), (G_N))$$

$$\tilde{b}'_N = b_N - \{V^N(v'_N) - {}^t_{\Omega}{}^N(v'_N) [U^N(b_N, G)]^{-1} \Omega^N(v'_N)\}^{-1} k^N(v'_N)$$

$$\tilde{a}'_N = \hat{a}(\tilde{b}'_N, G_N)$$

vérifient $\sqrt{N} \tilde{a}'_N \sim \sqrt{N} a_N, \sqrt{N} \tilde{b}'_N \sim \sqrt{N} \tilde{b}_N$ où $(\tilde{a}_N, \tilde{b}_N)$ est donné par (3.9)"

Démonstration

D'après (3.10) on a :

$$U^N(v_N) (\tilde{a}_N - a_N) + \Omega^N(v_N) (\tilde{b}_N - b_N) + \ell^N(v_N) = 0$$

$${}^t_{\Omega}{}^N(v_N) (\tilde{a}_N - a_N) + V^N(v_N) (\tilde{b}_N - b_N) + k^N(v_N) = 0$$

Eliminant $\tilde{a}_N - a_N$ du système précédent, il vient :

$$-\{V^N(v_N) - {}^t\Omega^N(v_N)[U^N(v_N)]^{-1} \Omega^N(v_N)\} (b_N - b_N) = {}^t\Omega^N(v_N)[U^N(v_N)]^{-1} \cdot \ell^N(v_N) - k^N(v_N)$$

D'après (3.11) et (3.13) le second membre est :

$$-[\Omega^N(v_N)(a_N' - a_N) + k^N(v_N)]$$

Or la matrice des dérivées de k^N par rapport à a , calculées en $v_N^{(1)}$ où $v_N^{(1)}$ est un estimateur consistant de v et ${}^t\Omega^N(v_N)$ convergent toutes les deux en probabilité quand $N \rightarrow \infty$ vers la même limite ${}^t\Omega$. L'expression précédente peut donc s'écrire :

$$-k^N(v_N') + \varepsilon_N(a_N' - a_N) \quad ; \quad v_N' = ((a_N', b_N), G_N)$$

où $\varepsilon_N \sim 0$. On en déduit alors :

$$\begin{aligned} \sqrt{N}(b_N - b_N) &\sim -\sqrt{N}\{V^N(v_N) - {}^t\Omega^N(v_N)[U^N(v_N)]^{-1} \Omega^N(v_N)\}^{-1} k^N(v_N') \\ &\sim \sqrt{N}(\tilde{b}_N' - b_N) \end{aligned}$$

Maintenant, la première équation du système précédent donne :

$$a_N - a_N = -[U^N(v_N)]^{-1} [\ell^N(v_N) + \Omega^N(v_N) - (b_N - b_N)]$$

Un raisonnement analogue au précédent montre que le deuxième crochet peut s'écrire (compte-tenu de (3.11)).

$$U^N(b_N, G_N) \cdot a_N + U_0^N(b_N, G) + \varepsilon_N \cdot (\tilde{b}_N - b_N)$$

où $\varepsilon_N \sim 0$. Ce qui entraîne que $\sqrt{N}(\tilde{a}_N - a_N) \sim \sqrt{N}(\tilde{a}_N' - a_N)$.

Remarque :

Dans le cas univariable, G (qui est alors un scalaire) peut être éliminé dans les formules de U_N , V_N , Ω_N et k_N , ℓ_N .

Nous terminons ce paragraphe en indiquant un choix possible des estimateurs initiaux a_N et b_N . En fait en utilisant l'algorithme basé sur le théorème 3.1, on n'a pas besoin de a_N . Maintenant la fonction de covariance R du processus vérifie (théorème 1.5 du chapitre 1).

$$(3.14) \quad \sum_{j=0}^q A(j) R(j-k) = 0 \quad k = r+1, \dots, r+q$$

$$(3.15) \quad \sum_{j=0}^q \sum_{k=0}^q A(j) R(u+j-k)^\dagger A(k) = \sum_{j=0}^{r-u} B(j) G^\dagger B(j+u) \quad u = 0, 1, \dots, r$$

Si on remplace R par la fonction de covariance empirique R^N , le système (3.14) fournit un estimateur $N^{1/2}$ -consistant $(A^N(1), \dots, A^N(q))$ de $A(1), \dots, A(q)$. Par suite, un estimateur $N^{1/2}$ -consistant de $\tilde{R}(u)$ où $\tilde{R}(u)$ désigne la valeur commune des deux membres de l'égalité (3.15) est :

$$(3.16) \quad \tilde{R}^N(u) = \sum_{j=0}^q \sum_{k=0}^q A^N(j) R^N(u-j-k)^\dagger A^N(k)$$

Alors l'unique solution du système :

$$(3.17) \quad \tilde{R}^N(u) = \sum_{j=0}^{r-u} B^N(j) G^N B^N(j+u) \quad B^N(0) = I_p \quad ; \quad u = 0, \dots, r$$

vérifiant $\det \left[\sum_{j=0}^r B^N(j) z^j \right] \neq 0 \quad \forall z : |z| < 1$, est un estimateur

$N^{1/2}$ -consistant de $(B(j), j = 1, \dots, r, G)$. On peut en déduire un estimateur $N^{1/2}$ -consistant b_N de b .

La résolution du système (3.17) est difficile. On peut l'éviter en utilisant la procédure suivante :

Soit $\tilde{f}(\lambda)$ la densité spectrale correspondant à \tilde{R} définie par (3.15) :

$$\tilde{f}(\lambda) = \sum_{|u| < q} \tilde{R}(u) e^{-iu\lambda} = \frac{1}{2\pi} \psi(e^{i\lambda}) G \psi^*(e^{i\lambda})$$

La dérivée de $-N\tilde{L}_N$ par rapport à b_j , notée $\tilde{k}_N(v)$ s'écrit :

$$\begin{aligned}\tilde{k}_N(\lambda) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr} G^{-1}(\psi^{-1}\psi_j\psi^{-1}\phi)(e^{i\lambda}) I^N(\lambda) (\phi^*\psi^{*-1})(e^{i\lambda}) d\lambda \\ &= \int_{-\pi}^{\pi} \text{Tr}(\psi_j\psi^{-1}\phi)(e^{i\lambda}) I^N(\lambda) \phi^*(e^{i\lambda}) f^{-1}(\lambda) d\lambda\end{aligned}$$

comme $\psi^{-1}(e^{i\lambda}) = (2\pi)^{-1} \sum b_k [G\psi_k(e^{i\lambda}) \tilde{f}^{-1}(\lambda)]$, on a :

$$(3.19) \quad \tilde{k}_j^N(v) = \sum_{k=0}^{\beta} \tilde{V}_{jk}^N(v) b_k$$

où :

$$(3.20) \quad \tilde{V}_{jk}^N(v) = \frac{1}{2} \int_{-\pi}^{\pi} \text{Tr}[\psi_j(e^{i\lambda}) G \psi_k^*(e^{i\lambda}) \tilde{f}^{-1}(\lambda) \phi^*(e^{i\lambda}) I^N(\lambda) (e^{i\lambda}) \tilde{f}^{-1}(\lambda)] d\lambda$$

Nous noterons respectivement $\tilde{V}^N(v)$ et $\tilde{V}_0^N(v)$ la matrice de termes $\tilde{V}_{jk}^N(v)$, $j, k = 1, \dots, \beta$ et le vecteur de composantes $V_{0k}^N(v)$, $k = 1, \dots, \beta$, et nous posons :

$$(3.21) \quad b_N' = - [V^N(v_N)]^{-1} V_0^N(v_N)$$

où $v_N = ((a_N, b_N), G_N)$ est l'estimateur $N^{1/2}$ -consistant de v calculé par la méthode précédente. D'après (3.19) :

$$b_N' - b_N = [V^N(v_N)]^{-1} \tilde{k}_N^N(v_N)$$

D'après le lemme 2.1 et le lemme 4.2 du chapitre I, la suite $\{\tilde{V}^N(v_N)\}$ est convergente en probabilité (la limite est d'ailleurs la même que celle de $V^N(v_N)$). Comme les dérivées de \tilde{k}_N calculées en v_N' , où v_N' un estimateur consistant de v , tend aussi en probabilité selon N vers une limite, on obtient d'après le théorème des valeurs moyennes et le fait que v_N est $N^{1/2}$ -consistant :

$$\sqrt{N} \tilde{k}^N(v_N) = \sqrt{N} \tilde{k}^N(v^*) + o(N)$$

où $o(N)$ est bornée en probabilité quand $N \rightarrow \infty$.

On en déduit alors que b_N^1 est un estimateur $N^{1/2}$ -consistant de b .
Son calcul ne fait intervenir que (voir (3.21) et (3.20)) :

$$\phi_N = \phi_0 + \sum_{j=1}^{\alpha} (a_N)_j \phi_j$$

$$\tilde{f}_N(\lambda) = \sum_{|u| < a} \tilde{R}^N(u) e^{-iu\lambda}$$

Ainsi on évite la résolution de (3.17).

D'après Hannan [14], on peut remplacer l'intégrale (3.20) par une somme de Riemann calculée au point $\lambda = 2\pi m/N$, $m = 1, \dots, N$. On peut montrer que l'erreur commise sur k_N est de l'ordre $N^{-1} o(N)$ où $o(N)$ est bornée en probabilité quand $N \rightarrow \infty$. Cette approximation ne détruit pas la propriété $N^{1/2}$ -consistance de b_N^1 .

CHAPITRE III

EXTENSION DE LA METHODE DU CHAPITRE I AU CAS DU MODELE DE REGRESSION A RESIDUS STATIONNAIRES

§ 1. LE MODELE

Nous allons étendre les résultats du chapitre I au modèle de régression du type :

$$(1.1) \quad X(t) = m(t) + Y(t)$$

où $\{Y(t), t \in \mathbb{Z}\}$ est un processus stationnaire centré appelé processus des résidus et où $m(t)$ dépend d'un paramètre vectoriel $\mu \in \mathbb{R}^k$ de la façon suivante :

$$(1.2) \quad m(t) = \sum_{j=0}^{\infty} B_{\mu}(j) Z(t-j)$$

$\{Z(t), t \in \mathbb{Z}\}$ étant une suite donnée de vecteurs de \mathbb{R}^q et $B_{\mu}(j)$ étant des matrices d'ordre (p, q) .

Nous supposons en plus que les résidus $Y(t), t \in \mathbb{Z}$ suivent le modèle général du chapitre I, § 2, à savoir :

$$(1.3) \quad Y(t) = \sum_{j=0}^{\infty} A_{\theta}(j) \varepsilon(t-j)$$

où les matrices $A(j)$ vérifient les hypothèses $H1$, $H2$ (ou $H'2$) et où $\varepsilon(t)$, $t \in \mathbb{Z}$ sont des vecteurs aléatoires indépendants centrés, de même loi ayant une matrice de covariance G .

Nous notons θ^*, μ^*, G^* les vraies valeurs de θ, μ, G . D'autre part ν désigne le couple (θ, G) et ω désigne le couple (ν, μ) ; leurs vraies valeurs sont notées ν^* et ω^* . Conformément aux notations du chapitre I, §2, nous posons :

$$(1.4) \quad h_{\theta}(z) = \sum_{j=0}^{\infty} A_{\theta}(j) z^j$$

$$(1.5) \quad h_{\theta}^{-1}(z) = \sum_{j=0}^{\infty} C_{\theta}(j) z^j$$

$$(1.6) \quad f_{\nu}(\lambda) = \frac{1}{2\pi} h_{\theta}(e^{i\lambda}) G h_{\theta}^*(e^{i\lambda})$$

Nous faisons les hypothèses suivantes relatives aux $B_{\mu}(j)$ et aux $Z(t)$. (Ces hypothèses assurent en particulier que la série (1.2) est convergente) :

H3 : La série :

$$(1.7) \quad g_{\mu}(z) = \sum_{j=0}^{\infty} B_{\mu}(j) z^j$$

est absolument convergente pour tout $\mu \in V(\mu^)$, un voisinage de μ^* , et la fonction $(\lambda, \mu) \rightarrow g_{\mu}(e^{i\lambda})$ est continue dans $-\pi, \pi \times V(\mu)$ et admet des dérivées partielles jusqu'à l'ordre 2 par rapport à θ , continues dans le même domaine.*

H4 : Les matrices $B_{\mu}(j)$ sont deux fois dérivables par rapport à μ et vérifient :

$$(1.8) \quad \sum j \|B_{\mu^*}(j)\| < +\infty$$

$$(1.9) \quad \sum j \|(\frac{\partial}{\partial \mu \ell} B_{\mu}(j))_{\mu=\mu^*}\| < +\infty$$

$$(1.10) \quad \sum_j j \left\| \left(\frac{\partial^2}{\partial \mu_\ell \partial \mu_h} B_\mu(j) \right)_{\mu=\mu^*} \right\| < +\infty$$

On a souvent besoin de l'hypothèse suivante, plus forte :

H4' : C'est H4 avec en plus :

$$(1.11) \quad \sum_j j \sup_{\mu \in V(\mu^*)} \left\| \frac{\partial^2}{\partial \mu_\ell \partial \mu_h} B_\mu(j) \right\| < +\infty$$

où $V(\mu^*)$ est un voisinage convenable de μ "

Remarque :

1.- L'hypothèse H2 implique (voir la remarque 2 du § 2, chapitre 1) :

$$(1.12) \quad \sum_j j \sup_{\mu \in V(\mu^*)} \left\| \frac{\partial}{\partial \mu_\ell} B_\mu(j) \right\| < +\infty$$

$$(1.13) \quad \sum_j j \sup_{\mu \in V(\mu^*)} \|B_\mu(j)\| < +\infty$$

2.- Les hypothèses H3 et H4' sont vérifiées si $g_\mu(z)$ est une fraction rationnelle en z , dont les coefficients sont des fonctions deux fois continûment dérivables de μ et dont les pôles sont de module strictement supérieur à l'unité.

Finalement la série $Z(t)$ sera supposée vérifier une condition analogue de celle de Grenander et Rosenblatt [12].

H5 : Il existe une constante A telle que $\|Z(t)\| < A$, $t \in \mathbb{Z}$. D'autre part quand $N \rightarrow \infty$, la matrice :

$$\frac{1}{N} \sum_{t=1}^N Z(t)^+ Z(t+u)$$

converge vers une limite notée $\rho(u)$ "

§ 2. LES APPROXIMATIONS DE LA FONCTION DE VRAISEMBLANCE

Soient X (resp. Y) le vecteur aléatoire de R^{pN} de composantes $X_j(t)$ (resp. $Y_j(t)$), $j = 1, \dots, p$, $t = 1, \dots, N$ et m le vecteur de R^{pN} de composantes $m_j(t)$, $j = 1, \dots, p$, $t = 1, \dots, N$. Si P_ω^X (resp. P_ν^Y) désigne la loi de probabilité de X (resp. Y) et λ désigne la mesure de Lebesgue, on a :

$$\left(\frac{dP^X}{d\lambda}\right)(x) = \left(\frac{dP^Y}{d\lambda}\right)(x-m) \quad x \in R^{pN}$$

On obtient donc d'après le paragraphe 3 du chapitre I deux approximations du logarithme de la fonction de vraisemblance :

$$(2.1) \quad L'_N(\omega) = \frac{N}{2} \log \det (2\pi G) - \frac{1}{2} \sum_{t=1}^N \varepsilon'(t, \theta, \mu) G^{-1} \varepsilon'(t, \theta, \mu)$$

où :

$$(2.2) \quad \varepsilon'(t, \theta, \mu) = \sum_{j=0}^{t-1} C_\theta(j) [X(t-j) - \sum_{k=0}^{\infty} B_\mu(k) Z(t-j-k)] \quad ; 1 < t < N$$

et :

$$(2.3) \quad \tilde{L}'_N(\omega) = -\frac{N}{2} \log \det (2\pi G) - \frac{1}{2} \sum_{t=1}^{\infty} \varepsilon'(t, \theta, \mu) G^{-1} \varepsilon'(t, \theta, \mu)$$

où :

$$(2.4) \quad \varepsilon'(t, \theta, \mu) = \sum_{j=t-N}^{t-1} C_\theta(j) [X(t-j) - \sum_{k=0}^{\infty} B_\mu(k) Z(t-j-k)] \quad ; t > N$$

Les expressions (2.2) et (2.4) font intervenir les valeurs de $Z(\tau)$ pour tout τ . On dispose seulement en général des observations de $Z(t)$ pour $1 \leq t \leq N$. Nous sommes alors amenés à remplacer chacune de ces expressions par :

$$(2.5) \quad \varepsilon(t, \theta, \mu) = \sum_{j=0}^{\infty} C_\theta(j) [X_N(t-j) - \sum_{k=0}^{\infty} B_\mu(k) Z_N(t-j-k)]$$

où $X_N(t) = X(t)$ si $1 < t < N$, $X_N(t) = 0$ sinon et où

$$Z_N(t) = Z(t) \text{ si } 1 < t < N, Z_N(t) = 0 \text{ sinon}$$

Ainsi nous obtenons les approximations suivantes du logarithme de la fonction de vraisemblance :

$$(2.6) \quad L_N(\omega) = -\frac{N}{2} \text{Log det } (2\pi G) - \frac{1}{2} \sum_{t=1}^N \epsilon(t, \theta, \mu) G^{-1}(t, \theta, \mu)$$

$$(2.7) \quad \tilde{L}_N(\omega) = -\frac{N}{2} \text{Log } (\text{det } 2\pi G) - \frac{1}{2} \sum_{t=1}^{\infty} \epsilon(t, \theta, \mu) G^{-1}(t, \theta, \mu)$$

La fonction \tilde{L}_N admet une deuxième écriture :

$$(2.8) \quad L_N(\omega) = -\frac{N}{2} \text{Log } (\text{det } 2\pi G) - \frac{N}{4\pi} \int_{-\pi}^{\pi} \text{Tr}[\varphi_{\omega}(\lambda) J^N(\lambda)] d\lambda$$

où :

$$(2.9) \quad J^N(\lambda) = \frac{1}{2\pi N} \begin{pmatrix} \tilde{X}_N(\lambda) \cdot \tilde{X}_N^*(\lambda) & \tilde{X}_N(\lambda) \cdot \tilde{Z}_N^*(\lambda) \\ \tilde{Z}_N(\lambda) \cdot \tilde{X}_N^*(\lambda) & \tilde{Z}_N(\lambda) \cdot \tilde{Z}_N^*(\lambda) \end{pmatrix}$$

$$(2.10) \quad \tilde{X}_N(\lambda) = \sum_{t=1}^N e^{i\lambda t} X(t)$$

$$(2.11) \quad \tilde{Z}_N(\lambda) = \sum_{t=1}^N e^{i\lambda t} Z(t)$$

et où :

$$(2.12) \quad \varphi_{\omega}(X) = \begin{pmatrix} f_v^{-1}(\lambda) & f_v^{-1}(\lambda) g_{\mu}(e^{i\lambda}) \\ g_{\mu}^*(e^{i\lambda}) f_v^{-1}(\lambda) & g_{\mu}^*(e^{i\lambda}) f_v^{-1}(\lambda) g_{\mu}(e^{i\lambda}) \end{pmatrix}$$

En effet, la série de Fourier $\sum_{t=1}^N \epsilon(t, \theta, \mu) e^{i\lambda t} = \tilde{\epsilon}(\lambda; \theta, \mu)$ est :

$$\begin{aligned} \tilde{\epsilon}(\lambda; \theta, \mu) &= h_{\theta}^{-1}(e^{i\lambda}) \tilde{X}_N(\lambda) + h_{\theta}^{-1}(e^{i\lambda}) g_{\mu}(e^{i\lambda}) \tilde{Z}_N(\lambda) \\ &= \begin{pmatrix} h_{\theta}^{-1}(e^{i\lambda}) & h_{\theta}^{-1}(e^{i\lambda}) g_{\mu}(e^{i\lambda}) \end{pmatrix} \cdot \begin{pmatrix} \tilde{X}_N(\lambda) \\ \tilde{Z}_N(\lambda) \end{pmatrix} \end{aligned}$$

On en déduit :

$$\begin{aligned} \sum_{t=1}^{\infty} {}^t \varepsilon(t, \theta, \mu) G^{-1} \varepsilon(t, \theta, \mu) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{\varepsilon}^*(\lambda; \theta, \mu) G^{-1} \tilde{\varepsilon}(\lambda; \theta, \mu) d\lambda \\ &= \frac{N}{2\pi} \int_{-\pi}^{\pi} \text{Tr}[\varphi_{\omega}(\lambda) J^N(\lambda)] d\lambda \end{aligned}$$

Nous montrons maintenant que L_N (resp. \tilde{L}_N) ne diffère de L'_N (resp. \tilde{L}'_N) que par une quantité négligeable.

Lemme 2.1.

"Soient

$$D_{\theta, \mu}(\ell) = \sum_{j=0}^{\ell} |C_{\theta}(j)| |B_{\mu}(\ell-j)| \quad \ell = 0, 1, \dots$$

alors :

$$\begin{aligned} \left| \varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) \right| &< \sum_{\ell=t}^{\infty} D_{\theta, \mu}(\ell) |Z(t-\ell)| \quad 1 < t < N \\ &< \sum_{\ell=t-N}^{\infty} D_{\theta, \mu}(\ell) |Z(t-\ell)| \quad t > N \end{aligned}$$

Démonstration

Si $1 < t < N$:

$$\varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) = \sum_{j=0}^{t-1} C_{\theta}(j) \sum_{k=t-j}^{\infty} B_{\mu}(k) Z(t-j-k)$$

En posant $\ell = k-j$, il vient :

$$\varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) = \sum_{\ell=t}^{\infty} \sum_{j=0}^{t-1} [C_{\theta}(j) B_{\mu}(\ell-j)] \cdot Z(t-\ell)$$

D'où :

$$\left| \varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) \right| < \sum_{\ell=t}^{\infty} D_{\theta, \mu}(\ell) |Z(t-\ell)|$$

Si $t > N$:

$$\begin{aligned} \varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) &= \sum_{j=t-N}^{t-1} C_{\theta}(j) \left[\sum_{k=t-j}^{t-1} B_{\mu}(k) Z(t-j-k) \right] \\ &\quad - \sum_{j=0}^{t-N-1} C_{\theta}(j) \left[\sum_{k=t-N-j}^{t-1-j} B_{\mu}(k) Z(t-j-k) \right] \end{aligned}$$

En posant $\ell = k+j$, il vient :

$$\begin{aligned} \varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) &= \sum_{\ell=t}^{\infty} \left[\sum_{j=t-N}^{t-1} C_{\theta}(j) B_{\mu}(\ell-j) \right] Z(t-\ell) \\ &\quad - \sum_{\ell=t-N}^{t-1} \left[\sum_{j=0}^{t-N-1} C_{\theta}(j) B_{\mu}(\ell-j) \right] Z(t-\ell) \end{aligned}$$

D'où

$$\left| \varepsilon(t, \theta, \mu) - \varepsilon'(t, \theta, \mu) \right| < \sum_{\ell=t-N}^{\infty} D_{\theta, \mu}(\ell) \left| Z(t-\ell) \right|$$

Lemme 2.2.

Si $B(j)$ et $C(j)$, $j = 0, 1, \dots$, vérifient :

$$\sum_{j=1}^{\infty} j \|B(j)\| < +\infty \quad ; \quad \sum_{j=1}^{\infty} j \|C(j)\| < +\infty$$

alors les $D(j)$ définis par :

$$D(j) = \sum_{k=0}^j \|C(k)\| \cdot \|B(j-k)\|$$

vérifient aussi :

$$\sum_{j=0}^{\infty} j D(j) < +\infty "$$

Démonstration

On a :

$$\begin{aligned} \sum_{j=1}^{\infty} j D(j) &= \sum_{j=0}^{\infty} \sum_{k=0}^j j \|B(j-k)\| \|C(k)\| \\ &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} (k+\ell) \|B(\ell)\| \|C(k)\| \end{aligned}$$

Comme $k+l \leq 2k.l$ si $k \geq 1, l \geq 1$, l'expression précédente est majorée par :

$$||B(0)|| \cdot \sum_{k=1}^{\infty} k ||C(k)|| + ||C(0)|| \cdot \sum_{l=1}^{\infty} ||B(l)|| + \left[\sum_{k=1}^{\infty} k ||B(k)|| \right] \left[\sum_{l=1}^{\infty} l ||B(l)|| \right]$$

D'où le lemme.

Le théorème suivant se démontre d'une manière analogue à celle dont on démontre les théorèmes 3.2 et 3.3 du chapitre 1, compte-tenu des lemmes 3.1 et 3.2 du même chapitre.

Théorème 2.1.

"Si $\mathcal{L}_N(\omega)$ est le logarithme de la fonction de vraisemblance du modèle, alors la différence $\delta_N(\omega)$ entre $L_N(\omega)$ (resp. $\tilde{L}_N(\omega)$) et $\mathcal{L}_N(\omega)$ est deux fois dérivable par rapport à ω et ses dérivées, calculées en ω^* ainsi $\delta_N(\omega^*)$ sont bornées en probabilité quand $N \rightarrow \infty$.

Si les hypothèses H2' et H4' sont satisfaites, il existe un voisinage $V(\omega^*)$ de ω^* tel que les variables aléatoires :

$$\sup_{\omega \in V(\omega^*)} |\tilde{\delta}_N(\omega)|, \quad \sup_{\omega \in V(\omega^*)} \left| \frac{\partial}{\partial \omega_j} \tilde{\delta}_N(\omega) \right|, \quad \sup_{\omega \in V(\omega^*)} \left| \frac{\partial^2}{\partial \omega_j \partial \omega_l} \tilde{\delta}_N(\omega) \right|$$

où $\tilde{\delta}_N(\omega) = \tilde{L}_N(\omega) - L_N(\omega)$, sont bornées en probabilité quand $N \rightarrow \infty$."

§ 3. PROPRIÉTÉS ASYMPTOTIQUES DE LA FONCTION DE VRAISEMBLANCE

Les propriétés asymptotiques de la fonction L_N (resp. \tilde{L}_N) du chapitre I, paragraphe 4 s'étendent dans ce cas. Pour les obtenir, nous commençons par remplacer, dans l'expression de L_N , $\varepsilon(t, \theta, \mu)$ par :

$$\bar{\varepsilon}(t, \theta, \mu) = \sum_{j=0}^{\infty} C_{\theta}(j) X(t-j) - \sum_{k=0}^{\infty} B_{\mu}(k) Z(t-j-k)$$

Le même raisonnement que celui du chapitre I montre que l'erreur commise, ainsi que ses dérivées premières et secondes calculées en ω^* sont toutes bornées en probabilité quand $N \rightarrow \infty$.

Lemme 3.1.

"Soit $\tilde{Z}_N(\lambda) = \sum_{t=1}^N e^{i\lambda t} Z(t)$ alors la mesure $(2\pi N)^{-1} \tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda) d\lambda$

converge faiblement quand $N \rightarrow \infty$ vers la mesure $dM(\lambda)$ définie de manière unique par (voir l'hypothèse H5) :

$$\rho(u) = \int_{-\pi}^{\pi} e^{iu\lambda} dM(\lambda) "$$

Démonstration

On a :

$$\tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda) = \sum_{|u| < N} \sum_{\substack{1 < t < N \\ 1 < t+u < N}} Z(t)^{\dagger} Z(t+u) e^{-i\lambda u}$$

On voit alors que les coefficients de Fourier de $(2\pi N)^{-1} \tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda)$ convergent vers $\rho(u)$ (d'après l'hypothèse H5).

Pour tout $\alpha \in \mathbb{C}^q$, le théorème de Lévy montre que la mesure positive :

$$(2\pi N)^{-1} \alpha^* \tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda) \alpha d\lambda$$

converge faiblement quand $N \rightarrow \infty$ vers la mesure $dM_{\alpha}(\lambda)$ définie par :

$$\alpha^* \rho(u) \alpha = \int_{-\pi}^{\pi} e^{i\lambda u} dM_{\alpha}(\lambda)$$

Or un calcul élémentaire montre que pour toute matrice A :

$$(\alpha+\beta)^* A(\alpha+\beta) + (\alpha-\beta)^* A(\alpha-\beta) - i[(\alpha+i\beta)^* A(\alpha+i\beta) - (\alpha-i\beta)^* A(\alpha-i\beta)] = 4\alpha^* A \beta$$

Par suite la mesure complexe :

$$(2\pi N)^{-1} \alpha^* \tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda) \beta d\lambda$$

converge faiblement quand $N \rightarrow \infty$ vers la mesure $dM_{\alpha\beta}(\lambda)$:

$$\alpha^* \rho(u) \beta = \int_{-\pi}^{\pi} e^{i\lambda u} dM_{\alpha\beta}(\lambda)$$

En choisissant α et β deux vecteurs quelconques de la base canonique de \mathbb{R}^q on obtient le lemme.

Nous pouvons maintenant étendre le théorème 4.1 du chapitre I en

Théorème 3.1.

"Quand $N \rightarrow \infty$, le vecteur des dérivées premières et la matrice des dérivées secondes de $N^{-1} L_N$ (resp. $N^{-1} \tilde{L}_N$, resp. $N^{-1} \mathcal{L}_N$) calculés en ω convergent en probabilité respectivement vers 0 et vers

$$-K = - \begin{pmatrix} S & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & H \end{pmatrix}$$

où S et H sont définies dans le théorème 4.1 du chapitre I et où T est la matrice de terme général :

$$(3.1) \quad T_{j\ell} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{Tr} f_v^{-1}(\lambda) \left(\frac{\partial}{\partial \mu_j} g_{\mu}(e^{i\lambda}) \right)_{\mu=\mu^*} dM(\lambda) \left(\frac{\partial}{\partial \mu_{\ell}} g_{\mu}(e^{i\lambda}) \right)_{\mu=\mu^*} d\lambda "$$

Démonstration

La convergence des dérivées de $N^{-1} L_N$ par rapport à v se démontre d'une manière analogue à celle dont on démontre le théorème 4.1. du chapitre I. Maintenant :

$$N^{-1} \left(\frac{\partial L_N}{\partial \mu_j} \right) (\omega^*) = N^{-1} \left[\sum_{t=1}^N t_{\varepsilon}(t) W_j(t) + O(N) \right]$$

où $O(N)$ désigne une variable aléatoire bornée en probabilité quand $N \rightarrow \infty$ et où :

$$W_j(t) = G^{-1} \sum_{k=0}^{\infty} \left[\sum_{u=0}^k C_{\theta}(u) \left(\frac{\partial}{\partial \mu_j} \right) B_{\mu}(k-u) \right]_{\mu=\mu^*} Z(t-k)$$

De la même manière :

$$N^{-1} \left(\frac{\partial^2 L_N}{\partial \mu_j \partial \theta_{\ell}} \right) (\omega^*) = N^{-1} \left[\sum_{t=1}^N t_{\varepsilon}(t) W_{j\ell}(t) + \sum_{t=1}^N t_{V_{\ell}}(t) W_j(t) + O(N) \right]$$

où :

$$W_j(t) = G^{-1} \left[\sum_{k=0}^{\infty} \sum_{u=0}^k \left(\frac{\partial}{\partial \theta_{\ell}} C_{\theta}(u) \right)_{\theta=\theta} \left(\frac{\partial}{\partial \mu_j} B_{\mu}(k-u) \right) \right]_{\mu=\mu^*} Z(t-k)$$

$$V_{\ell}(t) = \sum_{k=0}^{\infty} \left(\frac{\partial}{\partial \theta_{\ell}} C_{\theta}(k) \right)_{\theta=\theta^*} \left[X(t) - \sum_{u=0}^{\infty} B_{\theta}(u) Z(t-k-u) \right]$$

On peut montrer à l'aide du lemme 2.2 que $W_j(t)$ et $W_{j\ell}(t)$ sont bornés pour tout t . Par suite :

$$N^{-1} \sum_{t=1}^N t_{\varepsilon}(t) W_j(t) \quad \text{et} \quad N^{-1} \sum_{t=1}^N t_{\varepsilon}(t) W_{j\ell}(t)$$

convergent vers 0 en moyenne quadratique quand $N \rightarrow \infty$. D'où la convergence en probabilité de $N^{-1} \left(\frac{\partial L_N}{\partial \mu_j} \right) (\omega^*)$ vers 0 quand $N \rightarrow \infty$.

Calculons maintenant :

$$E \left[N^{-1} \sum_{t=1}^N t_{V_{\ell}}(t) W_j(t) \right]^2 = N^{-2} \sum_{s,t} t_{W_{\ell}}(s) E \left[t_{V_{\ell}}(s) V_j(t) \right] W_j(t)$$

Soit $\varphi_{\ell}(\lambda)$ la densité spectrale du processus stationnaire $V_j(t)$; l'expression précédente s'écrit :

$$N^{-2} \int_{-\pi}^{\pi} \sum_{s,t} t_{W_j}(s) \varphi_{\ell}(\lambda) e^{i\lambda(t-s)} W_j(t) d\lambda =$$

$$N^{-2} \int_{-\pi}^{\pi} \left[\sum_s W_j(s) e^{i\lambda s} \right] \varphi_{\ell}(\lambda) \left[\sum_t W_{\ell}(t) e^{i\lambda t} \right] d\lambda$$

Comme est continue, $||\varphi(\lambda)|| < K$, $\forall \lambda \in [-\pi, \pi]$ et l'expression précédente est majorée par :

$$N^{-2} K \int_{-\pi}^{\pi} ||\sum_{t=1}^N W_j(t) e^{i\lambda t}||^2 d\lambda = 2\pi K (N^{-2} \sum_{t=1}^N ||W_j(t)||^2)$$

et converge donc vers 0 quand $N \rightarrow \infty$. D'où la convergence en probabilité de $N^{-1} (\partial^2 L_N / \partial \mu_j \partial \theta_\ell)$ (ω^*) vers 0 quand $N \rightarrow \infty$.

D'autre part :

$$N^{-1} \left(\frac{\partial^2 L_N}{\partial \mu_j \partial \mu_\ell} \right) (\omega) = N^{-1} \left[\sum_{t=1}^N t \varepsilon(t) W'_{j\ell}(t) + \sum_{t=1}^N t W_\ell(t) G W_j(t) + O(N) \right]$$

où :

$$W'_{j\ell}(t) = G^{*-1} \sum_{k=0}^{\infty} \left[\sum_{u=0}^k C_\theta(k) \cdot \left(\frac{\partial^2}{\partial \mu_j \partial \mu_\ell} \right) B_{\mu}(k-u) \right]_{\mu=\mu^*} Z(t-k)$$

Le même raisonnement que précédemment montre que le premier terme du second membre de l'égalité précédente converge en moyenne quadratique vers 0 quand $N \rightarrow \infty$.

Considérons maintenant $W'_j(t)$ défini de la même façon que $W_j(t)$ avec $Z(t)$ remplacé par $Z_N(t)$ où $Z_N(t) = Z(t)$ si $1 < t < N$ et $Z_N(t) = 0$ sinon. Alors :

$$\tilde{W}'_j(\lambda) = \sum_{t=1}^{\infty} W'_j(t) e^{i\lambda t} = G^{*-1} h^{-1}(e^{i\lambda}) \left(\frac{\partial}{\partial \mu_j} g_\mu(e^{i\lambda}) \right)_{\mu=\mu^*} \tilde{Z}_N(\lambda)$$

où $\tilde{Z}_N(\lambda)$ est défini dans le lemme 3.1. D'après ce lemme :

$$N^{-1} \sum_{t=1}^{\infty} t W'_\ell(t) G W'_j(t) = (2\pi N)^{-1} \int_{-\pi}^{\pi} \tilde{W}_\ell^*(\lambda) G \tilde{W}_j(\lambda) d\lambda$$

converge quand $N \rightarrow \infty$ vers :

$$\text{Tr} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu_j} g_\mu(e^{i\lambda}) \right)_{\mu=\mu^*} h_{\theta^*}^{*-1}(e^{i\lambda}) G^{*-1} h_{\theta^*}^{-1}(e^{i\lambda}) \left(\frac{\partial}{\partial \mu_\ell} g_\mu(e^{i\lambda}) \right)_{\mu=\mu^*} dM(\lambda)$$

En utilisant les lemmes 2.2 et 3.2 du chapitre I et une démonstration analogue à celle des théorèmes 3.2 et 3.3 du chapitre I, on peut montrer que la différence :

$$\sum_{t=1}^N t_{W_\ell}(t) G W_j(t) - \sum_{t=1}^{\infty} t_{W'_\ell}(t) G W'_j(t)$$

est bornée en probabilité quand $N \rightarrow \infty$. D'où le théorème.

Théorème 3.2.

"Les conclusions du théorème 3.1 relatives à la fonction \tilde{L}_N restent vraies quand ω^* est remplacé par son estimateur consistant. Si les hypothèses H2' et H4' sont satisfaites, ce résultat s'applique aussi pour la fonction L_N ".

Le théorème précédent est une conséquence directe du lemme suivant dont la démonstration est analogue à celle du lemme 4.2 du chapitre 1.

Lemme 3.2.

"Soit la variable aléatoire :

$$Z_\omega^N = \int_{-\pi}^{\pi} \text{Tr}[p_\omega(\lambda) J^N(\lambda)] d\lambda$$

où $(\omega, \lambda) \rightarrow p_\omega(\lambda)$ est une fonction matricielle continue sur $V(\omega) \times [-\pi, \pi]$ $V(\omega)$ étant un voisinage de ω . Alors :

$$\forall \varepsilon > 0, \exists \eta > 0 : \sup_{\omega: ||\omega - \omega^*|| < \eta} |Z_\omega^N - Z_{\omega^*}^N| < \varepsilon M_N$$

où M_N tend en probabilité vers une constante quand $N \rightarrow \infty$."

Le théorème suivant est une extension du théorème 4.3 du chapitre 1.

Théorème 3.3.

"Le vecteur des dérivées de $N^{-1/2} L_N$ (resp. $N^{-1/2} \tilde{L}_N$, resp. $N^{-1/2} \mathcal{L}_N$) par rapport à (θ, μ) , calculé en ω^* converge en loi selon N vers la loi gaussienne centrée de matrice de covariance :

$$\begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix}$$

Si les $\varepsilon(t)$ possèdent des moments d'ordre 4, le vecteur des dérivées de $N^{-1/2} L_N$ (resp. $N^{-1/2} \tilde{L}_N$, resp. $N^{-1/2} \mathcal{L}_N$) calculé en ω converge en loi selon N vers la loi gaussienne centrée de matrice de covariance :

$$\begin{pmatrix} S & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & \tilde{H} \end{pmatrix}$$

où $\tilde{H} = H$ si les cumulants d'ordre 4 de $\varepsilon(t)$ sont nuls".

Démonstration

Nous allons montrer que toute combinaison linéaire des dérivées de $N^{-1/2} L_N$ par rapport à θ et à μ (resp. à θ , à G et à μ) est asymptotiquement normale.

Considérons d'abord une combinaison linéaire des dérivées de $N^{-1/2}$ par rapport à μ . Elle s'écrit d'après la démonstration du théorème 3.1.

$$N^{-1/2} \sum_{t=1}^N \varepsilon(t) W(t) + N^{-1/2} O(N) \quad ; \quad W(t) = \sum C_j W_j(t)$$

où $O(N)$ est borné en probabilité et où

$$W_j(t) = G^{-1} \sum_{k=0}^{\infty} \left[\sum_{u=0}^k C_{\theta}(u) \left(\frac{\partial}{\partial \mu_j} B_{\mu}(k-u)_{\mu=\mu^*} \right) \right] Z(t-k)$$

On peut appliquer le théorème central limite à la première somme de l'expression précédente si la condition de Lindeberg (voir par exemple [5], page 42) est vérifiée, i.e. :

$$\frac{1}{s_N^2} \sum_{t=1}^N E \left[\eta^2(t) \mathbb{1}_{\{|\eta(t)| > \varepsilon s_N\}} \right] \rightarrow 0 \quad N \rightarrow \infty$$

où $\eta(t) = \sum_{t=1}^t \varepsilon(t) W(t)$ et où $s_N = \sigma(\eta(1) + \dots + \eta(N))$

Comme $|\eta(t)| \leq ||W(t)|| |\epsilon(t)| \leq m_N s_N \cdot ||\epsilon(t)||$ où

$m_N = \text{Max}\{||W(t)|| / s_N, t = 1, \dots, N\}$ on a :

$$\frac{1}{s_N^2} \sum_{t=1}^N E[n^2(t) \mathbb{1}\{|n(t)| > \epsilon s_N\}] \leq$$

$$\left[\frac{1}{s_N^2} \sum_{t=1}^N ||W(t)||^2 \right] E[||\epsilon(t)||^2 \mathbb{1}\{||\epsilon(t)|| > \epsilon/m_N\}]$$

Or $s_N^2 = \sum_{t=1}^N t_{W(t)G} W(t)$; le premier crochet du second membre est donc

borné quand $N \rightarrow \infty$. D'autre part $m_N \rightarrow 0$ quand $N \rightarrow \infty$ car $W(t)$ reste borné pour tout t et $N^{-1} s_N^2$ tend vers $\sum C_j C_{j\ell} T_{j\ell}$ quand $N \rightarrow \infty$ (voir la démonstration du théorème 3.1). La condition de Lindeberg est satisfaite. Le vecteur des dérivées de $N^{-1} L_N$ par rapport à μ converge donc en loi vers la loi gaussienne centrée de matrice de covariance T .

Maintenant en se rapportant aux calculs de la démonstration du théorème 4.2 du chapitre 1, nous voyons que toute combinaison linéaire des dérivées de $N^{-1/2} L_N$ par rapport à θ et μ peut s'écrire :

$$N^{-1/2} \sum_{t=1}^N \eta(t) + N^{-1/2} \sum_{t=1}^N t_{\epsilon(t)G}^{-1} V(t) + N^{-1/2} O(N)$$

où $V(t) = \sum_j C_j' V_j(t)$ et :

$$V_j(t) = \sum_{k=1}^{\infty} \frac{\partial}{\partial \theta_j} C_{\theta}(k)_{\theta=\theta^*} \left[X(t-k) - \sum_{u=0}^{\infty} B_{\mu}(u) Z(t-k-u) \right]$$

La suite $\eta(t) + t_{\epsilon(t)G}^{-1} V(t)$ n'est ni à termes indépendants ni stationnaire. Nous sommes amenés à poser :

$$\xi_m(t) = t_{\epsilon(t)G}^{-1} V^{(m)}(t) ; V^{(m)}(t) = \sum_j C_j' V_j^{(m)}$$

où :

$$V_j^{(m)}(t) = \sum_{k=0}^m \frac{\partial}{\partial \theta_j} C_{\theta}(k)_{\theta=\theta} \left[X(t-k) - \sum_{u=0}^{\infty} B_{\mu}(u) Z(t-k-u) \right]$$

Les processus $\xi_m(t)$ est m-indépendant, c'est-à-dire $\xi_m(t)$ est indépendant de $\xi_m(s)$ si $|t-s| > m$. D'après un théorème de Orey (voir Rozen [26])

le théorème central limite s'applique à la somme $N^{-1/2} \sum_{t=1}^N [\eta(t) + \xi_m(t)]$ si :

(i) La condition de Linderberg est vérifiée

(ii) $\overline{\lim}_{N \rightarrow \infty} \sum_{t=1}^N \sigma^2[\eta(t) + \xi_m(t)] / s_N^2 < +\infty$ où s_N^2 est la variance

de $[\eta(1) + \xi_m(1)] + [\eta(2) + \xi_m(2)] + \dots + [\eta(N) + \xi_m(N)]$.

On montre (i) en utilisant le fait que la condition de Lindeberg est vérifiée pour les $\eta(t)$ et pour les $\xi_m(t)$ et en se servant du lemme suivant de Rozen [26]:

"Si T_1, \dots, T_k sont des variables aléatoires et $\alpha > 0$ alors :

$$E \left[\left(T_1 + \dots + T_k \right)^2 \mathbf{1}_{\{|T_1 + \dots + T_k| > \alpha\}} \right] < 2k^2 \sum_{n=1}^k E \left[T_n^2 \mathbf{1}_{\{|T_n| > \alpha\}} \right]$$

Quant au point (ii), il vient du fait que $(\eta(t), \xi_m(t))$ est non corrélé avec $(\eta(s), \xi_m(s))$, $t \neq s$ ce qui entraîne :

$$s_N^2 = \sum_{t=1}^N \sigma^2[\eta(t) + \xi_m(t)]$$

Il en découle que $N^{-1/2} \sum_{t=1}^N \eta(t) + \xi_m(t)$ converge en loi selon N

vers la loi normale centrée de variance $s^2 = \lim_{N \rightarrow \infty} N^{-1} s_N^2$. Un calcul

élémentaire montre que :

$$s^2 = \sum_{j,l} C_j C_l T_{jl} + \sum_{j,l} C_j' C_l' S_{jl}^m$$

où :

$$S_{jl}^m = E \left[{}^t V_j^{(m)}(t) G^{*-1} V_l^{(m)}(t) \right]$$

Quand $m \rightarrow \infty$, $V_j^{(m)}(t)$ converge en moyenne quadratique vers $V_j(t)$ et par suite $S_{j\ell}^m$ converge vers $S_{j\ell}$. Nous appliquons maintenant le lemme de Bernstein (Anderson [3] p. 425).

"Soit $Y_N = Y_{N,m} + \delta_{N,m}$ où $\delta_{N,m}$ converge vers 0 en probabilité quand $m \rightarrow \infty$, uniformément par rapport à N , i.e. :

$$\forall \alpha > 0, \exists \epsilon > 0 \quad \exists m_0 : m \geq m_0 \Rightarrow P\{|\delta_{N,m}| > \alpha\} < \epsilon$$

Si $Y_{N,m}$ converge en loi quand $N \rightarrow \infty$ vers la loi normale centrée de variance σ_m^2 tendant vers σ^2 quand $m \rightarrow \infty$, alors Y_N converge en loi selon N vers la loi normale de centrée de variance σ^2 ."

Il suffit donc de vérifier que $\delta_{N,m} = N^{-1/2} \sum_{t=1}^N \xi_m'(t)$ où

$$\xi_m'(t) = t_{\epsilon}(t) G^{-1} \sum_j C_j' [V_j(t) - V_j^{(m)}(t)]$$

converge vers 0 en probabilité quand $N \rightarrow +\infty$ uniformément par rapport à N . Or un calcul élémentaire montre que $E[\delta_{N,m}]^2 =$

$$\sum_{j,\ell} C_j' C_\ell' E\{[V_j(t) - V_j^{(m)}(t)] G^{-1} [V_\ell(t) - V_\ell^{(m)}(t)]\}. \text{ D'où le résultat.}$$

Dans le cas où les moments d'ordre 4 de $\epsilon(t)$ existe, la même démonstration que la précédente montre que toute combinaison linéaire des dérivées de $N^{-1/2} L_N$ par rapport à θ , à μ et à G est asymptotiquement normale. D'où le théorème.

§ 4. CONSTRUCTION DE L'ESTIMATEUR

La méthode du chapitre I, paragraphe 5 reste valable; en particulier on a le résultat suivant :

"Soit ω_N un estimateur $N^{-1/2}$ -consistant de ω , K_N un estimateur consistant de K (K étant défini dans le théorème 3.1) et $k_N(\omega)$ le vecteur des dérivées de $N^{-1} L_N$, alors l'estimateur :

$$\tilde{\omega}_N = \omega_N - K_N^{-1} k_N(\omega_N)$$

vérifie : $\sqrt{N}(\tilde{\omega}_N - \omega^*) \sim -\sqrt{N} K^{-1} \cdot k_N(\omega^*)$. Si les cumulants d'ordre 4 de $\varepsilon(t)$ sont nuls, alors $\sqrt{N} \tilde{\omega}_N \sim \sqrt{N} \hat{\omega}_N$ où $\hat{\omega}_N$ est l'estimateur consistant de maximum de vraisemblance de ω .

A cause de la forme particulière de K , et le fait que pour (θ, μ) fixé L_N est maximisé quand :

$$\begin{aligned} (4.1) \quad G = \hat{G}(\theta, \mu) &= \frac{1}{N} \sum_{t=1}^{\infty} \varepsilon(t, \theta, \omega)^\dagger \varepsilon(t, \theta, \mu) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_{\theta, \mu}(\lambda) J^N(\lambda) \psi_{\theta, \mu}^*(\lambda) d\lambda \end{aligned}$$

où

$$(4.2) \quad \psi_{\theta, \mu}(\lambda) = (h_{\theta}^{-1}(e^{i\lambda}) \quad h_{\theta}^{-1}(e^{i\lambda}) \cdot g_{\mu}(e^{i\lambda}))$$

on est amené à considérer l'algorithme suivant pour construire les estimateurs de θ, μ, G :

$$(4.3) \quad \theta^{(n+1)} = \theta^{(n)} - [S_N^{(n)}]^{-1} s_N(\theta^{(n)}, \mu^{(n)}, G^{(n)})$$

$$(4.4) \quad \mu^{(n+1)} = \mu^{(n)} - [T_N^{(n)}]^{-1} t_N(\theta^{(n)}, \mu^{(n)}, G^{(n)})$$

$$(4.5) \quad G^{(n+1)} = \hat{G}(\theta^{(n)}, \mu^{(n)})$$

où $S_N^{(n)}$, $T_N^{(n)}$ et $s_N(\theta, \mu, G)$, $t_N(\theta, \mu, G)$ sont respectivement des estimateurs consistants de S , T et les vecteurs des dérivées de $N^{-1} L_N$ par rapport à θ et à μ .

On montre de la même manière que pour le théorème 5.2 du chapitre I que si $(\theta^{(0)}, \mu^{(0)})$ est un estimateur $N^{1/2}$ consistant de (θ, μ) et $G^{(0)}$ un estimateur consistant de G alors :

$$\sqrt{N}(\theta^{(1)} - \theta^*) \sim -\sqrt{N} S^{-1} s_N(\theta^*, \mu^*, G^*)$$

$$\sqrt{N}(\mu^{(1)} - \mu^*) \sim -\sqrt{N} T^{-1} t_N(\theta^*, \mu^*, G^*)$$

et $G^{(1)}$ est consistant. D'autre part si $(\hat{\theta}, \hat{\mu}, \hat{G})$ est l'estimateur consistant de maximum de vraisemblance de (θ, μ, G) alors :

$$\sqrt{N} \theta^{(1)} \sim \sqrt{N} \hat{\theta} ; \quad \sqrt{N} \mu^{(1)} \sim \sqrt{N} \hat{\mu} ; \quad \sqrt{N} G^{(1)} \sim \sqrt{N} \hat{G}$$

L'interprétation de (4.3)-(4.5) est clair : Pour estimer les coefficients de régression, on suppose que le processus résiduel :

$$Y(t) = \sum_{j=0}^{\infty} A_{\theta}(j) \varepsilon(t-j)$$

a une structure de covariance connue, déterminée à l'aide des estimateurs $\theta^{(n)}$, $G^{(n)}$. De la même façon pour estimer les coefficients de structure θ, G du processus, on suppose que les coefficients de régression sont connus et donnés par $\mu^{(n)}$. On travaille alors sur le processus résiduel.

Le problème d'estimation simultanée des coefficients de régression et des coefficients de structure se décompose donc en deux problèmes : estimation des coefficients de régression, la structure de covariance du processus étant connue et estimation des coefficients de structure de covariance d'un processus centré.

§ 5. SUR L'ESTIMATION DES COEFFICIENTS DE REGRESSION

Considérons le modèle :

$$X(t) = B Z(t) + Y(t)$$

où $Y(t)$ est un processus centré donné par :

$$Y(t) = \sum_{j=0}^{\infty} A(j) \varepsilon(t-j)$$

On suppose que $Z(t)$ vérifie l'hypothèse H5 du § 1 et que la fonction :

$$h(z) = \sum_{j=0}^{\infty} A(j) z^j$$

vérifie :

" $h^{-1}(e^{i\lambda})$ est définie et continue sur $[-\pi, \pi]$; ses coefficients de Fourier notés $C(j)$ sont nuls pour $j < 0$ et satisfont : $\sum_j j ||C(j)|| < +\infty$ "

Nous nous intéressons à l'estimation de B , les matrices $A(j)$ étant supposées connues. On dispose deux méthodes :

5.1. Méthode temporelle

On maximise la fonction L_N du § 2

$$L_N = \text{Log}(\det 2 G) - \frac{1}{2} \sum_{t=1}^N \varepsilon(t, B) G^{-1} \varepsilon(t, B)$$

où :

$$\varepsilon(t, B) = \sum_{j=0}^{t-1} C_{\theta}(j) [X(t-j) - BZ(t-j)]$$

En notant $E_{\alpha\beta}$ la matrice dont tous les éléments sont nuls sauf celui sur la α -ième ligne et β -ième colonne qui vaut 1, on a :

$$\varepsilon(t, B) = \sum_{j=0}^{t-1} C(j) X(t-j) - \sum_{\alpha \beta} B_{\alpha\beta} \left[\sum_{j=0}^{t-1} C(j) E_{\alpha\beta} Z(t-j) \right]$$

La maximisation de L_N équivaut à la minimisation de :

$$\sum_{t=1}^N \varepsilon(t, B) G^{-1} \varepsilon(t, B)$$

ce qui revient à faire une régression de la série $(\mathcal{F}X)(t)$ sur les séries $(\mathcal{F} E_{\alpha\beta} Z)(t)$ où \mathcal{F} est le filtre linéaire défini par :

$$(\mathcal{F}X)(t) = \sum_{j=0}^{t-1} C(j) X(t-j)$$

5.2. Méthode spectrale

On maximise la fonction \tilde{L}_N du § 2

$$\tilde{L}_N = -\frac{N}{2} \text{Log det } (2\pi G) - \frac{N}{4\pi} \int_{-\pi}^{\pi} \text{Tr } \varphi(\lambda) J^N(\lambda) d\lambda$$

où :

$$\varphi(\lambda) = \begin{pmatrix} f^{-1}(\lambda) & f^{-1}(\lambda)B \\ {}^t_B f^{-1}(\lambda) & {}^t_B f^{-1}(\lambda) B \end{pmatrix}$$

$$J^N(\lambda) = \frac{1}{2\pi N} \begin{pmatrix} \tilde{X}_N(\lambda) \tilde{X}_N^*(\lambda) & \tilde{X}_N(\lambda) \tilde{Z}_N^*(\lambda) \\ \tilde{Z}_N(\lambda) \tilde{X}_N^*(\lambda) & \tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda) \end{pmatrix}$$

$$\tilde{X}_N(\lambda) = \sum_{t=1}^N e^{i\lambda t} X(t) \quad ; \quad \tilde{Z}_N(\lambda) = \sum_{t=1}^N e^{i\lambda t} Z(t)$$

On voit que \tilde{L}_N est une forme quadratique en $B_{\alpha\beta}$, soit :

$$L_N = C^{te} + \frac{1}{4\pi^2} \text{Tr } {}^t_B \int_{-\pi}^{\pi} f^{-1}(\lambda) \tilde{X}_N(\lambda) \tilde{Z}_N^*(\lambda) d\lambda - \frac{1}{2} \cdot \frac{1}{4\pi^2} \text{Tr } \int_{-\pi}^{\pi} {}^t_B f^{-1}(\lambda) B \tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda) d\lambda$$

On notera β (resp. t) le vecteur de \mathbb{R}^q de composantes $B_{\alpha\beta}$

$$\text{(resp. } \frac{1}{4\pi^{2N}} \left\{ \int_{-\pi}^{\pi} f^{-1}(\lambda) X_N(\lambda) Z_N(\lambda) d\lambda \right\}_{\alpha\beta} \text{)}$$

et T la matrice de terme général :

$$T_{(\alpha,\beta)(\gamma,\delta)} = \frac{1}{4\pi^{2N}} \int_{-\pi}^{\pi} f_{\alpha\gamma}^{-1}(\lambda) [\tilde{Z}_N(\lambda) \tilde{Z}_N^*(\lambda)]_{\delta\beta} d\lambda$$

L'estimateur $\hat{\beta}$ de β est alors donné par : $\hat{\beta} = T^{-1} t$

La présence des intégrales n'est pas commode au point de vue calcul numérique. On est conduit à remplacer t, T par :

$$(5.1) \quad \hat{t}_{\alpha\beta} = \frac{1}{N} \left[\sum_{j=1}^N f^{-1}(2\pi j/N) I_{XZ}^N(2\pi j/N) \right]_{\alpha\beta}$$

$$(5.2) \quad \hat{T}_{(\alpha,\beta)(\gamma,\delta)} = \frac{1}{N} \sum_{j=1}^N f_{\alpha\gamma}^{-1}(2\pi j/N) [I_Z^N(2\pi j/N)]_{\delta\beta}$$

où :

$$I_{XZ}^N(\lambda) = \frac{1}{2\pi N} X_N(\lambda) Z_N(\lambda)$$

$$I_Z^N(\lambda) = \frac{1}{2\pi N} Z_N(\lambda) Z_N(\lambda)$$

La justification de cette approximation consiste à montrer que la différence entre le vecteur $d = N(t - T\beta)$ des dérivées de L_N et son approximation $\hat{d} = N(\hat{t} - \hat{T}\beta)$ est borné en probabilité quand $N \rightarrow \infty$. Comme \hat{T} et T convergent vers une même matrice quand $N \rightarrow \infty$, on a que :

$$\hat{T}^{-1} \hat{d} - T^{-1} d = N(\hat{T}^{-1} \hat{t} - T^{-1} t)$$

est borné en probabilité quand $N \rightarrow \infty$. L'erreur d'approximation converge donc vers 0 en probabilité comme $1/N$.

Considérons maintenant le cas où la structure de covariance du processus dépend des paramètres θ, G . Ces paramètres seront remplacés par leurs estimateurs. Or les formules (5.1) et (5.2) ne font intervenir que $f(\lambda)$ et non (θ, G) . On peut penser à généraliser la méthode au cas où la structure de covariance du processus résiduel est non paramétrique mais sa densité spectrale peut être estimée à l'aide de la procédure habituelle du lissage du périodogramme. Hannan [14], p. 442 a résolu ce problème et sa méthode est effectivement analogue à la précédente. L'estimateur de β est donné par :

$$\hat{\beta} = -\tilde{T}^{-1} \tilde{t}$$

où :

$$\tilde{t}_{\alpha\beta} = \frac{1}{2M} \left[\sum_{j=-M+1}^M f_X^{-1}(\pi j/M) f_{XZ}(\pi j/M) \right]_{\alpha\beta}$$

$$\tilde{T}_{(\alpha,\beta)(\gamma,\delta)} = \frac{1}{2M} \sum_{j=-M+1}^M \left[f_X^{-1}(\pi j/M) \right]_{\alpha\gamma} \left[f_Z^{-1}(\pi j/M) \right]_{\delta\beta}$$

où f_X, f_{XZ}, f_Z sont les densités spectrales estimées calculées par la même méthode de lissage du périodogramme avec M comme point de troncation (Les séries $X(t), Y(t), Z(t)$ étant considérées comme des réalisations de processus stationnaires).

C H A P I T R E I V

ESTIMATION DES PARAMETRES D'UN PROCESSUS GAUSSIEN EN TEMPS CONTINU, STATIONNAIRE CENTRE DE DENSITE SPECTRALE RATIONNELLE

Dans les chapitres précédents, nous avons considéré le problème d'estimation des paramètres dans les modèles de processus en temps discret, en particulier les modèles auto-régressifs, en moyenne mobile et mixte. Nous considérons maintenant l'analogie de ce problème pour le cas du temps continu.

Nous savons (théorème 1.4 du chapitre I) que le processus mixte a la densité spectrale :

$$f(\lambda) = \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1} \left[\sum_{k=0}^r B(k) e^{ik\lambda} \right] \frac{G}{2\pi} \left[\sum_{k=0}^r B(k) e^{ik\lambda} \right]^* \left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1}$$

et admet la représentation :

$$X(t) = \sum_{j=0}^{\infty} D(j) \varepsilon(t-j) = \sum_{u=-\infty}^t D(t-u) \varepsilon(u)$$

où $\varepsilon(t)$, $t \in \mathbb{Z}$ sont des vecteurs aléatoires indépendants, centrés, de même loi ayant une matrice de covariance G , et où les matrices $D(j)$ sont les coefficients de Fourier de :

$$\left[\sum_{j=0}^q A(j) e^{ij\lambda} \right]^{-1} \left[\sum_{k=0}^r B(k) e^{ik\lambda} \right]$$

L'analogie en temps continu de ce processus sera un processus stationnaire centré admettant la représentation :

$$X(t) = \int_{-\infty}^t D(t-u) d\xi(u) \quad ; \quad t \in \mathbb{R}$$

où $\{\xi(u), u \in \mathbb{R}\}$ est un processus à accroissements indépendants, dont les accroissements $\xi(u+h) - \xi(u)$, $u \in \mathbb{R}$ (h fixé) sont de même loi ayant une matrice de covariance $G \cdot h$ et où la fonction $D(\cdot)$ est la transformée de Fourier d'une fonction de la forme :

$$\left[\sum_{j=0}^q A(j) (i\lambda)^j \right]^{-1} \left[\sum_{k=0}^r B(k) (i\lambda)^k \right]$$

D'après un théorème de Doob [9] si presque-sûrement les trajectoires du processus $\{\xi(u), u \in \mathbb{R}\}$ sont continues alors le processus $\{\xi(u) - \xi(0), u \in \mathbb{R}\}$ est gaussien. Par suite $\{X(t), t \in \mathbb{R}\}$ est aussi gaussien. L'analogie en temps continu du processus mixte serait donc le processus gaussien stationnaire centré de densité spectrale de la forme :

$$f(\lambda) = \left[\sum_{j=0}^q A(j) (i\lambda)^j \right]^{-1} \left[\sum_{k=0}^r B(k) (i\lambda)^k \right] \frac{G}{2\pi} \left[\sum_{k=0}^r B(k) (i\lambda)^k \right] \left[\sum_{j=0}^q A(j) (i\lambda)^j \right]^{-1}$$

Dans le cas univariable cela équivaut à dire que la densité spectrale du processus est une fraction rationnelle en λ (Rozanov [24] p.44). L'estimation des paramètres d'un tel processus est l'objet de ce chapitre.

Nous ne considérerons pas le modèle multivariable le plus général. Toutefois, dans le chapitre suivant, nous étudierons le cas d'un processus gaussien stationnaire centré de densité spectrale :

$$f(\lambda) = \left[\sum_{j=0}^q A(j) (i\lambda)^j \right]^{-1} \frac{G}{2\pi} \left[\sum_{j=0}^q A(j) (i\lambda)^j \right]^*^{-1} \quad A(q) = I_p$$

qui représente l'état stationnaire d'un système dynamique régi par l'équation différentielle stochastique :

$$dX^{(q-1)}(t) + \sum_{j=0}^{q-1} A(j) X^{(j)}(t) dt = G^{1/2} \xi(t)$$

§ 1. DÉFINITION DU PROBLÈME DE NOTATIONS

Considérons un processus $\{X(t), t \in \mathbb{R}\}$ gaussien stationnaire centré de densité spectrale :

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{p_\theta(i\lambda)}{q_\theta(i\lambda)} \right|^2$$

où p_θ et q_θ sont deux polynômes de degrés respectifs n et m ($m > n$), de coefficients du plus haut degré unité, dépendant d'un paramètre $\theta \in \mathbb{C}^k$.

Nous nous intéressons au problème d'estimation de θ au vu de l'observation d'une trajectoire du processus sur $[0, T]$, le paramètre σ^2 étant supposé connu car sa valeur peut être déterminée avec probabilité 1 de la façon suivante :

D'après le théorème de Baxter [4], si $\{Y(t), t \in \mathbb{R}\}$ est un processus gaussien stationnaire de fonction de covariance R_Y :

$$\lim_{p \rightarrow \infty} \sum_{k=1}^{2^p} \left(Y\left(\frac{k}{2^p}\right) - Y\left(\frac{k-1}{2^p}\right) \right)^2 = R_Y'(0-) - R_Y'(0+) \text{ presque-sûrement}$$

D'autre part le processus $\{X(t), t \in \mathbb{R}\}$ est $m-n-1$ fois dérivable en moyenne quadratique et possède une version dont les trajectoires sont $m-n-1$ fois dérivables (Cramer et Leadbetter [8]). Posons $Y(t) = X^{(m-n-1)}(t, T)$, on a :

$$R_Y(u) = \int_{-\infty}^{+\infty} e^{i\lambda u T} \frac{\sigma^2}{2\pi} \left| \frac{(i\lambda)^{m-n-1} p_\theta(i\lambda)}{q_\theta(i\lambda)} \right|^2 d\lambda$$

ce qui entraîne (Doob [9], p.544) $R_Y'(0-) - R_Y'(0+) = \sigma^2 T$. Par suite :

$$\sigma^2 = \frac{1}{T} \lim_{p \rightarrow \infty} \sum_{k=1}^{2^p} \left\{ X^{(m-n-1)}\left(\frac{kT}{2^p}\right) - X^{(m-n-1)}\left[\frac{(k-1)T}{2^p}\right] \right\}^2 \text{ p.s.}$$

Nous noterons $C_k([0, T])$ l'espace des fonctions k fois continûment dérivables sur $[0, T]$, et P^X la loi de probabilité image sur

$(C_{m-n-1}([0,T]), \mathcal{A})$ du processus $\{X(t), t \in [0,T]\}$, \mathcal{A} étant la tribu engendrée par les projections $\tilde{X}(t) : x(\cdot) \rightarrow x(t), t \in [0,T]$. D'autre part θ^* désignera la vraie valeur de θ .

Dans la suite nous faisons les hypothèses suivantes :

H0 : Pour tout $\theta \in \Theta$, les parties réelles des zéros de p_θ et de q_θ sont strictement négatives.

H1 : Les coefficients des polynômes p_θ et q_θ sont trois fois continûment dérivables par rapport à θ .

§ 2. CALCUL DE LA FONCTION DE VRAISEMBLANCE

Grâce à un théorème de Girsanov [13], nous pouvons donner une expression explicite de la dérivée de Radon-Nikodym de P_θ^X par rapport à une mesure dominante. Pour ce faire, nous redéfinissons P_θ^X à l'aide de la construction suivante :

Le processus $\{X(t), t \in R\}$ admet la représentation (Hannan [14])

$$(2.1) \quad X(t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \frac{p_\theta(-i\lambda)}{q_\theta(-i\lambda)} \sigma dZ(\lambda)$$

où $\{Z(\lambda), \lambda \in R\}$ est un processus gaussien centré à accroissements indépendants avec $E |Z(\lambda_2) - Z(\lambda_1)|^2 = |\lambda_2 - \lambda_1|$. Par suite, si on pose :

$$(2.2) \quad \xi(t) = \int_{-\infty}^{\infty} \frac{e^{-i\lambda t} - 1}{-i\lambda} dZ(\lambda)$$

$$(2.3) \quad Y(t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \frac{\sigma}{q_\theta(-i\lambda)} dZ(\lambda)$$

$\{\xi(t), t > 0\}$ est un mouvement brownien et :

$$(2.4) \quad X(t) = p_{\theta} \left(\frac{d}{dt} \right) Y(t)$$

$$(2.5) \quad Y^{(m-1)}(t) - Y^{(m-1)}(0) + \int_0^t q_{\theta} \left(\frac{d}{ds} \right) Y(s) ds = \sigma \xi(t)$$

où $\tilde{q}_{\theta}(u) = q_{\theta}(u) - u^m$ et où, pour $j < m$, $(d/dt)^j Y(t) = Y^{(j)}(t)$ est la dérivée j -ième en moyenne quadratique de $Y(t)$ (comme ce processus est gaussien, c'est aussi la dérivée j -ième de ses trajectoires. Cramer et Leadbetter [8]).

Posons maintenant :

$$Y_0 = {}^t(Y^{(0)}(0) \dots Y^{(n-1)}(0))$$

$$X_0 = {}^t(X^{(0)}(0) \dots X^{(m-n-1)}(0))$$

On montre que ces vecteurs aléatoires sont indépendants de $\xi(t)$, $t \geq 0$. En effet, si α et β sont les fonctions dont les transformées de Fourier sont $\sigma p(-i\lambda)/q(-i\lambda)$ et $\sigma/q(-i\lambda)$, on a (Hannan [16], p. 69) :

$$X(t) = \int_{-\infty}^{\infty} \alpha(t-s) d\xi(s) ; Y(t) = \int_{-\infty}^{\infty} \beta(t-s) d\xi(s)$$

Or les fonctions $p(-i\lambda)$ et $q(-i\lambda)$ sont analytiques dans le demi-plan supérieur du plan complexe ce qui implique que $\alpha(t)$ et $\beta(t)$ sont nuls pour $t < 0$. Par suite X_0 et Y_0 ne dépendent que de $\xi(s)$, $s < 0$. D'où le résultat.

Comme $Y^{(j)}(0)$, $j = n, \dots, m-1$ peuvent être déterminés à partir de Y_0 et X_0 par (2.4), la résolution de l'équation différentielle stochastique (2.5), avec la donnée de la loi $P_{\theta}^{Y_0 X_0}$ du couple (Y_0, X_0) détermine complètement la loi $P_{\theta}^{Y_0 X}$ de $(Y_0, X(t))$, $t \in [0, T]$. La loi P_{θ}^X apparaît comme la loi marginale de $P_{\theta}^{Y_0 X}$ sur $C_{m-n-1}(0, T)$.

Remarquons que $P_{\theta}^{Y_0 X}$ est une loi de probabilité sur $\tilde{\Omega} = \mathbb{R}^n \times C_{m-n-1}([0, T])$, muni de la tribu $\mathcal{B}_{\mathbb{R}}^{\otimes n} \otimes \mathcal{A}$. D'après sa définition on a :

Lemme 2.1.

"Soient $W, \tilde{\phi}_\theta$ les applications de $\tilde{\Omega} \times [0, T]$ dans \mathbb{R} définies par :

$$\tilde{\omega} = (y, x(\cdot)) \rightarrow W(\tilde{\omega}, t) = \frac{1}{\sigma} [x^{(m-n-1)}(t) - x^{(m-n-1)}(0)]$$

$$\tilde{\omega} = (y, x(\cdot)) \rightarrow \tilde{\phi}_\theta(\tilde{\omega}, t) = \frac{1}{\sigma} \left[\left(\frac{d}{dt}\right)^{m-n} \tilde{p}_\theta\left(\frac{d}{dt}\right) - \tilde{q}_\theta\left(\frac{d}{dt}\right) \right] y(t)$$

où $\tilde{p}_\theta(u) = p_\theta(u) - u^n$, $\tilde{q}_\theta(u) = q_\theta(u) - u^m$ et où $y(\cdot)$ est la solution de $p_\theta(d/dt) y(t) = x(t)$ de conditions initiales $y^{(j)}(0) = y_{j+1}$, $j = 1, \dots, n-1$

Alors le processus $\{\xi_\theta(t) = W(t) - \int_0^t \phi_\theta(s) ds, t > 0$ est un mouvement brownien sur $(\Omega, \mathcal{A}, P_{\theta, Y_0, X})$ "

Nous utilisons maintenant le théorème de Girsanov que nous énonçons ici sous la forme qui nous convient le mieux :

Théorème 2.1. (Girsanov [13])

" Soient $\{\eta(t), t \in [0, T]\}$ un mouvement brownien sur un espace probabilisé (Ω, \mathcal{A}, P) et \mathcal{A}_0 une sous-tribu de \mathcal{A} , P -indépendante de $\eta(t), t \in [0, T]$ et ϕ une application $\mathcal{B}_{[0, T]}$ -mesurable de $\Omega \times [0, T]$ dans \mathbb{R} , vérifiant :

(i) $\forall t \in [0, T] : \phi(\cdot, t)$ est \mathcal{A}_t -mesurable où \mathcal{A}_t est la sous-tribu engendrée par \mathcal{A}_0 et $\eta(s), s \leq t$.

$$(ii) \int_0^T |\phi(\cdot, t)|^2 dt < +\infty \quad P\text{-presque-sûrement.}$$

Si on pose :

$$\tilde{\eta}(t) = \eta(t) - \int_0^t \phi(s) ds$$

$$\xi(\phi) = \int_0^T \phi(t) d\eta(t) - \frac{1}{2} \int_0^T |\phi(t)|^2 dt$$

et \mathcal{Q} la mesure de densité $\exp \xi(\phi)$ par rapport à \mathbb{P} . Alors sous réserve que \mathcal{Q} soit une probabilité, le processus $\{\tilde{\eta}(t), t \in [0, T]\}$ est un mouvement brownien sur $(\Omega, \mathcal{A}, \mathcal{Q})$, \mathcal{Q} -indépendant de \mathcal{A}_0 et \mathbb{P} et \mathcal{Q} coïncident sur \mathcal{A}_0 "

A partir de ce théorème et du lemme 2.1, on montre :

Théorème 2.2.

"Soient $W, \tilde{\phi}_\theta$ définis comme dans le lemme 2.1 et :

$$Y_0 : (y, x(\cdot)) \rightarrow y \in \mathbb{R}^n$$

$$X_0 : (y, x(\cdot)) \rightarrow (x^{(0)}(0) \dots x^{(m-n-1)}(0)) \in \mathbb{R}^{m-n}$$

alors il existe une probabilité unique $P_{\theta,0}$ sur $\tilde{\mathcal{A}}$ telle que sur $(\tilde{\Omega}, \tilde{\mathcal{A}}, P_{\theta,0})$, $\{W(t), t \in [0, T]\}$ est un mouvement brownien, indépendant du couple (Y_0, X_0)

dont la loi image est $P_{\theta,0}^{Y_0 X_0}$ et :

$$\frac{dP_{\theta,0}^{Y_0 X_0}}{dP_{\theta,0}} = \exp \int_0^T \tilde{\phi}_\theta(t) dW(t) - \frac{1}{2} \int_0^T [\tilde{\phi}_\theta(t)]^2 dW(t)$$

où $\int_0^T \tilde{\phi}_\theta(t) dW(t)$ est l'intégrale stochastique de Ito par rapport au mouvement brownien $(\tilde{\Omega}, \mathcal{A}, P_{\theta,0}, \{W(t), t \in [0, T]\})$; cette intégrale peut être calculée dans l'espace $(\tilde{\Omega}, \tilde{\mathcal{A}}, P_{\theta^*,0}^{Y_0 X_0})$ où $\theta^* \in \Theta$ en décomposant (selon le lemme 2.1) $W(t)$ en une somme d'un mouvement brownien $\xi_{\theta^*}(t)$ et d'un processus à trajectoires à variation bornée $\int_0^t \tilde{\phi}_{\theta^*}(s) ds$ ".

Corollaire

Il existe une mesure $\tilde{\mu}$ unique dont l'image par $(Y_0, X_0, W(t), t \in [0, T])$ est le produit de la mesure de Lebesgue et de la mesure associée au mouvement brownien, telle que :

$$\frac{dP_{\theta}^{Y_0 X}}{d\tilde{\mu}} = f_{\theta}^{Y_0 X_0}(Y_0, X_0) \exp\left\{ \int_0^T \phi_{\theta}(t) dW(t) - \frac{1}{2} \int_0^T |\phi_{\theta}(t)|^2 dt \right\}$$

où $f_{\theta}^{Y_0 X_0}$ est la densité de $P_{\theta}^{Y_0 X_0}$ par rapport à la mesure de Lebesgue".

L'intégration de la fonction $\frac{dP_{\theta}^{Y_0 X}}{d\tilde{\mu}}$ définie sur $R^n \times C_{m-n-1}([0, T])$ par rapport à sa première variable conduit à :

Théorème 2.3.

"Il existe une mesure μ définie sur $C_{m-n-1}([0, T])$ dont l'image par $X_0, W(t), t \in [0, T]$ est le produit de la mesure de Lebesgue et de la mesure associée au mouvement brownien, telle que :

$$\begin{aligned} \text{Log} \frac{dP_{\theta}^{Y_0 X}}{d\mu} &= \text{Log} f_{\theta}^{Y_0 X_0} [E_{\theta}(\tilde{Y}_0 | \tilde{X}), X_0] + \frac{1}{2} \text{Log} \det(2\pi\Lambda_{\theta}) \\ &\quad + E_{\theta} \left[\int_0^T \phi_{\theta}(t) dW(t) | \tilde{X} \right] - \frac{1}{2} \int_0^T \{E_{\theta} [\phi_{\theta}(t) | \tilde{X}]\}^2 dt \end{aligned}$$

où les notations sont celles du théorème 2.2., où \tilde{X} est la projection de $\tilde{\Omega}$ sur $C_{m-n-1}([0, T])$, où $E_{\theta}(Z | X)$ est, relativement à la loi $P_{\theta}^{Y_0 X}$ l'espérance conditionnelle de Z à \tilde{X} , considérée comme une fonction sur $C_{m-n-1}([0, T])$ et où Λ_{θ} est, relativement à la loi $P_{\theta}^{Y_0 X}$, la matrice de covariances conditionnelles de \tilde{Y}_0 à \tilde{X} "

La démonstration du théorème faire appel au lemme suivant :

Lemme 2.2.

"Soit ϕ_{θ} l'application de $C_{m-n-1}([0, T]) \times [0, T]$ dans R définie par :

$$\phi_{\theta}(y, x(\cdot), t) = \tilde{\phi}_{\theta}[(0, x(\cdot)), t]$$

où ϕ_{θ} est définie dans le lemme 2.1, alors on a :

$$\tilde{\phi}_{\theta}[(y, x(\cdot)), t] = \phi_{\theta}(x(\cdot), t) + {}^t h_{\theta}(t) \cdot y$$

où la fonction h_{θ} appartient à $[L^2(R^+, \mathcal{B}_{R^+}, dt)]^n$ "

Pour appliquer la méthode du maximum de vraisemblance à l'estimation de θ , nous devons montrer l'existence d'une version $\mathcal{L}(\cdot, \theta)$ de $\log(dP_\theta^X/d\mu)$ telle que tout $x(\cdot) \in C_{m-n-1}([0, T])$, la fonction $\mathcal{L}(x(\cdot), \cdot)$ soit deux fois continûment dérivable. Une telle version, si elle existe, est nécessairement unique et est appelée version régulière de la fonction log-vraisemblance. Nous montrerons son existence dans le paragraphe suivant en nous appuyant sur l'hypothèse H1.

§ 3. FONCTION DE VRAISEMBLANCE ASYMPTOTIQUE

L'expression de $dP_\theta^X/d\mu$ donnée dans le théorème 2.3 est trop compliquée pour être utilisable. Nous sommes amenés à la remplacer par son approximation asymptotique ($T \rightarrow \infty$).

L'application du lemme 2.2 permet d'écrire :

$$\text{Log} \frac{dP_\theta^X}{d\mu} = L_T(\theta) + \varepsilon_T(\theta)$$

où :

$$(3.1) \quad L_T(\theta) = \int_0^T \phi_\theta(t) dW(t) - \frac{1}{2} \int_0^T |\phi_\theta(t)|^2 dt$$

$$(3.2) \quad \varepsilon_T(\theta) = \text{Log} f_{\theta}^{Y_0 X_0} [E_\theta(\tilde{Y}_0 | \tilde{X}), X_0] + \frac{1}{2} \text{Log} \det(2\pi\Lambda_\theta) \\ + {}^t E_\theta(Y_0 | \tilde{X}) \int_0^T h_\theta(t) [dW(t) - \phi_\theta(t) dt] \\ + {}^t E_\theta(Y_0 | \tilde{X}) \left[\int_0^T h_\theta(t) {}^t h_\theta(t) dt \right] E_\theta(Y_0 | \tilde{X}).$$

Nous allons montrer que ε_T , ainsi que ses dérivées par rapport à θ , sont bornées quand $T \rightarrow \infty$.

Lemme 3.1.

"L'application $\theta \rightarrow h_\theta \in [L^2(\mathbb{R}^+, \mathbb{R}^+, dt)]^n$ où h_θ est définie dans le lemme 2.2 est trois fois continûment dérivable".

Lemme 3.2.

"L'application $\theta \rightarrow E(\tilde{Y}_0 | \tilde{X}) \in [L^2(C_{m-n-1}([0, T]), \mathcal{A}, P_{\theta^*}^X)]^n$ est trois fois continûment dérivable; ses dérivées jusqu'à l'ordre 2 calculées au point θ^* sont bornées en norme L^2 quand $T \rightarrow \infty$ ".

Les résultats précédents permettent de montrer le :

Théorème 3.1.

"Les variables aléatoires $\varepsilon_T(\theta)$, $\theta \in \Theta$, définies par (3.2) possèdent des versions $\varepsilon_T(\cdot, \theta)$ telles que pour tout $x(\cdot) \in C_{m-n-1}([0, T])$, la fonction $\varepsilon_T(x(\cdot), \cdot)$ est deux fois continûment dérivable et les variables aléatoires définies par $\varepsilon_T(\cdot, \theta)$ et ses dérivées jusqu'à l'ordre 2 par rapport à θ calculées pour $\theta = \theta^*$ sont bornées en probabilité $P_{\theta^*}^X$ quand $T \rightarrow \infty$ ".

Le théorème 3.1 montre que $\log(dP_\theta^X/d\mu)$ possède une version régulière si L_T en possède une. Ceci est assuré par :

Théorème 3.2.

"Il existe une version $L_T(\cdot, \theta)$ de $L_T(\theta)$ telle que pour tout $x(\cdot) \in C_{m-n-1}([0, T])$ la fonction $L_T(x(\cdot), \cdot)$ est deux fois continûment dérivable et on a presque sûrement :

$$(3.3) \quad \frac{\partial}{\partial \theta_j} L_T(\cdot, \theta) = \int_0^T \frac{\partial}{\partial \theta_j} \phi_\theta(t) d\varepsilon_\theta(t) ; d\varepsilon_\theta(t) = dW(t) - \phi_\theta(t) dt$$

$$(3.4) \quad \frac{\partial^2}{\partial \theta_j \partial \theta_\ell} L_T(\cdot, \theta) = \int_0^T \frac{\partial^2}{\partial \theta_j \partial \theta_\ell} \phi_\theta(t) d\varepsilon_\theta(t) - \int_0^T \frac{\partial}{\partial \theta_j} \phi_\theta(t) \frac{\partial}{\partial \theta_\ell} \phi_\theta(t) dt$$

Nous allons donner une deuxième approximation de $\text{Log}(dP_\theta^X/d\mu)$, comparable à celle de Dzharidze [10].

Théorème 3.3.

"Ecrivons :

$$q_\theta(i\lambda) q_\theta(-i\lambda) = \sum_{j=0}^{m-n-1} d_j \lambda^{2j} + \lambda^{2(m-n)} p_\theta(i\lambda) p_\theta(-i\lambda) + R_\theta(\lambda)$$

où R_θ est un polynôme de degré strictement inférieur à $2n$; Alors :

$$(3.5) \quad L_T(x(\cdot), \theta) = -\frac{1}{2\sigma^2} \sum_{j=0}^{m-n-1} d_j \int_0^T [x^{(j)}(t)]^2 dt + \sigma^2 T(b_{m-1} - a_{n-1})$$

$$+ T \int_{-\infty}^{\infty} \frac{R_\theta(\lambda)}{|P_\theta(i\lambda)|^2} I_T(\lambda) d\lambda + \eta_T(x(\cdot), \theta)$$

où $I^T(\lambda) = (2\pi T)^{-1} \left| \int_0^T e^{i\lambda t} x(t) dt \right|^2$, où a_{n-1} et b_{m-1} sont les coefficients des termes de degré $n-1$ de p_θ et de degré $m-1$ de q_θ et où η_T vérifie :

(i) $\eta_T(x(\cdot), \cdot)$ est deux fois dérivable

(ii) Il existe un voisinage $V(\theta^*)$ de θ tel que les bornes supérieures des variables aléatoires $|\eta_T(\theta)|$, $|\partial \eta_T / \partial \theta_j|$ et $|\partial^2 \eta_T / \partial \theta_j \partial \theta_l|$ pour $\theta \in V(\theta^*)$ sont bornées en probabilité quand $T \rightarrow \infty$

Le résultat (ii) du théorème assure que η_T et ses dérivées calculées pour θ_T où θ_T est un estimateur consistant de θ sont bornées en probabilité quand $T \rightarrow \infty$. Ce résultat serait utile dans la suite.

La démonstration du théorème précédent s'appuie sur :

Lemme 3.3.

"Soit $\{Y(t), t \in [0, T]\}$ un processus presque sûrement à trajectoires $m-1$ fois dérivables tel que sa dérivée $(m-1)$ -ième soit de la forme $\sigma \xi(t) + V(t)$ où $\xi(t)$ est un mouvement brownien et où $V(t)$ est un processus presque sûrement à trajectoires à variations bornées; alors quels que soient les polynômes P_1, P_2 de degrés respectifs $m-1$ et m , on a :

$$\int_0^T P_1(D) Y(t) P_2(D) Y(t) dt = \sum_{j=0}^{m-1} \alpha_{2j} \int_0^T [Y^{(j)}(t)]^2 dt - i \alpha_{2m-1} \frac{\sigma^2 T}{2} + Q(T) - Q(0)$$

où $D^j Y(t) = Y^{(j)}(t)$, ($j < m$) est la dérivée j -ième de $Y(t)$, où $D^m Y(t)$ dt désigne $dY^{(m-1)}(t)$ et où les α_j sont les coefficients de polynôme $P_1(iD) P_2(-iD)$ et où enfin $Q(t)$ est une forme quadratique en $Y(t)$, $Y^{(1)}(t)$, ..., $Y^{(m-1)}(t)$

Remarque :

On peut montrer que l'approximation de $L_T(\theta)$ donnée dans le théorème 3.3 est, à un terme additif indépendant de θ près, l'approximation de la fonction log-vraisemblance de Dzhaparidze [10].

§ 4. PROPRIETES ASYMPTOTIQUES DE LA FONCTION L_T

Nous allons montrer que les dérivées de L_T possèdent des propriétés de convergence analogue à celles des dérivées d'une fonction log-vraisemblance classique.

La technique de démonstration consiste à plonger l'espace $C_{m-n-1}([0, T])$ dans $C_{m-n-1}(R)$ muni de la loi de probabilité \bar{P}_{θ^*} , induite par un processus stationnaire gaussien centré de densité spectrale :

$f_{\theta}(\lambda) = (\sigma^2/2\pi) |p_{\theta^*}(i\lambda)q_{\theta^*}(i\lambda)|^2$, θ^* étant la vraie valeur de θ . Dans cet espace, le processus $\phi_{\theta}(t)$, $t > 0$ est approché par un processus stationnaire de la façon suivante :

Soit la représentation spectrale du processus "coordonnée"
 $\{\bar{X}(t) : x(\cdot) \rightarrow x(t), t \in R\}$:

$$(4.1) \quad \bar{X}(t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \frac{p_{\theta}(-i\lambda)}{q_{\theta}(-i\lambda)} \sigma dZ(\lambda)$$

On pose :

$$(4.2) \quad \bar{Y}_\theta(t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \frac{1}{p_\theta(-i\lambda)} \frac{p_{\theta^*}(-i\lambda)}{q_{\theta^*}(-i\lambda)} \sigma dZ(\lambda)$$

$$(4.3) \quad \begin{aligned} \bar{\phi}_\theta(t) &= \frac{1}{\sigma} \left(\frac{d}{dt}\right)^{m-n} \tilde{p}_\theta\left(\frac{d}{dt}\right) - \tilde{q}_\theta\left(\frac{d}{dt}\right) \bar{Y}_\theta(t) \\ &= \int_{-\infty}^{\infty} e^{-i\lambda t} \left[(-i\lambda)^{m-n} - \frac{q_\theta(-i\lambda)}{p_\theta(-i\lambda)} \right] \frac{p_{\theta^*}(-i\lambda)}{q_{\theta^*}(-i\lambda)} dZ(\lambda) \end{aligned}$$

Il est facile de voir que les lemmes 2.1 et 2.2 admettent les extensions suivantes :

Lemme 4.1.

"Le processus $\{\xi(t) = W(t) - \int_0^t \phi_{\theta^*}(s) ds, t \geq 0$ est un mouvement brownien relativement à \bar{P}_{θ^*} ."

Lemme 4.2.

"On a :

$$\phi_\theta(t) = \phi_\theta(t) + {}^t h_\theta(t) \bar{Y}_{\theta,0}, \quad t > 0$$

où h_θ est définie dans le lemme 1.2 et où $\bar{Y}_{\theta,0}$ est le vecteur aléatoire de composantes $\bar{Y}^{(0)}(0), \dots, \bar{Y}^{(n-1)}(0)$."

Nous noterons $s^T(\theta)$ et $S^T(\theta)$ respectivement le vecteur des dérivées premières et la matrice des dérivées secondes de $-T^{-1} L_T$. Les lemmes 4.1 et 3.1 permettent de remplacer pour le calcul de $s^T(\theta)$ et $S^T(\theta)$ le processus $\phi_\theta(t)$ et ses dérivées, par $\bar{\phi}_\theta(t)$ et ses dérivées en ce sens que :

$$(4.4) \quad s_j^T(\theta^*) = - \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} d\bar{\xi}(t) + O(T)$$

$$(4.5) \quad s_j^T(\theta^*) = - \int_0^T \left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} d\bar{\xi}(t) + \frac{1}{2} \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} \frac{\partial}{\partial \theta_k} \phi_\theta(t)_{\theta=\theta^*} dt + O(T)$$

dt + O(T)

Lemme 4.3.

"On a au sens de la convergence en probabilité :

$$\lim_{T \rightarrow \infty} T^{-1} \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} d\bar{\xi}(t) = 0$$

$$\lim_{T \rightarrow \infty} T^{-1} \int_0^T \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} d\bar{\xi}(t) = 0$$

$$\lim_{T \rightarrow \infty} T^{-1} \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} \left(\frac{\partial}{\partial \theta_l} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} dt = S_{jl}$$

où :

$$(4.6) \quad S_{jl} = \frac{1}{\sigma^2} \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial \theta_j} \frac{q_\theta(i\lambda)}{p_\theta(i\lambda)} \right)_{\theta=\theta^*} f_\theta(\lambda) \left(\frac{\partial}{\partial \theta_l} \frac{q_\theta(i\lambda)}{p_\theta(i\lambda)} \right)_{\theta=\theta^*} d\lambda "$$

Théorème 4.1.

"Quand $T \rightarrow \infty$, $s^T(\theta^*)$ et $S^T(\theta^*)$ convergent en probabilité respectivement vers 0 et vers S. Le même résultat reste valable si θ est remplacé par un estimateur consistant".

La deuxième partie du théorème est une conséquence facile du :

Lemme 4.4.

"Pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que :

$$\sup_{\theta: \|\theta-\theta^*\| < \eta} \|s^T(\theta) - s^T(\theta^*)\| < \epsilon M_T ; \quad \sup_{\theta: \|\theta-\theta^*\| < \eta} \|S^T(\theta) - S^T(\theta^*)\| < \epsilon M_T$$

où M_T et M_T' tendent en probabilité vers des constantes quand $T \rightarrow \infty$

Théorème 4.2.

"Le vecteur aléatoire $T^{1/2} s^T(\theta^*)$ converge en loi quand $T \rightarrow \infty$ vers la loi gaussienne centrée de matrice de covariance S.

§ 5. CONSTRUCTION DE L'ESTIMATEUR

Pour estimer θ , on pourra maximiser la fonction $L_T(\theta)$ qui est une approximation de la fonction de vraisemblance. Remarquons que $L_T(\theta)$ peut s'écrire d'une manière formelle.

$$L_T(\theta) = - \int_0^T \left[\frac{dW}{dt} - \phi_\theta(t) \right]^2 dt + \int_0^T \left[\frac{dW}{dt} \right]^2 dt$$

De cette façon la maximisation de $L_T(\theta)$ apparaît comme la minimisation de la somme des résidus :

$$\int_0^T \left[\frac{d\xi_\theta}{dt} \right]^2 dt \quad d\xi_\theta(t) = dW(t) - \phi_\theta(t) dt$$

En fait l'intégrale précédente est divergente et nous utilisons à la place la fonction $-L_T(\theta)$. Notre méthode généralise donc la méthode de Box et Jenkins [6] dans le cas discret.

On peut montrer avec un raisonnement classique (Aitchinson et Silvey [1]), compte-tenu du lemme 4.1 et du théorème 4.1, que :

"Il existe pour tout T un estimateur θ_T , fonction de $x(t), t \in [0, T]$ tel que quand $T \rightarrow \infty, \theta_T$ converge en probabilité vers θ et est solution de : $h^T(\theta) = 0$ avec probabilité tendant vers 1".

Cet estimateur θ_T sera appelé estimateur consistant de maximum de vraisemblance.

La méthode du chapitre I, paragraphe 5 permet de construire un estimateur approchant l'estimateur consistant de maximum de vraisemblance. Le résultat suivant se démontre de la même manière que le théorème 5.1 du chapitre I :

Théorème 5.1.

"Soient θ_T un estimateur $T^{1/2}$ -consistant de θ (i.e. $T^{1/2}(\theta_T - \theta)$ est bornée en probabilité quand $T \rightarrow \infty$) et S^T un estimateur consistant de S , alors l'estimateur :

$$\tilde{\theta}_T = \theta_T - (S^T)^{-1} s^T(\theta_T)$$

est tel que $\sqrt{T}(\tilde{\theta}_T - \theta)$ converge en loi quand $T \rightarrow \infty$ vers un vecteur gaussien centré de matrice de covariance S . D'autre part si $\hat{\theta}_T$ est l'estimateur consistant de maximum de vraisemblance de θ , alors $\sqrt{T}(\tilde{\theta}_T - \hat{\theta}_T) \rightarrow 0$ en probabilité quand $T \rightarrow \infty$ "

Pour appliquer ce théorème, on peut prendre $S^T = S^T(\theta_T)$, mais pour le calcul numérique, il est plus simple de prendre $S^T = U^T(\theta_T)$ où :

$$(5.1) \quad U_{j\ell}^T(\theta) = -\frac{1}{T} \int_0^T \frac{\partial}{\partial \theta_j} \phi_\theta(t) \frac{\partial}{\partial \theta_\ell} \phi_\theta(t) dt$$

Comme $U^T(\theta)$ converge en probabilité vers S quand $T \rightarrow \infty$ (voir lemmes 4.2 et 4.3), $U^T(\theta_T)$ est bien un estimateur consistant de S d'après le résultat suivant : (voir la démonstration du théorème 4.2)

Lemme 5.1.

"Pour tout $\epsilon > 0$, il existe $\eta \rightarrow 0$ tel que :

$$\sup_{\theta : \|\theta - \theta^*\| \leq \eta} \|U^T(\theta) - U^T(\theta^*)\| < \epsilon M_T$$

où M_T converge en probabilité vers une constante quand $T \rightarrow \infty$."

La démonstration du lemme 5.1 est la même que celle du lemme 4.4, compte-tenu de :

Lemme 5.2.

Ecrivons :

$$\frac{\partial}{\partial \theta_j} \left(\frac{q_\theta(i\lambda)}{p_\theta(i\lambda)} \right) \frac{\partial}{\partial \theta_\ell} \left(\frac{q_\theta(i\lambda)}{p_\theta(i\lambda)} \right) = \sum_{j=0}^{2(m-n-1)} C_j \lambda^j + \frac{g_\theta(\lambda)}{|p_\theta(i\lambda)|^4}$$

où g_θ est un polynôme de degré strictement inférieur à $2n$; Alors :

$$U^T(\theta) = -\frac{1}{2\sigma} \sum_{j=0}^{m-n-1} C_{2j} \int_0^T [x^{(j)}(t)]^2 dt + \int_{-\infty}^{\infty} \frac{\operatorname{Re}(g_{\theta}(\lambda))}{p(i)^4} \mathbb{I}^T(\lambda) d\lambda + \xi_T(x(\cdot), \theta)$$

où la borne supérieure de $\xi_T(\cdot, \theta)$ sur un voisinage convenable $V(\theta^*)$ de θ^* tend vers 0 en probabilité quand $T \rightarrow \infty$.

Cas de l'estimation des coefficients de p_{θ} et de q_{θ} .

Considérons le cas où le paramètre θ se compose de deux vecteurs $a = {}^t(a_1, \dots, a_{\alpha})$ et $b = {}^t(b_1, \dots, b_{\beta})$ et où :

$$p_{\theta}(z) = p_0(z) + \sum_{j=1}^{\alpha} a_j p_j(z) ; q_{\theta}(z) = q_0(z) + \sum_{j=1}^{\beta} b_j q_j(z)$$

Les polynômes p_j (resp. q_j) sont évidemment de degrés au plus $n-1$ (resp. $m-1$). Le cas intéressant est celui où $\alpha = n$, $p_j(z) = z^{n-j}$, $j=0, \dots, \alpha$ et $\beta = m$, $q_j(z) = z^{m-j}$, $j=0, \dots, m$. Dans ce cas les a_j et b_j sont les coefficients des polynômes p_{θ} et q_{θ} .

Le lemme suivant donne une approximation utile des dérivées de ϕ_{θ} :

Lemme 5.3.

"Soient $y_{\theta}(\cdot)$ et $z_{\theta}(\cdot)$ respectivement les solutions de conditions initiales nulles de :

$$(5.2) \quad p_{\theta} \left(\frac{d}{dt} \right) y(t) = x(t)$$

$$(5.3) \quad \int_0^t dz^{(n-1)}(s) + p_{\theta} \left(\frac{d}{ds} \right) z(s) ds = \int_0^t dy^{(m-1)}(s) + \tilde{q}_{\theta} \left(\frac{d}{ds} \right) y(s) ds$$

où $\tilde{p}_\theta(u) = p_\theta(u) - u^n$, $\tilde{q}_\theta(u) = q_\theta(u) - u^m$; Alors :

$$\frac{\partial}{\partial b_j} \phi_\theta(x(\cdot), t) = -\frac{1}{\sigma} q_j\left(\frac{d}{dt}\right) y_\theta(t)$$

$$\frac{\partial}{\partial a_j} \phi_\theta(x(\cdot), t) = \frac{1}{\sigma} p_j\left(\frac{d}{dt}\right) z_\theta(t) + {}^t h_{\theta,j}(t) x$$

où $x_0 = (x^{(0)}(0) \dots x^{(m-n-1)}(0))$ et où $h_{\theta,j}(t)$ est une combinaison linéaire des fonctions $\gamma_r(\theta) \exp. \lambda_r(\theta), t$, les $\lambda_r(\theta)$ étant les zéros de p_θ et les γ_r étant des fonctions continues de θ ."

Nous décomposons, suivant qu'il s'agit de dérivées par rapport à a_j ou à b_j , le vecteur $s^T(\theta)$ en $(h^T(\theta) \ k^T(\theta))$ et la matrice $U^T(\theta)$ en

$$\begin{pmatrix} H^T(\theta) & W^T(\theta) \\ {}^t W^T(\theta) & K^T(\theta) \end{pmatrix}$$

On en déduit alors du lemme 5.2 le

Théorème 5.2.

"Si θ_T est un estimateur consistant de θ , alors :

$$(5.4) \quad \sqrt{T} h_j^T(\theta_T) \sim -\frac{1}{\sigma^2 \sqrt{T}} \int_0^T p_j\left(\frac{d}{dt}\right) z_{\theta_T} \sigma d\xi_{\theta_T}(t)$$

$$(5.5) \quad \sqrt{T} k_j^T(\theta_T) = -\frac{1}{\sigma^2 \sqrt{T}} \int_0^T q_j\left(\frac{d}{dt}\right) y_{\theta_T}(t) \sigma d\xi_{\theta_T}(t)$$

$$(5.6) \quad H_j^T(\theta_T) = \frac{1}{\sigma^2 \sqrt{T}} \int_0^T p_j\left(\frac{d}{dt}\right) z_{\theta_T}(t) p_k\left(\frac{d}{dt}\right) z_{\theta_T}(t) dt$$

$$(5.7) \quad w_{jl}^T(\theta_T) = \frac{1}{\sigma^2_T} \int_0^T p_j\left(\frac{d}{dt}\right) z_{\theta_T}(t) q_l\left(\frac{d}{dt}\right) y_{\theta_T}(t) dt$$

$$(5.8) \quad k_{jl}^T(\theta_T) = \frac{1}{\sigma^2_T} \int_0^T q_j\left(\frac{d}{dt}\right) y_{\theta_T}(t) q_l\left(\frac{d}{dt}\right) y_{\theta_T}(t) dt$$

où $u_T \sim v_T$ signifie que $(u_T - v_T) \rightarrow 0$ en probabilité quand $T \rightarrow \infty$

Remarque :

Les formules (5.4) et (5.5) font intervenir des intégrales du type :

$$\int_0^T f_{\theta_T}(t) d\xi_{\theta_T}(t)$$

Cette intégrale est interprétée comme suit : le processus :

$$\{I(\theta) = \int_0^T f_{\theta}(t) d\xi_{\theta}(t), \theta \in \Theta\}$$

est continu et possède donc une version séparable $I(\theta, x(\cdot))$ qui est mesurable. L'intégrale précédente sera alors définie par $I(\theta_T(x(\cdot)), x(\cdot))$.

On peut calculer l'intégrale en question par une intégration par parties. La fonction f_{θ} qui nous intéresse peut se mettre sous la forme :

$$f_{\theta}(t, x(\cdot)) = \lambda \xi_{\theta}(t, x(\cdot)) + \int_0^t g_{\theta}(s, x(\cdot)) ds$$

où la fonction $g(\cdot, x(\cdot))$ est continue pour tout $x(\cdot) \in C_{m-n-1}([0, T])$ et où $\{\xi_{\theta}(t), t \in [0, T]\}$ est un mouvement brownien relativement à P_{θ}^X . D'où

$$\int_0^T \xi_{\theta}(t) d\xi_{\theta}(t) = \frac{1}{2} (\xi_{\theta}^2(T) - T) \quad P_{\theta}^X \text{ presque-sûrement .}$$

Par suite on a P_{θ}^X presque-sûrement :

$$\int_0^T f_{\theta}(t) d\xi_{\theta}(t) = \left[\int_0^T g_{\theta}(t) dt \right] \xi_{\theta}(T) - \int_0^T g_{\theta}(t) \xi_{\theta}(t) dt + \frac{\lambda}{2} (\xi_{\theta}^2(T) - T)$$

Le second membre est défini pour tout $(\theta, x(\cdot))$ et est bien la version séparable du premier membre.

D'après le théorème 5.2 nous estimons les paramètres (a, b) selon l'algorithme :

"Soient $a^{(v)}, b^{(v)}$ les v -ièmes estimateurs de a, b ; on calcule les $(v+1)$ -ièmes estimateurs comme suit :

(i) On résoud les équations (5.2) et (5.3) avec $\theta = (a^{(v)}, b^{(v)})$ et on obtient des solutions de conditions initiales nulles notées $y(\cdot), x(\cdot)$.

(ii) On calcule :

$$\sigma\xi(t) = y^{(n-1)}(t) + \int_0^t q_\theta \left(\frac{d}{ds}\right) y(s) ds$$

puis :

$$h_j = - \int_0^T p_j \left(\frac{d}{dt}\right) z(t) \sigma d\xi(t)$$

$$k_j = \int_0^T q_j \left(\frac{d}{dt}\right) y(t) \sigma d\xi(t)$$

$$H_{j\ell} = \int_0^T p_j \left(\frac{d}{dt}\right) z(t) p_\ell \left(\frac{d}{dt}\right) z(t) dt$$

$$W_{j\ell} = \int_0^T p_j \left(\frac{d}{dt}\right) z(t) q_\ell \left(\frac{d}{dt}\right) y(t) dt$$

$$K_{j\ell} = \int_0^T q_j \left(\frac{d}{dt}\right) y(t) q_\ell \left(\frac{d}{dt}\right) y(t) dt$$

alors les $(v+1)$ estimateurs de a, b sont donnés par :

$$\begin{pmatrix} a^{(v+1)} \\ b^{(v+1)} \end{pmatrix} = \begin{pmatrix} a^{(v)} \\ b^{(v)} \end{pmatrix} - \begin{pmatrix} H & W \\ t_W & K \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}$$

Remarques :

1.- On doit prendre les précautions décrites dans la remarque précédente pour le calcul de h et k .

2.- Comme k_j est une forme linéaire en b : $k_j = \sum_{\ell} K_{j\ell} b_{\ell} + K_{j0}$,

où :

$$K_{j0} = \int_0^T q_j \left(\frac{d}{dt} \right) y(t) dy^{(m-1)}(t) + q_0 \left(\frac{d}{dt} \right) y(t) dt$$

on peut séparer l'estimation de b de celle de a en procédant comme au chapitre II, § 3:

"Soit $a^{(v)}$ le v -ième estimateur de a :

(i) A l'aide de $a^{(v)}$, on calcule K et K_0 et puis le v -ième estimateur $b^{(v)} = K^{-1} K_0$ de b .

(ii) On calcule maintenant les matrices H , W et le vecteur h , ce qui permet de calculer le $(v+1)$ -ième estimateur de a par :

$$a^{(v+1)} = a^{(v)} - (H - t_W K^{-1} W)^{-1} h.$$

On revient à l'étape (i) où $a^{(v)}$ est remplacé par $a^{(v+1)}$.

A N N E X E

DEMONSTRATIONS DES THEOREMES ET LEMMES

Démonstration du théorème 2.2.

Nous allons appliquer le théorème de Girsanov à l'espace

$(\tilde{\Omega}, \tilde{\mathcal{A}}, P_{\theta}^{Y_0^X})$. Il est clair que les applications \tilde{Y}_0, \tilde{X}_0 et $\tilde{\xi}_{\theta}(s), s \leq t$ et les applications $(y, s(\cdot)) \rightarrow y(s), s \leq t$ engendrent une même sous-tribu notée $\tilde{\mathcal{A}}_t$ de $\tilde{\mathcal{A}}$. On peut vérifier que la fonction $-\tilde{\phi}_{\theta}$ du lemme 2.1 satisfait aux conditions du théorème de Girsanov. Ainsi, si la mesure Q , de densité par rapport à $P_{\theta}^{Y_0^X}$:

$$\exp\left[\int_0^T -\tilde{\phi}_{\theta}(t) d\tilde{\xi}_{\theta}(t) - \frac{1}{2} \int_0^T [\tilde{\phi}_{\theta}(t)]^2 dt\right]$$

est une probabilité, elle possède toutes les propriétés demandées pour être la probabilité $P_{\theta,0}$ du théorème. Comme les applications $\tilde{Y}_0, \tilde{X}_0, W(t), t \in [0, T]$ mettent $\tilde{\Omega}$ en bijection avec $R^m \times C_0([0, T])$ où $C_0([0, T])$ est l'espace des fonctions continues sur $[0, T]$ nulles en 0, on voit que $P_{\theta,0}$ est l'image d'une loi de probabilité bien définie sur $R^m \times C_0([0, T])$ et par conséquent est unique.

Pour montrer que Q est une probabilité, on utilise le lemme 7 de [13] et la remarque qui suit. Ce qui revient à vérifier que :

$$\phi_{\theta}(\tilde{\omega}, t) = f(Z(\cdot), t)$$

où $Z(t), t \in [0, T]$ est un processus de Ito, et f vérifie :

$$|f(z(\cdot), t)| < f_0(\|z(\cdot)\|), f_0 \text{ étant une fonction non décroissante.}$$

Nous prenons comme $Z(t)$ le vecteur aléatoire $\tilde{\omega} \rightarrow (y^{(0)}(t), \dots, y^{(m-1)}(t))$, alors on voit facilement que $|\phi_{\theta}(\tilde{\omega}, t)| < K \cdot \max\{|y^{(i)}(t)|, i = 0, \dots, m-1, t \in [0, T]\}$, ce qui montre que Q est une probabilité.

Maintenant la densité de $P_{\theta}^{Y_0 X}$ par rapport à $P_{\theta,0}$ est :

$$\exp \int_0^T \phi_{\theta}(t) d\tilde{\xi}_{\phi}(t) + \frac{1}{2} \int_0^T [\tilde{\phi}_{\theta}(t)]^2 dt$$

où la première intégrale est l'intégrale stochastique par rapport au mouvement

brownien $(\tilde{\Omega}, \tilde{\mathcal{A}}, P_{\theta}^{Y_0 X}, \{\tilde{\xi}_{\phi}(t), t \in [0, T]\})$. Nous allons montrer que cette intégrale

est $P_{\theta,0}$ (ou $P_{\theta}^{Y_0 X}$)-presque-sûrement égale à :

$$\int_0^T \tilde{\phi}_{\theta}(t) dW(t) - \int_0^T [\tilde{\phi}_{\theta}(t)]^2 dt$$

où le premier terme est l'intégrale stochastique par rapport au mouvement

brownien $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}_{\phi,0}, \{W(t), t \in [0, T]\})$. En effet, le processus $\tilde{\phi}_{\theta}(t), t \in [0, T]$ étant gaussien continu, la suite :

$$\begin{aligned} I_n &= \sum_{j=0}^{n-1} \tilde{\phi}_{\theta}(t_j) [\tilde{\xi}_{\theta}(t_{j+1}) - \tilde{\xi}_{\theta}(t_j)] & t_j &= j \frac{T}{n} \\ &= \sum_{j=0}^{n-1} \tilde{\phi}_{\theta}(t_j) [W(t_{j+1}) - W(t_j)] - \sum_{j=0}^{n-1} \tilde{\phi}_{\theta}(t_j) \int_{t_j}^{t_{j+1}} \tilde{\phi}_{\theta}(t) dt \end{aligned}$$

converge en probabilité $P_{\theta}^{Y_0 X}$ (resp. $P_{\theta,0}$) vers $\int_0^T \phi_{\theta}(t) d\tilde{\xi}_{\theta}(t)$

(resp. $\int_0^T \phi_{\theta}(t) dW(t) - \int_0^T [\tilde{\phi}_{\theta}(t)]^2 dt$).

Le même raisonnement montre que $\int_0^T \tilde{\phi}_{\theta}(t) dW(t)$ est presque-sûrement

égale à $\int_0^T \phi_{\theta}(t) d\xi_{\theta^{**}}(t) + \int_0^T \tilde{\phi}_{\theta}(t) \tilde{\phi}_{\theta^{**}}(t) dt$ où le premier terme est l'in-

tégrale stochastique par rapport au mouvement brownien

$\{\xi_{\theta^{**}}(t) = W(t) - \int_0^t \phi_{\theta^{**}}(s) ds, t \in [0, T]\}$ sur $(\tilde{\Omega}, \tilde{\mathcal{A}}, P_{\theta^{**}}^{Y_0 X_0})$.

Démonstration du lemme 2.2 et du lemme 3.1

Si $y^*(.)$ désigne la solution de l'équation $p_\theta(\frac{d}{dt})y(t) = x(t)$, $t \in [0, T]$ de conditions initiales nulles, on a : $y(t) = y^*(t) + z(t)$ où $z(t)$ est la solution de $p_\theta(\frac{d}{dt})z(t) = 0$ de conditions initiales $z^{(j)}(0) = y_{j+1}$, $j = 0, \dots, n-1$.

Posons $Z(t) = {}^t(z^{(0)}(t) \dots z^{(n-1)}(t))$; la fonction vectorielle Z vérifie l'équation différentielle : $Z'(t) + A_\theta Z(t) = 0$ dont la solution est :

$$Z(t) = e^{-A_\theta \cdot t} \cdot Z(0) = e^{-A_\theta t} \cdot y.$$

Il est clair que :

$$\begin{aligned} \phi_\theta [y, x(.), t] &= \phi_\theta(x(.), t) + [(\frac{d}{dt})^{m-n} \tilde{p}_\theta(\frac{d}{dt}) - \tilde{q}_\theta(\frac{d}{dt})] z(t) \\ &= \phi_\theta(x(.), t) + h_\theta(t) \cdot y \end{aligned}$$

où ${}^t h_\theta(t)$ est la première ligne de :

$$[(\frac{d}{dt})^{m-n} p_\theta(\frac{d}{dt}) - q_\theta(\frac{d}{dt})] e^{-A_\theta t} = B_\theta e^{-A_\theta t}$$

Un calcul explicite montre que $\det(A_\theta - zI) = (-1)^n p_\theta(-z)$. Par suite les valeurs propres de A_θ sont de parties réelles strictement positives, ce qui entraîne que $h_\theta \in [L^2(\mathbb{R}^+, \mathbb{R}^+, dt)]^n$.

Montrons maintenant que l'application $\theta \rightarrow h_\theta \in [L^2(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, dt)]^n$ est trois fois continûment dérivable. Pour cela, il suffit de montrer que l'application $\theta \rightarrow h_\theta \in [L^2(\mathbb{R}, \mathbb{R}, d\lambda/2\pi)]^n$ où h_θ est la transformée de Fourier de h_θ^* , est trois fois continûment dérivable.

Or $\tilde{t}_h(\lambda)$ est la première ligne de la matrice :

$$\int_0^{\infty} e^{i\lambda t} B_{\theta} e^{-A_{\theta} t} dt = B_{\theta} (A - i\lambda I)^{-1}$$

Comme $\det(A_{\theta} - i\lambda I) = (-1)^n p_{\theta}(-i\lambda)$, $\tilde{h}_{\theta,j}(\lambda)$ est une fraction rationnelle de la forme $r_{\theta}(\lambda)/p_{\theta}(-i\lambda)$ où r_{θ} est un polynôme de degré strictement inférieur à n car $h_{\theta}(\lambda) \rightarrow 0$ quand $\lambda \rightarrow \infty$.

Les dérivées partielles de $\tilde{h}_{\theta,j}(\lambda)$ par rapport à θ jusqu'à l'ordre 3 sont aussi des fractions rationnelles de la forme $u_{\theta}(\lambda)/v_{\theta}(\lambda)$ où le polynôme u_{θ} est de degré strictement inférieur à celui du polynôme v_{θ} et où les coefficients de u_{θ} et de v_{θ} sont des fonctions continues de θ .

Soient a_{θ} et b_{θ} respectivement les coefficients du plus haut degré de u_{θ} et de v_{θ} . Il n'est pas difficile de montrer que :

$$((\lambda, \theta') \rightarrow (\infty, \theta)) \rightarrow (\sqrt{1+\lambda^2} \cdot \frac{u_{\theta'}(\lambda)}{v_{\theta'}(\lambda)} \rightarrow \frac{a_{\theta}}{b_{\theta}})$$

La fonction $(\lambda, \theta') \rightarrow \sqrt{1+\lambda^2} u_{\theta}(\lambda)/v_{\theta}(\lambda)$ est donc continue sur le compact $\bar{R} \times V(\theta)$ où \bar{R} est la droite achevée et où $v(\theta)$ est un voisinage compact de θ . On peut alors montrer que :

$$(\theta' \rightarrow \theta) \quad \sqrt{1+\lambda^2} \frac{u_{\theta'}(\lambda)}{v_{\theta'}(\lambda)} \rightarrow \sqrt{1+\lambda^2} \frac{u_{\theta}(\lambda)}{v_{\theta}(\lambda)}$$

uniformément pour tout $\lambda \in R$.

Considérons maintenant $\tilde{h}_{\theta+\delta\theta}(\lambda) - \tilde{h}_{\theta}(\lambda)$ où $\delta\theta$ est de la forme $\delta\theta_j = \delta$, $\delta\theta_{\ell} = 0$, $\ell \neq j$. On a :

$$\delta^{-1} [\tilde{h}_{\theta+\delta\theta}(\lambda) - \tilde{h}_{\theta}(\lambda)] = \left[\frac{\partial}{\partial \theta_j} \tilde{h}_{\theta}(\lambda) \right]_{\theta = \theta_{\lambda} \in [\theta, \theta+\delta\theta]}$$

Or chaque $(\partial \tilde{h}_{\theta}(\lambda) / \partial \theta_j)_{\lambda}$ est de la forme $u_{\theta}(\lambda)/v_{\theta}(\lambda)$ précédente.

Par suite si δ est choisi suffisamment petit, on a :

$$||\delta^{-1} [\tilde{h}_{\theta+\delta\theta}(\lambda) - \tilde{h}_{\theta}(\lambda)] - \frac{\partial}{\partial \theta_j} \tilde{h}_{\theta}(\lambda)|| < \frac{\varepsilon}{\sqrt{1+\lambda^2}}$$

La norme dans $L^2(R, \mathcal{B}_R, d\lambda/2\pi)^n$ de la fonction de λ du premier membre est donc inférieure à $\varepsilon/2$. D'où :

$$\lim_{\delta \rightarrow 0} \delta^{-1} [\tilde{h}_{\theta+\delta\theta}(\cdot) - \tilde{h}_{\theta}(\cdot)] = \frac{\partial}{\partial \theta_j} \tilde{h}_{\theta}(\cdot)$$

au sens de $L^2(R, \mathcal{B}_R, d\lambda/2\pi)^n$. On démontre de la même façon que les dérivées d'ordre 2 et 3 de l'application $\theta \rightarrow \tilde{h}_{\theta} \in L^2(R, \mathcal{B}_R, d\lambda/2\pi)^n$ sont aussi les dérivées partielles correspondantes de $\tilde{h}_{\theta}(\lambda)$ par rapport à θ . D'où le lemme 3.1.

Démonstration du théorème 2.3.

Comme $\text{Log } f_{\theta}^{Y_0 X_0}(y, x)$ est une forme quadratique en y , d'après le corollaire du théorème 2.2 et le lemme 2.2, on peut écrire :

$$\begin{aligned} \frac{dP_{\theta}^{Y_0 X}}{d\tilde{\mu}}(y, x(\cdot)) &= K_{\theta} \exp\left\{-\frac{1}{2} {}^t [y A_{\theta} y + 2 {}^t b_{\theta}(x(\cdot)) \cdot y + C(x(\cdot))]\right\} \\ &= K_{\theta} \exp\left\{-\frac{1}{2} {}^t [y - y^*(x(\cdot))] A_{\theta} [y - y^*(x(\cdot))] - \frac{1}{2} D_{\theta}(x(\cdot))\right\} \end{aligned}$$

On reconnaît facilement que $y^*(x(\cdot))$ est relativement à $P_{\theta}^{Y_0 X}$

l'espérance conditionnelle de Y_0 à \tilde{X} , définie comme une fonction sur $C_{m-n-1}([0, T])$ et que A_{θ}^{-1} est relativement à $P_{\theta}^{Y_0 X}$ la matrice de covariances conditionnelles de \tilde{Y}_0 à \tilde{X} .

En intégrant la densité précédente par rapport à la variable y , on a :

$$\frac{dP_{\theta}^X}{d\mu}(x(\cdot)) = \det(2\pi A^{-1}) \cdot K_{\theta} \exp\left[-\frac{1}{2} D_{\theta}(x(\cdot))\right].$$

Or $K_{\theta} \exp\left[-\frac{1}{2} D_{\theta}(x(\cdot))\right]$ n'est autre chose que $\frac{dP_{\theta}^{Y_0 X}}{d\mu}[y^*(x(\cdot)), x(\cdot)]$, ce qui entraîne le théorème.

Démonstration du lemme 3.2.

Notons H_θ le sous-espace fermé de $L^2(\Omega, \mathcal{A}, P_0^{Y_0^X})$ engendré par les variables aléatoires $\tilde{X}(t) : (y, x(\cdot)) \rightarrow x(t), t \in [0, T]$. Nous savons que la correspondance $\tilde{X}(t) \leftrightarrow (\lambda \sim e^{-i\lambda t})$ établit un isomorphisme de H_θ sur un sous-espace fermé de $L^2(\mathbb{R}, \mathcal{B}_R, F_\theta)$ noté $L_T^2(F_\theta)$, F_θ étant la mesure spectrale du processus $\tilde{X}(t)$ (voir Rozanov [24]).

Comme le vecteur aléatoire $E_\theta(\tilde{Y}_0 | \tilde{X})$ a pour composantes les projections sur H_θ des composantes de \tilde{Y}_0 , il correspond par l'isomorphisme précédent à un élément χ_θ de $[L_T^2(F_\theta)]^n$ vérifiant :

$$\forall t \in [0, T] : \int_{-\infty}^{\infty} \chi_\theta(\lambda) e^{i\lambda t} f_\theta(\lambda) d\lambda = E_\theta [\tilde{Y}_0 \tilde{X}(t)]$$

$$= \int_{-\infty}^{\infty} \eta_\theta(\lambda) e^{i\lambda t} f_\theta(\lambda) d\lambda$$

où $f_\theta(\lambda) = (\sigma^2/2\pi) |p_\theta(i\lambda)/q_\theta(i\lambda)|^2$ est la densité spectrale du processus et où η_θ est la fonction intervenant dans la représentation spectrale :

$$Y_0 = \int_{-\infty}^{\infty} \eta_\theta(\lambda) \frac{p(-i\lambda)}{q(-i\lambda)} \sigma dZ(\lambda)$$

Remarquons que les métriques des espaces $L^2(\mathbb{R}, \mathcal{B}_R, F_\theta)$ sont équivalentes car pour tout $(\theta, \theta') \in \mathbb{C}^2$ on a $0 < C_1 < f_\theta(\lambda)/f_{\theta'}(\lambda) < C_2, \forall \lambda \in \mathbb{R}$. Par suite les espaces $L_T^2(F_\theta)$ ont même support. Nous allons montrer que si l'application $\theta \rightarrow k_\theta \in L^2(\mathbb{R}, \mathcal{B}_R, F_{\theta'})$ est continue (resp. continûment dérivable) alors l'application $\theta \rightarrow \pi_\theta^T k_\theta$ où π_θ^T désigne le projecteur de $L^2(\mathbb{R}, \mathcal{B}_R, F_\theta)$ sur $L_T^2(F_\theta)$, est aussi continue (resp. continûment dérivable) dont les dérivées sont :

$$\frac{\partial}{\partial \theta_j} \pi_\theta^T k_\theta = \pi_\theta^T \frac{\partial}{\partial \theta_j} k_\theta + (k_\theta - \pi_\theta^T k_\theta) \left(\frac{\partial}{\partial \theta_j} f_\theta \right) \frac{1}{f_\theta}.$$

En effet, soit $\delta\theta$ un accroissement θ et $\delta k_\theta, \delta f_\theta, \delta \pi_\theta^T k_\theta$ les accrois-

sements correspondants de k_θ , f_θ et $\pi_\theta^T k_\theta$. Nous avons :

$$\forall t \in [0, T] : \int_{-\infty}^{\infty} (\delta \pi_\theta^T k_\theta)(\lambda) e^{i\lambda t} f_\theta(\lambda) d\lambda = \int_{-\infty}^{\infty} \delta k_\theta(\lambda) e^{i\lambda t} f_\theta(\lambda) d\lambda \\ + \int_{-\infty}^{\infty} [k_{\theta+\delta\theta}(\lambda) - \pi_{\theta+\delta\theta}^T k_{\theta+\delta\theta}(\lambda)] e^{i\lambda t} \delta f_\theta(\lambda) d\lambda$$

ce qui entraîne que :

$$\delta \pi_\theta^T k_\theta = \pi_\theta^T [\delta k_\theta + (k_{\theta+\delta\theta} - \pi_{\theta+\delta\theta}^T k_{\theta+\delta\theta}) \delta f_\theta / f_\theta]$$

Faisons $\delta\theta \rightarrow 0$, il est facile de voir que $\delta f_\theta(\lambda)/f_\theta(\lambda)$ tend vers zéro uniformément. D'autre part, dans $L^2(\mathbb{R}, \mathbb{R}, F_{\theta+\delta\theta})$, la norme de

$\pi_{\theta+\delta\theta}^T k_{\theta+\delta\theta}$ est inférieure à celle de $k_{\theta+\delta\theta}$; par suite, dans $L^2(\mathbb{R}, \mathbb{R}, F_\theta)$,

la norme du premier est inférieure à K fois celle du dernier (K étant une constante) et reste donc bornée quand $\delta\theta \rightarrow 0$. Ceci entraîne que $\delta \pi_\theta^T k_\theta \rightarrow 0$ quand $\delta\theta \rightarrow 0$. D'où la continuité de l'application $\theta \rightarrow \pi_\theta^T k_\theta$.

Prenant maintenant $\delta\theta$ de la forme $\delta\theta_j = h$, $\delta\theta_\ell = 0$, $\ell \neq j$ et calculant la limite quand $h \rightarrow 0$ de $h^{-1} \delta \pi_\theta^T k_\theta$, on obtient le résultat annoncé.

Nous appliquons ce résultat pour $k_\theta = \eta_{\theta, i}$, puis pour

$$k_\theta = \eta_\theta^1 = \frac{\partial}{\partial \theta_j} \eta_{\theta, i} + (\eta_{\theta, i} - \pi_\theta^T \eta_{\theta, i}) \left(\frac{\partial}{\partial \theta_j} f_\theta \right) \frac{1}{f_\theta}$$

et enfin pour $k_\theta = \eta_\theta^2$ défini par la même formule que la précédente où η_θ est remplacé par η_θ^1 . Chaque fois l'application $\theta \rightarrow k_\theta$ est continûment dérivable d'après les remarques suivantes :

(i) L'application $\theta \rightarrow \eta_\theta \in L^2(\mathbb{R}, \mathbb{R}, F_{\theta::})$ est trois fois continûment dérivable. Ceci résulte d'une démonstration analogue à celle du lemme 3.1.

(ii) La fonction $(\partial f_\theta / \partial \theta_j) / f_\theta$ est une fraction rationnelle dont les coefficients sont deux fois continûment dérivables par rapport à θ et dont le degré du numérateur est strictement inférieur à celui du dénominateur et par conséquent appartient à l'espace $\mathcal{L}^\infty(\mathbb{R})$ des fonctions mesurables bornées sur \mathbb{R} muni de la norme du sup. Une démonstration analogue à celle du lemme 3.1 montre que l'application $\theta \rightarrow (\partial f_\theta / \partial \theta_j) / f_\theta \in \mathcal{L}^\infty(\mathbb{R})$ est deux fois continûment dérivable.

(iii) Si $\theta \rightarrow h_\theta \in L^2(\mathbb{R}, \mathcal{B}_R, F_{\theta^*})$ et $\theta \rightarrow g_\theta \in \mathcal{L}^\infty(\mathbb{R})$ sont des applications continues (resp. continûment dérivables) alors l'application $\theta \rightarrow h_\theta g_\theta \in L^2(\mathbb{R}, \mathcal{B}_R, F_{\theta^*})$ est aussi continue (resp. continûment dérivable de dérivées $\partial(h_\theta g_\theta) / \partial \theta_j = (\partial h_\theta / \partial \theta_j) g_\theta + h_\theta \partial g_\theta / \partial \theta_j$).

Nous avons donc montré que l'application $\theta \rightarrow \pi_{\theta}^T [L^2(\mathbb{R}, \mathcal{B}_R, F_{\theta^*})]^n$ est trois fois continûment dérivable. Un calcul explicite montrera que ses dérivées calculées pour $\theta = \theta^*$ sont bornées en norme quand $T \rightarrow \infty$. D'où le lemme.

Démonstration du théorème 3.1 :

Pour monter l'existence d'une "version régulière" de $\varepsilon_T(\theta)$, il nous suffit de montrer l'existence des "versions régulières" de $F_\theta(\tilde{Y}_0 | \tilde{X})$ et de $\int_0^T h_\theta(\lambda) dW(\lambda)$. Pour ce faire nous utilisons une généralisation facile d'un résultat de Kolmogoroff (Neveu [20], p. 92) :

"Soient X_1, \dots, X_n , n applications de θ dans $L^2(\Omega, \mathcal{A}, P)$ où \mathcal{O} est un compact de \mathbb{R}^k et supposons qu'il existe des constantes $c > 0$, $p > 1$, $q > 1$ telles que :

$$E\{|X_i(\theta+h) - X_i(\theta)|^p\} < c \|h\|^q$$

alors il existe des versions $X_1(\cdot, \theta), \dots, X_n(\cdot, \theta)$ de $X_1(\theta), \dots, X_n(\theta)$ telles que les fonctions $X_i(\omega, \cdot)$ sont continues sur \mathcal{O} ".

Comme l'application $\theta \in \mathcal{O} \rightarrow E_\theta(Y_0 | X) \in [L^2(P_{\theta^*}^X)]^n$ est trois fois continûment dérivable, si \mathcal{O} est compact, ses dérivées jusqu'à l'ordre 2 vérifient la condition précédente (avec $p = q = 2$). Si \mathcal{O} n'est pas compact, il

est contenu dans l'union dénombrable de compacts et on peut se ramener au cas précédent. Il existe donc des versions $\alpha_{j\ell}(\cdot, \theta)$ de $\partial^2 E_\theta(\tilde{Y}_0 | \tilde{X}) / \partial \theta_j \partial \theta_\ell$ qui, pour chaque $x(\cdot)$, sont des fonctions continues sur \mathbb{D} . On définit ensuite des versions des dérivées premières de $E_\theta(\tilde{Y}_0 | \tilde{X})$, puis de $E_\theta(\tilde{Y}_0 | \tilde{X})$ par l'intégration des fonctions $\alpha_{j\ell}(x(\cdot), \cdot)$. De cette façon, on obtient une version régulière de $E_\theta(\tilde{Y}_0 | \tilde{X})$ ayant les propriétés demandées.

D'autre part, d'après le théorème 2.3, il existe une probabilité P équivalente à $P_{\theta^*}^X$ telle que $\{W(t), t \in [0, T]\}$ soit un mouvement brownien sur $(C_{m-n-1}([0, T]), \mathcal{A}, P)$. Par suite l'application $\theta \rightarrow \int_0^T h_\theta(t) dW(t) \in L^2(P)$ est trois fois continûment dérivable. Le procédé précédent fournit une version régulière de $\int_0^T h_\theta(t) dW(t)$.

Finalement, les lemmes 2.1 et 2.2 montrent que les dérivées de $\epsilon_T(\theta)$, calculées pour la valeur θ^* sont majorées en norme $L^1(P_{\theta^*}^X)$ quand $T \rightarrow \infty$.

Démonstration du théorème 3.2 :

L'équation différentielle $p_\theta(d/dt) y(t) = x(t)$ peut s'écrire sous la forme :

$$\frac{dZ}{dt} + A_\theta Z(t) = V(t)$$

où :

$$Z(t) = {}^t(y^{(0)}(t), y^{(1)}(t), \dots, y^{(n-1)}(t))$$

$$V(t) = {}^t(0, \dots, x(t))$$

La solution de cette équation de conditions initiales nulles est :

$$Z(t) = \int_0^t e^{-A_\theta(t-s)} dV(s)$$

Maintenant la relation $p_\theta(d/dt) y(t) = x(t)$ permet d'exprimer $y^{(n)}(t)$, puis $y^{(n+1)}(t), \dots, y^{(m-1)}(t)$ comme combinaisons linéaires de $y^{(i)}(t)$, $j = 0, \dots, n-1$ et de $x^{(j)}(t)$, $j = 0, \dots, m-n-1$. Ce qui entraîne :

$$\phi_{\theta}(x(\cdot), t) = \sum_{j=0}^{m-n-1} \alpha_j(\theta) x^{(j)}(t) + \int_0^t \beta_{\theta}(t-s) x(s) ds$$

où les fonctions α_j sont trois fois continûment dérivables et où la fonction $(\theta, t) \rightarrow \beta_{\theta}(t)$ admet des dérivées partielles jusqu'à l'ordre 3 par rapport à θ , continues sur $\mathbb{W} \times [0, T]$.

Soit P une probabilité relativement à laquelle $W(t), t \in [0, T]$ est un mouvement brownien. (voir démonstration du théorème 3.1). L'application :

$$\phi_{\theta} \rightarrow \int_0^T \phi_{\theta}(t) dW(t)$$

est un isomorphisme de $E = L^2(C_{m-n-1}([0, T]) \times [0, T], \mathcal{A} \otimes \mathcal{B}_{[0, T]}, P \times dt)$ sur un sous-espace de $L^2(C_{m-n-1}([0, T]), \mathcal{A}, P)$ (Métivier [18]).

Nous allons montrer que l'application $\theta \rightarrow \phi_{\theta}(\cdot, \cdot)$ à valeurs dans E est trois fois continûment dérivable dont les dérivées premières et secondes sont :

$$\frac{\partial}{\partial \theta_j} \phi_{\theta}(\cdot, \cdot) \quad \text{et} \quad \frac{\partial^2}{\partial \theta_j \partial \theta_k} \phi_{\theta}(\cdot, \cdot)$$

On obtient alors le théorème avec une démonstration analogue à celle du théorème 3.1.

Pour montrer l'assertion précédente, notons que les dérivées partielles jusqu'à l'ordre 3 de $\phi_{\theta}(x(\cdot), t)$ par rapport à θ sont de la forme :

$$g_{\theta}(x(\cdot), t) = \sum_{j=0}^{m-n-1} \gamma_j(\theta) x^{(j)}(t) + \int_0^t \psi_{\theta}(t-s) x(s) ds$$

où $g_{\theta} \in \mathcal{L}^{\infty}([0, T])$ ($\mathcal{L}^{\infty}[0, T]$ est l'espace des fonctions mesurables bornées sur $[0, T]$ muni de la norme du sup) et où les fonctions γ_j et l'application $\theta \rightarrow \psi_{\theta} \in \mathcal{L}^{\infty}([0, T])$ sont continues. Par suite :

$$\forall \epsilon > 0, \quad \exists \delta > 0 : \|\theta' - \theta\| < \delta$$

$$|g_{\theta'}(x(\cdot), t) - g_{\theta}(x(\cdot), t)| < c \left[\sum_{j=0}^{m-n-1} |x^{(j)}(t)| + \int_0^t |x(s)| ds \right]$$

Le crochet du membre droite est un élément de E, par suite :

$$\|g_{\theta}(\dots) - g_{\theta}(\dots)\|_E < \varepsilon K$$

Une démonstration analogue à celle du lemme 3.1 montre alors que l'application $\theta \rightarrow \phi_{\theta}(\dots) \in E$ possède les propriétés énoncées ci-dessus. D'où le théorème.

Démonstration du lemme 3.3 :

D'après la formule d'Ito (voir [15] ou [18]) :

$$Y^{(j)}(T) Y^{(l)}(T) - Y^{(j)}(0) Y^{(l)}(0) = \int_0^T Y^{(j)}(t) dY^{(l)}(t) + \int_0^T Y^{(j+1)}(t) Y^{(l)}(t) dt$$

si $j < m-1, l < m$

$$[Y^{(m-1)}(T)]^2 - [Y^{(m-1)}(0)]^2 = 2 \int_0^T Y^{(m-1)}(t) dY^{(m-1)}(t) + \frac{\sigma^2 T}{2}$$

L'application répétée de ces formules donne : (on note $Q(t)$ une forme quadratique en $Y^{(0)}(t), \dots, Y^{(m-1)}(t)$).

(i) $j < l < m$ et j et l sont de même parité :

$$\int_0^T Y^{(j)}(t) dY^{(l-1)}(t) = (-1)^{(l-j)/2} \int_0^T [Y^{(j+l)/2}(t)]^2 dt + Q(T) - Q(0)$$

(ii) $j < l < m-1$ et j et l de parités différentes :

$$\int_0^T Y^{(j)}(t) dY^{(l-1)}(t) = Q(T) - Q(0)$$

$$(iii) \int_0^T Y^{(m-1)}(t) dY^{(m-1)}(t) = \frac{-\sigma^2 T}{2} + Q(T) - Q(0)$$

Or, si j et l sont de même parité $(-1)^{(l-j)/2} = (i)^l (-i)^j = (-i)^l (i)^j$

est le coefficient du monôme $(iD)^l (-iD)^j$. On en déduit le lemme dans le cas où les polynômes P_1 et P_2 sont de degrés inférieurs ou égaux à $m-1$

Dans le cas où P_2 est de degré m . L'intégrale :

$$\int_0^T P_1(D) Y(t) P_2(D) Y(t) dt$$

contient le terme :

$$\int_0^T a_{m-1} b_m Y^{(m-1)}(t) dY^{(m-1)}(t) = -a_{m-1} b_m \frac{\sigma^2 T}{2} + Q(T) - Q(0)$$

où a_{m-1} et b_m sont les coefficients de plus haut degré de P_1 et P_2 . Comme le coefficient de plus haut degré de $P_1(iD) P_2(-iD)$ est $\alpha_{2m-1} = ia_{m-1}b_m$, le lemme est démontré.

Démonstration du théorème 3.3 :

Par définition :

$$L_T(\theta) = \int_0^T \phi_\theta(t) dW(t) - \frac{1}{2} \int_0^T [\phi_\theta(t)]^2 dt$$

où :

$$\phi_\theta(t) = \frac{1}{\sigma} [D^{m-n} p_\theta(D) - q_\theta(D)] Y_\theta(t)$$

$$dW(t) = \frac{1}{\sigma} d [D^{m-n-1} p_\theta(D)] Y_\theta(t)$$

($D = d/dt$ et $Y_\theta(t)$ étant la solution de $p_\theta(D) Y(t) = x(t)$ de conditions initiales nulles).

D'après le lemme 3.3 :

$$L_T(\theta) = -\frac{1}{2\sigma^2} \left\{ \sum_{j=0}^{m-1} \alpha_{2j} \int_0^T [Y_\theta^{(j)}(t)]^2 dt - i\alpha_{2m-1} \frac{\sigma^2 T}{2} \right\} + Q_\theta(T)$$

où $Q_\theta(T)$ est une forme quadratique en $Y_\theta^{(0)}(T), \dots, Y_\theta^{(m-1)}(T)$ et où les α_j sont les coefficients du polynôme :

$$[(iD)^{m-n} p_\theta(iD) - q_\theta(iD)][(-iD)^{m-n} p_\theta(-iD) - q_\theta(-iD)] \\ - 2[(iD)^{m-n} p_\theta(iD) - q_\theta(iD)](-iD)^{m-n} p_\theta(-iD)$$

Un calcul simple montre que le polynôme précédent est égal à :

$$[q_\theta(iD)q_\theta(-iD) - D^{2(m-n)} p_\theta(iD)p_\theta(-iD)] \\ + [(iD)^{m-n} p_\theta(iD) q_\theta(-iD) - (-iD)^{m-n} p_\theta(-iD) q_\theta(iD)]$$

Le deuxième crochet est un polynôme impair de coefficient de plus haut degré $\alpha_{2m-1} = 2i(b_{m-1} - a_{n-1})$. Par suite les α_{2j} sont les coefficients du premier crochet.

Une nouvelle application du lemme 3.3 montre que :

$$- \frac{1}{2\sigma^2} \left\{ \sum_{j=0}^{m-n-1} d_j \int_0^T [D^j p_\theta(D) Y_\theta(t)]^2 dt + \sum_{j=0}^{n-1} \beta_{2j} \int_0^T [Y_\theta^{(j)}(t)]^2 dt \right. \\ \left. + \sigma^2 T(b_{m-1} - a_{n-1}) \right\} = L_T(\theta) + Q_\theta(T).$$

où les β_{2j} sont les coefficients du polynôme R_θ (R_θ ainsi que les d_j étant définis dans le théorème), et où $Q_\theta(T)$ est une forme quadratique en $Y_\theta^{(0)}(T), \dots, Y_\theta^{(m-1)}(T)$.

Nous allons maintenant montrer que :

$$\sum_{j=0}^{n-1} \beta_{2j} \int_0^T [Y_\theta^{(j)}(t)]^2 dt = T \int_{-\infty}^{\infty} \frac{R_\theta(\lambda)}{|p_\theta(i\lambda)|^2} I^T(\lambda) d\lambda + Q'_\theta(T)$$

où $Q'_\theta(T)$ est une forme quadratique en $Y_\theta^{(0)}(T), \dots, Y_\theta^{(n-1)}(T)$

Pour ce faire nous définissons $Y_\theta(t)$ pour $t > T$ par la résolution de l'équation différentielle $p_\theta(d/dt) Y_\theta(t) = 0$, $t > T$, de conditions initiales $Y_\theta^{(j)}(T+0) = Y_\theta^{(j)}(T-0)$, $j = 0, \dots, n-1$, et nous posons $Y_\theta(t) = 0$, $t < 0$. La transformée de Fourier de Y_θ est alors :

$$\tilde{Y}_\theta(\lambda) = \frac{1}{p(-i\lambda)} \int_0^T e^{i\lambda t} x(t) dt$$

Comme la transformée de Fourier de $Y_\theta^{(j)}(t)$ est $(i\lambda)^j \tilde{Y}_\theta(\lambda)$, on a :

$$\begin{aligned} \sum_{j=0}^{n-1} \beta_{2j} \int_0^\infty [Y_\theta^{(j)}(t)]^2 dt &= \sum_{j=0}^{n-1} \beta_{2j} \int_{-\infty}^\infty |\lambda^j \tilde{Y}_\theta(\lambda)|^2 d\lambda \\ &= T \int_{-\infty}^\infty \frac{R_\theta(\lambda)}{|p_\theta(i\lambda)|^2} I^T(\lambda) d\lambda \end{aligned}$$

Une démonstration analogue à celle des lemmes 2.2 et 2.3 montre que :

$$Y_\theta^{(j)}(T+t) = {}^t Z_\theta(T) h_{\theta,j}(t)$$

où $Z_\theta(T) = (Y_\theta^{(0)}(T) \dots Y_\theta^{(n-1)}(T))$ et où $h_{\theta,j} \in [L^2(\mathbb{R}^+, \mathbb{R}^+, dt)]^n$ et l'application $\theta \rightarrow h_{\theta,j}$ est deux fois dérivable. Par suite :

$$Q'_\theta(T) = - \sum_{j=0}^{n-1} \beta_{2j} \int_T^\infty [Y_\theta^{(j)}(t)]^2 dt$$

est une forme quadratique en $Y_\theta^{(j)}(T)$, $j=0, \dots, n-1$, dont les coefficients sont deux fois dérivables par rapport à θ . Un examen de la démonstration du lemme 3.3 montre que les coefficients de $Q_\theta(T)$ sont aussi deux fois dérivables en θ .

On obtient donc le résultat du théorème avec $\eta_T(\theta) = Q_\theta(T) + Q'_\theta(T)$.

Il nous reste à montrer que η_T possède les propriétés énoncées.

Pour ce faire, il suffit de montrer que les dérivées jusqu'à l'ordre 2 de $Y_\theta^{(j)}(x(\cdot), T)$ par rapport à θ existent et que les bornes supérieures de leurs absolues pour θ parcourant un voisinage convenable $V(\theta^*)$ de θ^* , sont des variables aléatoires bornées en probabilité quand $T \rightarrow \infty$.

Or d'après la démonstration du théorème 3.2, $Y_\theta^{(j)}(T)$ est de la forme :

$$\sum_{\ell=0}^{m-n-1} \alpha_\ell(\theta) x^{(\ell)}(T) + \int_0^T \beta_\theta(T-s) x(s) ds$$

où les fonctions α_ℓ sont deux fois continûment dérivables et où β_θ est une fonction du type : $\sum_{r=1}^n \gamma_r(\theta) \exp \lambda_r(\theta) \cdot t$, $\lambda_r(\theta)$ étant les zéros du polynôme p_θ et les fonctions γ_r étant deux fois continûment dérivables.

Il est clair que $Y_\theta^{(j)}(T)$ est deux fois dérivable par rapport à θ et que ses dérivées sont de la forme précédente sauf que les fonctions α_ℓ , γ_r sont seulement continues. Soit $V(\theta^*)$ un voisinage compact de θ^* et $\lambda_0 < 0$ la borne supérieure des parties réelles de $\lambda_r(\theta)$, $r = 1, 2, \dots, n$, $\theta \in V(\theta^*)$, alors on a :

$$\sup_{\theta \in V(\theta^*)} |\beta_\theta(t)| < K e^{\lambda_0 t}$$

D'où la majoration :

$$\theta \in V(\theta^*) : |Y_\theta^{(j)}(T)| < \sum_{\ell=0}^{m-n-1} K_\ell |x^{(\ell)}(T)| + \int_0^T e^{\lambda_0(T-s)} |x(s)| ds$$

et des majorations analogues pour les dérivées de $Y_\theta^{(j)}(T)$ par rapport à θ . Comme le second membre de l'inégalité précédente est de norme L^2 bornée quand $T \rightarrow \infty$, on obtient le résultat cherché.

Démonstration du lemme 4.3 :

Nous avons (Doob [9], Métivier [18]) :

$$\begin{aligned} E_{\theta::} \left[\int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta} d\xi(t) \right]^2 &= \int_0^T E_{\theta::} \left[\left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta::} \right]^2 dt \\ &= T E_{\theta::} \left[\left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(0) \right)_{\theta=\theta::} \right]^2 \end{aligned}$$

D'où la convergence en moyenne quadratique vers 0 quand $T \rightarrow \infty$ de :

$$T^{-1} \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} d\xi(t)$$

De même façon, on obtient la convergence en moyenne quadratique vers 0 quand $T \rightarrow \infty$ de :

$$T^{-1} \int_0^T \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} d\xi(t)$$

Finalement le dernier résultat du lemme est une conséquence du théorème ergodique.

Démonstration du lemme 4.4 :

D'après le théorème 3.3 les variables aléatoires $s_j^T(\theta)$ et $S_{jl}^T(\theta)$ sont de la forme :

$$\sum_{j=0}^{m-n-1} \alpha_j(\theta) \frac{1}{T} \int_0^T [x^{(j)}(t)]^2 dt + \alpha(\theta) + \int_{-\infty}^{\infty} g_\theta(\lambda) I^T(\lambda) d\lambda + \frac{1}{T} \xi_T(\theta)$$

où α et α_j , $j = 0, \dots, m-n-1$ sont des fonctions continues, et où g_θ est une fraction rationnelle dont les coefficients sont des fonctions continues de θ et telle que la différence des degrés du dénominateur et du numérateur est au moins égale à 2, et où enfin $\xi_T(\theta)$ est telle que pour un voisinage convenable $V(\theta^{**})$ de θ^{**} , $\text{Sup}\{|\xi_T(\theta)|, \theta \in V(\theta^{**})\}$ est borné en probabilité quand $T \rightarrow \infty$.

On peut prendre $V(\theta^{**})$ compact, l'application $(\theta, \lambda) \rightarrow g_\theta(\lambda)$ est continue sur le compact $V(\theta^{**}) \times \bar{R}$, où \bar{R} est la droite achevée. On peut montrer que :

$$\begin{aligned} \forall \varepsilon > 0, \exists \eta > 0 : & \|\theta - \theta^{**}\| < \eta \Rightarrow \theta \in V(\theta^{**}) \text{ et :} \\ & |\alpha_j(\theta) - \alpha_j(\theta^{**})| > \varepsilon, \quad j = 0, \dots, m-n-1 ; \quad |\alpha(\theta) - \alpha(\theta^{**})| < \varepsilon \\ & |g_\theta(\lambda) - g_{\theta^{**}}(\lambda)| < \varepsilon \quad \forall \lambda \in R. \end{aligned}$$

Par suite pour $\theta: ||\theta - \theta^{**}|| < \varepsilon$, la différence $S_j^T(\theta)$ et $s_j^T(\theta^{**})$ est majoré en valeur absolue par :

$$\varepsilon \left[1 + \sum_{j=0}^{m-n-1} \frac{1}{T} \int_0^T [x^{(j)}(t)]^2 dt + \int_{-\infty}^{\infty} I^T(\lambda) d\lambda \right] + \frac{1}{T} \sup_{V(\theta^{**})} |\xi_T(\theta) - \xi_T(\theta^{**})|$$

Quand $T \rightarrow \infty$, le crochet et le dernier terme de l'expression précédente convergent en probabilité respectivement vers une constante et vers 0.

On obtient de la même façon des majorations analogues pour la différence $S_{j\ell}^T(\theta) - S_{j\ell}^T(\theta^{**})$. D'où le lemme.

Démonstration du théorème 4.1

La première partie de ce théorème est une conséquence directe du lemme 4.3; la seconde partie découle du fait que si θ_T est un estimateur consistant de θ , $s^T(\theta_T) - s^T(\theta^{**})$ et $S^T(\theta_T) - S^T(\theta^{**})$ tendent en probabilité vers 0 quand $T \rightarrow \infty$. En effet, d'après le lemme 4.4 :

$$\begin{aligned} P\{|s^T(\theta_T) - s^T(\theta^{**})| > \varepsilon A\} &< P\{\{||\theta_T - \theta^{**}|| > \eta\} \cup \{M_T > A\}\} \\ &< P\{||\theta_T - \theta^{**}|| > \eta\} + P\{M_T > A\} \end{aligned}$$

On peut choisir A de façon que $P\{M_T > A\} \rightarrow 0$ quand $T \rightarrow \infty$. D'où le résultat.

Démonstration du théorème 4.2

Pour alléger les écritures, on pose $\psi_j(t) = (\partial \bar{\phi} / \partial \theta_j)_{\theta = \theta^{**}}$. D'après les lemmes 4.1 et 4.2 :

$$\sqrt{T} s^T(\theta^{**}) = T^{-1/2} \int_0^T \psi(t) d\bar{\xi}(t) + o(T)$$

où $o(T) \rightarrow 0$ en probabilité quand $T \rightarrow \infty$.

Il suffit donc d'étudier la convergence de $T^{-1/2} \int_0^T \psi(t) d\bar{\xi}(t)$.

Ce dernier vecteur aléatoire peut être approché par :

$$T^{-1/2} \sum_{j=0}^{[nT]-1} \psi\left(\frac{j}{n}\right) \left[\bar{\xi}\left(\frac{j+1}{n}\right) - \bar{\xi}\left(\frac{j}{n}\right) \right] = \left(\frac{[nT]}{nT}\right)^{1/2} S_{[nT],n}$$

où $[nT]$ est la partie entière de nT et où :

$$S_{N,m} = N^{-1/2} \sum_{j=0}^{N-1} \psi\left(\frac{j}{n}\right) \left[\bar{\xi}\left(\frac{j+1}{n}\right) - \bar{\xi}\left(\frac{j}{n}\right) \right]$$

L'erreur d'approximation est :

$$R_{T,n} = T^{-1/2} \left\{ \sum_{j=0}^{[nT]-1} \int_0^{1/n} \left[\psi\left(\frac{j}{n} + u\right) - \psi\left(\frac{j}{n}\right) \right] d\bar{\xi}(u) + \int_{[nT]/n}^T \psi(t) d\bar{\xi}(t) \right\}$$

Nous allons montrer que :

$$(i) \quad T_0 : \forall \varepsilon > 0, \forall n < 0, \quad n_0 : T > T_0, n > n_0 \quad P(|R_{T,n}| > \varepsilon) < n^{-1}$$

(ii) Quand $T \rightarrow \infty$, $S_{[nT],n}$ converge en loi vers un vecteur aléatoire gaussien centré de matrice de covariance S .

Le résultat du théorème découlera alors du lemme de Bernstein (voir p. 99).

Montrons (i). On a en utilisant la stationnarité de $\{\psi(t), t \in \mathbb{R}\}$:

$$\begin{aligned} E |R_{T,n}|^2 &= T^{-1} \left\{ [nT] \int_0^{1/n} E_{\theta} \left| \psi(u) - \psi(0) \right|^2 du + \frac{nT - [nT]}{E} E_{\theta} \left| \psi(0) \right|^2 \right\} \\ &\leq \frac{[nT]}{nT} K_n + \frac{nT - [nT]}{nT} E_{\theta} \left| \psi(0) \right|^2 \end{aligned}$$

où $K_n = \text{Sup}\{E_{\theta} \left| \psi(u) - \psi(0) \right|^2 ; 0 < u < 1/n\}$. Comme $K_n \rightarrow 0$ quand $n \rightarrow \infty$ car le processus $\{\psi(t), t \in \mathbb{R}\}$ est continue en moyenne quadratique,

$E_{\theta} \left| R_{T,n} \right|^2 \rightarrow 0$ quand $n \rightarrow \infty$, uniformément pour tout $T > T_0$. D'où le résultat.

Montrons maintenant (ii). Le processus $\{\psi(t), t \in \mathbb{R}\}$ admet la représentation spectrale.

$$\psi_j(t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \left[\left(\frac{\partial}{\partial \theta_j} \frac{q_\theta(-i\lambda)}{p_\theta(-i\lambda)} \right)_{\theta=\theta^*} \frac{p_{\theta^*}(-i\lambda)}{q_{\theta^*}(-i\lambda)} \right] dZ(\lambda)$$

où la fonction dans le crochet est analytique dans le demi-plan supérieur du plan complexe. Par conséquent $\psi(t)$ peut être exprimé en fonction de $\bar{\xi}(s)$, $s \leq t$ (voir les arguments du § 3). D'où :

$$E_\theta \left\{ \psi\left(\frac{j}{n}\right) \left[\bar{\xi}\left(\frac{j+1}{n}\right) - \bar{\xi}\left(\frac{j}{n}\right) \right] \mid \bar{\xi}(s), s \leq \frac{j}{n} \right\} = 0$$

Si on pose :

$$V_j = n^{1/2} \psi\left(\frac{j}{n}\right) \left[\bar{\xi}\left(\frac{j+1}{n}\right) - \bar{\xi}\left(\frac{j}{n}\right) \right]$$

$$\text{alors : } E_\theta (V_j \mid V_{j-1}, V_{j-2}, \dots) = E_\theta [E_\theta (V_j \mid \bar{\xi}(s), s < j/n) \mid V_{j-1}, \dots] = 0$$

comme la suite $\{V_j\}$ est stationnaire ergodique, le théorème de Billingsley [5], p. 206) montre que :

$$S_{N,n} = N^{-1/2} \sum_{j=0}^{N-1} V_j$$

converge en loi quand $N \rightarrow \infty$ vers un vecteur aléatoire gaussien centré de matrice de covariance :

$$E_\theta [V_0 {}^t V_0] = E_\theta [\psi(0) {}^t \psi(0)] = S$$

Démonstration du lemme 5.2 :

Nous avons :

$$\phi_\theta(t) = \frac{1}{\sigma} \left[\left(\frac{d}{dt} \right)^{m-n} p_\theta \left(\frac{d}{dt} \right) - q_\theta \left(\frac{d}{dt} \right) \right] y_\theta(t)$$

où $y_\theta(t)$ est la solution, de conditions initiales nulles, de $p_\theta(d/dt) y(t) = x(t)$

Par suite :

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \phi_\theta(t) &= \frac{1}{\sigma} \left[\left(\frac{d}{dt}\right)^{m-n} \frac{\partial}{\partial \theta_j} p_\theta \left(\frac{d}{dt}\right) - \frac{\partial}{\partial \theta_j} q_\theta \left(\frac{d}{dt}\right) \right] y_\theta(t) \\ &+ \frac{1}{\sigma} \left[\left(\frac{d}{dt}\right)^{m-n} p_\theta \left(\frac{d}{dt}\right) - q_\theta \left(\frac{d}{dt}\right) \frac{\partial}{\partial \theta_j} \right] y(t) \end{aligned}$$

Or, il est facile de voir que $\partial y_\theta(t)/\partial \theta_j$ est la solution, de conditions initiales nulles, de

$$p_\theta \left(\frac{d}{dt}\right) u(t) + \frac{\partial}{\partial \theta_j} p_\theta \left(\frac{d}{dt}\right) y_\theta(t) = 0$$

Par suite si z_θ est la solution, de conditions initiales nulles, de :

$$p_\theta \left(\frac{d}{dt}\right) z(t) = y_\theta(t) \quad \text{i.e.} \quad [p_\theta \left(\frac{d}{dt}\right)]^2 z(t) = x(t)$$

alors on a :

$$\frac{\partial}{\partial \theta_j} y_\theta(t) = - \frac{\partial}{\partial \theta_j} p_\theta \left(\frac{d}{dt}\right) z_\theta(t)$$

Exprimons maintenant $\partial \phi_\theta(t)/\partial \theta_j$ en fonction de z_θ :

$$\frac{\partial}{\partial \theta_j} \phi_\theta(t) = \frac{1}{\sigma} \left[q_\theta \left(\frac{d}{dt}\right) \frac{\partial}{\partial \theta_j} p_\theta \left(\frac{d}{dt}\right) - \frac{\partial}{\partial \theta_j} q_\theta \left(\frac{d}{dt}\right) p_\theta \left(\frac{d}{dt}\right) \right] z_\theta(t)$$

D'après le lemme 3.3, on a :

$$\int_0^T \frac{\partial}{\partial \theta_j} \phi_\theta(t) \frac{\partial}{\partial \theta_\ell} \phi_\theta(t) dt = \frac{1}{\sigma^2} \sum_{j=0}^{n+m-1} \alpha_{2j} \int_0^T [z_\theta^{(j)}(t)]^2 dt + Q_\theta(T)$$

où $Q_\theta(T)$ une forme quadratique en $z_\theta^{(0)}(T), \dots, z_\theta^{(m+n-1)}(T)$ et où les α_k sont les coefficients du polynôme :

$$[p_\theta(i\lambda)]^2 \frac{\partial}{\partial \theta_j} \left(\frac{q_\theta(i\lambda)}{p_\theta(i\lambda)} \right) \times [p_\theta(-i\lambda)]^2 \frac{\partial}{\partial \theta_\ell} \left(\frac{q_\theta(-i\lambda)}{p_\theta(-i\lambda)} \right)$$

En écrivant le polynôme précédent sous la forme :

$$\sum_{j=0}^{2(m-n-1)} C_j \lambda^j |p_\theta(i\lambda)|^4 + \sum_{j=0}^{2n-1} \beta_j \lambda^j$$

Une nouvelle application du lemme 3.3. montre que :

$$\sum_{j=0}^{m-n-1} C_{2j} \int_0^T [x^{(j)}(t)]^2 dt + \sum_{j=0}^{n-1} \beta_{2j} \int_0^T [z_\theta^{(j)}(t)]^2 dt =$$

$$\sum_{j=0}^{m-n-1} \alpha_{2j} \int_0^T z_\theta^{(j)}(t)^2 dt + Q'_\theta(T)$$

où $Q'_\theta(T)$ est encore une forme quadratique en $z_\theta^{(j)}(T)$.

En répétant la même démonstration que celle du théorème 3.3, et en remarquant que $\sum_{j=0}^{n-1} \beta_{2j} \lambda^{2j} = \mathcal{R}e g_\theta(\lambda)$, on obtient le résultat.

Démonstration du lemme 5.3.

Nous répétons les mêmes calculs que ceux du début de la démonstration du lemme 5.2 en remarquant que :

(i) p_θ ne dépend que de a avec $\partial p_\theta(\cdot)/\partial a_j = p_j(\cdot)$

(ii) q_θ ne dépend que de b avec $\partial q_\theta(\cdot)/\partial b_j = q_j(\cdot)$

(iii) $y(\cdot)$ est la solution, de conditions initiales nulles, de $p(d/dt)y(t) = x(t)$ et par suite ne dépend que de a .

Ainsi on obtient :

$$\frac{\partial}{\partial a_j} \phi_\theta(t) = \frac{1}{\sigma} \left(\frac{d}{dt}\right)^{m-n} p_j\left(\frac{d}{dt}\right) y_\theta(t) + \frac{1}{\sigma} \left[\left(\frac{d}{dt}\right)^{m-n} p_\theta\left(\frac{d}{dt}\right) - q_\theta\left(\frac{d}{dt}\right)\right] \cdot \frac{\partial}{\partial a_j} y_\theta(t)$$

$$\frac{\partial}{\partial b_j} \phi_\theta(t) = -\frac{1}{\sigma} q_j\left(\frac{d}{dt}\right) y_\theta(t)$$

et :

$$\frac{\partial}{\partial a_j} y_\theta(t) = - p_j \left(\frac{d}{dt} \right) v_\theta(t)$$

où v_θ est la solution, de conditions initiales nulles, de $p_\theta(d/dt) v_\theta(t) = y_\theta(t)$

Par suite :

$$\frac{\partial}{\partial a_j} \phi_\theta(t) = - q_\theta \left(\frac{d}{dt} \right) p_j \left(\frac{d}{dt} \right) v_\theta(t)$$

Posons $\tilde{z}_\theta(t) = q_\theta(d/dt) v_\theta(t)$. La fonction \tilde{z}_θ est $n-1$ fois dérivable. En tenant compte de la relation $y_\theta(t) = p_\theta(d/dt) v_\theta(t)$, un calcul élémentaire montre que :

$$\begin{aligned} \tilde{z}_\theta^{(n-1)}(t) - \tilde{z}_\theta^{(n-1)}(0) + \int_0^t \tilde{p}_\theta \left(\frac{d}{ds} \right) z_\theta(s) ds &= y_\theta^{(m-1)}(t) - y_\theta^{(m-1)}(0) + \\ &\int_0^t \tilde{q}_\theta \left(\frac{d}{ds} \right) y_\theta(s) ds \end{aligned}$$

Les $z_\theta^{(j)}(0)$, $j = 0, 1, \dots, n-1$ ne sont pas tous nuls. En effet, d'après la relation $p_\theta(d/dt) v_\theta(t) = y_\theta(t)$, les $v_\theta^{(n+j)}(0)$, $j = 0, \dots, m-1$ sont des combinaisons linéaires de $y_\theta^{(\ell)}(0)$, $\ell < j$, et d'après la relation $p_\theta(d/dt) y_\theta(t) = x(t)$, $y_\theta^{(\ell)}(0)$ est nul si $\ell < n$ et est une combinaison linéaire de $x^{(0)}(0), \dots, x^{(\ell-n)}(0)$ sinon. Par suite $\tilde{z}_\theta^{(j)}(0)$ est nul si $j+m-n < n$ et est une combinaison linéaire de $x^{(h)}(0)$ sinon.

Par conséquent, si z_θ est la solution, de conditions initiales nulles,

de :

$$\int_0^t [dz_\theta^{(n-1)}(s) + \tilde{p}_\theta \left(\frac{d}{ds} \right) z_\theta(s) ds] = \int_0^t [dy_\theta^{(m-1)}(s) + \tilde{q}_\theta \left(\frac{d}{ds} \right) y_\theta(s) ds]$$

alors :

$$\tilde{z}_\theta(t) = z_\theta(t) + h_\theta(t) \cdot x_0 \quad \text{où } x_0 = {}^t(x^{(0)}(0) \dots x^{(m-n-1)}(0)) \text{ et où}$$

les composantes de $h_\theta(\cdot)$ sont des solutions, de conditions initiales données, de $p_\theta(d/dt) z(t) = 0$

La démonstration du théorème 3.2 montre que les $h_{\theta,j}(\cdot)$ sont de la forme $\sum \gamma_r(\theta) \exp \lambda_r(\theta)$ où $\lambda_r(\theta)$ sont les zéros de p_θ et où γ_r sont des fonctions continues de θ .

Démonstration du théorème 5.2

L'assertion (5.4) est prouvée si on montre que :

$$\begin{aligned} \frac{1}{\sqrt{T}} \int_0^T h_{\theta_T, j}(t) d\xi_{\theta_T}(t) &= \frac{1}{\sqrt{T}} \int_0^T h_{\theta_T, j}(t) d\xi_{\theta^*}(t) + \\ &+ \frac{1}{\sqrt{T}} \int_0^T h_{\theta_T, j}(t) [\phi_{\theta^*}(t) - \phi_{\theta_T}(t)] dt \end{aligned}$$

tend vers 0 quand $T \rightarrow \infty$.

Nous commençons par montrer qu'il existe un voisinage $V(\theta^*)$ de θ tel que :

$$\theta \in V(\theta^*) : \left| \int_0^T h_{\theta, j}(t) d\xi_{\theta^*}(t) \right| < Z_T$$

où Z_T est une variable aléatoire bornée en norme L^2 pour tout T .

On a :

$$\int_0^T h_{\theta, j}(t) d\xi_{\theta^*}(t) = h_{\theta, j}(T) \xi_{\theta^*}(T) - \int_0^T h'_{\theta, j}(t) \xi_{\theta^*}(t) dt$$

D'après le résultat du lemme 5.3, il existe un voisinage $V(\theta^*)$ de θ^* tel que :

$$\theta \in V(\theta^*) : |h_{\theta, j}(t)| < K e^{\lambda_0 t} ; |h'_{\theta, j}(t)| < K' e^{\lambda_0 t}$$

où $\lambda_0 < 0$ et K, K' sont des constantes. Par suite, le second membre de l'inégalité précédente est majoré par :

$$Z_T = K(\exp \lambda_0 T) |\xi_{\theta^*}(T)| - \int_0^T (K' \exp \lambda_0 t) \xi_{\theta^*}(t) dt$$

En utilisant le fait que l'intégrale du deuxième terme est aussi définie au sens de la moyenne quadratique, on obtient :

$$\begin{aligned} \{E \left[\int_0^T (\exp \lambda_0 t) |\xi_{\theta^*}(t)| dt \right]^2\}^{1/2} &\leq \int_0^T (\exp \lambda_0 t) E |\xi_{\theta^*}(t)|^2 dt \\ &\leq \int_0^T t e^{\lambda_0 t} dt < C \end{aligned}$$

Par suite Z_T est bornée en norme L^2 .

Maintenant, on a :

$$\frac{1}{T} \left| \int_0^T h_{\theta,j}(t) |\phi_{\theta^*}(t) - \phi_{\theta}(t)| dt \right|^2 < \frac{1}{T} \int_0^T h_{\theta,j}^2(t) dt \int_0^T [\phi_{\theta^*}(t) - \phi_{\theta}(t)]^2 dt$$

Or :

$$\frac{1}{T} \int_0^T |\phi_{\theta^*}(t) - \phi_{\theta_T}(t)|^2 dt = \sum_{j,l} (\theta_{T,j} - \theta_j)(\theta_{T,l} - \theta_l) \frac{1}{T} \int_0^T \left(\frac{\partial}{\partial \theta_j} \phi_{\theta}(t) \frac{\partial}{\partial \theta_l} \phi_{\theta}(t) \right)_{\theta=\theta_T} dt$$

où θ_T est un point sur le segment joignant θ_T et θ . Le lemme 5.1 montre alors que le second membre de l'égalité précédente tend vers 0 quand $T \rightarrow \infty$. L'assertion (5.4) est donc démontrée.

Maintenant pour démontrer (5.6) et (5.7), il suffit de montrer que :

$$\frac{1}{T} \int_0^T h_{\theta_T,j}(t) h_{\theta_T,l}(t) dt \quad \text{et} \quad \frac{1}{T} \int_0^T h_{\theta_T,j}(t) \left(\frac{\partial}{\partial \theta_l} \phi_{\theta}(t) \right)_{\theta=\theta_T} dt$$

convergent en probabilité vers 0 quand $T \rightarrow \infty$. Or ceci résulte du fait que l'intégrale $\int_0^T h_{\theta,j}^2(t) dt$ est majorée par une fonction continue de θ et que :

$$\frac{1}{T} \int_0^T \left(\frac{\partial}{\partial \theta_l} \phi_{\theta}(t) \right)_{\theta=\theta_T}^2 dt$$

converge vers une constante quand $T \rightarrow \infty$ (lemme 5.1). D'où le résultat.

CHAPITRE V

EXTENSION DE LA METHODE D'ESTIMATION DU CHAPITRE PRECEDENT

Nous considérons ici deux extensions du modèle de processus gaussien stationnaire centré de densité spectrale rationnelle. Le premier est le modèle multivariable et le second le modèle de régression à résidus stationnaires.

§ 1. ESTIMATION DANS LE MODELE D'EVOLUTION STOCHASTIQUE LINEAIRE

Un modèle de processus couramment employé pour décrire le comportement d'un système dynamique stochastique est défini par l'équation différentielle stochastique linéaire :

$$(1.1) \quad dX(t) = AX(t) dt + Bd\xi(t)$$

où $X(t)$ est un vecteur aléatoire à valeurs dans R^p décrivant l'état du système et où $\{\xi(t), t > 0\}$ est un mouvement brownien standard.

Etant donnée la loi $P^{X(0)}$ du vecteur $X(0)$ l'équation (1.1) détermine complètement la loi du processus.

Nous allons montrer que le modèle précédent contient comme cas particulier le modèle de processus gaussien stationnaire centré de densité spectrale :

$$(1.2) \quad f(\lambda) = \frac{1}{2\pi} (A+i\lambda I)^{-1} B^t B(A+i\lambda I)^{-1}$$

En effet, un tel processus admet la représentation spectrale :

$$X(t) = \int_{-\infty}^{+\infty} e^{-i\lambda t} (A+i I)^{-1} B dZ(\lambda)$$

où $\{Z(\lambda); \lambda \in \mathbb{R}\}$ est un processus à accroissements orthogonaux avec

$$E\{[Z(\lambda_2)-Z(\lambda_1)][Z(\lambda_2)-Z(\lambda_1)]^*\} = (\lambda_2-\lambda_1)I. \text{ Par suite, le processus :}$$

$$\xi(t) = \int_{-\infty}^{+\infty} \frac{e^{-i\lambda t}-1}{i\lambda} dZ(\lambda)$$

est un mouvement brownien standard. Comme :

$$X(t) - X(0) - \int_0^t AX(s) ds = \int_{-\infty}^{+\infty} \frac{e^{-i\lambda t}-1}{i\lambda} B dZ(\lambda) = B\xi(t)$$

On voit que le processus $X(t)$, $t \geq 0$ vérifie (1.1). Si donc la loi du vecteur initial $P^{X(0)}$ est convenablement donnée, la résolution de (1.1) fournit un processus ayant même loi que le processus $X(t)$, $t \geq 0$ précédent.

1.1. Structure statistique

Considérons l'équation (1.1) où la matrice A dépend d'un paramètre $\theta \in \Theta \subset \mathbb{R}^k$ que l'on notera maintenant A_θ . La résolution de cette équation avec la loi du vecteur initial $P_\theta^{X(0)}$ (qui dépend éventuellement de θ) fournit une loi de probabilité sur $(C([0,T], \mathbb{R}^P), \mathcal{A})$ où $C([0,T], \mathbb{R}^P)$ est l'espace des fonctions continues de $[0,T]$ à valeurs dans \mathbb{R}^P et \mathcal{A} la tribu trace de la tribu $\mathcal{B}_{\mathbb{R}^P} \otimes [0,T]$ ($\mathcal{B}_{\mathbb{R}^P}$ étant la tribu des boreliens de \mathbb{R}^P). Cette loi ainsi définie dépend de deux paramètres θ et $G = B^t B$ et sera notée $P_{\theta,G}^X$. Notre problème est l'estimation de θ et G dans la structure statistique :

$$(C([0,T], \mathbb{R}^P), \mathcal{A}, \{P_{\theta,G}^X, \theta \in \Theta, G \in \mathcal{Y}\})$$

où \mathcal{Y} est l'espace des matrices symétriques définies positives.

En fait, on peut déterminer G avec probabilité 1 au vu d'une seule observation. Cela vient du fait que pour tout mouvement brownien scalaire standard $\{W(t), t \geq 0\}$ on a :

$$\frac{1}{T} \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left[W\left(\frac{kT}{2^n}\right) - W\left(\frac{(k-1)T}{2^n}\right) \right]^2 = 1 \quad \text{presque sûrement.}$$

En utilisant l'égalité $4a.b = (a+b)^2 - (a-b)^2$ nous déduisons que si $\{\xi(t), t \geq 0\}$ est un mouvement brownien vectoriel standard :

$$\frac{1}{T} \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left[\xi\left(\frac{kT}{2^n}\right) - \xi\left(\frac{(k-1)T}{2^n}\right) \right]^t \left[\xi\left(\frac{kT}{2^n}\right) - \xi\left(\frac{(k-1)T}{2^n}\right) \right] = I \quad \text{p.s}$$

D'autre part si $X(t), t \geq 0$ est le processus solution de (1.1) dont les trajectoires sont continues, alors $X(t) = V(t) + \xi(t)$ où $V(t)$ est un processus à trajectoires continûment dérivables, ce qui entraîne :

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left\| V\left(\frac{k}{2^n} T\right) - V\left(\frac{(k-1)}{2^n} T\right) \right\|^2 = 0 \quad \text{p.s}$$

On en déduit, en se servant de l'inégalité de Schartz :

$$\frac{1}{T} \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} \left[X\left(\frac{k}{2^n} T\right) - X\left(\frac{(k-1)}{2^n} T\right) \right]^t \left[X\left(\frac{kT}{2^n}\right) - X\left(\frac{(k-1)}{2^n} T\right) \right] = G \quad \text{p.s}$$

Nous pouvons donc supposer G connue et nous écrivons P_{θ}^X à la place de $P_{\theta, G}^X$. La structure statistique considérée devient :

$$(\mathcal{C}([0, T], \mathbb{R}^p), \mathcal{A}, \{P_{\theta}^X, \theta \in \Theta\})$$

Nous faisons dans toute la suite l'hypothèse que pour tout $\theta \in \Theta$, les parties réelles des valeurs propres de A_{θ} sont strictement négatives et que l'application $\theta \rightarrow A_{\theta}$ est deux fois continûment dérivable.

1.2. Fonction de vraisemblance et estimation de paramètre

Rappelons d'abord que le théorème de Girsanov [13] est aussi valable dans le cas vectoriel. Il s'énonce de même façon que le théorème 2.1 du chapitre IV avec des modifications évidents, à savoir :

- (i) $\eta(t)$, $t \in [0, T]$ est un mouvement brownien à valeurs dans R^p ,
- (ii) ϕ_θ est une application de $\Omega \times [0, T]$ dans R^p ,
- (iii) $||\phi_\theta(t)||^2$ remplace $|\phi_\theta(t)|^2$ et $\int_0^T \phi_\theta(t) d\eta(t)$ remplace $\int_0^T \phi_\theta(t) d\eta(t)$.

Nous noterons $\tilde{X}(t) : x(\cdot) \rightarrow x(t)$, $t \geq 0$ le processus des "coordonnées" et $\phi_\theta(t) = G^{-1/2} A X(t)$, $W(t) = G^{-1/2} [X(t) - X(0)]$; alors il est clair que le processus :

$$(1.3) \quad \xi_\theta(t) = W(t) - \int_0^t \phi_\theta(s) ds$$

est un mouvement brownien sur $(C([0, T], R^p), \mathcal{A}, P_\theta^X)$

Avec des arguments analogues à ceux de la démonstration du théorème 2.2 du chapitre IV, on déduit :

Théorème 1.1 :

"Il existe une probabilité unique $P_{0, \theta}$ sur $(C([0, T], R^p), \mathcal{A})$ telle que relativement à cette probabilité $X(0)$ est de loi $P_{X(0)}^{\theta}$ et $\{W(t), t \in [0, T]\}$ est un mouvement brownien indépendant de $X(0)$ et :

$$\frac{dP_\theta^X}{dP_{0, \theta}} = \exp \int_0^T \phi_\theta(t) dW(t) - \frac{1}{2} \int_0^T ||\phi_\theta(t)||^2 dt$$

où $\int_0^T \phi_\theta(t) dW(t)$ est l'intégrale stochastique de Ito par rapport au mouvement brownien $(\{W(t), t > 0\}, P_{0, \theta})$; cette intégrale peut être calculée relativement à la loi $P_{\theta^{**}}$ où $\theta^{**} \in \Theta$ en décomposant d'après (1.3), $W(t)$ en une somme d'un mouvement brownien $\xi_{\theta^{**}}(t)$ et un processus à trajectoires à variation bornée $\int_0^t \phi_{\theta^{**}}(s) ds$ "

Supposons maintenant que les lois $P_\theta^{X(0)}$, $\theta \in \Theta$, admettent des densités f_θ par rapport à une mesure ν sur R^P . En remarquant que $C([0, T], R^P)$ est en bijection avec $R^P \times C_0([0, T], R^P)$ par la correspondance $x(\cdot) \rightarrow (x(0), x(\cdot) - x(0))$ où $C_0([0, T], R^P)$ est le sous-espace de $C([0, T], R^P)$ formé par les fonctions nulles en 0, il existe une mesure unique μ sur $C([0, T], R^P)$ telle que son image par la bijection précédente soit le produit de ν avec la probabilité image du mouvement brownien. Le théorème précédent montre que :

Corollaire :

"Le logarithme de la densité de P_θ^X par rapport à μ s'écrit :

$$\text{Log } f_\theta + \text{Tr } G^{-1} \left[\int_0^T d\tilde{X}(t) {}^t\tilde{X}(t) \right] {}^t A_\theta - \frac{1}{2} G^{-1} A_\theta \left[\int_0^T \tilde{X}(t) {}^t\tilde{X}(t) dt \right] {}^t A_\theta "$$

Comme f_θ ainsi que ses dérivées jusqu'à l'ordre 2 par rapport à θ , supposées exister, ne dépendent pas de T , une approximation asymptotique de la fonction log-vraisemblance est :

$$L_T = T \text{Tr} [G^{-1} R_T' {}^t A_\theta - \frac{1}{2} G^{-1} A_\theta R_T] {}^t A_\theta$$

où :

$$R_T' = T^{-1} \int_0^T d\tilde{X}(t) {}^t\tilde{X}(t) ; R_T = T^{-1} \int_0^T X(t) X(t) dt.$$

L'estimateur de θ est défini par la résolution du système :

$$(1.4) \quad T^{-1} \frac{\partial L_T}{\partial \theta_j} = \text{Tr} [G^{-1} (R_T' - A_\theta R_T) \cdot (-\frac{\partial}{\partial \theta_j} {}^t A_\theta)] = 0$$

Dans le cas où θ est l'ensemble des termes de la matrice A , l'estimateur A de A est alors :

$$(1.5) \quad A = R_T' R_T^{-1}$$

1.3. Propriétés asymptotiques de l'estimateur

Nous commençons par montrer que les dérivées de L_T possèdent les propriétés asymptotiques habituelles d'une fonction de vraisemblance. Pour

cela nous construisons un espace probabilisé de référence (Ω, \mathcal{F}, P) sur lequel est défini un processus gaussien stationnaire centré $\{Y(t), t \in \mathbb{R}\}$ de densité spectrale :

$$\frac{1}{2\pi} (A_{\theta^{**}} + i\lambda I)^{-1} G(A_{\theta^{**}} + i\lambda I)^{-1}$$

où θ^{**} est la vraie valeur de θ . Ce processus vérifie l'équation différentielle :

$$(1.6) \quad dY(t) = A_{\theta^{**}} Y(t) dt + G^{1/2} d\xi(t) \quad t > 0$$

où $\xi(t), t \geq 0$ est un mouvement brownien standard.

En augmentant éventuellement l'espace (Ω, \mathcal{F}, P) , nous pouvons construire un vecteur aléatoire $Z(0)$ sur (Ω, \mathcal{F}, P) , indépendant de $\xi(t), t \geq 0$ et $Y(0)$ et de loi $P_{\theta^{**}}^{X(0)}$. Nous définissons ensuite le processus $Z(t), t \geq 0$ comme étant la solution de (1.6) de condition initiale $Z(0)$. Maintenant, pour étudier le comportement asymptotique des dérivées de L_T , nous devons plonger l'espace $(C(0, T, \mathbb{R}^p), \mathcal{A}, P_{\theta^{**}})$ dans $(C(\mathbb{R}^+, \mathbb{R}^p), \tilde{\mathcal{A}}, \tilde{P})$ où $\tilde{\mathcal{A}}$ est la tribu trace sur $C(\mathbb{R}^+, \mathbb{R}^p)$ de $\mathcal{B}_{\mathbb{R}^p}^{\mathbb{R}^+}$ et \tilde{P} est précisément la loi de $Z(t), t \geq 0$. Par suite, la loi de probabilité de la famille des matrices aléatoires $\{R'_T, R_T, T \geq 0\}$ est la même que celle de la famille : $\{\tilde{R}'_T, \tilde{R}_T, T \geq 0\}$ où

$$\tilde{R}'_T = T^{-1} \int_0^T dZ(t) {}^t Z(t) \quad ; \quad \tilde{R}_T = T^{-1} \int_0^T Z(t) Z(t) dt$$

En ce qui concerne les propriétés de convergence en probabilité ou en loi, on peut travailler sur $(\tilde{R}'_T, \tilde{R}_T)$ à la place de (R'_T, R_T) .

Maintenant comme $\{Y(t), t \geq 0\}$ et $\{Z(t), t \geq 0\}$ vérifie la même équation différentielle (1.6), on a :

$$Z(t) = Y(t) + [Z(0) - Y(0)] e^{A_{\theta^{**}} t}$$

En remarquant que les parties réelles des valeurs propres de A_{θ} sont strictement négatives, on a :

Lemme 1.1.

$$\tilde{R}_T = T^{-1} \left[\int_0^T Y(t) {}^t Y(t) dt + O(T) \right]$$

$$\begin{aligned} \tilde{R}'_T - A_{\theta} \tilde{R}_T &= T^{-1} \left[\int_0^T dY(t) {}^t Y(t) - A_{\theta} \int_0^T Y(t) {}^t Y(t) dt + O(T) \right] \\ &= T^{-1} \left[\int_0^T G^{1/2} d\xi(t) {}^t Y(t) dt + O(T) \right] \end{aligned}$$

où $O(T)$ est une matrice aléatoire bornée en probabilité quand $T \rightarrow \infty$.

Une démonstration détaillée de ce lemme se trouve dans Le Breton [17]. A l'aide de ce lemme et avec une technique analogue à celle des démonstrations du lemme 4.3 et du théorème 4.2 du chapitre IV, on montre :

Théorème 1.2

"Quand $T \rightarrow \infty$:

- (i) R_T converge en probabilité vers $R(0)$ où $R(\cdot)$ est la fonction de covariance du processus stationnaire $\{Y(t), t \in R\}$.
- (ii) $R'_T - A_{\theta} R_T$ converge en probabilité vers 0.
- (iii) $T(R'_T - A_{\theta} R_T)$ converge en loi vers une matrice aléatoire gaussienne centrée M de structure de covariance $E(M_{ij} M_{kl}) = G_{ik} R(0)_{jl}$ "

Maintenant les dérivées premières de $T^{-1} L_T$ sont données par (1.4) et les dérivées secondes de $T^{-1} L_T$ sont :

$$\begin{aligned} (1.7) \quad T^{-1} \frac{\partial^2}{\partial \theta_j \partial \theta_l} L_T &= \text{Tr} \left[G^{-1} (R'_T - A_{\theta} R_T) \left(\frac{\partial^2}{\partial \theta_j \partial \theta_l} {}^t A_{\theta} \right) \right] \\ &\quad - \text{Tr} \left[G^{-1} \left(\frac{\partial}{\partial \theta_l} A_{\theta} \right) R_T \left(\frac{\partial}{\partial \theta_j} {}^t A_{\theta} \right) \right] \end{aligned}$$

En tenant compte de (1.4) et (1.7) le théorème 1.2 montre que :

Théorème 1.3 :

"Le vecteur $s^T(\theta)$ des dérivées premières et la matrice $S^T(\theta)$ des dérivées secondes de $-T^{-1}L_T$ vérifient :

(i) Quand $T \rightarrow \infty$, $s^T(\theta^{**})$ et $S^T(\theta^{**})$ convergent respectivement en probabilité vers 0 et vers la matrice S de terme général :

$$S_{j\ell} = \text{Tr} \left[G^{-1} \left(\frac{\partial}{\partial \theta_\ell} A_\theta \right)_{\theta=\theta^{**}} R(0) \left(\frac{\partial}{\partial \theta_j} {}^t A_\theta \right)_{\theta=\theta^{**}} \right]$$

(ii) Pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que

$$\sup_{\theta: \|\theta - \theta^{**}\| < \eta} \|s^T(\theta) - s^T(\theta^{**})\| \leq \epsilon \cdot M_T; \quad \sup_{\theta: \|\theta - \theta^{**}\| < \eta} \|S^T(\theta) - S^T(\theta^{**})\| \leq \epsilon M'_T$$

où M_T, M'_T convergent en probabilité vers des constantes quand $T \rightarrow \infty$ et par suite le point (i) reste vrai quand θ^{**} est remplacé par un estimateur consistant θ_T .

(iii) Le vecteur aléatoire $T^{1/2} s^T(\theta^{**})$ converge en loi quand $T \rightarrow \infty$ vers la loi gaussienne centrée de matrice de covariance S'' .

On peut donc utiliser la technique de construction de l'estimateur décrite dans les chapitres I au IV qui fournit un estimateur asymptotiquement normal et efficace de θ .

Dans le cas où θ est l'ensemble des termes de la matrice A , on déduit directement du théorème 1.2 que l'estimateur de maximum de vraisemblance $\hat{A} = R'_T \cdot R_T^{-1}$ est : (i) consistant, (ii) asymptôtiquement normal de matrice de covariance Λ de terme général :

$$\Lambda_{(i,j)(k,\ell)} = G_{ik} [R^{-1}(0)]_{j\ell}$$

§ 2. ESTIMATION DANS LE MODELE DE REGRESSION A RESIDUS STATIONNAIRES

2.1. Définition du modèle et structure statistique

Le modèle qui nous intéresse est de la forme :

(2.1) $X(t) = m(t) + U(t)$

où $\{U(t), t \in \mathbb{R}\}$ est un processus stationnaire centré appelé processus résiduel et où $m(\cdot)$ est une fonction de la forme :

$$m(t) = \sum_{j=1}^{\ell} v_j z_j(t)$$

On suppose en plus que le processus résiduel $\{U(t), t \in \mathbb{R}\}$ est gaussien centré stationnaire de densité spectrale :

$$f_{\theta}(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{p_{\theta}(i\lambda)}{q_{\theta}(i\lambda)} \right|^2$$

où p_{θ}, q_{θ} sont deux polynômes de coefficients de plus haut degré unité, de degrés respectifs n et m ($m > n$), dépendant d'un paramètre $\theta \in \mathbb{C} \subset \mathbb{R}^k$ dont les coefficients sont des fonctions trois fois continûment dérivables de θ et dont les zéros ont des parties réelles strictement négatives pour tout $\theta \in \mathbb{C}$

Nous supposons d'autre part que les fonctions $z_j(\cdot)$ possèdent presque partout des dérivées d'ordre $m-n$ qui sont des fonctions de carrés intégrables sur tout intervalle fini. Cette hypothèse implique que les supports des lois image $P_{\sigma^2, \theta, v}^X$ du processus $\{X(t), t \in [0, T]\}$ sont dans $C_{m-n-1}([0, T])$

et d'après Rozanov [25], elle est nécessaire et suffisante pour que les lois $P_{\sigma^2, \theta, v}^X$ pour σ^2 fixé soient toutes équivalentes.

Le paramètre σ^2 sera supposé connu car sa valeur peut être déterminée avec probabilité 1 par :

$$\sigma^2 = \lim_{p \rightarrow \infty} \frac{1}{T} \sum_{k=1}^{2^p} \left\{ X^{(m-n-1)}\left(\frac{kT}{2^p}\right) - X^{(m-n-1)}\left[\frac{(k-1)T}{2^p}\right] \right\}^2 \quad \text{p.s}$$

Ce résultat vient du fait que :

$$\begin{aligned} \sum_{k=1}^{2^p} \left\{ m^{(m-n-1)}\left(\frac{kT}{2^p}\right) - m^{(m-n-1)}\left[\frac{(k-1)T}{2^p}\right] \right\}^2 &= \sum_{k=1}^{2^p} \int_{(k-1)T/2^p}^{kT/2^p} [m^{(m-n)}(t)]^2 dt \\ &\leq \frac{1}{2^p} \int_0^T [m^{(m-n)}(t)]^2 dt \rightarrow 0 \end{aligned}$$

et des arguments analogues à ceux du chapitre IV, § 1.

Nous noterons alors $P_{\theta, \nu}^X$ à la place de $P_{\sigma^2, \theta, \nu}^X$. Notre problème est d'estimer (θ, ν) dans la structure $(C_{m-n-1}(0, T), \mathcal{A}, P_{\theta, \nu}^X, \theta \in \Theta, \nu \in \mathbb{R}^k)$

2.2. Fonction de vraisemblance asymptotique

Considérons la bijection de $C_{m-n-1}(0, T)$ sur lui-même :

$$\tau_\nu : x(\cdot) \rightarrow x(\cdot) - m(\cdot) \quad (m(t) = \sum_{j=1}^k \nu_j z_j(t))$$

Il est clair que la loi image de $P_{\theta, \nu}^X$ par τ_ν est $P_{\theta, 0}$. D'après le résultat du chapitre IV, § 2, il existe une mesure μ telle que :

$$(2.2) \quad \begin{aligned} \text{Log} \frac{dP_{\theta, 0}^X}{d\mu} &= \log f_{\theta}^{Y_0 X_0} [E_\theta(\tilde{Y}_0 | \tilde{X}), \tilde{X}_0] + \frac{1}{2} \text{Log det}(2\pi\Lambda_\theta) \\ &+ E_\theta \left[\int_0^T \phi_\theta(t) dW(t) | \tilde{X} \right] - \frac{1}{2} \int_0^T E_\theta(\phi_\theta(t) | \tilde{X})^2 dt \end{aligned}$$

Lemme 2.1.

"On a :

$$\text{Log} \frac{dP_{\theta, \nu}^X}{d\mu} = \left(\text{Log} \frac{dP_{\theta, 0}^X}{d\mu} \right) \circ \tau_\nu + \int_0^T \frac{1}{\sigma} m^{(m-n)}(t) dW(t) - \frac{1}{2} \int_0^T \frac{m^{(m-n)}(t)^2}{\sigma} dt"$$

Démonstration

D'après le théorème de changement de variable, pour toute fonction mesurable bornée h , on a :

$$\begin{aligned} \int h dP_{\theta, \nu}^X &= \int (h \circ \tau_\nu^{-1}) dP_{\theta, 0}^X \\ &= \int (h \circ \tau_\nu^{-1}) \cdot \frac{dP_{\theta, 0}^X}{d\mu} d\mu \\ &= \int h \cdot \left(\frac{dP_{\theta, 0}^X}{d\mu} \circ \tau_\nu \right) d(\mu \circ \tau_\nu) \end{aligned}$$

où $\mu \circ \tau_v$ est la mesure image de μ par τ_v^{-1} . Par suite :

$$\frac{dP_{\theta,v}^X}{d(\mu \circ \tau_v)} = \frac{dP_{\theta,0}^X}{d\mu} \circ \tau_v$$

D'autre part si T désigne la bijection de $C_{m-n-1}([0,T])$ sur $R^{m-n} \times C_0([0,T])$:

$$T : x(\cdot) \rightarrow (x^{(j)}(0), j=0, \dots, m-n-1, \frac{1}{\sigma} [x^{(m-n-1)}(\cdot) - x^{(m-n-1)}(0)])$$

la mesure image de μ (resp $\mu \circ \tau_v$) par T est le produit de la mesure de Lebesgue et de la loi d'un mouvement brownien centré (resp. de moyenne $[m^{(m-n)}(\cdot) - m^{(m-n)}(0)]/\sigma$). D'après un résultat de Rozanov [25], p. 40 on a :

$$\frac{d(\mu \circ \tau_v)}{d\mu} = \exp \int_0^T \frac{1}{\sigma} m^{(m-n)}(t) dW(t) - \frac{1}{2} \int_0^T \left[\frac{m^{(m-n)}(t)}{\sigma} \right]^2 dt$$

d'où le résultat.

Théorème 2.2.

On a :

$$\text{Log} \frac{dP_{\theta,v}^X}{d\mu} = L_T(\theta,v) + \varepsilon_T(\theta,v)$$

où :

$$(2.3) \quad L_T(\theta,v) = \int_0^T [\phi_\theta(t) + \frac{1}{\sigma} m^{(m-n)}(t) - \phi_\theta(m(\cdot),t)] dW(t) - \frac{1}{2} \int_0^T [\phi_\theta(t) + \frac{1}{\sigma} m^{(m-n)}(t) - \phi_\theta(m(\cdot),t)]^2 dt$$

et où $\varepsilon_T(\theta,v)$ possède la propriété suivante :

Il existe une version $\varepsilon_T(\cdot, \theta, v)$ de $\varepsilon_T(\theta, v)$ telle que pour tout $x(\cdot) \in C_{m-n-1}([0,T])$, la fonction $\varepsilon_T(x(\cdot), \dots)$ est deux fois dérivable et les variables aléatoires :

$$\varepsilon_T(\cdot, \theta^*, v^*), \frac{\partial}{\partial \theta_j} \varepsilon_T(\cdot, \theta^*, v^*), \frac{\partial}{\partial \theta_j} \varepsilon_T(\cdot, \theta^*, v^*)$$

$$\frac{\partial^2}{\partial \theta_j \partial \theta_l} \varepsilon_T(\cdot, \theta^*, v^*), \frac{\partial^2}{\partial \theta_j \partial \theta_l} \varepsilon_T(\cdot, \theta^*, v^*) \text{ et } \frac{\partial}{\partial \theta_j} \varepsilon_T(\cdot, \theta^*, v^*)$$

définies sur $(C_{m-n-1}([0, T]), P_{\theta, \nu})$ sont bornées en probabilité quand $T \rightarrow \infty$ "

Démonstration :

Nous avons d'après le théorème 3.1 du chapitre IV :

$$\text{Log} \frac{dP_{\theta, 0}}{d\mu} = L_T(\theta) + \varepsilon_T(\theta)$$

Nous posons : $\varepsilon_T(\theta, \nu) = \varepsilon_T(\theta) \circ \tau_\nu$ et $L_T(\theta, \nu) = L_T(\theta) \circ \tau_\nu$.

On a :

$$\begin{aligned} L_T(\theta, \nu) &= \left[\int_0^T \phi_\theta(t) dW(t) \right] \circ \tau_\nu - \frac{1}{2} \int_0^T [\phi_\theta(t) \circ \tau_\nu]^2 dt \\ &\quad + \int_0^T \frac{1}{\sigma} m^{(m-n)}(t) dW(t) - \frac{1}{2} \int_0^T \left[\frac{1}{\sigma} m^{(m-n)}(t) \right]^2 dt \end{aligned}$$

Comme l'intégrale :

$$\int_0^T \phi_\theta(t) dW(t)$$

peut être évaluée presque-sûrement par :

$$\lim_{k \rightarrow \infty} \sum_{j=1}^{n_k} \phi_\theta \left(\frac{j-1}{n_k} T \right) \left[W \left(\frac{j}{n_k} T \right) - W \left(\frac{j-1}{n_k} T \right) \right]$$

où $\{n_k\}$ est une suite d'entiers tendant vers l'infini avec k , on a presque sûrement :

$$\left[\int_0^T \phi_\theta(t) dW(t) \right] \circ \tau_\nu = \int_0^T [\phi_\theta(t) \circ \tau_\nu] dW(t) - \int_0^T [\phi_\theta(t) \circ \tau_\nu] \frac{m^{(m-n)}(t)}{\sigma} dt$$

En remarquant que $\phi_\theta(t) \circ \tau_\nu = \phi_\theta(t) - \phi_\theta(m(\cdot), t)$, on obtient l'expression de L_T donnée dans le théorème.

Maintenant $\varepsilon_T(\theta)$ est donné par :

$$\begin{aligned} \varepsilon_T(\theta) &= \text{Log } f_{\theta}^{Y_0 X} [E_{\theta}(\tilde{Y}_0 | \tilde{X}), X_0] + \frac{1}{2} \text{Log det}(2\pi\Lambda_{\theta}) \\ &+ {}^t E_{\theta}(\tilde{Y}_0 | \tilde{X}) \int_0^T h_{\theta} [dW(t) - \phi_{\theta}(t) dt] \\ &+ {}^t E_{\theta}(\tilde{Y}_0 | \tilde{X}) \left[\int_0^T h_{\theta}(t) {}^t h_{\theta}(t) dt \right] E_{\theta}(\tilde{Y}_0 | \tilde{X}) \end{aligned}$$

Il est clair que $\varepsilon_T(\theta, \nu)$ est donné par la même expression avec $E_{\theta}(\tilde{Y}_0 | \tilde{X}), \phi_{\theta}$ et $\int_0^T h_{\theta}(t) dW(t)$ remplacées par $E_{\theta}(\tilde{Y}_0 | \tilde{X}) \circ \tau_{\nu}, \phi_{\theta}(t) \circ \tau_{\nu}$ et $\left[\int_0^T h_{\theta}(t) dW(t) \right] \circ \tau_{\nu}$

Calculons $E_{\theta}(\tilde{Y}_0 | \tilde{X}) \circ \tau_{\nu}$. Nous savons que la correspondance $X(t) \leftrightarrow (\lambda \rightsquigarrow e^{-i\lambda t})$ définit un isomorphisme entre l'espace de Hilbert engendré par $X(t), t \in [0, T]$ et le sous-espace vectoriel $L_T^2(f_{\theta})$ de $L^2(\mathbb{R}, \mathbb{R}, f_{\theta}(\lambda) d\lambda)$ engendré par les fonctions $\lambda \rightsquigarrow e^{i\lambda t}, t \in [0, T]$. Soient $k_{\theta, j}(\cdot)$ l'image de $(E_{\theta}(\tilde{Y}_0 | \tilde{X}))_j$ par cette isomorphisme et k_{θ} le vecteur $(k_{\theta, 1}, \dots, k_{\theta, n})$.

Si :

$$k_{\theta}^{(p)} = \sum_{\ell=1}^{m_p} \alpha_{\ell}^{(p)}(\theta) e^{-i\lambda_{\ell} t}^{(p)}$$

est une suite de $[L_T(f_{\theta})]^n$ tendant vers k_{θ} , on a au sens de la convergence en moyenne quadratique :

$$E_{\theta}(\tilde{Y}_0 | \tilde{X}) = \lim_{p \rightarrow \infty} \sum_{\ell=1}^{m_p} \alpha_{\ell}^{(p)} \tilde{X}(t_{\ell}^{(p)})$$

En remplaçant la suite du second membre par une sous-suite convenable, on peut supposer que la convergence a lieu aussi presque-sûrement. D'où

$$E_{\theta}(\tilde{Y}_0 | \tilde{X}) \circ \tau_{\nu} = E_{\theta}(\tilde{Y}_0 | \tilde{X}) - \lim_{p \rightarrow \infty} \sum_{\ell=1}^{m_p} \alpha_{\ell}^{(p)} m(t_{\ell}^{(p)})$$

D'après un résultat de Rozanov [25] p. 44, les fonctions $z_j(\cdot)$ admettent les représentations :

$$z_j(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \tilde{z}_{j,\theta}(\lambda) f_{\theta}(\lambda) d\lambda$$

D'où :

$$m(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \tilde{m}_{\theta}(\lambda) f_{\theta}(\lambda) d\lambda \quad ; \quad \tilde{m}_{\theta}(\lambda) = \sum v_j \tilde{z}_{j,\theta}(\lambda)$$

On en déduit :

$$E_{\theta}(\tilde{Y}_0 | \tilde{X}) \circ \tau_v = E_{\theta}(\tilde{Y}_0 | \tilde{X}) - \int_{-\infty}^{\infty} \bar{k}_{\theta}(\lambda) \tilde{m}_{\theta}(\lambda) f_{\theta}(\lambda) d\lambda$$

On montre sans difficulté que les applications $\theta \rightarrow \tilde{z}_{j,\theta}$ sont trois fois continûment dérivables dans $L_T^2(f_{\theta})$ dont les dérivées jusqu'à l'ordre 2 calculées en θ sont bornées en norme $L_T^2(f_{\theta})$ quand $T \rightarrow \infty$. D'après le lemme 3.1 du chapitre IV, il existe donc une version de $E_{\theta}(\tilde{Y}_0 | \tilde{X}) \circ \tau_v$ qui pour $x(\cdot)$ fixé est une forme affine en v et une fonction deux fois dérivable en θ .

D'autre part un raisonnement analogue à la première partie de la démonstration de ce théorème montre que :

$$\int_0^T h_{\theta}(t) dW(t) \circ \tau_v = \int_0^T h_{\theta}(t) dW(t) - \int_0^T h_{\theta}(t) \frac{m^{(m-n)}(t)}{\sigma} dt$$

D'après la démonstration du théorème 3.1 du chapitre IV, il existe une version de $\int_0^T h_{\theta}(t) dW(t)$ qui pour $x(\cdot)$ fixé est une fonction deux fois dérivable de θ . La relation précédente permet alors de définir une version de

$$\int_0^T h_{\theta}(t) dW(t) \circ \tau_v$$

qui pour $x(\cdot)$ fixé est une forme affine en v et deux

fois dérivable en θ .

Finalement :

$$\phi_{\theta}(t) \circ \tau_v = \phi_{\theta}(t) - \phi_{\theta}(m(\cdot), t) = \phi_{\theta}(t) - \sum_{j=1}^k v_j \phi_{\theta}(z_j(\cdot), t)$$

et $\phi_{\theta}(z_j(\cdot), t)$ sont des fonctions deux fois dérivables en θ .

Nous déduisons alors l'existence d'une version de $\varepsilon_T(\theta, \nu)$ qui pour chaque $x(\cdot)$ est une fonction deux fois dérivable en θ et ν . Les résultats précédents, compte-tenu des lemmes 2.1 et 3.1 du chapitre IV, montrent que cette version et ses dérivées jusqu'au deuxième ordre, calculées en θ^*, ν^* définissent des variables aléatoires sur $(C_{m-n}([0, T]), P_{\theta^*, \nu^*})$ bornées en probabilité quand $T \rightarrow \infty$. (On remarque que la loi image de $P_{\theta, \nu}$ par une variable aléatoire Z est la même que celle de $P_{\theta, \nu}$ par la variable $Z \circ \tau_\nu$). D'où le théorème.

Une deuxième approximation de $\text{Log}(dP_{\theta, \nu}^X/d\mu)$ est basée sur le théorème 3.3 du chapitre IV. Posons :

$$(2.4) \quad \tilde{L}_T(\theta) = -\frac{1}{\sigma^2} \left\{ \sum_{j=1}^{m-n-1} d_j \int_0^T [x^{(j)}(t)]^2 dt + \sigma^2 T (b_{m-1} - a_{n-1}) \right. \\ \left. + T \int_0^T \frac{R_\theta(i\lambda)}{|p_\theta(i\lambda)|^2} I_T(\lambda) d\lambda \right\}$$

où les notations sont celles du théorème 3.3 du chapitre IV :

Théorème 2.3.

On a :

$$(2.5) \quad L_T(\cdot, \theta, \nu) = \tilde{L}_T(\cdot, \theta) \circ \tau_\nu + \int_0^T \frac{m^{(m-n)}(t)}{\sigma} dW(t) \\ - \frac{1}{2} \int_0^T \left[\frac{m^{(m-n)}(t)}{\sigma} \right]^2 dt + \eta_T(\cdot, \theta, \nu)$$

où $\eta_T(\cdot, \theta, \nu)$ vérifie :

- (i) $\eta_T(x(\cdot), \dots)$ est deux fois dérivable
- (ii) Il existe un voisinage $V(\theta^*, \nu^*)$ de θ^*, ν^* tel que les bornes supérieures de $\eta_T(\cdot, \theta, \nu)$ et de ses dérivées jusqu'à l'ordre 2 par rapport à (θ, ν) , quand ce dernier parcourt $V(\theta^*, \nu^*)$, définissent des variables aléatoires bornées en norme $L^1(P_{\theta^*, \nu^*})$ quand $T \rightarrow \infty$.

Démonstration

Il est clair que $\eta_T(\cdot, \theta, \nu) = \eta_T(\theta) \circ \tau_\nu$ où $\eta_T(\theta)$ (définie dans le théorème 3.3 du chapitre IV) est une forme quadratique en $Y_\theta^{(j)}(x(\cdot))$, $j = 0, \dots, m-1$ dont les coefficients sont des fonctions deux fois dérivables en \cdot . Les variables $Y_\theta^{(j)}(x(\cdot))$ sont définies dans la démonstration du théorème 3.3 du chapitre IV et il est clair que :

$$Y_\theta^{(j)}(x(\cdot)) \circ \tau_\nu = Y_\theta^{(j)}(x(\cdot)) - \sum_{\ell=1}^k \nu_\ell Y_\theta^{(j)}(z_\ell(\cdot)).$$

On en déduit les propriétés de $\eta_T(\cdot, \theta, \nu)$.

2.3. Propriétés asymptotiques de la fonction L_T

Nous allons étendre les propriétés asymptotiques de la fonction $L_T(\theta)$ du chapitre IV. Pour cela nous aurons besoin de la notion de mesure spectrale d'une fonction de R^+ dans R^n .

Définition 2.1.

On dit que la fonction $h : R^+ \rightarrow R^n$ possède une mesure spectrale si :

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T h(t+r) {}^t h(t+s) = \rho(s-r)$$

La mesure M uniquement définie par :

$$\rho(u) = \int_{-\infty}^{+\infty} e^{i\lambda u} dM(\lambda)$$

est appelée mesure spectrale de h ".

Le résultat suivant est l'analogie en temps continu du lemme 3.1 du chapitre II.

Lemme 2.2.

"Soit

$$h_T(\lambda) = \int_0^T e^{i\lambda t} h(t) dt$$

alors la mesure $(2\pi T)^{-1} h_T(\lambda) h_T^*(\lambda) d\lambda$ converge faiblement quand $T \rightarrow \infty$ vers la mesure spectrale de h ".

Le résultat suivant généralise celui de Kholevo [16].

Lemme 2.3

"Soit h une fonction vérifiant :

$$\lim_{T \rightarrow \infty} T^{-1/2} h^{(j)}(T) = 0 \quad j = 0, 1, \dots, r-1$$

et supposons qu'il existe un polynôme stable (c'est-à-dire les parties réelles de ses zéros sont strictement négatives). R , de degré r tel que la fonction $k(t) = R(d/dt) h(t)$ possède une mesure spectrale M ; alors quels que soient les polynômes stables P et Q de degrés respectifs v_1 et v_2 avec $v_2 - v_1 < r$, la fonction :

$$l(t) = Q\left(\frac{d}{dt}\right) y(t)$$

où, $y(\cdot)$ est la solution de $P(d/dt) y(t) = h(t)$ de conditions initiales nulles, possède une mesure spectrale N donnée par :

$$dN(\lambda) = \left| \frac{Q(i\lambda)}{P(i\lambda) R(i\lambda)} \right|^2 dM(\lambda)$$

Démonstration

Nous avons $R(d/dt) P(d/dt) y(t) = k(t)$. Définissons la fonction y' comme solution de :

$$\begin{aligned} R(d/dt) P(d/dt) y'(t) &= k(t) & 0 < t < T \\ &= 0 & t < T \end{aligned}$$

de condition initiale nulle et posons $l'(t) = Q(d/dt) y'(t)$. La fonction l' est bien définie car y' est v_1+r fois dérivable et il est clair que :

$$l(t) = l'(t) + \sum_{j=0}^{v_1+r-1} y^{(j)}(0) g_j(t) \quad 0 < t < T$$

où g_j sont des fonctions tendant exponentiellement vers 0 à l'infini (P et Q étant stables). Par suite :

$$\lim_{T \rightarrow \infty} T^{-1} \int_0^T [l(t) - l'(t)]^t [l(t-u) - l'(t-u)] = 0 \quad u > 0$$

La mesure spectrale de l est donc égale à celle de l' (si cette dernière existe). D'autre part, pour $t > T$:

$$l'(t) = \sum_{j=0}^{v_1+r-1} y^{(j)}(T) g_j(t-T)$$

et comme $T^{-1/2} y^{(j)}(T) \rightarrow 0$ quand $T \rightarrow \infty$ (cela résulte de l'hypothèse faite sur h), on a :

$$\lim T^{-1} \int_T^\infty l'(t) {}^t l'(t+u) = 0$$

On déduit alors compte-tenu du lemme 2.2 que la mesure spectrale de l est la limite faible quand $T \rightarrow \infty$ de la mesure $(2\pi T)^{-1} \tilde{l}'(\lambda) [\tilde{l}'(\lambda)]^* d\lambda$ où \tilde{l}' est la transformée de Fourier de l' , c'est-à-dire :

$$\tilde{l}'(\lambda) = \frac{Q(i\lambda)}{P(i\lambda)R(i\lambda)} \int_0^T k(t) e^{i\lambda t}$$

D'où le résultat.

Nous faisons dans la suite l'hypothèse suivante :

H : La fonction $z(\cdot) = {}^t(z_1(\cdot) \dots z_k(\cdot))$ vérifie les conditions du lemme 2.4 relative à un polynôme stable de degré $m-n$.

Il résulte de cette hypothèse que la fonction z possède une mesure spectrale notée M .

Maintenant, pour obtenir les propriétés asymptotiques de la fonction L_T , nous plongeons l'espace $C_{m-n-1}([0, T])$ dans $C_{m-n-1}(R)$ muni de la loi $\bar{P}_{\theta^*, \nu^*}$ d'un processus gaussien stationnaire de moyenne $\sum \nu_j^* z_j(\cdot) = m^*(\cdot)$ et de densité spectrale f_{θ^*} (les fonctions z_j étant prolongées sur R en posant $z(t) = 0, t < 0$).

Nous notons $\{\tilde{X}(t) : x(\cdot) \rightarrow x(t), t \in \mathbb{R}\}$ le processus des "coordonnées". Il admet la représentation spectrale :

$$\tilde{X}(t) = m^{**}(t) + \int_{-\infty}^{\infty} e^{-i\lambda t} \frac{p_{\theta^{**}}(i\lambda)}{q_{\theta^{**}}(i\lambda)} \sigma dZ(\lambda)$$

On pose :

$$\bar{Y}_{\theta}(t) = \int_{-\infty}^{\infty} e^{-i\lambda t} \frac{1}{p_{\theta}(-i\lambda)} \frac{p_{\theta^{**}}(-i\lambda)}{q_{\theta^{**}}(-i\lambda)} \sigma dZ(\lambda)$$

$$(2.6) \quad \begin{aligned} \bar{\phi}_{\theta}(t) &= \frac{1}{\sigma} \left[\left(\frac{d}{dt} \right)^{m-n} p_{\theta} \left(\frac{d}{dt} \right) - q_{\theta} \left(\frac{d}{dt} \right) \right] \bar{Y}_{\theta}(t) \\ &= \int_{-\infty}^{\infty} e^{-i\lambda t} \left[(-i\lambda)^{m-n} - \frac{q_{\theta}(-i\lambda)}{p_{\theta}(-i\lambda)} \right] \frac{p_{\theta^{**}}(-i\lambda)}{q_{\theta^{**}}(-i\lambda)} dZ(\lambda) \end{aligned}$$

Les lemmes 4.1 et 4.2 du chapitre IV admettent alors les extensions suivantes :

Lemme 2.4.

"Le processus $\{\bar{\xi}(t), t \geq 0\}$ où

$$\bar{\xi}(t) = W(t) - \frac{1}{\sigma} [m^{**^{(m-n-1)}}(t) - m^{**^{(m-n-1)}}(0)] - \int_0^t \bar{\phi}_{\theta^{**}}(s) ds$$

est un mouvement brownien relativement à $\bar{P}_{\theta^{**}, \nu^{**}}$."

Lemme 2.5.

On a :

$$\bar{\phi}_{\theta}(t) = \phi_{\theta}(t) - \phi_{\theta}(m(\cdot), t) + {}^t h_{\theta} \bar{Y}_{\theta, 0}$$

où $\bar{Y}_{\theta, 0} = ({}^t \bar{Y}_{\theta}(0), \bar{Y}_{\theta}^{(1)}(0), \dots, \bar{Y}_{\theta}^{(n-1)}(0))$ et où $h_{\theta}, \phi_{\theta}$ sont définies dans le lemme 1.2 du chapitre IV".

Compte-tenu de ces résultats, un calcul analogues à celui du chapitre IV, § 4 montre que :

$$(2.6) \quad \left(\frac{\partial L_T}{\partial \theta_j} \right) (\theta^{**}, v^{**}) = \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} d\bar{\xi}(t) + O(T)$$

$$(2.7) \quad \left(\frac{\partial L_T}{\partial v_j} \right) (\theta^{**}, v^{**}) = \int_0^T x_{j, \theta^{**}}(t) d\bar{\xi}(t)$$

$$(2.8) \quad \left(\frac{\partial^2 L_T}{\partial v_j \partial \theta_\ell} \right) (\theta^{**}, v^{**}) = \int_0^T \left(\frac{\partial}{\partial \theta_\ell} x_{j, \theta}(t) \right)_{\theta=\theta^{**}} d\bar{\xi}(t) \\ - \int_0^T \left(\frac{\partial}{\partial \theta_\ell} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} j_{j, \theta^{**}}(t) dt + O(T)$$

$$(2.9) \quad \left(\frac{\partial^2 L_T}{\partial \theta_j \partial \theta_\ell} \right) (\theta^{**}, v^{**}) = \int_0^T \left(\frac{\partial^2}{\partial \theta_j \partial \theta_\ell} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} d\bar{\xi}(t) \\ - \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} \left(\frac{\partial}{\partial \theta_\ell} \bar{\phi}_\theta(t) \right)_{\theta=\theta^{**}} dt + O(T)$$

$$(2.10) \quad \left(\frac{\partial^2 L_T}{\partial v_j \partial v_\ell} \right) (\theta^{**}, v^{**}) = - \int_0^T j_{j, \theta^{**}}(t) \ell_{\ell, \theta^{**}}(t) dt$$

où $x_{\theta, j}(t) = \frac{1}{\sigma} z_j(t) - \phi_\theta(z_j(\cdot), t)$ et où $O(T)$ est une variable aléatoire bornée en probabilité quand $T \rightarrow \infty$.

Remarque :

La fonction $x_{\theta, j}$ peut être obtenue par : $x_{\theta, j}(t) = \frac{1}{\sigma} q_\theta(d/dt) \lambda_j(t)$

où $\lambda_j(\cdot)$ est la solution de $p_\theta(d/dt) \lambda_j(t) = z_j(t)$ de conditions initiales nulles.

Théorème 2.4

"Quand $T \rightarrow \infty$, le vecteur des dérivées premières et la matrice des dérivées secondes de $T^{-1}L_T$ convergent en probabilité respectivement vers 0 et vers :

$$-K = - \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix}$$

où S est donnée par (4.6) du chapitre IV et où :

$$T = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \frac{q_{\theta^{**}}(i\lambda)}{p_{\theta^{**}}(i\lambda)} dM(\lambda)"$$

Démonstration

La convergence en probabilité quand $T \rightarrow \infty$ des dérivées de $T^{-1}L_T$ par rapport à θ se démontre d'une manière analogue à celle de la démonstration du théorème 4.1 du chapitre IV.

Montrons maintenant que :

$$\therefore T^{-1} \left(\frac{\partial L_T}{\partial v_j} \right) (\theta^{**}, v^{**}) \rightarrow 0 \text{ en probabilité quand } T \rightarrow \infty$$

D'après (2.7) il suffit de montrer que :

$$\lim_{T \rightarrow \infty} T^{-1} \int_0^T \chi_{j, \theta^{**}}(t) d\bar{\xi}(t) = 0$$

en moyenne quadratique. Or la norme de l'expression dont on cherche la limite est :

$$T^{-2} \int_0^T [\chi_{j, \theta^{**}}(t)]^2 dt$$

D'après le lemme 2.3 et la remarque qui précède ce théorème et compte-tenu de l'hypothèse H, la fonction $\chi_{\theta^{**}} = {}^t(\chi_{1, \theta^{**}}, \dots, \chi_{k, \theta^{**}})$ possède une mesure spectrale ayant la densité $\sigma^{-2} |q_{\theta^{**}}(i\lambda)/p_{\theta^{**}}(i\lambda)|^2$ par rapport à M . L'expression précédente converge donc vers 0 comme T^{-1} quand $T \rightarrow \infty$.

Maintenant, compte-tenu de (2.8) pour montrer la convergence vers 0 en probabilité quand $T \rightarrow \infty$ des dérivées de $T^{-1}L_T$ par rapport à v_j, θ_l il suffit de montrer que :

$$(i) \quad T^{-1} \int_0^T \left(\frac{\partial}{\partial \theta_\ell} x_{j, \theta^*}(t) \right)_{\theta = \theta^*} d\bar{\xi}(t)$$

et

$$(ii) \quad T^{-1} \int_0^T \left(\frac{\partial}{\partial \theta_\ell} \bar{\phi}_\theta(t) \right)_{\theta = \theta^*} x_{j, \theta^*}(t) dt$$

convergent vers 0 en moyenne quadratique quand $T \rightarrow \infty$. Or la norme en moyenne quadratique de ces expressions sont respectivement :

$$T^{-2} \int_0^T \left(\frac{\partial}{\partial \theta_\ell} x_{\ell, \theta}(t) \right)_{\theta = \theta^*}^2 dt$$

et

$$T^{-2} \int_0^T \psi_\ell(\lambda) |x_{j, \theta^*}(\lambda)|^2 d\lambda$$

où $\psi_\ell(\lambda)$ est la densité spectrale du processus $\{(\partial \bar{\phi}_\theta(t)/\partial \theta_\ell)_{\theta = \theta^*}, t \in \mathbb{R}\}$ et où x_{j, θ^*} est la transformée de Fourier de $x_{j, \theta}$. Le lemme 2.3 permet également de montrer que les fonctions $(\partial x_{j, \theta^*}(t)/\partial \theta_j)_{\theta = \theta^*}$ possèdent une mesure spectrale. On voit alors que les deux expressions précédentes convergent vers 0 quand $T \rightarrow \infty$.

Finalement, le dernier résultat du théorème résulte de (2.10) et de ce que :

$$\lim_{T \rightarrow \infty} T^{-1} \int_0^T x_{j, \theta^*}(t) x_{\ell, \theta^*}(t) dt = \int_{-\infty}^{\infty} \frac{1}{\sigma} \frac{q_{\theta^*}(i\lambda)}{p_{\theta^*}(i\lambda)} dM_{j\ell}(\lambda)$$

qui est une conséquence du lemme 2.3.

Maintenant, avec une démonstration analogue à celle du lemme 4.4 du chapitre IV et en tenant compte du théorème 2.3, on a :

Théorème 2.5

"Les résultats du théorème 2.4 restent vraies si (θ^*, ν^*) est remplacé par un estimateur consistant".

Théorème 2.6

"Supposons que les fonctions $z_j(\cdot)$ sont bornées, alors le vecteur des dérivées de $T^{-1/2} L_T$, calculé en (θ^*, ν^*) converge en loi quand $T \rightarrow \infty$ vers la loi gaussienne centrée de matrice de covariance

$$\begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix}$$

Démonstration

La démonstration du théorème est une synthèse des démonstrations du théorème 3.3 du chapitre III et du théorème 4.2 du chapitre IV.

D'après (2.6) et (2.7) il nous suffit de montrer que toute combinaison linéaire :

$$T^{-1/2} \left[\sum_j C_j \int_0^T \left(\frac{\partial}{\partial \theta_j} \bar{\phi}_\theta(t) \right)_{\theta=\theta^*} d\bar{\xi}(t) + \sum_\ell d_\ell \int_0^T x_{\ell, \theta}(t) d\bar{\xi}(t) \right]$$

converge en loi vers une loi normale centrée quand $T \rightarrow \infty$.

Nous posons $\psi_j(t) = (\partial \bar{\phi}_\theta / \partial \theta_j)_{\theta=\theta^*}$ et nous considérons l'expression discrétisée :

$$T^{-1/2} \sum_{p=0}^{[nT]-1} \left[\sum_j C_j \psi_j \left(\frac{p}{n} \right) \left[\bar{\xi} \left(\frac{p+1}{n} \right) - \bar{\xi} \left(\frac{p}{n} \right) \right] + \sum_\ell d_\ell \int_{p/n}^{(p+1)/n} x_{\ell, \theta}(t) d\bar{\xi}(t) \right]$$

$$= \left(\frac{[nT]}{nT} \right)^{1/2} \cdot [nT]^{-1/2} \sum_{p=0}^{[nT]-1} \left[\varepsilon_p^{(n)} + \eta_p^{(n)} \right]$$

D'après la démonstration du théorème 4.2 du chapitre IV, il nous suffit de montrer que :

$$N^{-1/2} \sum_{p=0}^{N-1} \left[\varepsilon_p^{(n)} + \eta_p^{(n)} \right]$$

converge en loi vers une loi normale centrée de variance σ_n^2 telle que

$\lim_{p \rightarrow \infty} \frac{\sigma_n^2}{n}$ existe. D'autre part, la même démonstration montre que $\psi_j(t)$ peut

s'écrire sous la forme :

$$\psi_j(t) = \int_{-\infty}^t \alpha_j(s) d\bar{\xi}(s) = \int_{t-m/n}^t \alpha_j(s) d\bar{\xi}(s) + \int_{-\infty}^{t-m/n} \alpha_j(s) d\bar{\xi}(s)$$

On décompose alors $\varepsilon_p^{(n)}$ sous la forme $\varepsilon_p^{(n,m)} + r_p^{(n,m)}$ où

$$n^{1/2} \varepsilon_p^{(n,m)} = \sum c_j \psi_j^{(m)} \left(\frac{p}{n} \right) \left[\bar{\xi} \left(\frac{p+1}{n} \right) - \bar{\xi} \left(\frac{p}{n} \right) \right] ; \psi_j^{(m)}(t) = \int_{t-m/n}^t \alpha_j(s) d\bar{\xi}(s)$$

On vérifie facilement que $N^{-1/2} \sum_{p=0}^{N-1} r_p^{(n,m)}$ converge en probabilité

vers 0 uniformément en N quand $m \rightarrow \infty$ (n étant fixe). L'application du lemme de Bernstein (voir la démonstration du théorème 3. du chapitre III) nous conduit à monter que :

$$N^{-1/2} \sum_{p=0}^{N-1} [\varepsilon_p^{(n,m)} + \eta_p^{(n)}]$$

converge en loi vers une loi normale de variance $\sigma_{n,m}^2$ telle que

$$\lim_{m \rightarrow \infty} \sigma_{n,m}^2 = \sigma_n^2$$

Nous pouvons alors appliquer la même démonstration que celle du théorème 3.3 du chapitre III. Pour cela, il suffit de montrer que la condition de Lindeberg est vérifiée pour la série $\sum \eta_p^{(n)}$. Or les variables aléatoires $\eta_p^{(n)}$ sont normales centrées de variances :

$$\alpha_p^2 = \int_{p/n}^{(p+1)/n} |\sum d_j - \chi_{j,\theta}(t)|^2 dt$$

Par suite :

$$E[(\eta_p^{(n)})^2 \mathbf{1}_{\{|\eta_p^{(n)}| > \lambda\}}] = \alpha_p^2 E[X^2 \mathbf{1}_{\{|X| > \lambda/\alpha_p\}}]$$

où X est une variable normale standard et on en déduit que la condition de Lindeberg est vérifiée si :

$$\text{Max}_{j=1, \dots, N} \frac{\alpha_j^2}{\alpha_1^2 + \dots + \alpha_N^2}$$

tend vers 0 quand $N \rightarrow \infty$. Comme les fonctions z_j sont bornées, il en est de même les fonctions $x_{j,\theta}$, et par suite la dernière condition est satisfaite.

Finalement, un calcul élémentaire montre que :

$$\lim_{N \rightarrow \infty} \sigma^2 \{ N^{-1/2} \sum_{p=0}^{N-1} \{ \epsilon_p^{(n,m)} + \eta_p^{(n)} \} \} = \sum_{j,l} C_j C_l S_{jl}^m + \sum_{j,l} d_j d_l^T S_{jl}$$

où

$$S_{jl}^{(m)} = E \left[\sum_j^{(m)}(0) \sum_l^{(m)}(0) \right] \rightarrow S_{jl} \text{ quand } n \rightarrow \infty$$

Le théorème est démontré.

2.4. Construction de l'estimateur

La méthode exposée au chapitre III, § 4 s'adapte sans difficulté dans ce problème. En particulier le problème d'estimation simultanée des coefficients de régression et des coefficients de structure se décompose en deux problèmes: estimation des coefficients de régression, la densité spectrale du processus résiduel étant connue et estimation des coefficients de la densité spectrale d'un processus stationnaire gaussien centré.

Nous terminons ce paragraphe en considérant le problème d'estimation des coefficients de régression connaissant la densité spectrale du processus résiduel, soit :

$$f(\lambda) = \frac{\sigma}{2\pi} \left| \frac{p(i\lambda)}{q(i\lambda)} \right|^2$$

D'après le théorème 2.2, la fonction log-vraisemblance approchée s'écrit :

$$L_T = \int_0^T \sum v_j x_j(t) [dW(t) - \phi(t)dt] - \frac{1}{2} \int_0^T \sum v_j x_j(t)^2 dt + C^{te}$$

où $\chi_j(\cdot)$ est définie par : $\chi_j(t) = \sigma^{-1} q(d/dt) \lambda_j(t)$, $\lambda_j(\cdot)$ étant la solution de $p(d/dt) \lambda_j(t) = z_j(t)$, de conditions initiales nulles, et où $\phi(t)$ est définie par :

$$\phi(t) = \frac{1}{\sigma} \left[\left(\frac{d}{dt} \right)^{m-n} p\left(\frac{d}{dt} \right) - q\left(\frac{d}{dt} \right) \right] y(t)$$

$y(\cdot)$ étant la solution de $p(d/dt) y(t) = x(t)$ de conditions initiales nulles.

Posons $\xi(t) = X(t) - \int_0^t \phi(s) ds$, on a :

$$d\xi(t) = dy^{(m-1)}(t) + \tilde{q}\left(\frac{d}{dt} \right) y(t) ; \tilde{q}(u) = q(u) - u^m$$

On voit que $d\xi(t)$ est définie comme $\chi_j(t) dt$ avec $z_j(\cdot)$ remplacée par $x(\cdot)$.

L'estimateur de maximum de vraisemblance v est alors solution de :

$$(2.11) \quad \int_0^T \chi_j(t) d\xi(t) = \sum_{\ell} \left[\int_0^T \chi_j(t) \chi_{\ell}(t) dt \right] v_{\ell} = 0$$

L'interprétation de (2.11) est claire : le processus $\xi(t)$ admet la fonction moyenne $\int_0^t \sum v_j \chi_j(s) ds$; le système (2.11) n'est autre que le système des équations normales pour l'estimation des coefficients de régression, sous l'hypothèse que $\{ \xi(t), t \in [0, T] \}$ est un mouvement brownien.

B I B L I O G R A P H I E

- [1] AITCHINSON J., SILVEY S.D. : Maximum likelihood estimation of parameters subject to constraint.
Annales of Math. Stat., Vol. 29 (1955), pp. 813-828
- [2] AKAIKE H.: Maximum likelihood identification of gaussian auto-regressive moving average model.
Biometrika, vol. 60, (1973), pp. 255-265
- [3] ANDERSON T.W. : The statistical analysis of time series.
Wiley, New-York, (1971)
- [4] BAXTER G. : A strong limit theorem for gaussian processes.
Proc. Amer. Soc., vol. 7 (1956), n° 3, pp. 522-527
- [5] BILLINGSLEY P. : Convergence of probability measures.
Wiley, New-York (1968)
- [6] BOX G.P.E., JENKINS G.M. : Time series analysis, forecasting and Control.
Holdenday, San Francisco (1970).
- [7] CRAMER H. : Mathematical methods of statistics.
Princeton University Press (1946)
- [8] CRAMER H., LEADBETTER M.R. : Stationary and related processes.
Wiley, New-York (1970)
- [9] DOOB J.L. : Stochastic processes.
Wiley, New-York (1953)
- [10] DZHAPARADZE K.O. : On the estimation of the spectral parameters of a gaussian stationary process with rational spectral density
Theory of Prob. and Appl., vol. 16, (1971), pp.531-537
- [11] DZHAPARIDZE K.O. : On methods for obtaining asymptotically efficient spectral parameter estimates for a stationary gaussian process with rational spectral density.
Theory of Prob. and Appl., vol. 16 (1971),pp. 550-554
- [12] GRENANDER U, ROSENBLATT M. : Statistical analysis of stationary time series.
Wiley, New-York (1957)
- [13] GIRSANOV I.V. : On transforming of a certain class of stochastic process by absolutely continuous substitution of mesures.
Theory of Prob. and Appl., vol. 5 (1960), pp.285-301

- [14] HANNAN E.J. : Multiple time series.
Wiley, New York (1970)
- [15] ITO K. : On stochastic differential equations.
Memoires Amer. Math. Soc., vol. 4 (1951)
- [16] KHOLEVU A.S. : On the estimate of regression coefficients.
Theory of Prob. and Appl., vol. 16 (1969), pp. 79-103
- [17] LE BRETON A. : Parameters estimation in a linear stochastic differential equation.
A paraître dans Transac. of 1974 Europ. meeting Stat. and the 7th Prague Conference (Prague août 1974)
- [18] METIVIER M. : Processus de Markov, chap. IV : intégrale stochastique.
Cours photocopié, Université de Rennes (1972)
- [19] MORAN P.A. : On asymptotically optimal tests of composite hypotheses.
Biometrika, vol. 57 (1970) pp. 47-55
- [20] NEVEU J. : Bases mathématiques du calcul des probabilités.
Masson & C^o, Paris (1964)
- [20 b] NEVEU J. : Processus aléatoires gaussiens.
Séminaire de Mathématique Supérieure, Montréal (1968)
- [21] NEYMAN J. : Optimal asymptotic tests of composite statistical hypotheses.
Dans Probability and Statistics, ed. U. Grenander, pp. 213-234, Wiley, New-York (1959)
- [22] PHAM-DINH T. : Sur le calcul de la fonction de vraisemblance liée à l'estimation des paramètres d'un processus gaussien stationnaire centré de densité spectrale rationnelle.
C.R.A.S., Série A, tome 278 (1974) pp. 1441-1444
- [23] PIERCE D.A. : Least squares estimation in the regression model with auto-regressive moving average errors.
Biometrika, Vol. 58 (1971) pp. 299-312
- [24] ROZANOV Yu. A. : Stationary random processes (English translation)
Holden day, San-Fransico (1967)
- [25] ROZANOV Yu. A. : Infinite dimensional gaussian distributions (English Stranslation).
American Mathematical Society, Providence Rhode Island, (1971)

- [26] ROZEN B. : On the central limit theorem for sum of dependent random variables.
Z. Wahrscheinlichkeitstheorie, vol. 7 (1967) pp. 48-82
- [27] WILSON G.T. : The estimation of parameters in multivariate time series models.
J.R.S.S. serie B, vol. 35 (1973) pp. 76-85
- [28] WHITTLE P. : On the fitting of multivariate auto-regression and the approximate canonical factorisation of a spectral density matrix.
Biometrika, vol. 50, (1963), pp. 129-134
- [29] WHITTLE P. : Gaussian estimation in stationary time series.
Bull. Int. Stat. Inst. Vol. 33 (1961) pp. 105-129
- [30] WALKER A.M. : Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary non deterministic time series.
Journal Anstr. Math. Soc. Vol. 4 (1964), pp. 363-384

