

REPUBLIQUE ALGERIENNE DEMOCRATIQUE et POPULAIRE.
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU
Faculté des Sciences
Département de Mathématiques

THESE DE DOCTORAT en Mathématiques

Option : Statistique

Influence des valeurs manquantes sur la prédiction de séries chronologiques

présenté par

Abdelghani Hamaz

Devant le jury composé de

Mohamed Aidene	Professeur	U.M.M.T.O	Président
Mohamed Morsli	Professeur	U.M.M.T.O	Rapporteur
Mohamed Ibazizen	MC "B"	U.M.M.T.O	Co-directeur
Mohamed Bentarzi	Professeur	U.S.T.H.B	Examineur
Djamal Hamadouche	Professeur	U.M.M.T.O	Examineur
Abdelhakim Aknouche	MC "A"	U.S.T.H.B	Examineur

soutenu le 25 / 02 / 2010

Remerciements

Il m'est agréable d'adresser mes premiers remerciements à mon promoteur, Mr. Ibazizen Mohamed. Je me vois encore, "débarquant" dans son bureau à Oued Aissi, un après-midi de novembre, il y a maintenant près de cinq ans. C'était le début de mon parcours scientifique à ses cotés avec autant d'années de collaboration et de nombreuses discussions qui ont abouti à cette thèse. Un grand scientifique, mais aussi un homme que j'ai appris à connaître, avec son charisme, son humour et ses éclats de rire. Il a toujours montré de l'intérêt pour mes travaux malgré son stress permanent à force de gérer des dizaines de questions et sa fatigue accumulée suite aux nombreuses heures de travail, notamment pour relire l'un ou l'autre article de ses doctorants. Et au-delà de tout cela, une disponibilité et une grande liberté offertes à ses jeunes chercheurs.

Je voudrais ensuite remercier Mr Aidene Mohammed, Professeur à l'université de Tizi-Ouzou, pour l'honneur qu'il me fait en présidant ce jury.

Mes vifs remerciements vont aussi à Mr Bentarzi Mohamed, Professeur à l'université de l'USTHB, Mr Hamadouche Djamel, Professeur à l'université de Tizi-Ouzou et Mr Aknouche Abdelhakim, Maître de conférence à l'université de l'USTHB, pour l'intérêt qu'ils accordent à ce travail en acceptant de faire partie du jury.

Je ne sais pas comment exprimer ma gratitude envers Mr Morsli Mohammed, Professeur à l'université de Tizi-ouzou, autrement qu'en lui promettant d'agir comme lui avec des étudiants dans ma situation, si un jour l'occasion m'en sera donnée.

Je tiens également à remercier Fazia Khellas, le couple Merakeb, kader et farida ainsi que tous les membres du laboratoire LMPA. Le nombre élevé de fois où je suis venu frapper à leurs portes est au moins aussi grand que la disponibilité et la patience avec lesquelles ils ont, à chaque fois, pris le temps de me répondre.

Ces remerciements ne seraient pas complets sans mentionner mouh Sbihi. Il m'a évité, par ses multiples lectures, de rendre un manuscrit désagréablement rempli de fautes d'orthographe, de fautes de style et de fautes tout court.

Last but not least, je remercie du fond du cœur ma femme Sabrina. Merci à toi, la noblesse de ton nom n'est rien à côté de la noblesse de ta personne, de ta gentillesse et de ta générosité.

Table des matières

Introduction	3
1 Prédiction de séries chronologiques	6
1.1 Méthodes de calcul des coefficients des prédicteurs	7
1.1.1 Equations de prédiction	8
1.1.2 Algorithme de Durbin-Levinson	9
1.1.3 Algorithme d'innovation et applications	11
1.2 Prédiction à h pas	17
1.2.1 Relations de récurrence entre les coefficients des prédicteurs	18
1.2.2 Algorithmes récursifs	20
1.3 Prédiction d'une combinaison linéaire d'observations futures.	26
1.4 Factorisation de la densité spectrale et application à la prévision	28
1.4.1 Préliminaires sur la densité spectrale	28
1.4.2 Lien entre prédiction et densité spectrale	30
1.5 Processus p -stationnaire	33
1.5.1 Décomposition prédictive et représentation de Wold finie	35
1.5.2 Décomposition prédictive et décomposition de Wold	40
2 Estimation et interpolation des données manquantes d'une série chronologique stationnaire	41
2.1 Introduction	41
2.2 Préliminaires	42
2.3 Interpolation d'une seule valeur manquante	44
2.4 Interpolation de plusieurs valeurs manquantes	47
2.5 Algorithme EM	49
2.6 Problème d'interpolation	51
2.7 Interpolation suboptimale	54

2.8	Proximité de Pitman	57
3	Prédiction à passé incomplet	60
3.1	Introduction	60
3.2	Généralisation de l'algorithme d'innovation	61
3.3	Erreur de prévision à passé incomplet	64
3.4	Représentation AR du prédicteur à passé infini incomplet	72
3.5	Influence de valeurs manquantes sur la prédiction de séries chronologiques .	74
	Conclusion	78
	Bibliographie	80

Introduction

La prédiction à plusieurs pas apparaît dans de nombreuses applications telles que la prédiction de données économiques ou géophysiques, la prédiction de la consommation électrique ou la commande de processus. La plupart du temps, les prédicteurs à plusieurs pas sont calculés à partir d'un modèle initial ajusté aux données disponibles en remplaçant les données futures apparaissant dans le modèle par leur propre prédiction. Cette méthode utilise le modèle initial comme s'il s'agissait du « véritable » modèle ayant généré les données. Comme, en pratique, le modèle ajusté peut être incorrect, une méthode directe pour la prédiction à plusieurs pas consistant à ajuster un modèle différent pour chaque pas peut être plus appropriée. L'algorithme de Levinson-Durbin permet de calculer les paramètres de modèles autorégressifs d'ordre croissant minimisant une erreur quadratique moyenne de prédiction correspondant à un pas donné, et ceci quel que soit ce pas. Par conséquent, il est possible d'utiliser cet algorithme pour ajuster chaque modèle autorégressif indépendamment des autres. Lorsque les données à prédire sont consécutives dans le temps, il est naturel de se demander si les coefficients des prédicteurs à plusieurs pas peuvent être calculés d'une manière plus efficace. L'idée est alors d'utiliser des relations entre ces coefficients qui soient non seulement récursives sur l'ordre comme dans l'algorithme de Levinson-Durbin, mais aussi récursives sur le pas de prédiction. Des relations récursives sur le pas existent dans la littérature lorsque la prédiction est basée sur le passé infini, mais n'existaient pas lorsque la taille du passé est finie. Ces relations permettent de calculer les prédicteurs à plusieurs pas plus efficacement qu'en utilisant l'algorithme de Levinson-Durbin ou l'algorithme des innovations. Enfin, elles peuvent être facilement adaptées au cas d'une série chronologique vectorielle.

Dans de nombreuses situations pratiques, les données ne sont pas observées à des intervalles de temps réguliers, soit parce que certaines données sont manquantes, soit parce qu'une analyse statistique prudente conduit à éliminer certaines données que l'on juge

peu fiables. Dans ces cas, un modèle paramétrique peut être ajusté aux données soit en maximisant la fonction de vraisemblance exacte des données observées, soit en utilisant un algorithme de type EM. Lorsque le modèle paramétrique admet une représentation par espace d'état de dimension finie comme c'est le cas par exemple pour un modèle ARMA, cette représentation peut être modifiée pour tenir compte des données manquantes et la fonction de vraisemblance exacte peut être calculée au moyen des récursions de Kalman associées. Ensuite, des estimées des données manquantes peuvent être obtenues en utilisant un algorithme de lissage appliqué à la représentation d'état modifiée. Une autre approche consiste à utiliser des formules explicites des estimées des données manquantes dans lesquelles les paramètres inconnus sont remplacés par leurs estimations. Cette approche directe est plus efficace, particulièrement si le nombre de données manquantes est faible. De plus, des formules explicites sont utiles pour analyser théoriquement l'influence des positions des données manquantes sur les variances d'erreur d'interpolation. Bondon (2003), s'est intéressé au problème de la prédiction d'une série chronologique dont le passé est altéré par des données manquantes. Le nombre d'observations manquantes est supposé fini mais les indices temporels de ces données sont quelconques. Cette hypothèse est conforme à ce que l'on rencontre dans la pratique. Son travail s'est soldé par le résultat suivant : si le prédicteur basé sur le passé complet admet une représentation autorégressive, alors le meilleur prédicteur linéaire en moyenne quadratique basé sur le passé incomplet admet aussi une représentation autorégressive.

La présente thèse porte sur la prédiction de séries chronologiques dans le cas de valeurs manquantes. Elle apporte à la fois une nouvelle contribution et présente un tour d'horizon sur les divers résultats dans le domaine. Notre contribution est une étude comparative entre deux méthodes d'estimation de données manquantes par rapport au critère de proximité de Pitman. Elle est présentée au chapitre 2. Le reste de la thèse est organisé comme suit : dans le premier chapitre, nous nous sommes intéressés au problème de prédiction de séries chronologiques stationnaires. Dans un premier temps, nous exposons des relations de récurrence entre les coefficients des prédicteurs. Nous présentons ensuite, une méthode permettant d'effectuer des prévisions d'une combinaison linéaire d'observations futures. Ce chapitre s'achève par la prédiction de processus p -stationnaires.

Le deuxième chapitre est consacré à l'estimation des données manquantes d'une série chronologique. En premier lieu, nous résolvons le problème d'une seule donnée manquante en nous appuyant sur la décomposition de Wold. Cette procédure se généralise facilement au cas de plusieurs données manquantes. En second lieu, nous présentons une méthode

d'estimation dite interpolation suboptimale. Cette méthode consiste à construire deux estimations de la donnée manquante basées sur les observations passées et futures, ensuite nous chercherons la meilleure combinaison linéaire des deux estimations obtenues. En adaptant le critère de proximité de Pitman à la comparaison des estimateurs de données manquantes d'une série chronologique, nous montrons la supériorité de la méthode d'interpolation sur la méthode classique de prédiction linéaire.

Dans le troisième chapitre, nous nous intéressons au problème de la prédiction de X_t lorsque le passé infini de X_t est altéré par des données manquantes. Après quelques outils préliminaires, nous présentons l'algorithme d'innovation généralisé. Une formule explicite de la variance de l'erreur de prédiction est donnée dans la section 3. Dans la section 4, Nous nous intéressons à la représentation autorégressive des prédicteurs à passé incomplet. La dernière section de ce chapitre porte sur l'influence des positions des données manquantes sur la variance de l'erreur de prédiction. L'étude du comportement asymptotique du prédicteur à passé incomplet quand les indices des valeurs manquantes tendent vers l'infini est ensuite présentée. Dans ce cas, le prédicteur à passé incomplet tend vers le prédicteur à passé complet. Nous terminons par une caractérisation de la vitesse de convergence des processus à courte dépendance de type ARMA et pour des processus à longue dépendance de type ARMA fractionnaire (ARFIMA).

Chapitre 1

Prédiction de séries chronologiques

Ce chapitre porte sur la prédiction de séries chronologiques. Il s'agit de développer différentes approches et algorithmes permettant d'atteindre l'objectif en question.

Nous nous limitons à l'étude de prédicteurs linéaires. Le critère retenu pour mesurer la qualité d'un prédicteur \hat{X} de X est l'erreur quadratique $\mathbf{E}[(\hat{X} - X)^2]$. Pratiquement, il permet de pénaliser plus fortement les grandes erreurs que les petites. Mathématiquement, il permet de se placer dans l'espace de Hilbert $L^2(\Omega, \mathcal{F}, P)$ des variables aléatoires réelles de carrés intégrables et de pouvoir utiliser tous les outils afférents notamment pour expliciter \hat{X} . On obtient ainsi un prédicteur ponctuel et une estimation de l'erreur quadratique commise.

Il est important de noter que le prédicteur de Wiener-Kolmogorov, noté \hat{X}_{n+h} est un prédicteur de référence pour lequel l'erreur quadratique de prédiction est minimum. Plus précisément, \hat{X}_{n+h} est le meilleur prédicteur linéaire à l'horizon h , au sens de L^2 , connaissant le passé infini $\{X_{n+1-j}, j \geq 1\}$, c'est-à-dire, \hat{X}_{n+h} est la projection de X_{n+h} sur le sous-espace engendré par la famille de variables aléatoires $\{X_{n+1-j}, j \geq 1\}$.

Le prédicteur de Wiener-Kolmogorov s'écrit :

$$\hat{X}_{n+h} = P_n X_{n+h} = \sum_{j=0}^{\infty} a_j X_{n-j}$$

où P_n est le projecteur sur le sous espace engendré par la famille $\{X_j, j \leq n\}$.

Les $a_j, j \in \mathbb{N}$ s'obtiennent en minimisant l'espérance :

$$\mathbf{E}[(\hat{X}_{n+h} - X_{n+h})^2].$$

En pratique on ne dispose jamais d'un nombre infini d'observations, on ne peut donc pas

utiliser le prédicteur de Wiener-Kolmogorov. On a alors deux possibilités : soit recourir au prédicteur de Wiener-Kolmogorov tronqué soit utiliser le prédicteur projeté sur les n dernières observations disponibles.

Le prédicteur projeté sur les n dernières observations fera l'objet des deux premières sections en supposant que le processus est stationnaire. Dans ce contexte, des algorithmes et des relations de récurrence entre les coefficients des prédicteurs sont établies en s'appuyant sur le domaine des temps. Il est également possible d'obtenir d'autres relations récurrentes en utilisant le domaine des fréquences, ce qui fera l'objet de la section 3. Dans la section 4, nous nous intéressons aux prévisions de combinaisons linéaires d'observations futures. Le chapitre sera achevé par un exposé synthétique de certains résultats portant sur les processus p -stationnaires.

1.1 Méthodes de calcul des coefficients des prédicteurs

Considérons une série chronologique $\{X_t, t \in \mathbb{Z}\}$, à valeurs réelles et de variance finie. On suppose ce processus stationnaire dans le sens où sa fonction de covariance $Cov(X_s, X_t)$ ne dépend que de $t - s$. Cette hypothèse est fondamentale et correspond à une régularité dans le temps de certaines propriétés statistiques du processus. En pratique, la stationnarité n'est pas nécessairement satisfaite et devra donc être vérifiée au moyen de tests statistiques adéquats. La stationnarité permet d'obtenir des informations sur des données futures en utilisant des statistiques calculées au moyen de données passées. Le problème de la prédiction linéaire à un pas de X_t consiste à trouver une fonction linéaire de X_{t-1}, X_{t-2}, \dots qui approxime X_t au sens de l'erreur quadratique moyenne. En pratique, on dispose seulement d'un nombre fini d'observations $X_{t-1}, X_{t-2}, \dots, X_{t-n}$ à partir desquelles on construit le prédicteur $\hat{X}_{t,n}$ sous la forme $\hat{X}_{t,n} = \sum_{k=1}^n a_{k,n,t} X_{t-k}$. Les coefficients $a_{k,n,t}$ sont obtenus en régressant X_t sur $X_{t-1}, X_{t-2}, \dots, X_{t-n}$. L'hypothèse de stationnarité permet de démontrer que les coefficients $a_{k,n,t}$ sont indépendants de t . On pose alors $a_{k,n,t} = a_{k,n}$, et le calcul des paramètres $a_{k,n}$ nécessite l'inversion d'une matrice de covariance de taille n . Quand n est grand, il est souvent plus commode de résoudre le problème de prédiction en supposant que l'on dispose d'un nombre infini d'observations ($n = \infty$). On note \hat{X}_t le prédicteur linéaire de X_t correspondant, et on suppose que \hat{X}_t peut s'écrire sous la forme $\hat{X}_t = \sum_{k=1}^{\infty} a_k X_{t-k}$ où la série converge en moyenne quadratique. On dit alors que X_t admet une représentation autorégressive, et les coefficients a_k s'appellent les paramètres autorégressifs. Des conditions assez faibles sur la densité spectrale f de $\{X_t\}$ garantissant l'existence et l'unicité

d'une représentation autorégressive. Un prédicteur de X_t à l'échantillon de taille finie n est obtenu par la série tronquée $S_{t,n} = \sum_{k=1}^n a_k X_{t-k}$. L'intérêt de cette approche est que les paramètres a_k ne dépendent pas de la taille du passé n , ils dépendent uniquement des coefficients de Fourier de $\ln f$ et peuvent être calculés facilement.

1.1.1 Equations de prédiction

Soit \hat{X}_{n+1} le prédicteur de la variable X_{n+1} défini par

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{si } n = 0, \\ P_{1,n} X_{n+1} & \text{si } n \geq 1, \end{cases}$$

où $P_{k,l}$ est le projecteur sur $\mathcal{H}_{k,l} = \overline{\text{sp}}\{X_j, k \leq j \leq l\}$.

Comme $\hat{X}_{n+1} \in \mathcal{H}_{1,n}$, il s'écrit

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}, \quad n \geq 1, \quad (1.1)$$

où les $\phi_{n1}, \phi_{n2}, \dots, \phi_{nn}$ vérifient les équations de prédiction suivantes :

$$\left\langle \sum_{i=1}^n \phi_{ni} X_{n+1-i}, X_{n+1-j} \right\rangle = \langle X_{n+1}, X_{n+1-j} \rangle \quad j = 1, \dots, n \quad (1.2)$$

En utilisant la linéarité du produit scalaire, ces équations peuvent s'écrire sous la forme

$$\sum_{i=1}^n \phi_{ni} \gamma(i-j) = \gamma(j), \quad j = 1, \dots, n. \quad (1.3)$$

ou d'une manière plus compacte

$$\Gamma_n \Phi_n = \gamma_n, \quad (1.4)$$

où $\Gamma_n = [\gamma(i-j)]_{i,j=1,\dots,n}$, $\gamma_n = (\gamma(1), \dots, \gamma(n))'$ et $\Phi_n = (\phi_{n1}, \phi_{n2}, \dots, \phi_{nn})'$. Si Γ_n est non sigulière, alors (1.4) admet une seule solution donnée par

$$\Phi_n = \Gamma_n^{-1} \gamma_n. \quad (1.5)$$

C'est notamment le cas sous les hypothèses de la proposition suivante.

Proposition 1.1. *Si $\gamma(0) > 0$ et $\gamma(h) \rightarrow 0$ quand $h \rightarrow \infty$ alors la matrice de covariance $\Gamma_n = [\gamma(i-j)]_{i,j=1,\dots,n}$ de (X_1, \dots, X_n) est non-singulière pour tout n .*

Corollaire 1.1. *Sous les hypothèses de la proposition 1.1, le meilleur prédicteur linéaire \hat{X}_{n+1} de X_{n+1} en fonction de X_1, \dots, X_n est*

$$\hat{X}_{n+1} = \sum_{i=1}^n \phi_{ni} X_{n+1-i}, \quad n = 1, 2, \dots, \quad (1.6)$$

où $\Phi = (\phi_{n1}, \dots, \phi_{nn})' = \Gamma_n^{-1} \gamma_n$, $\gamma_n = (\gamma(1), \dots, \gamma(n))'$ et $\Gamma_n = [\gamma(i-j)]_{i,j=1, \dots, n}$. De plus, l'erreur quadratique moyenne est $\nu_n = \gamma(0) - \gamma_n \Gamma_n^{-1} \gamma_n$.

Calcul des coefficients des prédicteurs

L'équation (1.5) donne l'expression du prédicteur mais nécessite le calcul de l'inverse de la matrice Γ . Dans ce qui suit, nous décrivons deux algorithmes récursifs simples et efficaces permettant d'obtenir l'expression du meilleur prédicteur linéaire $\hat{X}_{n+1}, n \geq 1$ défini dans (1.6), sans avoir à calculer explicitement l'inverse de Γ . Nous montrons également comment ce prédicteur peut être utilisé pour trouver l'expression de $\hat{X}_{n+h}, h \geq 2$.

1.1.2 Algorithme de Durbin-Levinson

L'algorithme de Durbin-Levinson donne un schéma récursif pour calculer les coefficients $\Phi_n = (\phi_{n1}, \dots, \phi_{nn})'$ du prédicteur \hat{X}_{n+1} défini dans (1.6) ainsi que les erreurs quadratiques moyennes de la prédiction définies par

$$\nu_n = \mathbf{E}[X_{n+1} - \hat{X}_{n+1}]^2, \quad n \geq 1. \quad (1.7)$$

Il est clair que $\nu_0 = 0$.

Proposition 1.2. *Si $\{X_t\}$ est stationnaire, centré de fonction d'autocovariance $\gamma(\cdot)$ telle que $\gamma(0) > 0$ et $\gamma(h) \rightarrow 0$ quand $h \rightarrow \infty$, alors les coefficients ϕ_{nj} et l'erreur quadratique moyenne ν_n définie dans (1.6) et (1.7) vérifient $\phi_{11} = \gamma(1)/\gamma(0)$, $\nu_0 = \gamma(0)$,*

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right] \nu_{n-1}^{-1} \quad (1.8)$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \quad (1.9)$$

et

$$\nu_n = \nu_{n-1} [1 - \phi_{nn}^2]. \quad (1.10)$$

Démonstration. Par définition, $\mathcal{H}_{2,n} = \overline{\text{sp}}\{X_2, \dots, X_n\}$ et $\mathcal{H}_{1,1}^\perp = \overline{\text{sp}}\{X_1 - P_{\mathcal{H}_{2,n}}X_1\}$ sont deux sous espaces orthogonaux de $\mathcal{H}_{1,n}$. Ce qui implique $\forall Y \in L^2(\Omega, \mathcal{F}, p)$, $P_{1,n}Y = P_{2,n}Y + P_{1,1}^\perp Y$ où $P_{1,1}^\perp$ est le projecteur sur $\mathcal{H}_{1,1}^\perp = \overline{\text{sp}}\{X_1 - P_{\mathcal{H}_{2,n}}X_1\}$.

Donc

$$\hat{X}_{n+1} = P_{2,n}X_{n+1} + P_{1,1}^\perp X_{n+1} = P_{2,n}X_{n+1} + a(X_1 - P_{2,n}X_1), \quad (1.11)$$

où

$$a = \langle X_{n+1}, X_1 - P_{2,n}X_1 \rangle / \|X_1 - P_{2,n}X_1\|^2. \quad (1.12)$$

La stationnarité du processus entraîne que les vecteurs $(X_1, \dots, X_n)'$ et $(X_2, \dots, X_{n+1})'$ ont une même matrice d'autocovariance, et pour la même raison, les vecteurs $(X_n, X_{n-1}, \dots, X_1)'$ et $(X_2, \dots, X_{n+1})'$ ont également la même matrice de covariance, par conséquent

$$P_{2,n}X_1 = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{j+1} \quad (1.13)$$

$$P_{2,n}X_{n+1} = \sum_{j=1}^{n-1} \phi_{n-1,j} X_{n+1-j} \quad (1.14)$$

et

$$\|X_1 - P_{2,n}X_1\|^2 = \|X_{n+1} - P_{2,n}X_{n+1}\|^2 = \|X_n - \hat{X}_n\|^2 = \nu_{n-1}. \quad (1.15)$$

En utilisant (1.11), (1.13) et (1.14), on obtient

$$\hat{X}_{n+1} = aX_1 + \sum_{j=1}^{n-1} [\phi_{n-1,j} - a\phi_{n-1,j-j}] X_{n+1-j}. \quad (1.16)$$

De (1.11)

$$\begin{aligned} a &= \left(\langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} \phi_{n-1,j} \langle X_{n+1}, X_{j+1} \rangle \right) \nu_{n-1}^{-1} \\ &= \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right] \nu_{n-1}^{-1}. \end{aligned} \quad (1.17)$$

La proposition (1.1) nous assure l'unicité de la représentation

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}. \quad (1.18)$$

En identifiant, les paramètres dans (1.16) et (1.18), on déduit

$$\phi_{nn} = a \quad (1.19)$$

et

$$\phi_{nj} = \phi_{n-1,j} - a\phi_{n-1,n-j}, \quad j = 1, \dots, n-1. \quad (1.20)$$

De (1.8) et (1.9) découle

$$\begin{aligned}
 \nu_n &= \|X_{n+1} - \hat{X}_{n+1}\|^2 \\
 &= \|X_{n+1} - P_{2,n}X_{n+1}\|^2 + \|P_{1,1}^\perp X_{n+1}\|^2 - \langle X_{n+1} - P_{2,n}X_{n+1}, P_{1,1}^\perp X_{n+1} \rangle \\
 &= \nu_{n-1} + a^2\nu_{n-1} - 2a \langle X_{n+1}, X_1 - P_{2,n}X_1 \rangle.
 \end{aligned} \tag{1.21}$$

En utilisant (1.15) et l'orthogonalité de $\mathcal{H}_{2,n}$ avec $\mathcal{H}_{1,1}^\perp$ on obtient

$$\nu_n = \nu_{n-1}(1 - a^2). \tag{1.22}$$

□

1.1.3 Algorithme d'innovation et applications

L'idée principale de l'algorithme de Durbin-Levinson est de décomposer le sous espace $\mathcal{H}_{1,n}$ en deux sous espaces orthogonaux $\mathcal{H}_{2,n}$ et $\mathcal{H}_{1,1}^\perp$. L'idée de l'algorithme d'innovation est de décomposer le sous espace $\mathcal{H}_{1,n}$ en n sous espaces orthogonaux. L'algorithme qui suit permet d'étendre l'expression du prédicteur à un cadre plus général qui inclut le cas non stationnaire.

Proposition 1.3. (*Brockwell et Davis, 1991*)

Si $\{X_t\}$ est centré et $\mathbf{E}[X_i X_j] = \kappa(i, j)$, où la matrice $[\kappa(i, j)]_{i, j=1}^n$ est non-singulière pour tout $n = 1, 2, \dots$, alors les coefficients du prédicteur $\hat{X}_{n+1, n} \geq 0$ défini par

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{si } n = 0 \\ \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) & \text{si } n \geq 1 \end{cases} \tag{1.23}$$

et les erreurs quadratiques ν_n s'obtiennent par les relations de récurrence suivantes :

$$\begin{cases} \nu_0 &= \kappa(1, 1). \\ \theta_{n, n-k} &= \nu_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} \nu_j \right), \quad k = 0, 1, \dots, n-1, \\ \nu_n &= \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n, n-j} \nu_j. \end{cases} \tag{1.24}$$

Prévision d'un ARMA

Rappelons qu'un processus aléatoire $\{X_t\}$ du second ordre ($\mathbf{E}[X_t^2] < \infty$) admet une représentation ARMA d'ordre p et q , s'il est solution d'une équation aux différentielles stochastiques de la forme

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}, \quad t \in \mathbb{Z}$$

ou encore

$$\Phi(B)X_t = \Theta(B)\epsilon_t, \quad t \in \mathbb{Z}$$

où $\{\epsilon_t\}$ est un processus bruit blanc, i.e. une suite de variables aléatoires non corrélées ($\mathbf{E}[\epsilon_t \epsilon_{t-h}] = 0, h \neq 0$) de moyenne $\mathbf{E}[\epsilon_t]$ nulle et de variance σ^2 , et $\Phi(B) = -\sum_{j=1}^p \phi_j B^j$ et $\Theta(B) = \sum_{j=1}^q \theta_j B^j$, $\theta_j \in \mathbb{R}, \forall j < p$, $\phi_j \in \mathbb{R}, \forall j < q$, $\phi_0 = \theta_0 = 1$ et $(\phi_p, \theta_q) \in \mathbb{R}^* \times \mathbb{R}^*$.

Introduisons un processus transformé, en notant $m = \max(p, q)$,

$$\begin{cases} W_t = \frac{1}{\sigma} X_t, & t = 1, \dots, m \\ W_t = \frac{1}{\sigma} (X_t - a_1 X_{t-1} - \dots - a_p X_{t-p}), & t > m. \end{cases}$$

De plus, on définit $b_0 = 0$ pour $j > q$. On supposera aussi que $p \geq 1$ et $q \geq 1$ quitte à choisir des coefficients égaux à 0. Supposons aussi qu'on ait calculé la fonction d'auto-covariance de $\{X_t, t \in \mathbb{Z}\}$. L'auto-covariance $\kappa(i, j) = \mathbf{E}[W_i W_j]$, $i, j \geq 1$ s'écrit alors :

$$\begin{cases} \frac{1}{\sigma^2} \gamma_X(i-j), & 1 \leq i, j \leq m \\ \frac{1}{\sigma^2} (\gamma_X(i-j) - \sum_{r=1}^p a_r \gamma_X(r - |i-j|)), & \min(i, j) \leq m \leq \max(i, j) \leq 2m \\ \sum_{r=0}^q b_r b_{r+|i-j|}, & \min(i, j) > m \\ 0 & \text{sinon} \end{cases}$$

En appliquant l'algorithme d'innovation à W_t , on obtient

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^n \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \leq n < m \\ \hat{W}_{n+1} = \sum_{j=1}^q \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}), & n \geq m \end{cases}$$

où les coefficients θ_{nj} et les erreurs quadratiques moyennes $\nu_n = \mathbf{E} [W_{n+1} - \hat{W}_{n+1}]^2$ sont calculés récursivement par l'algorithme d'innovation. On notera que $\theta_{nj} = 0$ si $n \geq m$ et $j > q$. C'est la conséquence de $\kappa(r, s) = 0$ si $r > m$ et $|r - s| > q$.

Maintenant, comme $X_t - \hat{X}_t = \sigma(W_t - \hat{W}_t)$, on aura finalement, pour tout $t \geq 1$

$$\begin{cases} \hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < m \\ \hat{X}_{n+1} = a_1 X_{n+1} + \dots + a_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m. \end{cases}$$

Prévision d'un PARMA

Les techniques standards d'analyse de séries chronologiques ont longtemps reposé sur les propriétés fondamentales de linéarité et de stationnarité. Eu égard à leurs coefficients constants et à leur dynamique linéaire, les modèles autorégressifs moyennes mobiles (ARMA) ont ainsi fait l'objet d'un intérêt croissant depuis le début des années 1980. Cependant, de nombreuses recherches ont démontré que les hypothèses de linéarité et de stationnarité n'étaient qu'un confort appréciable dans l'étude probabiliste et statistique du modèle. Le recours à des modèles plus souples est alors vivement apparu comme une nécessité en séries chronologiques, parmi ces modèles, nous citons les modèles à structure périodique. Notons que ces modèles apparaissent dans plusieurs domaines, la climatologie (Hannan, 1955; Jones et Brelsford 1967, Lund et al., 1995), l'hydrologie (McLeod, 1993), économétrie (Parzen et Pagano, 1979).

Définition 1.1. Un processus du second ordre $\{Y_t; t \in \mathbb{Z}\}$ de moyenne μ_t et de fonction d'autocovariance $\gamma(t,s)$ est dit périodiquement corrélé (ou périodique au sens faible), s'il existe un entier positif d , tel que pour tout $t,s \in \mathbb{Z}$, on ait

$$\begin{aligned}\mu_{t+sd} &= \mu_t \\ \gamma(t+d\tau, s+d\tau) &= \gamma(t,s), \quad \forall \tau \in \mathbb{Z}.\end{aligned}$$

Le plus petit positif d satisfaisant à ces deux conditions est appelé la période du processus périodiquement corrélé Y_t . Gladyshev (1961) a établi un lien entre un processus périodiquement corrélé de période d et un processus d -varié stationnaire, c'est à dire, on peut représenter un processus périodique, de période d ($d \geq 2$), par un processus d -varié stationnaire.

Définition 1.2. On dit qu'un processus périodiquement corrélé $\{Y_t; t \in \mathbb{Z}\}$ admet une représentation autorégressive moyenne mobile d'ordre p_t et q_t , périodique de période d , noté ARMA(p_t, q_t), s'il est solution d'une équation aux différences stochastique de la forme

$$Y_t - \sum_{i=1}^{p_t} \phi_{t,i} Y_{t-i} = \epsilon_t - \sum_{j=1}^{q_t} \theta_{t,j} \epsilon_{t-j}, \quad t \in \mathbb{Z} \quad (1.25)$$

qui s'écrit encore sous la forme :

$$\Phi_t(L)Y_t = \Theta_t(L)\epsilon_t, \quad t \in \mathbb{Z},$$

où $\{\epsilon_t; t \in \mathbb{Z}\}$ est un processus non corrélé de moyenne nulle et de variance σ_t^2 .

Les coefficients autorégressifs $\phi_{t,i}$, $i = 1, \dots, p_t$, les coefficients moyenne mobile $\theta_{t,j}$, $j =$

$1, \dots, q_t$, l'ordre (p_t, q_t) ainsi que la variance du bruit blanc σ_t^2 sont des fonctions périodiques dans le temps de période d , c'est à dire :

$$\begin{aligned} \phi_{t+dk} &= \phi_{t,i} \quad \text{pour } i = 1, \dots, p_t \\ \theta_{t+dk,i} &= \theta_{t,i}, \quad \text{pour } j = 1, \dots, q_t \\ \sigma_{t+dk,j}^2 &= \sigma_t^2 \\ p_{t+dk} &= p_t \\ q_t &= q_{t+dk}, \quad \forall k, t \in \mathbb{Z} \end{aligned}$$

et les deux polynômes périodiques $\Phi_t(L)$ et $\Theta_t(L)$ sont donnés par :

$$\Phi_t(L) = - \sum_{i=1}^{p_t} \phi_{t,i} L^i \quad \text{et} \quad \Theta_t(L) = 1 - \sum_{j=1}^{q_t} \theta_{t,j} L^j, \quad t \in \mathbb{Z}$$

où L est l'opérateur retard ($L^j Y_t = Y_{t-j}, t \in \mathbb{Z}, j \in \mathbb{N}$).

L'équation (1.25) peut s'écrire sous forme d'un modèle ARMA d -varié stationnaire (voir Vacchia 1985a)

$$\Phi_0 X_{n-k} - \sum_{k=1}^{p^*} \Phi(k) X_{n-k} = \Theta_0 \eta_0 + \sum_{k=1}^{q^*} \Theta_k \eta_{n-k}, \quad (1.26)$$

où $X_n = (Y_{Nd+1}, \dots, Y_{nd+d})'$ et $\eta = (\epsilon_{nd+1}, \dots, \epsilon_{nd+d})'$, l'ordre (p^*, q^*) est donné par

$$\begin{aligned} p^* &= \max_{1 \leq i \leq d} \left(\frac{p_i - i}{d} + 1 \right) \\ q^* &= \max_{1 \leq i \leq d} \left(\frac{q_i - i}{d} + 1 \right). \end{aligned}$$

Les matrices carrées $(d \times d)$ des paramètres autorégressifs Φ_0 et $\Phi_j, j = 1, \dots, p^*$ et des paramètres de la représentation moyenne mobile Θ_0 et $\Theta_i, i = 1, \dots, q^*$ sont données par

$$(\Phi_0)_{i,j} = \begin{cases} 1 & i = j \\ 0 & i < j, \\ \phi_{i-j}(i) & i > j \end{cases} \quad (\Phi_k)_{i,j} = \phi_{kd+i-j}(i) \quad 1 \leq k \leq p^*. \quad (1.27)$$

La solution de l'équation (1.25) est unique et peut être représentée par une forme périodique causale donnée par

$$Y_{nd+h} = \sum_{k=1}^{\infty} \psi_k(h) \epsilon_{nd+h-k}, \quad (1.28)$$

où les coefficients $\psi_k(h)$ sont absolument sommables, i.e.

$$\sum_{k=1}^{\infty} |\psi_k(k)| < \infty. \quad (1.29)$$

Pour tout $1 \leq h \leq d$, la représentation causale donnée par (1.28) est vérifiée si

$$\det \left[\Phi_0 - \sum_{k=1}^{\infty} \Phi_k z^k \right] \neq 0 \quad (1.30)$$

pour tout $|z| \leq 1$, (Vacchia, 1985a). Les coefficients $\psi_k(h)$ s'obtiennent en itérant la récurrence

$$\psi_k(h) = \theta_k(h) 1_{k \leq q} + \sum_{j=1}^{\min(k,p)} \phi_j(h) \psi_{k-j}(h-j) \quad k \geq 1, 1 \leq h \leq d. \quad (1.31)$$

La représentation périodique inversible

$$\epsilon_{nd+h} = \sum_{k=0}^{\infty} \pi_k(h) Y_{nd+h-k} \quad (1.32)$$

où

$$\sum_{k=1}^{\infty} |\pi_k(h)| < \infty \quad (1.33)$$

pour tout $1 \leq h \leq d$, la représentation (1.32) est vérifiée si

$$\det \left[\Theta_0 + \sum_{k=1}^{q^*} \Theta_k z^k \right] \neq 0 \quad (1.34)$$

pour tout $|z| \leq 1$. Les coefficients de cette représentation s'obtiennent par la relation

$$\pi_k(h) = -\phi_k(h) 1_{[k \leq p]} - \sum_{j=1}^{\min(k,p)} \theta_j(h) \pi_{k-j}(h-j) \quad k \geq 1, 1 \leq h \leq T. \quad (1.35)$$

Bentarzi et Hallin (1993) donnent une condition nécessaire et suffisante pour qu'un processus PARMA soit causal et inversible.

Dans ce qui suit, en posant $m = \max(p, q)$, nous allons vérifier que l'expression du meilleur prédicteur linéaire de la forme

$$\hat{Y}_{t+1} = P_t Y_{t+1} \quad (1.36)$$

est

$$\hat{Y}_{nd+h} = \sum_{j=1}^{nd+h-1} \vartheta_{nd+h-1,j} (Y_{nd+h-j} - \hat{Y}_{nd+h-j}) \quad 2 \leq nd+h \leq m. \quad (1.37)$$

et

$$\hat{Y}_{nd+h} = \sum_{k=1}^p \phi_k(h) Y_{nd+h-k} + \sum_{k=1}^q \vartheta_{nd+h-1} (Y_{nd+h-k} - \hat{Y}_{nd+h-k}) \quad nd+h > m \quad (1.38)$$

Pour établir (1.37), introduisons le processus transformé $\{W_t\}$ défini par

$$W_{nd+h} = \begin{cases} Y_{nd+h} & nd+h \leq m, \\ Y_{nd+h} - \sum_{k=1}^p \phi_k(h) Y_{nd+h-k} & nd+h > m. \end{cases} \quad (1.39)$$

Comme $\overline{\text{sp}}\{Y_1, \dots, Y_n\} = \overline{\text{sp}}\{W_1, \dots, W_n\}$ pour tout $t \in \mathbb{Z}$, il s'en suit que, $\hat{W}_{nd+h} = P_t(W_{t+1})$ et

$$\hat{W}_{nd+h} = \hat{Y}_{nd+h} - \sum_{k=1}^p \phi_k(h) Y_{nd+h-k} \quad nd+h > m \quad (1.40)$$

d'où

$$Y_t - \hat{Y}_t = W_t - \hat{W}_t, \quad t \geq 1 \quad (1.41)$$

et $\nu_t = \mathbf{E}\{W_{t+1} - \hat{W}_{t+1}\}$ pour tout $t \geq 0$.

En utilisant la notation $\kappa_W(t, s) = \text{Cov}(W_t, W_s)$ et en appliquant l'algorithme d'innovation au processus $\{W_t\}$ en se donnant $\hat{W}_1 = 0$, $\nu_0 = \kappa_W(1, 1)$, on obtient

$$\hat{W}_{t+1} = \sum_{j=1}^t \vartheta_{t,j} (W_{t+1-j} - \hat{W}_{t+1-j}) \quad t \geq 1 \quad (1.42)$$

$$\vartheta_{t,t-k} = \nu_k^{-1} \left\{ \kappa_W(t+1, k+1) - \sum_{j=0}^{k-1} \vartheta_{k,k-j} \vartheta_{t,t-j} \nu_j \right\} \quad 0 \leq k \leq t-1 \quad (1.43)$$

et

$$\nu_t = \kappa_W(t+1, t+1) - \sum_{j=0}^{t-1} \vartheta_{t,t-j} \nu_j \quad t \geq 1. \quad (1.44)$$

En utilisant $\hat{W}_j = \hat{Y}_j$ pour $j \leq m$ et (1.41) dans (1.42), on obtient (1.37).

En multipliant chaque membre de (1.42) par $W_{k+1} - \hat{W}_{k+1}$ pour $0 \leq k \leq t-1$, on obtient

$$\begin{aligned} \nu_k \vartheta_{t,t-k} &= \mathbf{E}\{\hat{W}_{t+1} (W_{k+1} - \hat{W}_{k+1})\} \\ &= \mathbf{E}\{W_{t+1} (W_{k+1} - \hat{W}_{k+1})\} \end{aligned} \quad (1.45)$$

De (1.39), il s'en suit que, $\kappa_W(t, j) = 0$ pour $t > m$ et $|t-j| > q$. Du fait que $W_{t+1} - \hat{W}_{t+1}$ est une combinaison linéaire de Y_1, \dots, Y_{t+1-k} et chacun d'eux est non corrélé avec W_{t+1} alors

$$\vartheta_{t,k} = 0 \quad t \geq m, k > q. \quad (1.46)$$

En remplaçant (1.46) dans (1.42), on obtient (1.38).

Quelques manipulations directes montrent que (1.43) et (1.44) se réduisent à

$$\vartheta_{t,r} = \nu_{t-1}^{-1} \left\{ \kappa_W(t+1, t-r+1) - \sum_{j=r+1}^q \vartheta_{t-r, j-r} \vartheta_{t,j} \nu_{t-j} \right\} \quad t \geq m, 1 \leq r \leq q \quad (1.47)$$

et

$$\nu_t = \kappa_W(t+1, t+1) - \sum_{j=1}^q \vartheta_{t,j}^2 \nu_{t-j} \quad t \geq m. \quad (1.48)$$

Pour $1 \leq j \leq t$ et $1 \leq t \leq m-1$, les coefficients $\{\vartheta_{t,j}\}$ et ν_t s'obtiennent par (1.43) et (1.44) dans l'ordre suivant : $\nu_0; \vartheta_{1,1}, \nu_1; \vartheta_{2,2}, \vartheta_{2,1}, \nu_2; \dots; \vartheta_{m-1, m-1}, \dots, \vartheta_{m-1, 1}, \nu_{m-1}$. Pour $t \geq m$ et $1 \leq j \leq q$, on utilise (1.47) et (1.48) pour les calcul des $\vartheta_{t,j}$ et ν_t . Ces derniers s'obtiennent dans l'ordre $\vartheta_{m,q}, \vartheta_{m,q-1}, \dots, \vartheta_{m,1}, \nu_m; \vartheta_{m+1,q-1}, \dots, \vartheta_{m+1,1}, \nu_{m+1}$; etc.

La mise en œuvre de cet algorithme nécessite le calcul des covariances du processus $\{W_t\}$. En notant $\kappa_X(t, s) = \text{cov}(X_t, X_s)$, il est facile de montrer que

$$\kappa_W(t, s) = \begin{cases} \kappa_X(t, s) & t \leq m \\ \kappa_X(t, s) - \sum_{k=1}^p \phi_k(t) \kappa_X(t-k, s) & s \leq m \leq 2m \\ \sum_{k=0}^q \theta_k(s) \theta_{t-s+k}(t) \sigma^2(s-k) \mathbf{1}_{[0 \leq t-s+k \leq q]} & s > m \\ 0 & \text{ailleurs.} \end{cases} \quad (1.49)$$

1.2 Prédiction à h pas

Dans cette section, nous allons nous intéresser à la construction du meilleur prédicteur linéaire à h pas sur celle à 1 pas. En d'autres termes, des algorithmes et des relations de récurrence entre les coefficients des prédicteurs $X_{n+h}, h \geq 1$, d'une série chronologique sont présentés.

Supposons que $\{X_n\}$ soit purement non déterministe, de représentation MA(∞) donnée par

$$X_n = \sum_{i=0}^{\infty} c_i \epsilon_{n-i}, \quad (1.50)$$

où $\epsilon_n = X_n - P_{n-1}X_n$, $c_0 = 1$, et $\sum_{i=0}^{\infty} c_i^2 < \infty$. La représentation autorégressive AR(∞) de $\{X_n\}$ est

$$X_n = \epsilon_n + \sum_{i=1}^{\infty} a_i X_{n-i}. \quad (1.51)$$

Il est bien connu que les paramètres de la représentation autorégressive et de la représentation moyenne mobile sont liés par les relations

$$\begin{aligned} a_0 &= -1 \\ a_i &= -\sum_{j=0}^{\infty} a_j c_{i-j} \quad i \geq 1. \end{aligned}$$

Comme $\{X_n\}$ admet une représentation autorégressive, il en résulte de Bloomfield (1985) que, pour tout $h \geq 1$,

$$P_n X_{n+h} = \sum_{i=1}^{\infty} a_i^h X_{n+1-i},$$

où

$$a_i^h = \sum_{j=0}^{i+h-1} c_j a_{i+h-1-j} = -\sum_{j=h}^{i+h-1} c_j a_{i+h-1-j} \quad i \geq 1. \quad (1.52)$$

Il est à noter, que les coefficients des prédicteurs à passé infini vérifient les relations de récurrence :

$$a_i = a_{i+h-1} + \sum_{j=1}^{h-1} a_j a_i^{h-j}, \quad (1.53)$$

$$a_i^h = a_{i+1}^{h-1} + c_{h-1} a_i. \quad (1.54)$$

1.2.1 Relations de récurrence entre les coefficients des prédicteurs

Dans ce qui suit, deux relations de récurrence entre les coefficients des prédicteurs d'horizon $h, h \geq 2$ et les coefficients du prédicteurs d'horizon 1 sont présentées, dans le cas d'un passé fini.

Proposition 1.4. (Bondon, 2001).

Pour tout $h \geq 1$ et pour tout $n \geq 1$,

$$a_{n,i}^h = a_{n+h-1,i+h-1}^1 + \sum_{j=1}^{h-1} a_{n+h-1,j}^1 a_{n,i}^{h-j} \quad i = 1, \dots, n. \quad (1.55)$$

Démonstration. Pour $h \geq 1$

$$\mathcal{H}_{1,n} \subset \mathcal{H}_{1,n+h-1}$$

et

$$P_{1,n} X_{n+h} = P_{1,n} \circ P_{1,n+h-1} X_{n+h}$$

par conséquent

$$\begin{aligned}
 P_{1,n}X_{n+h} &= P_{1,n} \sum_{j=1}^{n+h-1} a_{n+h-1}^1 X_{n+h-j} \\
 &= \sum_{j=1}^{h-1} a_{n+h-1,j} P_{1,n}X_{n+h-j} + \sum_{j=h}^{n+h-1} a_{n+h-1}^1 X_{n+h-j} \\
 &= \sum_{j=1}^{h-1} a_{n+h-1,j}^1 \sum_{i=1}^n a_{n,i}^{h-j} X_{n+1-i} + \sum_{i=1}^n a_{n+h-1,i+h-1}^1 X_{n+1-i}
 \end{aligned}$$

ce qui est équivalent à (1.55) car $\{X_n\}$ est non déterministe. \square

Proposition 1.5. (Bondon, 2001).

Pour un horizon de prévision $h \geq 2$ et pour tout $n \geq 1$

$$a_{n,i}^h = a_{n+1,i+1}^{h-1} + a_{n+1,1}^{h-1} a_{n,i}^1 \quad i = 1, \dots, n \quad (1.56)$$

$$\nu_n^h = \nu_{n+1}^{h-1} + (a_{n+1,1}^{h-1})^2 \nu_n^1. \quad (1.57)$$

Démonstration. Comme

$$\mathcal{H}_{1,n+1} = \mathcal{H}_{1,n} \oplus \overline{\text{sp}} \{X_{n+1} - P_{1,n}X_{n+1}\}$$

donc

$$P_{1,n+1}X_{n+h} = P_{1,n}X_{n+h} + \lambda(X_{n+1} - P_{1,n}X_{n+1}) \quad (1.58)$$

où $\lambda \in \mathbb{R}$. L'équation (1.58) est équivalente à

$$\sum_{i=1}^{n+1} a_{n+1,i}^{h-1} X_{n+2-i} = \sum_{i=1}^n a_{n,i}^h X_{n+1-i} + \lambda X_{n+1} - \lambda \sum_{i=1}^n a_{n,i}^1 X_{n+1-i}.$$

On en déduit que

$$a_{n+1,1}^{h-1} = \lambda$$

et

$$a_{n+1,i+1}^{h-1} = a_{n,i}^h - \lambda a_{n,i}^1 \quad \text{pour } 1 \leq i \leq n.$$

Ce qui prouve (1.56).

Il en résulte aussi de (1.58) que

$$\|P_{1,n+1}X_{n+h}\|^2 = \|P_{1,n}X_{n+h}\|^2 + \lambda^2 \nu_n^1.$$

Or

$$\nu_n^1 = \|X_{n+h}\|^2 - \|P_{1,n}X_{n+h}\|^2$$

et

$$\nu_{n+1}^{h-1} = \|X_{n+h}\|^2 - \|P_{1,n+1}X_{n+h}\|^2.$$

Par conséquent,

$$\nu_n^h = \nu_{n+1}^{h-1} + \lambda^2 \nu_n^1,$$

ce qui donne (1.57). □

Remarque 1.1. Si $\{X_t\}$ est un processus autorégressif causal d'ordre p , alors $a_i = 0$ pour $i > p$ dans (1.51). Par conséquent, les relations de récurrence données par (1.53) et (1.54) peuvent être utilisées pour calculer $P_{1,n}X_{n+h}$ pourvu que $n \geq p$.

1.2.2 Algorithmes récursifs

Dans cette partie, le problème posé est un problème de réduction numérique. Autrement dit, un problème de minimisation du nombre d'opérations nécessaires pour le calcul des coefficients des prédicteurs. Visant cet objectif, nous montrons comment les relations (1.55), (1.56) et (1.57) peuvent être utilisées pour le calcul des coefficients $a_{p,i}^h$ et des ν_p^h pour $1 \leq h \leq s$, où p est un ordre donné et s est l'horizon final de la prévision. Signalons au passage, pour $h = 1$, on peut utiliser l'algorithme de Levinson.

Proposition 1.6. (Bondon, 2001).

Pour tout $h \geq 1$ et pour tout $n \geq 1$, nous avons

$$a_{n,n}^h = \left[\gamma_{n+h-1} - \sum_{i=1}^{n-1} a_{n-1,i}^1 \gamma_{n+h-i-1} \right] (\nu_{n-1}^1)^{-1}, \quad (1.59)$$

$$a_{n,i}^h = a_{n-1,i}^h - a_{n,n}^h a_{n-1,n-i}^1 \quad i = 1, \dots, n-1, \quad (1.60)$$

$$\nu_n^h = \nu_{n-1}^h - (a_{n,n}^h)^2 \nu_{n-1}^1, \quad (1.61)$$

avec $\nu_0^h = \gamma_0$.

Démonstration. De la décomposition orthogonale

$$\mathcal{H}_{1,n} = \mathcal{H}_{2,n} \oplus \overline{\text{sp}}\{X_1 - P_{2,n}X_1\}$$

découle

$$P_{1,n}X_{n+h} = P_{2,n}X_{n+h} + \theta(X_1 - P_{2,n}X_1), \quad (1.62)$$

où

$$\theta = \frac{\langle X_{n+h}, X_1 - P_{2,n}X_1 \rangle}{\|X_1 - P_{2,n}X_1\|^2}. \quad (1.63)$$

Notons que $\|X_1 - P_{2,n}X_1\| \neq 0$ car $\{X_n\}$ est non déterministe.

En utilisant la stationnarité de $\{X_n\}$, on déduit :

$$\begin{aligned} P_{2,n}X_{n+h} &= \sum_{i=1}^{n-1} a_{n-1,i}^h X_{n+1-i}, \\ P_{2,n}X_1 &= \sum_{i=1}^{n-1} a_{n-1,i}^1 X_{i+1}, \\ \|X_1 - P_{2,n}X_1\|^2 &= \nu_{n-1}^1. \end{aligned}$$

Cependant, (1.62) peut s'écrire

$$\sum_{i=1}^n a_{n,i}^h X_{n+a-1} = \sum_{i=1}^{n-1} a_{n-1,i}^h X_{n+1-i} + \theta X_1 - \theta \sum_{i=1}^{n-1} a_{n-1,i}^1 X_{i+1}$$

ce qui est équivalent à $a_{n,n}^h = \theta$ et

$$a_{n,i}^h = a_{n-1}^h - \theta a_{n-1,n-1}^1 \quad 1 \leq i \leq n-1$$

d' où (1.60).

On déduit de (1.63) que

$$a_{n,n}^h = \left\langle X_{n+h}, X_1 - \sum_{i=1}^{n-1} a_{n-1,i}^1 X_{i+1} \right\rangle (\nu_{n-1}^1)^{-1}$$

ce qui est équivalent à (1.59). D'une part, il en résulte de (1.62) que

$$\|P_{1,n}X_{n+h}\|^2 = \|X_{n+h}\|^2 - \|P_{1,n}X_{n+h}\|^2 + \theta^2 \nu_{n-1}^1.$$

D'autre part, nous avons

$$\nu_n^h = \|X_{n+h}\|^2 - \|P_{1,n}X_{n+h}\|^2$$

et

$$\nu_{n-1}^h = \|X_{n+h}\|^2 - \|P_{2,n}X_{n+h}\|^2$$

par conséquent

$$\nu_n^h = \nu_{n-1}^h - \theta^2 \nu_{n-1}^1$$

d' où (1.61). □

Pour tout n et h , le calcul des $(a_{n,i}^h)$ et ν_n^h dans (1.59)-(1.61) nécessite $2n + 1$ multiplications. Ainsi, lorsque $(a_{p,i}^h)$ et ν_p^h sont calculés pour $1 \leq h \leq s$ en utilisant la proposition (1.6) (Algorithme A_1), le coût numérique totale est

$$N_1 = s \sum_{n=1}^p (2n + 1) = p^2 s + 2ps.$$

Une autre alternative (Algorithme A_2), consiste à caculer, en un premier temps, les coefficients $(a_{n,i}^1)$ pour $1 \leq n \leq p + s - 1$ en utilisant (1.59)-(1.61), ensuite, à caculer les $(a_{p,i}^h)$ pour $1 \leq h \leq s$, en utilisant (1.55) quand $n = p$, et les ν_p^h s'obtiennent par la relation

$$\nu_p^h = \langle X_{p+h} - P_{1,p} X_{p+h} \ X_{p+h} \rangle = \gamma_0 - \sum_{i=1}^p a_{p,i}^h \gamma_{h+i-1}.$$

Le coût numérique de l'algorithme A_2 est

$$\begin{aligned} N_2 &= (p + s - 1)^2 + 2(p + s - 1) - 2 + p \sum_{h=2}^s (h - 1) + p(s - 1) \\ &= p^2 + \frac{1}{2}p(s^2 + 5s - 2) + s^2 - 3. \end{aligned}$$

Pour comparer N_1 et N_2 , on pose $\tau = s/p$. Alors

$$N_2 - N_1 = p^3 \tau \left(\frac{1}{2} \tau - 1 \right) - p^2 \left(\tau^2 + \frac{1}{2} \tau + 1 \right) - p - 3.$$

Si $\tau = \frac{1}{2}$,

$$N_2 - N_1 \geq 2p^2 + (p - 1)(4p + 3) > 0 \text{ pour } p \geq 1$$

Donc, si $\tau = \frac{1}{2}$, $N_2 - N_1 > 0$ pour $p \geq 1$. En conclusion, le choix entre l'algorithme A_1 et A_2 dépend de la paire (p, s) .

Nous proposons un autre algorithme dont la complexité numérique est inférieure à $\min(N_1, N_2)$, pour toute paire (p, s) telle que $p, s > 1$. Cet algorithme est basé sur la proposition suivante :

Proposition 1.7. (Bondon, 2001).

Pour tout $h \geq 1$ et pour tout $n \geq 1$, on a

$$a_{n,i}^h = a_{n,i+1}^{h-1} + a_{n,1}^{h-1} a_{n-1,i}^1 - a_{n,n}^h a_{n-1,n-i}^1, \quad i = 1, 2, \dots, n - 1, \quad (1.64)$$

$$\nu_n^h = \nu_n^{h-1} + [(a_{n,1}^{h-1})^2 - (a_{n,n}^h)^2] \nu_{n-1}^1. \quad (1.65)$$

L'algorithme A_3 consiste à calculer les coefficients $(a_{n,i}^1)$ et ν_n^1 pour $1 \leq n \leq p$ en utilisant (1.59)-(1.61), ensuite, calculer $(a_{p,i}^h)$ et ν_p^h pour $2 \leq h \leq s$ en utilisant (1.59), (1.64) et (1.65) quand $n = p$. La complexité numérique de cet algorithme est

$$N_3 = p^2 + 2p + (s-1)(3p+1) = p^2 + p(3s-1) + s - 1.$$

Nous avons

$$N_1 - N_3 = (s-1)(p^2 - p - 1) > 0 \quad \text{si } p > 1$$

et

$$N_2 - N_3 = \frac{1}{2}ps(s-1) + (s-2)(s+1) > 0 \quad \text{si } s > 1.$$

Par conséquent, $N_3 < \min(N_1, N_2)$, pour toute paire (p, s) avec $p, s > 1$.

Remarque 1.2. Une autre alternative (Algorithme A_4), consiste à calculer les $(a_{n,i}^1)$ et ν_n^1 pour $1 \leq n \leq p + s - 1$ en utilisant (1.59)-(1.61), $(a_{n,i}^h)$ et ν_n^h pour $2 \leq h \leq s$ et $p \leq n \leq p + s - h$ en utilisant (1.56) et (1.57). La complexité numérique de cet algorithme est alors

$$\begin{aligned} N_4 &= (p+s-1)^2 + 2(p+s-1) + \sum_{h=2}^s \sum_{n=p}^{p+s-h} (n+2) \\ &= p^2 + \frac{1}{2}ps(s+3) + \frac{1}{6}(s-1)(s^2 + 10s + 6). \end{aligned}$$

Nous avons

$$N_4 - N_3 = (s-1) \left[\frac{1}{2}p(s-2) + \frac{1}{6}s(s+10) \right] > 0 \quad \text{si } s > 1.$$

En d'autres termes, l'algorithme A_4 calcule les coefficients $(a_{q-h+1,i}^h)$ et les variances ν_{q-h+1}^h , $1 \leq h \leq s$ pour q entier donné, $q \geq s$. Les paramètres $a_{q-h+1,i}^h$ sont les coefficients du prédicteur $P_{h,q}X_{q+h}$ et utilise seulement les covariances $\gamma_k, 0 \leq k \leq q$. Donc, pour calculer les coefficients $(a_{q-h+1,i}^h)$ et ν_{q-h+1}^h pour $2 \leq h \leq s$, il est suffisant d'utiliser (1.56) et (1.57) pour $n = q - h + 1$. Par conséquent, L'algorithme A_3 est meilleur que A_4 , dans le sens où, il nécessite moins d'opérations numériques.

Remarque 1.3. Une autre approche différente pour calculer les coefficients $(a_{n,i}^h)$ du prédicteur d'horizon h , consiste à calculer les coefficients $(a_{n,i}^1)$ et ν_n^1 pour $1 \leq n \leq p$ en utilisant (1.59)-(1.61) et de décomposer ensuite le sous espace des observations $\mathcal{H}_{1,p}$ en somme de p sous espaces orthogonaux. Ainsi, pour un horizon de prévision $h > 1$, $P_{1,p}X_{p+h}$ et ν_n^h s'obtiennent par les relations

$$P_{1,p}X_{p+h} = \sum_{i=1}^p c_i^h (X_i - P_{1,i-1}X_i),$$

$$\nu_n^p = \gamma_0 - \sum_{i=1}^p (c_i^h)^2 \nu_{i-1}^1,$$

où $P_{1,0}X_1 = 0$ et

$$c_i^h = \frac{\langle X_{p+h}, X_i - P_{1,i-1}X_i \rangle}{\|X_i - P_{1,i-1}X_i\|} = \left[\gamma_{p+h-i} - \sum_{j=1}^{i-1} a_{i-1,j}^1 \gamma_{p+h-i+j} \right] (\nu_{i-1}^1)^{-1}.$$

Le nombre de multiplications générées par cet algorithme (A_5) est $p^2 + 2p$ dans l'application de l'algorithme de Levinson, $i - 1$ pour $P_{1,i-1}$, i pour c_i^h et $2p$ pour ν_p^h . La complexité numérique totale est donc

$$N_5 = p^2 \left(\frac{1}{2}s + 1 \right) + p^2 \left(\frac{5}{2}s - 1 \right).$$

Nous avons

$$N_5 - N_3 = \frac{1}{2}s(p-2)(p+1) + 1 > 0 \quad p > 1$$

donc l'algorithme A_3 est moins complexe que l'algorithme A_5 . D'autre part, nous avons

$$N_1 - N_5 = p(p-1) \left(\frac{1}{2}s - 1 \right) > 0 \quad \text{si } p > 1 \text{ et } s > 2.$$

Remarque 1.4. L'algorithme de Levinson donne la décomposition de Cholesky de l'inverse de la matrice de covariance $[\Gamma_p(i,j) = \gamma_{i-j}, i,j = 1, \dots, p]$,

$$\Gamma_p^{-1} = A_p' \Sigma_p^{-2} A_p$$

où A_p est une matrice triangulaire avec $A_p(i,i) = 1$ et $A_p(i,j) = -a_{i-1,i}$, pour $i > j$, et Σ_p^{-2} est diagonale avec $\Sigma_p^{-2}(i,i) = \nu_{i-1}^1$.

Pour un horizon de prévision $h > 1$, les coefficients $(a_{p,i}^h)$ peuvent être calculés directement par la relation

$$(a_{p,1}^h, \dots, a_{p,p}^h) = \Gamma_p^{-1}(\gamma_h, \dots, \gamma_{h+p-1})' = A_p' \Sigma_p^{-2} A_p(\gamma_h, \dots, \gamma_{h+p-1})'$$

et la variance ν_p^h s'obtient par la relation

$$\nu_p^h = \langle X_{p+h} - P_{1,p}X_{p+h}, X_{p+h} \rangle = \gamma_0 - \sum_{i=1}^p a_{p,i}^h \gamma_{h+i-1}.$$

Ainsi, la complexité numérique globale pour le calcul des $(a_{p,i}^h)$ et ν_p^h pour $1 \leq h \leq s$ est

$$N_6 = p^2 + 2p + (s-1)(p^2 + p) = p^2s + p(s+1).$$

Nous avons

$$N_6 - N_3 = (s-1)(p^2 - 2p - 1) > 0 \quad \text{si } s > 1 \text{ et } p > 2$$

d'où, à l'exception de $p = 1$ ou $p = 2$, l'algorithme A_3 est préférable à A_6 .

Remarque 1.5. Il résulte de (1.61) que $\nu_n^h \leq \nu_{n-1}^h$. Cette inégalité est également une conséquence immédiate de la stationnarité du processus $\{X_t\}$ et de l'inclusion $\mathcal{H}_{2,n} \subset \mathcal{H}_{1,n}$. Par conséquent, pour $h = 1$ dans (1.61), on déduit que $|a_{n,n}^1| < 1$. Cette même inégalité n'est pas vérifiée pour $h > 1$. En effet, considérons un processus autorégressif causal

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \epsilon_n \quad (1.66)$$

où $\epsilon_n = X_n - P_{n-1}X_n$, $\|\epsilon_n\|^2 = \sigma_\epsilon^2$. Nous avons

$$P_{1,2}X_4 = (a_2 + a_1^2)X_2 + a_1 a_2 X_1.$$

Pour $a_2 = -\frac{1}{4}a_1^2$, (1.66) est causal si $|a_1| < 2$, et nous obtenons $a_{2,2}^2 = -\frac{1}{4}a_1^3$. D'où, si $4 < |a_1^3| < 8$, $|a_{2,2}^2| > 1$, et si $|a_1^3| \leq 4$, $|a_{2,2}^2| \leq 1$. Quand le passé est infini, l'inclusion $\mathcal{H}_n \subset \mathcal{H}_{n+1}$ implique que la variance de l'erreur de la prédiction d'horizon h est décroissante en fonction de h . Quand le passé est fini, il n'existe pas d'inégalité entre ν_n^{h-1} et ν_n^h . En effet, pour $a_2 = 0$ dans (1.66), nous avons

$$\begin{aligned} P_{1,1}X_2 &= a_1 X_1 \\ \nu_1^1 &= \sigma_\epsilon^2 \\ P_{1,1}X_3 &= \sigma_\epsilon^2(1 + a_1^2) \\ \nu_1^2 &= \sigma_\epsilon^2(1 + a_1^2) \end{aligned}$$

et donc $\nu_1^1 < \nu_1^2$. Cependant, quand $a_1 = 0$ dans (1.66), nous avons

$$\begin{aligned} P_{1,1}X_2 &= 0 \\ \nu_1^1 &= \sigma_\epsilon^2(1 - a_2^2)^{-1} \\ P_{1,1}X_3 &= a_2 X_1 \\ \nu_1^2 &= \sigma_\epsilon^2 \end{aligned}$$

et donc $\nu_1^1 > \nu_1^2$. Finalement, on déduit de (1.57) que, $\nu_{n+1}^h \leq \nu_n^h$. Cette inégalité découle de $\mathcal{H}_{1,n} \subset \mathcal{H}_{1,n+1}$.

Remarque 1.6. Les algorithmes présentés ci-dessus se généralisent facilement pour un processus $\{X_t\}$ multivarié. Les coefficients $(a_{n,i}^h)$ sont des matrices carrés d'ordre m , et en définissant

$$\gamma_n = \mathbf{E}X_n X_n'$$

et

$$\nu_n^h = \mathbf{E}(X_{n+h} - P_{1,n}X_{n+h})(X_{n+h} - P_{1,n}X_{n+h})',$$

les propositions 1.5 et 1.6 se généralisent en remplaçant (1.57) par

$$\nu_n^h = \nu_{n+1}^{h-1} + a_{n+1,1}^{h-1} \nu_n^1 (a_{n+1,1}^{h-1})'.$$

1.3 Prédiction d'une combinaison linéaire d'observations futures.

Dans cette section, nous nous intéressons aux prédicteurs de la forme

$$X = \sum_{r=1}^m c_r X_r \quad (1.67)$$

où m est un entier et c_1, c_2, \dots, c_m sont des constantes. Notons que la prévision d'une combinaison linéaire d'observations futures est importante aussi bien en pratique qu'en théorie. Par exemple, pour des ventes enregistrées mensuellement, les pronostiqueurs peuvent s'intéresser aux ventes de l'année suivante ($m = 12$).

Soit $\mathcal{H}_0 = \overline{\text{sp}}\{X_t, t \leq 0\}$. De (1.67) et en utilisant la linéarité de la projection, il s'en suit :

$$\begin{cases} \hat{X} = P_{\mathcal{H}_0} X = \sum_{r=1}^m c_r \hat{X}_r \\ X - \hat{X} = \sum_{r=1}^m c_r (X_r - \hat{X}_r), \\ \text{Var}(X - \hat{X}) = \sum_{r=1}^m \sum_{s=1}^m g_{rs}^m c_r c_s = \mathbf{c}' G_m \mathbf{c}, \end{cases} \quad (1.68)$$

où $g_{rs}^m = \sigma^2 \sum_{k=0}^{m-1} b_{r+k-m} b_{s+k-m}$, $G_m = (g_{rs}^m)_{r,s=1,m}$ et $\mathbf{c} = (c_1, \dots, c_m)$. La matrice G_m dans (1.68) est la matrice de covariance du vecteur des erreurs de prédiction $(X_1 - \hat{X}_1, \dots, X_m - \hat{X}_m)$. Elle peut s'identifier également à une sous matrice de dimension $(m \times m)$ de la matrice infinie

$$G = \sigma^2 T T', \quad (1.69)$$

où $T = (b_{i-j})_{i,j=0}^{\infty}$.

Le "baromètre" de la prédictibilité de X peut être défini comme le rapport de la variance de l'erreur de la prévision sur sa variance. Plus spécifiquement, nous utilisons la quantité $\lambda(X) = 1 - \frac{\text{Var}(X - \hat{X})}{\text{Var}(X)}$ comme mesure de la prédictibilité de X (Yaglom, 1963). Notons que, plus $\lambda(X)$ est grand, meilleure est la prédiction.

Comme

$$\text{Var}(X) = \sum_{r=1}^m \sum_{s=1}^m c_r c_s \gamma_{r-s} = \mathbf{c}' \Gamma_m \mathbf{c},$$

nous avons

$$\lambda(X) = 1 - \frac{\mathbf{c}' G_m \mathbf{c}}{\mathbf{c}' \Gamma_m \mathbf{c}}. \quad (1.70)$$

Pour un processus $\{X_t\}$, nous nous intéressons à la combinaison linéaire qui maximise $\lambda(X)$.

Théorème 1.1. (Pourahmadi, 2001).

Soit $\{X_t\}$ un processus stationnaire non déterministe et $X = \sum_{r=1}^m c_r X_r$, où $m \geq 1$ est fixé.

Soit $\lambda_1 < \lambda_2 < \dots < \lambda_k$, ($k \leq m$) les racines de l'équation

$$\det(G_m - \lambda \Gamma_m) = 0,$$

et c_1, c_2, \dots, c_m les vecteurs propres correspondants, i.e.

$$\mathbf{c}'_i \Gamma_m \mathbf{c}_j = \delta_{i,j}, \quad i, j = 1, 2, \dots, m.$$

Alors $U_1 = \mathbf{c}_1 X_m = \sum_{r=1}^m c_r X_r$ est la meilleure prévision linéaire avec la mesure de prédictibilité

$$\lambda(U_1) = 1 - \lambda_1,$$

et, en général, la mesure de prédictibilité de $U_k = \mathbf{c}'_k X_m$, est $1 - \lambda_k$.

Exemple 1. Pour un processus de représentation moyenne mobile

$$X_t = \epsilon_t + \epsilon_{t-1}, \quad \text{Var}(\epsilon_t) = 1,$$

et pour $m = 2$, nous avons

$$\Gamma_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

$$\det(G_2 - \lambda \Gamma_2) = (1 - \lambda)(1 - 3\lambda) = 0.$$

Donc $\lambda_1 = \frac{1}{3}, \lambda_2 = 1$, et les vecteurs propres correspondants sont

$$\mathbf{c}_1 = \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right), \quad \mathbf{c}_2 = \left(0, \frac{1}{\sqrt{2}} \right).$$

Donc

$$U_1 = -\frac{2}{\sqrt{6}} X_1 + \frac{1}{\sqrt{6}} X_2, \quad \lambda(U_1) = \frac{2}{3},$$

est la meilleure combinaison linéaire de valeurs futures X_{t+1} et X_{t+2} , et

$$U_2 = \frac{1}{\sqrt{2}} X_2, \quad \lambda(U_2) = 0,$$

est la seconde combinaison linéaire des mêmes valeurs futures. Notons que U_2 est non corrélé avec le passé du processus.

Pour m quelconque, en développant $\det(G_m - \lambda\Gamma_m)$ par rapport à la première ligne, on obtient

$$|G_m - \lambda\Gamma_m| = (1 - 2\lambda)D_m - (1 - \lambda^2)D_{m-1}, \quad (1.71)$$

où D_m est le déterminant de la $(m - 1) \times (m - 1)$ matrice

$$\begin{pmatrix} 2 - \lambda & 0 & \dots & \dots & 0 \\ 1 - \lambda & 2 - \lambda & 0 & \dots & 0 \\ 0 & 1 - \lambda & 0 & \dots & 0 \\ \vdots & 0 & 1 - \lambda & 2 - \lambda & 0 \\ 0 & 0 & 0 & 1 - \lambda & 2 - \lambda \end{pmatrix}.$$

1.4 Factorisation de la densité spectrale et application à la prévision

Dans les paragraphes précédents, nous nous sommes essentiellement intéressés à la prédiction de séries chronologiques par l'intermédiaire de la fonction d'autocovariance. L'idée était d'analyser les liaisons entre le passé et le futur en faisant jouer un rôle primordial à l'indice des temps t . Cette approche est dite approche dans le domaine des temps. une autre approches possible est l'approche spectrale. Le théorème de la représentation spectrale, permet en effet d'écrire les composantes du processus comme des combinaisons à coefficients aléatoires de fonction sinusoidales $\exp(it\omega)$. On peut alors étudier l'importance de chaque fréquence dans les valeurs prises par X_t , voir comment les corrélations se décomposent fréquence par fréquence..., une telle approche est dite approche dans le domaine des fréquences.

Cette section est consacrée à la détermination de relations de récurrence entre les coefficients des prédicteurs dans le domaine des fréquences. Cette approche est souvent plus riche en termes d'interprétations, mais nécessite un recours à des techniques mathématiques plus complexes.

1.4.1 Préliminaires sur la densité spectrale

Soit $\{X_t\}$ un processus stationnaire centré, de fonction d'autocovariance $\gamma(h), h \in \mathbb{Z}$ et possédant une densité spectrale $f(\lambda), -\pi < \lambda < \pi$. Les liens entre l'autocovariance et la

densité spectrale sont données par

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(i\lambda h),$$

$$\gamma(h) = \int_{-\pi}^{\pi} \exp(-i\lambda h) f(\lambda) d\lambda.$$

La densité spectrale du processus d'innovation $\{\epsilon_t\}$ est une fonction constante de valeur 1 sur $[-\pi, \pi]$.

Le processus $\{X_t\}$ est purement non déterministe si et seulement si

$$\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty.$$

En s'appuyant sur la décomposition de Wold, on obtient

$$X_n = \sum_{k=0}^{\infty} c_k \epsilon_{n-k} \tag{1.72}$$

avec

$$\epsilon_n = X_n - \hat{X}_n,$$

$$\sigma_{\epsilon_n}^2 = \exp \left\{ \int_{-\pi}^{\pi} \ln f(\lambda) \frac{d\lambda}{2\pi} \right\}. \tag{1.73}$$

Si la densité spectrale f du processus $\{X_t\}$, est telle que $f = |\phi|^2 = \phi \bar{\phi}$ avec

$$\phi(\lambda) = \sum_{k=0}^{\infty} c_k e^{ik\lambda}, \quad \sum_{k=0}^{\infty} c_k^2 < \infty, \sigma = c_0 > 0,$$

alors la fonction ϕ est dite factorisation optimale (maximale) de f .

Elle est unique et admet une représentation analytique à l'intérieur du disque unité \mathbb{D} donnée par

$$\phi_+(z) = \exp \left[\frac{1}{2} \int_{-\pi}^{\pi} \frac{e^{i\lambda+z}}{e^{i\lambda} - z} \ln f(\lambda) \frac{d\lambda}{2\pi} \right]. \tag{1.74}$$

La fonction $\phi_+(z)$ n'admet pas de zéro à l'intérieur de \mathbb{D} et $\phi_+(0) = c_0 > 0$. On a

$$\phi_+(z) = \exp \left\{ \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k z^k \right\}, \tag{1.75}$$

$$a_k = \int_{-\pi}^{\pi} e^{-ik\lambda} \ln f(\lambda) \frac{d\lambda}{2\pi}.$$

1.4.2 Lien entre prédiction et densité spectrale

Dans la théorie de la prédiction linéaire d'un processus purement non déterministe, la factorisation optimale de la densité spectrale joue un rôle crucial. En effet, si ϕ est la factorisation optimale de f et $\phi^{-1} = \sum_{k=0}^{\infty} d_k e^{ik\lambda}$ sa réciproque alors, le prédicteur d'horizon h , basé sur le passé $\{X_t; t \leq m\}$ est donné par

$$\hat{X}_{n+h} = \sum_{k=0}^{\infty} \alpha_{hk} X_{n-k} \quad (1.76)$$

avec

$$\alpha_{hk} = \sum_{j=0}^k c_{h+j} d_{k-j} \quad \text{et} \quad \sigma_h^2 = \sum_{k=0}^{h-1} c_k^2 < \infty.$$

En utilisant la relation (1.75), le théorème suivant établit deux relations de récurrence pour obtenir les coefficients c_k et d_k en fonction des a_k .

Théorème 1.2. (*Pourahmadi, 1983*).

Soit $\{X_n\}$ un processus purement non déterministe, stationnaire, de densité spectrale f . Soit ϕ la factorisation optimale de f . Alors;

(a)

$$c_{n+1} = \sum_{k=0}^n \left(1 - \frac{k}{n+1}\right) a_{n+1-k} c_k, \quad n = 0, 1, 2, \dots,$$

où $a_k = \int_{-\pi}^{\pi} e^{-ik\lambda} \ln f(\lambda) \frac{d\lambda}{2\pi}$ et $c_0 = \exp\{a_0/2\}$.

(b) Les coefficients de Fourier d_n de ϕ^{-1} s'obtiennent par la relation de récurrence suivante :

$$d_{n+1} = 1 - \frac{k}{n+1} a_{n+1-k} d_k, \quad n = 0, 1, 2, \dots$$

où $d_0 = c_0^{-1}$.

Démonstration. – (a)

En dérivant (1.75) par rapport à z et en remplaçant $\exp\left\{\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k z^k\right\}$ par

$\sum_{k=0}^{\infty} c_k z^k$ on obtient :

$$\sum_{n=0}^{\infty} n c_n z^{n-1} = \left(\sum_{n=1}^{\infty} n a_n z^{n-1} \right) \left(\sum_{k=0}^{\infty} c_k z^k \right)$$

ce qui est équivalent à

$$\begin{aligned} \sum_{n=0}^{\infty} (n+1)c_{n+1}z^n &= \left[\sum_{n=0}^{\infty} (n+1)a_{n+1}z^n \right] \left(\sum_{n=0}^{\infty} c_n z^n \right) \\ &= \sum_{n=0}^{\infty} \left[\sum_{k=0}^n (n+1-k)a_{n+1-k}c_k \right] z^n. \end{aligned}$$

Par identification suivant les puissances de z , on aura :

$$(n+1)c_{n+1} = \sum_{k=0}^n (n+1-k)a_{n+1-k}c_k,$$

où

$$c_{n+1} = \sum_{k=0}^n \left(1 - \frac{k}{n+1} \right) a_{n+1-k}c_k, \quad n = 0, 1, 2, \dots$$

– (b)

Comme $\phi_+(z)$ n'admet pas de zéro à l'intérieur de \mathbb{D} , il s'en suit que, ϕ_+^{-1} admet un développement en série entière; $\phi_+^{-1}(z) = \sum_{k=0}^{\infty} d_k z^k$. En égalisant avec (1.75), on obtient :

$$\sum_{k=0}^{\infty} d_k z^k = \exp \left\{ -\frac{a_0}{2} - \sum_{k=1}^{\infty} a_k z^k \right\}. \quad (1.77)$$

(b) s'obtient par l'application de (a) et le fait que $d_0 = \exp\{-a_0/2\} = c_0^{-1}$.

□

Remarque 1.7. Le théorème 1.2 dit que, pour trouver les coefficients c_k et d_k , il suffit de trouver les coefficients de Fourier de $\ln f$, i.e.

$$a_k = \int_{-\pi}^{\pi} e^{-ik\lambda} \ln f(\lambda) \frac{d\lambda}{2\pi} = \frac{1}{\pi} \int_0^{\pi} \cos k\lambda \ln f(\lambda) d\lambda, \quad k = 0, 1, 2, \dots$$

Ainsi, si f est connu, les coefficients a_k s'obtiennent par le calcul des intégrales précédentes. Toutefois, il n'est pas facile de calculer ces intégrales, par conséquent, les coefficients a_k sont approximés par des méthodes numériques d'intégrations.

Remarque 1.8. Les premiers coefficients c_k sont

$$\begin{aligned} c_0 &= \exp \left\{ \frac{1}{2\pi} \int_0^\pi \ln f(\lambda) d\lambda \right\}, \\ c_1 &= a_1 c_0, \\ c_2 &= \left(a_2 + \frac{1}{2} a_1^2 \right) c_0, \\ c_3 &= \left(a_3 + a_1 a_2 + \frac{1}{6} a_1^3 \right) c_0 \\ c_4 &= \left(a_4 + a_3 a_1 + \frac{1}{2} a_2 a_1^2 + \frac{1}{2} a_2^2 + \frac{1}{24} a_1^4 \right) c_0. \end{aligned}$$

Remarque 1.9. Les formules de récurrence des coefficients c_k peuvent s'écrire sous une formule plus pratique

$$c_{n+1} = \frac{1}{n+1} \sum_{k=0}^n (k+1) a_{k+1} c_{n-k}, \quad n = 0, 1, 2, \dots$$

Ainsi, pour une série chronologique stationnaire, le degré de croissance des ka_k , $k = 1, 2, \dots$ constitue une mesure de mémoire de la série, cf. Parzen (1984).

Le théorème qui suit fournit un algorithme récursif pour le calcul des coefficients du prédicteur \hat{X}_{n+h} . Bien que, l'idée est similaire au théorème 4.1 de Miamee et Salehi (1980), ce théorème permet de déterminer les coefficients des prédicteurs d'une façon plus commode. Aussi, il met en évidence le rôle important des coefficients c_k et d_k dans la détermination des coefficients du prédicteur.

Pour les besoins du théorème, introduisons la suite $\{Q_n\}_{n=0}^\infty$ définie par les relations de récurrence

$$\begin{aligned} Q_0 &= 0, \quad Q_1 = 1, \\ Q_n &= -c_0 \sum_{k=1}^{n-1} d_k Q_{n-k}, \quad n \geq 2. \end{aligned}$$

Théorème 1.3. (Pourahmadi, 1983).

Soit $\{X_t, t \in \mathbb{Z}\}$ un processus stochastique faiblement stationnaire. Alors,

- (a) $\hat{X}_{n+1+h} = -c_0 \sum_{k=1}^{\infty} d_k X_{m+1-k}$.
- (b) $\hat{X}_{m+h-1} = -c_0 \sum_{k=0}^{\infty} d_{k+1} \hat{X}_{m+h-k} - c_0 \sum_{k=h}^{\infty} d_{k+1} X_{m+h-k}$, $h \geq 1$.
- (c) $\hat{X}_{n+h} = \hat{X}_h + \sum_{k=0}^{m-1} Q_{m-k} [X_{k+1} - \hat{X}_{k+1}]$ $0 < n < h$.

1.5 Processus p -stationnaire

De nombreux travaux de prospection et d'analyse des processus stochastique à variance infinie ont constitué le centre d'intérêt de plusieurs chercheurs Bhansli (1987), Cambanis, S. et Soltani, A.R. (1988), Cambanis et al (1988), Cline, D.B.H. et Brockwell, P.J. (1985). Pour de tels processus, les notions de densité spectrale et de covariance ne peuvent être définies. De nombreuses difficultés sont dès lors rencontrées dans la construction du domaine spectral. Ceci entraîne également une difficulté majeure pour établir le lien entre le domaine spectrale et le domaine temporel.

Malgré cela, plusieurs tentatives ont été développées afin d'adapter la théorie de Wiener-Kolmogorov à la prédiction de processus à variance infinie. Pour palier à ces insuffisances, les notions de "pseudo" covariance et de "pseudo" densité spectrale ont été introduites, cf. Cambanis, S. et Soltani, A.R. (1988), Cambanis et al (1988).

Le succès de la théorie de Kolmogorov-Wiener dans la prédiction des processus faiblement stationnaire est essentiellement attribué à deux propriétés. La première est la structure hilbertienne du domaine temporel. Cette propriété assure que le processus d'innovation de $\{X_t\}$ soit un bruit blanc. De plus, elle permet d'établir la décomposition de Wold. La seconde permet de définir la fonction d'autocovariance et de densité spectrale.

Cette section a pour but d'introduire les processus p -stationnaire. Nous donnons pour de tels processus la décomposition de Wold (représentation moyenne mobile) et la décomposition prédictive finie (représentation autorégressive), sans recourir aux notions de covariance et de densité spectrale. Il est important de noter que ces processus englobent les processus faiblement stationnaires et les processus α -stable ($1 < \alpha \leq 2$).

Définition 1.3. Pour $0 < p < \infty$, le processus stochastique $\{X_t\}$ est dit p -stationnaire si

$$X_t \in L^p(\Omega), \quad \text{pour tout } t \in \mathbb{Z}, \quad (1.78)$$

et

$$\mathbf{E} \left| \sum_{k=1}^n a_k X_{t_k+h} \right|^p = \mathbf{E} \left| \sum_{k=1}^n a_k X_{t_k} \right|^p, \quad (1.79)$$

pour tout entier $n \geq 1, t_1, t_2, \dots, t_n, h$ et scalaires a_1, a_2, \dots, a_n

Remarque 1.10.

- (a) Notons que la p -stationnarité est une généralisation de la stationnarité au sens faible.
- (b) Pour $p = 2$, il est facile de montrer que la 2-stationnarité est confondue avec stationnarité au sens faible.

- (b) Tout processus strictement stationnaire est p -stationnaire.
 (d) Les processus ARMA de variance infinie étudiés notamment par Brockwell et Cline (1985) forment une sous classe des processus p -stationnaire avec $p < 2$.

Théorème 1.4. (Miamee et Pourahmadi, 1988).

Soit $\{\epsilon_j\}_{j=-\infty}^{\infty}$ un processus p -stationnaire et $\{c_j\}_{j=0}^{\infty}$ une suite de scalaires telle que la série $\sum_{j=-\infty}^{\infty} c_j \epsilon_{t-j}$ est convergente dans la métrique $L^p(\Omega)$. Alors, le processus $\{X_t\}$ défini par

$$X_t = \sum_{j=-\infty}^{\infty} c_j \epsilon_{t-j}, \quad \text{pour tout entier } t,$$

est p -stationnaire.

Corollaire 1.2. (Miamee et Pourahmadi, 1988).

Soit $\{\epsilon_j\}_{j=-\infty}^{\infty}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) avec $\epsilon_j \in L^p(\Omega)$, pour tout entier j , et $\{c_j\}_{j=-\infty}^{\infty}$ une suite de scalaires telle que $\sum_{j=-\infty}^{\infty} c_j \epsilon_j$ est convergente dans la métrique $L^p(\Omega)$. Alors, le processus $\{X_t\}$ donné par

$$X_t = \sum_{j=-\infty}^{\infty} c_j \epsilon_{t-j}, \quad \text{pour tout entier } t,$$

est p -stationnaire.

Pour $p > 2$, le problème de prédiction de X est similaire au cas $p = 2$. Par contre, pour $p \leq 1$, tout $X \in L^p(\Omega)$ n'admet pas une unique projection sur le sous espace fermé M de $L^p(\Omega)$. Dans ce qui suit, Nous restreignons l'étude au cas $1 < p \leq 2$.

Pour un processus $\{X_t\}$ p -stationnaire, $\mathcal{H}(X)$ est le domaine temporel, $\mathcal{H}_t(X)$ est le sous espace présent-passé. On note par \hat{X}_{t+h} , $h \geq 1$, le meilleur prédicteur linéaire de X_{t+h} basé sur le passé X_t, X_{t-1}, \dots , i.e. \hat{X}_{t+h} est tel que $\|X_{t+h} - \hat{X}_{t+h}\|_p \leq \|X_{t+h} - Y\|_p$ pour tout $Y \in \mathcal{H}_t(X)$. Pour $h \geq 1$ fixé, $\{\hat{X}_{t+h}\}$ est un processus stochastique. Nous associons au processus $\{X_t\}$ l'opérateur U qui est une isométrie de $\mathcal{H}(X)$ dans $\mathcal{H}(X)$ défini par

$$U \left(\sum_{k=1}^n a_k X_{t_k} \right) = \sum_{k=1}^n a_k X_{t_k+1}.$$

Le théorème qui suit est essentiel pour l'étude du problème de prédiction d'un processus p -stationnaire.

Théorème 1.5. (Miamee et Pourahmadi, 1988).

Soit $\{X_t\}$ un processus non déterministe p -stationnaire et U son opérateur inverse, $\{\epsilon_t\}$ son processus d'innovations et $\{\hat{X}_{t+h}\}$ le processus des prédicteurs. Alors,

- (a) le processus $\{\hat{X}_{t+1}\}$ est p -stationnaire d'opérateur inverse U ,
- (b) pour tout $h \geq 1$, le processus $\{\hat{X}_{t+h}\}$ est p -stationnaire d'opérateur inverse U ,
- (c) le processus d'innovations $\{\epsilon_t\}$ est p -stationnaire d'opérateur inverse U .

1.5.1 Décomposition prédictive et représentation de Wold finie

Soit $\{X_t\}$ un processus p -stationnaire, $1 < p < 2$, $\{\epsilon_t\}$ son processus d'innovation. Dans cette section, nous donnons la décomposition de Wold d'un processus p -stationnaire et nous montrons que cette décomposition de $\{X_t\}$ fournit de bonnes informations sur les prédicteurs et la structure de tels processus. Cette décomposition s'obtient en notant que pour tout processus non déterministe p -stationnaire d'innovations $\{\epsilon_t\}$ on a

$$\mathcal{H}_t = \overline{\text{sp}}\{X_t, \mathcal{H}_{t-1}\}, \quad X_t \notin \mathcal{H}_{t-1}, \quad \text{pour tout } t, \quad (1.80)$$

et aussi

$$\mathcal{H}_t = \overline{\text{sp}}\{\epsilon_t, \mathcal{H}_{t-1}\}, \quad \epsilon_t \notin \mathcal{H}_{t-1}, \quad \text{pour tout } t. \quad (1.81)$$

Le lemme qui suit joue un rôle crucial dans la révélation de la signification géométrique de (1.80) et (1.81) et permet d'obtenir la décomposition de Wold et la décomposition prédictive d'un processus p -stationnaire.

Lemme 1.1. *Soit M un sous espace d'un espace de Banach B et soit X un élément de B n'appartenant pas à M . Alors, pour tout $Y \in \overline{\text{sp}}\{X, M\}$, il existe un scalaire unique a et un vecteur $e \in M$ tel que*

$$Y = aX + e$$

En utilisant le lemme 1.1, le théorème suivant établit la décomposition prédictive et la représentation de Wold d'un processus p -stationnaire.

Théorème 1.6. *(Miamee et Pourahmadi, 1988).*

Soit $\{X_t\}$ un processus p -stationnaire d'innovation $\{\epsilon_t\}$. Alors, pour tout entier $n \geq 1$,

- (a) *Il existe des constantes uniques a_1, \dots, a_n et un processus p -stationnaire $\{e_{t,n}\}$ avec*

$e_{t,n} \in \mathcal{H}_{t-1-n}(X)$ tel que

$$\begin{cases} \hat{X}_t = \sum_{k=1}^n a_k X_{t-k} + e_{t,n}, \\ X_t = \epsilon_t + \sum_{k=1}^n a_k X_{t-k} + \epsilon_{t,n}. \end{cases} \quad (1.82)$$

(b) Il existe des constantes uniques c_1, c_2, \dots, c_n et un unique processus p -stationnaire $\{V_{t,n}\}$ avec $V_{t,n} \in \mathcal{H}_{t-n-1}$ tel que

$$\begin{cases} \hat{X}_t = \sum_{k=1}^n c_k \epsilon_{t-k} + V_{t,n}, \\ X_t = \epsilon_t + \sum_{k=1}^n c_k \epsilon_{t-k} + V_{t,n}. \end{cases} \quad (1.83)$$

Comme l'entier n du théorème 1.6 est arbitraire et que les coefficients a_k et c_k ne dépendent pas de n et de plus ils sont uniques alors les suites infinies $\{a_k\}_{k=0}^\infty$ et $\{c_k\}_{k=0}^\infty$ véhiculent toute l'information liée à la définition d'un processus non déterministe p -stationnaire. Il est aussi important de noter que les deux séries sont liées par des relations de récurrence, données par le corollaire suivant :

Corollaire 1.3. *les suites $\{a_k\}$ et $\{c_k\}$ définies dans le théorème 1.6 sont liées par les relations de récurrence suivantes*

$$\begin{cases} c_0 = 1, \\ \sum_{k=0}^{l-1} c_k a_{l-k} = c_l, \quad l = 1, 2, \dots \end{cases} \quad (1.84)$$

Quand la suite $\{c_k\}$ est connue, les coefficients $\{a_k\}$ s'obtiennent par la résolution du système matriciel suivant :

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ c_1 & 1 & 0 & 0 \\ c_2 & c_1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \vdots \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \vdots \end{bmatrix}. \quad (1.85)$$

Dans ce qui suit, nous étudions deux exemples dont l'objectif est de trouver $\{a_k\}$ en fonction de la suite connue $\{c_k\}$. Comme supposé, $\{\epsilon_k\}$ est l'innovation du processus p -stationnaire $\{X_t\}$. Notre point de départ est la décomposition de Wold finie de $\{X_t\}$, cf. Théorème 1.6 (b).

Exemple 2. *Soit $\{X_t\}$ un processus vérifiant*

$$X_t = \epsilon_t - \epsilon_{t-1}, \quad \text{pour tout } t,$$

i.e. $c_0 = 1, c_1 = 1, c_k = 0, k \geq 2$. Alors, en utilisant (1.84) on obtient

$$a_k = -1, \quad \text{pour tout } k \geq 1,$$

et

$$e_{t,n} = \hat{X}_t - \sum_{k=1}^n a_k X_{t-k} = -\epsilon_{t-n-1}.$$

Exemple 3. Soit $\{X_t\}$ un processus donné par

$$X_t = \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2}, \quad \text{pour tout } t,$$

i.e. $c_0 = 1, c_1 = -2, c_2 = 1, c_k = 0, \text{ pour tout } k \geq 3$. Alors, il s'ensuit de (1.84) que

$$a_k = -(k+1), \quad \text{pour } k \geq 1,$$

et

$$e_{t,n} = \hat{X}_t - \sum_{k=1}^n a_k X_{t-k} = -(n+2)\epsilon_{t-n-1} + (n+1)\epsilon_{t-n-2}.$$

Remarque 1.11. La représentation de $\hat{X}_t = \sum_{k=1}^n a_k X_{t-k} + e_{t,n}$ dans (1.82) fournit une formule pour approximer le prédicteur \hat{X}_t en fonction des n observations passées X_{t-1}, \dots, X_{t-n} . L'intérêt de cette approximation réside dans le fait que les coefficients $\{a_k\}$ ne dépendent pas de n , et par conséquent ne varient pas quand n augmente c'est à dire quand le passé devient important. C'est dans ce contexte que nous posons le problème de recherche du meilleur prédicteur linéaire en fonction du passé $\{X_{t-1}, \dots, X_{t-n}\}$ noté $X_{t,n}^*$ qui minimise la quantité

$$\mathbf{E}|X_t - X_{t,n}^*|^p = \mathbf{E}|X_t - \sum_{k=1}^n a_{n,k} X_{t-k}|^p.$$

Il est bien connu que les coefficients du prédicteur à h pas d'un processus faiblement stationnaire, de fonction d'autocovariance $\{\gamma_k\}$ s'obtiennent facilement, cependant pour $1 < p < 2$, il n'est pas facile de trouver ces coefficients. Néanmoins, dans ce cas, \hat{X}_t peut être approximé en résolvant le système d'équations donné par (1.85). Ceci montre l'importance des coefficients $\{c_k\}$ dans la détermination de \hat{X}_t et dans d'autres caractérisations des processus p -stationnaire.

Pour illustrer l'importance de la suite $\{c_k\}$ dans la caractérisation d'un processus p -stationnaire, il est important de noter que, pour un processus 2-stationnaire, les coefficients sont à carrés sommables et par conséquent la suite est bornée. Pour un processus p -stationnaire, tout ce qui est connu sur ces paramètres est la majoration $|c_k| < c^{2^k}$, $k = 1, 2, \dots$, cette dernière permet de démontrer le théorème qui suit.

Théorème 1.7. *Soit $\{X_t\}$ un processus non déterministe p -stationnaire, $\{\epsilon_t\}$ son processus d'innovation et $\{c_k\}$, $\{V_{t,n}\}$ sont définis dans le théorème 1.6(b). Alors, pour tout $n \geq 1$*

$$V_{t,n} = P_{\mathcal{H}_{t-n-1}} \left(X_t - \sum_{k=0}^{n-1} c_k \epsilon_{t-k} \right). \quad (1.86)$$

Maintenant, en utilisant la suite unique $\{c_k\}$ associée au processus p -stationnaire $\{X_t\}$, non déterministe, on définit la fonction ϕ

$$\phi(z) = 1 + \sum_{k=1}^{\infty} c_k z^k, \quad |z| < 1/2, \quad (1.87)$$

qui est analytique sur le disque ouvert de rayon $1/2$. De (1.84), il est immédiat que les coefficients $\{a_k\}$ du théorème 1.6 (b) peuvent s'obtenir par

$$a_k = -d_k, \quad k = 1, 2, \dots \quad (1.88)$$

où les termes de la suite $\{d_k\}$ sont les coefficients de Taylor de la fonction réciproque ϕ^{-1} , i.e.

$$\phi^{-1}(z) = 1 + \sum_{k=1}^{\infty} d_k z^k. \quad (1.89)$$

Ces arguments montrent qu'en se donnant la fonction de transfert ϕ d'un processus non déterministe p -stationnaire, il est possible de déterminer les coefficients $\{c_k\}$ et $\{a_k\}$. Dans ce qui suit, nous explorons la possibilité de trouver les coefficients du prédicteur à h pas de $\{X_t\}$, i.e. les coefficients de \hat{X}_{t+h} , en fonction des coefficients de ϕ et de ϕ^{-1} . Bien entendu, ce théorème est une généralisation du théorème 1.6.

Théorème 1.8. *(Miamee et Pourahmadi, 1988).*

Soit $\{X_t\}$ un processus non déterministe, p -stationnaire et $\{\epsilon_t\}$ son processus d'innovation. Alors, pour $h \geq 1$ et $n \geq 1$,

(a) *Il existe une suite de constantes $a_{1,h}, \dots, a_{n,h}$ et un unique processus stationnaire*

$\{e_{t,n,h}\}_{t=-\infty}^{\infty}$ avec $e_{t,n,h} \in \mathcal{H}_{t-n-1}(X)$ tels que

$$\hat{X}_{t+h-1} = \sum_{k=1}^n a_{k,h} X_{t-k} + e_{t,n,h}. \quad (1.90)$$

De plus, pour $h = 1$, $a_{k,h} = a_k$, pour $k \geq 1$.

- (b) Il existent des constantes uniques c_1, \dots, c_n et un processus p -stationnaire $\{V_{t,n,h}\}_{t=-\infty}^{\infty}$ avec $V_{t,n,h} \in \mathcal{H}_{t-n-1}(X)$ tels que

$$\hat{X}_{t+h-1} = \sum_{k=1}^n c_{k,h} \epsilon_{t-k} + V_{t,n,h}. \quad (1.91)$$

De plus, pour $h = 1$, $c_{k,h} = c_k$, pour $k \geq 1$.

- (c) Pour $h \geq 1$, les suites $a_{k,h}$ et $c_{k,h}$ citées dans (a) et (b), sont liées par les relations de récurrence

$$\begin{aligned} c_0 &= 1, \\ \sum_{k=0}^{l-1} c_k a_{l-k,h} &= c_{l-h} \quad l = 1, 2, \dots, \end{aligned} \quad (1.92)$$

où $\{c_k\}$ est la suite des coefficients de la fonction de transfert de X_t .

Ce théorème montre que pour $h \geq 1$ fixé, les deux suites $\{a_{k,h}\}$ et $\{c_{k,h}\}$ sont liées par les relations de récurrence données par (1.92). Par conséquent, il est possible de déterminer les coefficients $a_{k,h}$ du prédicteur \hat{X}_{t+h-1} en fonction des coefficients $\{c_{k,h}\}$. Cependant, il n'est pas facile de trouver les coefficients $\{c_{k,h}\}$, cf. Cambanis, S. et Soltani, A.R. (1984). Dans ce qui suit, nous donnons quelques situations où les coefficients peuvent être calculés facilement.

Corollaire 1.4. (Miamee et Pourahmadi, 1988).

Soit $\{X_t\}$ un processus non déterministe p -stationnaire et $\{\epsilon_t\}$ son processus d'innovation. Si (i) $p = 2$ ou (ii) la suite $\{\epsilon_t\}$ est une suite i.i.d de variables aléatoires, alors, pour $h \geq 1$ fixé, et pour tout entier $n \geq 1$,

(a)

$$\hat{X}_{t+h-1} = \sum_{k=1}^n c_{h+k-1} \epsilon_{t-k} + V_{t-1+h,n+h-1},$$

où

$$V_{t-1+h,n+h-1} \in \mathcal{H}_{t-n-1}.$$

(b)

$$\hat{X}_{t+h-1} = \sum_{k=1}^n a_{k,h} X_{t-k} + e_{t,n,h}, \quad e_{t,n,h} \in \mathcal{H}_{t-n-1}$$

et

$$a_{k,h} = \sum_{j=0}^{k-1} c_{h+j} d_{k-j}, \quad k = 1, 2, \dots (c_0 = d_0 = 1).$$

Ce résultat montre que la décomposition de Wold finie suffit pour le calcul des coefficients du prédicteur d'horizon h d'un processus p -stationnaire.

1.5.2 Décomposition prédictive et décomposition de Wold

Le rôle de la décomposition de Wold dans la prédiction et la caractérisation d'un processus faiblement stationnaire non déterministe est bien connu. En effet, l'implication puissante de cette décomposition est l'orthogonalité des variables aléatoires du processus d'innovation.

Dans le cas d'un processus p -stationnaire, l'orthogonalité du processus d'innovation n'est pas vérifiée. Il est donc naturel de s'intéresser à cette décomposition dans un tel processus. Cette raison, entre autres, fournit une grande motivation pour la recherche de critères de convergence des séries $\sum_{k=1}^{\infty} c_k \epsilon_{t-k}$ et $\sum_{k=1}^{\infty} a_k X_{t-k}$. Le théorème qui suit donne une condition suffisante de la convergence de ces deux séries.

Théorème 1.9. (Miamee et Pourahmadi, 1988).

Soit $\{X_t\}$ un processus non déterministe, p -stationnaire d'innovation $\{\epsilon_t\}$. Alors,

- (a) $\sum_{k=1}^{\infty} a_k X_{t-k}$ est convergente en moyenne d'ordre p , si $\sum_{k=1}^{\infty} a_k < \infty$.
- (b) $\sum_{k=1}^{\infty} c_k \epsilon_{t-k}$ est convergente en moyenne d'ordre p , si $\sum_{k=1}^{\infty} |c_k| < \infty$.

Théorème 1.10. Soit $\{X_t\}$ un processus non déterministe p -stationnaire, d'innovation $\{\epsilon_t\}$ et $\{a_k\}$, $\{c_k\}$ les deux suites définies dans le théorème 1.6.

- (a) Si $\sum_{k=1}^{\infty} a_k X_{t-k}$ est convergente en moyenne d'ordre p , alors il existe un unique processus p -stationnaire $\{e_t\}$ avec $e_t \in \mathcal{H}_{-\infty}(X)$ tel que, pour tout t ,

$$X_t = \epsilon_t + \sum_{k=1}^{\infty} a_k X_{t-k} + e_t. \quad (\text{décomposition prédictive}).$$

- (b) Si $\sum_{k=1}^{\infty} c_k \epsilon_{t-k}$ est convergente en moyenne d'ordre p , alors il existe un unique processus p -stationnaire $\{V_t\}$ avec $V_t \in \mathcal{H}_{-\infty}(X)$ tel que, pour tout t ,

$$X_t = \epsilon_t + \sum_{k=1}^{\infty} c_k \epsilon_{t-k} + V_t = U_t + V_t, \quad (\text{Décomposition de Wold}).$$

Chapitre 2

Estimation et interpolation des données manquantes d'une série chronologique stationnaire

2.1 Introduction

On traite dans ce chapitre le problème de données manquantes d'une série chronologique. L'intérêt de ce problème, est justifié par la rencontre fréquente de séries avec données manquantes dans la réalité des problèmes pratiques. En effet, on a souvent affaire à des séries que l'on n'observe que pendant quelques heures de la journée, voire quelques jours de la semaine, à cause des rythmes de la vie sociale ou économique. En général, les données manquantes peuvent être attribuées à des origines diverses tenant - par exemple - à la nature du phénomène observé, à l'appareil de mesure, etc ..., on peut citer des exemples de séries avec données manquantes plus systématique, comme dans Parzen (1983) : enregistrer l'écho réfléchi d'un signal radar émis vers la lune nécessite l'arrêt de l'émission durant la période de réception de l'écho - faute d'émission. D'autres exemples figurent dans l'expérience océanographique.

Il est important de souligner que l'analyse des données manquantes est en relation avec les méthodes de détection de valeurs aberrantes. En effet, dans bien de situations réelles, les observations d'une série sont observées avec erreurs, une analyse prudente des données exige, dans certaines situations, l'élimination de ces observations et de les considérer manquantes.

L'estimation des paramètres dans le cas de valeurs manquantes a été intensivement relaté dans la littérature. Parzen (1983) l'a étudiée dans un cas particulier, en préconisant une approche qui a donné lieu à certains développements concernant l'aspect non paramétrique

du problème, tels ceux de Dunsumir et Robinson (1981) consacré au domaine temporel et ceux de Bloomfield (1970) consacré au domaine spectral; quant à l'aspect paramétrique, il a diversement intéressé, dès 1979, des auteurs comme Jones et Tan (1980) et Dunsmuir et Robinson (1981). Ce problème a été aussi traité en introduisant les processus modulés en amplitude, dans son aspect non paramétrique par Ladjouze (1986). Shumway et Stoffer (2000, ch 4) suggèrent un modèle paramétrique pour ajuster les données, soit en maximisant la fonction de vraisemblance exacte des données observées, soit en utilisant un algorithme de type EM. Lorsque le modèle paramétrique admet une représentation par espace d'état de dimension finie comme c'est le cas par exemple pour un modèle ARMA, cette représentation peut être modifiée pour tenir compte des données manquantes et la fonction de vraisemblance exacte peut être calculée au moyen des récursions de Kalman associées (Brockwell and Davis 1992). Ensuite, des estimées des données manquantes peuvent être obtenues en utilisant un algorithme de lissage appliqué à la représentation d'état modifiée. Une autre approche consiste à utiliser des formules explicites des estimées des données manquantes dans lesquelles les paramètres inconnus sont remplacés par leurs estimations. Cette approche directe est plus efficace, particulièrement si le nombre de données manquantes est faible.

Ce chapitre est organisé comme suit : Dans la section 2, nous collectons quelques résultats basiques sur les processus non déterministes, décomposition de Wold et coefficients du meilleur prédicteur linéaire. Dans la section 3, nous résolvons le problème d'estimation d'une seule valeur manquante. Dans la section 4, une généralisation du résultat obtenu dans 3 est présentée pour le cas de plusieurs valeurs manquantes dont les indices ne sont pas nécessairement successifs. Dans la section 5, nous utilisons l'algorithme EM pour estimer simultanément les valeurs manquantes et les paramètres du modèle. Nous terminons ce chapitre par une étude comparative entre deux estimateurs de données manquantes d'une série chronologique. Cette étude comparative est effectuée en retenant la proximité de Pitman comme critère de comparaison.

2.2 Préliminaires

Soit $\{X_t\}$ une série chronologique centrée stationnaire de fonction d'autocovariance γ_k . Notons $L^2(\Omega, \mathcal{F}, P)$ l'espace des variables aléatoires réelles définies sur (Ω, \mathcal{F}, P) pour lesquelles $\mathbf{E}X^2 < \infty$, muni du produit scalaire $\langle X, Y \rangle = \mathbf{E}XY$.

Pour tout $t \in \mathbb{Z}$, nous définissons le sous espace

$$\mathcal{H}_t = \overline{\text{sp}}\{X_s, s \leq t\},$$

où $\overline{\text{sp}}\{A\}$ est le sous espace engendré par les éléments de A , qui est, un sous espace fermé de $L^2(\Omega, \mathcal{F}, P)$.

Le processus $\{X_t\}$ est non déterministe si $X_{t+1} \notin \mathcal{H}_t$.

Il est bien connu que tout processus non déterministe peut se décomposer d'une manière unique comme

$$X_t = \sum_{k=0}^{\infty} b_k \epsilon_{t-k} + V_t, \quad b_0 = 1, \quad \sum_{k=0}^{\infty} b_k^2 < \infty \quad (2.1)$$

où $\{\epsilon_t\}$ est une suite de v.a. non corrélées, avec $V(\epsilon_t) = \sigma^2$, ϵ_t et V_t sont non corrélées.

Les coefficients de la représentation autorégressive d'un tel processus s'obtiennent par récurrence :

$$\begin{aligned} b_0 &= 1 \\ b_l &= \sum_{k=0}^{l-1} b_k a_{l-k} \quad (l = 1, 2, \dots). \end{aligned} \quad (2.2)$$

Notons que la décomposition de Wold engendre la décomposition utile de la matrice de covariance $\Gamma = (\gamma_{k-l})$ du processus $\{X_t\}$:

$$\Gamma = \sigma^2 T T' + \Gamma_V \quad (2.3)$$

où

$$T = (b_{j-i})_{\substack{i=0, \infty \\ j=0, \infty}} \quad (2.4)$$

avec $b_0 = 1$, $b_k = 0$ pour $k < 0$, et Γ_V est la matrice de covariance du processus $\{V_t\}$.

Rappelons aussi que $\{X_t\}$ est purement non déterministe si $V_t \equiv 0$ ou $\Gamma_V = 0$.

Pour $k > 0$, le prédicteur de l'observation X_k basé sur $\mathcal{H}_{-1} = \overline{\text{sp}}\{X_{-1}, X_{-2}, \dots\}$ est noté par \hat{X}_k . Il est la projection orthogonale de X_k sur \mathcal{H}_{-1} . De (2.1), il suit

$$\begin{aligned} \hat{X}_k &= \sum_{i=k+1}^{\infty} b_i \epsilon_{k-i} + V_t \\ X_k - \hat{X}_k &= \sum_{i=0}^k b_i \epsilon_{k-i} \\ \text{var}(X_k - \hat{X}_k) &= \sigma^2 \sum_{i=0}^k b_i^2 \\ \text{cov}(X_k - \hat{X}_k, X_l - \hat{X}_l) &= \sigma^2 \sum_{i=0}^{\min(k,l)} b_i^2 + |l - k| \quad (l \geq 0). \end{aligned} \quad (2.5)$$

Il est important de signaler que $\text{cov}(X_k - \hat{X}_k, X_l - \hat{X}_l)$ est le (k,l) ème élément de la matrice $G = TT'$. Cette matrice joue un rôle important dans ce qui suit.

Il est bien connu que $\hat{X}_k, k \geq 0$, possède une représentation AR, en fonction des valeurs passées X_{-1}, X_{-2}, \dots :

$$\hat{X}_k \sim \sum_{i=1}^{\infty} a_{i,k} X_{-i} + V_k \quad (2.6)$$

et que les coefficients $\{a_{i,k}\}_{i=1}^{\infty}$ de la représentation AR dans (2.6) satisfont

$$\begin{aligned} b_0 &= 1 \\ \sum_{j=0}^{l-1} b_j a_{l-j,k} &= b_{l+k} \quad (l = 1, 2, \dots). \end{aligned} \quad (2.7)$$

2.3 Interpolation d'une seule valeur manquante

Pour la clarté de la présentation, nous commençons par le cas où une seule valeur est manquante. Soit $X'_{0,r}$ la projection orthogonale de X_0 sur

$$\mathcal{H}'_{0,r} = \overline{\text{sp}}\{X_t; t \leq r, t \neq 0\}.$$

Une difficulté essentielle dans la projection orthogonale sur le sous espace $\mathcal{H}'_{0,r}$, est que les sous espaces engendrés par le passé et le futur ne sont pas orthogonaux. Il convient donc d'écrire le sous espace $\mathcal{H}'_{0,r}$ sous forme d'une somme directe de deux espaces orthogonaux. Les deux sous espaces sont \mathcal{H}_{-1} et $S_r = \overline{\text{sp}}\{X_k - \hat{X}_k; 1 \leq k \leq r\}$.

Lemme 2.1. (*Pourahmadi, 1989*).

Soit $\{X_t\}$ un processus stationnaire non déterministe. Alors, pour tout $0 \leq r < \infty$,

- (a) $X_0 \notin \mathcal{H}'_{0,r}$
- (b) \mathcal{H}_{-1} et S_r sont orthogonaux, et

$$\mathcal{H}'_{0,r} = \mathcal{H}_{-1} \oplus S_r.$$

Démonstration. (a).

Supposons que $X_0 \in \mathcal{H}'_{0,r}$. Alors on peut l'écrire comme

$$X_0 = \sum_{k=1}^r c_k X_k + Y \quad (2.8)$$

avec $Y \in \mathcal{H}_{-1}$. Sans perte de généralité, nous supposons que $c_r \neq 0$. Ceci entraîne que

$$X_r = c_r^{-1}X_0 - \sum_{k=1}^{r-1} c_r^{-1}c_k X_k - c_r^{-1}Y \in \mathcal{H}_{r-1},$$

ce qui contredit que $\{X_t\}$ est non déterministe. Par conséquent $X_0 \notin \mathcal{H}'_{0,r}$.
(b).

Tout $X \in \mathcal{H}'_{0,r}$ peut s'écrire sous la forme

$$X = Y + \sum_{k=1}^r c_k X_k = \left(Y + \sum_{k=1}^r c_k \hat{X}_k \right) + \sum_{k=1}^r (X_k - \hat{X}_k) \in \mathcal{H}_{-1} \oplus S_r.$$

□

Il est maintenant possible d'écrire

$$\hat{X}'_{0,r} = P_{\mathcal{H}'_{0,r}} X_0 = P_{\mathcal{H}_{-1}} X_0 + P_{S_r} X_0 \quad (2.9)$$

Une représentation explicite de $P_{S_r} X_0$ est fournie par le lemme suivant :

Lemme 2.2. (Pourahmadi, 1989).

Si $\{X_t\}$ est un processus non déterministe alors

$$P_{S_r} X_0 = \sum_{i=1}^r c_{i,r} (X_i - \hat{X}_i),$$

où le vecteur $c = (c_1, \dots, c_r)'$ vérifie

$$A_r c = b$$

et $b = (b_1, \dots, b_r)'$ représente les premières composantes de la représentation moyenne mobile de $\{X_t\}$,

$$A_r = T_r' T_r + b b'$$

avec

$$T_r = (b_{j-i})_{i,j=0,\dots,r-1}.$$

Le théorème qui suit donne l'expression explicite de la meilleure interpolation linéaire d'une seule valeur manquante.

Théorème 2.1. (Pourahmadi, 1989).

Soit $\{X_t\}$ un processus stationnaire non déterministe de représentation autorégressive

(AR) de paramètres $\{a_k\}$, et soit $\hat{X}'_{0,r}$ la meilleure interpolation linéaire de X_0 basée sur $\{X_t, t \leq r, t \neq 0\}$, alors

(a)

$$\hat{X}'_{0,r} = \hat{X}_0 + \sum_{k=1}^r c_{k,r}(X_k - \hat{X}_k), \quad (2.10)$$

où

$$c_{k,r} = \left(1 + \sum_{i=1}^r a_i^2\right)^{-1} \left(a_k - \sum_{i=1}^{r-k} a_i a_{i+k}\right) \quad (k = 1, 2, \dots, r),$$

(b)

$$\text{var}(X_0 - \hat{X}'_{0,r}) = \sigma^2 \left(1 + \sum_{i=1}^r a_i^2\right)^{-1}. \quad (2.11)$$

Démonstration. (2.10) découle du lemme 2.2.

Pour démontrer (2.11), remarquons d'abord

$$X_0 - \hat{X}'_{0,r} = X_0 - \hat{X}_0 - \sum_{k=1}^r c_{k,r}(X_k - \hat{X}_k).$$

Comme $\hat{X}'_{0,r}$ est orthogonale à $(X_0 - \hat{X}'_{0,r})$, et

$$\text{var}(X_0 - \hat{X}'_{0,r}) = \sigma^2,$$

on en déduit

$$\begin{aligned} \text{var}(X_0 - \hat{X}'_{0,r}) &= \text{cov}(X_0 - \hat{X}'_{0,r}, X_0) \\ &= \text{cov}(X_0 - \hat{X}_0, X_0) - \sum_{k=1}^r c_{k,r} \text{cov}(X_k - \hat{X}_k, X_0) \\ &= \text{var}(X_0 - \hat{X}_0) - \sigma^2 \sum_{k=1}^r c_{k,r} b_k \\ &= \sigma^2(1 - b'A_r^{-1}b) \\ &= \sigma^2(1 + a'a). \end{aligned}$$

□

Exemple 2.1. Soit $\{X_t\}$ un processus autorégressif d'ordre 1, stationnaire, d'équation $X_t = a_1 X_{t-1} + \epsilon_t$ alors

$$\hat{X}'_{0,r} = \frac{a_1}{1 + a_1^2}(X_1 + X_{-1}).$$

2.4 Interpolation de plusieurs valeurs manquantes

Le schéma précédent s'applique tout à fait à plusieurs valeurs manquantes et fournit une procédure d'estimation de ces dernières en usant de la même technique. Ainsi, nous résoudrons ce problème sans restriction sur le modèle et nous distinguons le cas particulier où les indices des valeurs manquantes sont successifs du cas où les indices sont quelconques. Le cas particulier a été l'objet des travaux de Brubacker et Wilson (1976) et Abrahm (1981). Pourahmadi (1989) s'est intéressé en suite à la généralisation de ces résultats, d'autant plus c'est le cas le plus fréquent dans la pratique. Soit

$$M = \{n_1, \dots, n_m\} \quad n_1 = 0 < n_2 < \dots < n_m$$

les indices des valeurs manquantes du processus, et $m = \text{card } M$ est le nombre total des valeurs manquantes.

Soit

$$K = \{k_1, \dots, k_r\} \quad 0 < k_1 < k_2 < \dots < k_r < \infty$$

l'ensemble des indices des valeurs observées après l'insant $t = 0$.

Nous utilisons le vecteur

$$X_M = [X_{n_1}, \dots, X_{n_m}]'$$

pour caractériser les données manquantes et le vecteur

$$X_K = [X_{k_1}, \dots, X_{k_r}]'$$

pour caractériser les données observées.

Il est à noter que le lemme 2.2 est démontré d'une manière analogue quand plusieurs valeurs sont manquantes. L'équivalent de la matrice A_r dans le lemme 2.2 est

$$A_K = \sigma^{-2} \{ \text{cov}(X_{k_i} - \hat{X}_{k_i}, X_{k_j} - \hat{X}_{k_j}) \}_{i,j=1,\dots,r}. \quad (2.12)$$

L'équivalent du théorème 2.1 s'annonce alors :

Théorème 2.2. (Pourahmadi, 1989).

Soit $\{X_t\}$ un processus stationnaire non déterministe dont les indices des valeurs manquantes sont données par M . Alors, avec les notations précédentes, nous avons :

a)

Pour tout $1 \leq j \leq m$

$$\hat{X}'_{n_j} = \hat{X}_{n_j} + \sum_{i=1}^r c_{ij}(X_{k_i} - \hat{X}_{k_i}),$$

où les vecteurs

$$c_j = (c_{1j}, \dots, c_{rj})'$$

sont solutions des équations

$$A_K c_j = b_j.$$

L'interpolation des valeurs manquantes peut s'écrire sous la forme vectorielle suivante

$$\hat{X}'_M = \hat{X}_M + C'(X_K - \hat{X}_K),$$

où C est une $r \times m$ vérifiant

$$A_K C = B.$$

(b)

$$\text{cov}(X_M - \hat{X}'_M) = \text{Cov}(X_M - \hat{X}_M) - \sigma^2 B' A_K^{-1} B.$$

Exemple 4. Pour un processus autorégressif d'ordre 1 stationnaire

$$X_t = aX_{t-1} + \epsilon_t$$

nous avons, avec les mêmes notations,

$$A = \begin{bmatrix} a^m & a^{m-1} & \dots & a \\ 0 & 0 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Soit a_1 la première ligne de la matrice A . Alors

$$A'A = a'_1 a_1 = [a^{2m-i-j}]_{i,j=1,\dots,m}.$$

En utilisant le lemme 2.2, avec $\alpha = (1 + a_1 a'_1)^{-1}$, on obtient

$$(I_m + A'A)^{-1} = I_m - \alpha a'_1 a_1 = I_m - \alpha A'A.$$

D'où

$$C = A(I_m + A'A)^{-1} T_m = (I_r - \alpha A A') A T_m.$$

Notons que

$$AA' = \begin{bmatrix} \sum_{i=1}^m a^{2i} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

et

$$AT_m = \begin{bmatrix} a_1 T_m \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Par conséquent

$$X'_M = \hat{X}_M + \alpha(a_1 T_m)'(X_m - \hat{X}_m),$$

où

$$\hat{X}_M = X_{-1} \begin{bmatrix} a \\ a^2 \\ \vdots \\ a^m \end{bmatrix}, \quad \hat{X}_m = a^{m+1} X_{-1}.$$

Comme attendu, l'interpolation des valeurs manquantes X_0, X_1, \dots, X_{m-1} dépend seulement de X_{-1} et X_m .

2.5 Algorithme EM

L'algorithme d'espérance maximisation (EM) est une méthode itérative efficace pour calculer l'estimation de vraisemblance maximale en présence de données manquantes. Le but est d'estimer les paramètres du modèle pour lesquelles les données observées sont les plus vraisemblables tout en tenant compte de l'existence de données manquantes. Le problème de la maximisation de la vraisemblance consiste à trouver les paramètres qui maximisent la vraisemblance. Plus formellement, si on définit X_K comme les données incomplètes observées, on assume qu'il existe un ensemble complet de données $X = (X_K, X_M)$ où X_M représente l'ensemble des données manquantes. On peut donc écrire la fonction de densité jointe comme suit :

$$P(X \setminus \theta) = P(X_K, X_M \setminus \theta) = P(X_M \setminus X_K, \theta).P(X_K \setminus \theta)$$

et de la même façon la log-vraisemblance de l'ensemble complet de données :

$$L(\theta \setminus X) = L(\theta \setminus X_K, X_M) = \ln(X_K, X_M \setminus \theta)$$

On doit donc trouver la valeur espérée de la log-vraisemblance de l'ensemble complet de données X par rapport aux données manquantes X_M sachant les données observées X_K et les paramètres du modèle θ . Le problème est qu'on ne connaît pas les variables cachées, il faut donc utiliser les données et les paramètres d'une itération précédente X_K et $\theta^{(i-1)}$. La définition de cette espérance est :

$$Q(\theta, \theta^{(i-1)}) = \mathbf{E}_{X_M}[\ln(P(X_K, X_M \setminus \theta) \setminus X_K, \theta^{(i-1)})]$$

où \mathbf{E}_{X_M} est l'espérance par rapport à X_M , $\theta^{(i-1)}$ les paramètres utilisés pour évaluer l'espérance et θ les nouveaux paramètres à optimiser pour maximiser Q . C'est par le calcul de cette espérance conditionnelle qu'on mesure la log-vraisemblance. Dans cette expression, X_K et $\theta^{(i-1)}$ sont constants, θ est la variable ajustable et X_M est une variable aléatoire gouvernée par la distribution $f(X_M \setminus X_K, \theta^{(i-1)})$. On peut alors réécrire le terme de droite comme :

$$\mathbf{E}_{X_M}[\ln(P(X_K, X_M \setminus \theta) \setminus X, \theta^{(i-1)})] = \int_{x_m} \ln(P(X_K, x_m \setminus \theta)) f(x_m \setminus X_K, \theta^{(i-1)}) dx_m$$

Cette fonction est déterministe et peut donc être maximisée. Une fois ces quantités définies, tout est prêt pour l'optimisation.

Les étapes de l'algorithme EM sont donc :

1. Étape E: Évaluation de l'espérance $Q(\theta, \theta^{(i-1)})$ selon les données observées et les paramètres à notre disposition.
2. Étape M: Maximisation de cette espérance $\theta^{(i)} = \max_{\theta} Q(\theta, \theta^{(i-1)})$ selon θ .

Ces étapes sont répétées autant de fois que nécessaire. Chaque itération fait augmenter la log-vraisemblance, donc l'algorithme converge directement vers un maximum local de la fonction de log-vraisemblance.

En utilisant les notations de la section 4, dans ce qui suit, nous appliquons l'algorithme EM pour estimer les paramètres du modèles et les valeurs manquantes en s'appuyant sur le théorème 2.2.

- Etape E: Sachant θ , l'estimateur des valeurs manquantes s'obtiennent par le théorème 2.2

$$\hat{X}'_M = \hat{X}_M + C'(X_K - X_K),$$

où $\hat{X}_K = (\hat{X}_{n_1}, \dots, \hat{X}_{n_m})'$ et $\hat{X}_M = (\hat{X}_{k_1}, \dots, \hat{X}_{k_r})$ sont les prédicteurs de X_M et X_K basés sur le passé fini $X_1, X_2, \dots, X_{n_1-1}$ $\hat{X}'_M = (\hat{X}'_{k_1}, \dots, \hat{X}'_{k_r})'$ est le vecteur des interpolateurs des valeurs manquantes X_M basé sur $\{X_t; t \in M^c = \{1, 2, \dots, N\} \setminus M\}$

- Etape M : Après l'estimation des données manquantes, les données "complètes" peuvent être utilisées pour maximiser la fonction de vraisemblance. Soit $T(X_1, X_2, \dots, X_N)$ la solution de ce problème de maximisation.

Le calcul de ce nouvel estimateur permet de retourner à l'étape du calcul de l'espérance et ainsi de suite. Une fois le seuil de tolérance est atteint on possède l'estimation de la vraisemblance maximale de notre ensemble complet de données X .

Remarque 2.1. Comme tout procédé itératif, la mise en œuvre de cet algorithme nécessite une valeur initiale $\theta^{(0)}$. Le choix de cette dernière dépend de la nature des données et de la fonction d'autocovariance γ_k du processus. En effet, trois cas sont possibles :

- (a) La fonction γ_k est connue ;
- (b) On dispose d'un long segment d'observations ;
- (c) On ne dispose pas d'un long segment d'observations.

Dans le cas de la situation (a), une seule utilisation de l'étape E est nécessaire pour estimer les valeurs manquantes. Dans le cas de la situation (b), on utilise le segment pour estimer les paramètres du modèle (procédure de Box-Jenkins), la convergence de l'algorithme est obtenue en peu d'itérations. Dans le dernier cas, nous supposons que les données peuvent être modélisées par un ARMA(p,q) fini d'équation

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}.$$

Ainsi, zéro est la valeur initiale la plus appropriée ; i.e. nous supposons que X_1, \dots, X_N sont collectées suivant un processus bruit blanc. Dans ce cas, l'interpolation des valeurs manquantes est la moyenne des valeurs observées.

2.6 Problème d'interpolation

Dans cette section, nous discutons du rôle du théorème 2.1 dans la caractérisation d'un processus minimal et nous résolvons le problème d'interpolation ($r = \infty$) des valeurs manquantes. Cette approche fournit quelques nouvelles perspectives concernant le rapport entre la prédiction et l'interpolation. En outre, une relation de dualité entre la représentation autorégressive AR(∞) et la moyenne mobile MA(∞) est établie. Cette relation est une généralisation du résultat connu dans le cas d'ordre fini.

Afin de résoudre le problème d'interpolation de X_t basée sur $\{X_s, s \neq t\}$, nous utilisons les notations suivantes :

$$\begin{aligned}\mathcal{H}'_t &= \overline{\text{sp}}\{X_s; s \neq t\} \\ \hat{X}'_t &= P_{\mathcal{H}'_t} X_t \\ \eta_t &= X_t - \hat{X}'_t \\ \sigma'^2 &= \text{var}(\eta_t) \geq 0.\end{aligned}\tag{2.13}$$

Le processus stationnaire $\{\eta_t\}$ est associé à l'erreur d'interpolation. Notons que, contrairement à $\{\epsilon_t\}$, les erreurs d'interpolation $\{\eta_t\}$ ne sont pas orthogonales.

Le processus stationnaire $\{X_t\}$ est dit minimal si $X_t \notin \mathcal{H}'_t$, pour tout t , ce qui est équivalent à

$$\sigma'^2 \neq 0.\tag{2.14}$$

Comme $\mathcal{H}_{t-1} \subseteq \mathcal{H}'_t$, il s'en suit que tout processus stationnaire minimal est non déterministe, l'inverse n'est pas nécessairement vraie.

Le fait que $\mathcal{H}'_{0,r} \rightarrow \mathcal{H}'_0$, quand $r \rightarrow \infty$ nous assure que

$$\hat{X}'_0 = \lim_{r \rightarrow \infty} \hat{X}'_{0,r}\tag{2.15}$$

et

$$\sigma'^2 = \lim_{r \rightarrow \infty} \text{var}(X_0 - \hat{X}'_{0,r}).\tag{2.16}$$

En utilisant le théorème 2.1, (2.15), (2.16) et en supposant que

$$\sum_{i=1}^{\infty} a_i^2 < \infty,\tag{2.17}$$

il vient

$$\sigma'^2 = \sigma^2 \lim_{r \rightarrow \infty} \left(1 + \sum_{i=1}^r a_i^2\right)^{-1} = \sigma^2 (a_k - 1 + a_k^2)^{-1} \neq 0\tag{2.18}$$

et

$$\lim_{r \rightarrow \infty} c_{k,r} = \left(1 + \sum_{i=1}^{\infty} a_i^2\right)^{-1} \left(a_k - \sum_{i=1}^{\infty} a_i a_{i+k}\right) = c_k \quad (k = 1, 2, \dots).\tag{2.19}$$

Ces deux dernières équations fournissent une preuve du théorème important de Kolmogorov (1941).

Théorème 2.3. (Pourahmadi, 1989).

Soit $\{X_t\}$ un processus stationnaire non déterministe de représentation AR de paramètres $\{a_k\}$. Alors on a :

(a) Le processus $\{X_t\}$ est minimal si et seulement si

$$\sum_{i=1}^{\infty} a_i^2 < \infty.$$

(b) Pour un processus stationnaire minimal

$$\sigma'^2 = \sigma^2 \left(1 + \sum_{i=1}^{\infty} a_i^2 \right)^{-1}.$$

Notons que (2.15) et (2.19) suggèrent que \hat{X}'_t admet une représentation de la forme

$$\hat{X}'_t \sim \hat{X}_t + \sum_{k=1}^{\infty} c_k (X_{t+k} - \hat{X}_{t+k}), \quad (2.20)$$

où \hat{X}_{t+k} , $k = 0, 1, 2, \dots$, est le meilleur prédicteur linéaire de X_{t+k} basé sur \mathcal{H}_{t-1} .

Il s'en suit que

$$\eta_t = X_t - \hat{X}'_t \sim X_t - \hat{X}_t - \sum_{k=1}^{\infty} c_k (X_{t+k} - \hat{X}_{t+k}) \quad (2.21)$$

peut s'exprimer en fonction des $\epsilon_t, \epsilon_{t-1}, \dots$

Nous présentons maintenant une conséquence de (2.20) et (2.21) en précisant son intérêt dans l'analyse des séries chronologiques. Dans ce qui suit, nous utiliserons la version normalisée de $\{\eta_t\}$ notée par $\{X'_t\}$:

$$X'_t = \sigma'^{-2} \eta_t. \quad (2.22)$$

Le processus $\{X'_t\}$ est dit processus dual (inverse) associé à $\{X_t\}$. Pour simplifier les notations et, sans perte de généralités, nous supposons que $\sigma^2 = 1$. Il est trivial que $\{X'_t\}$ est un processus stationnaire. Nous utilisons $\{\gamma'_k\}$, $\{a'_k\}$ etc ... pour représenter sa fonction de covariance, les paramètres de la représentation AR etc Quelques propriétés importantes de $\{X'_t\}$ sont résumées dans le lemme suivant :

Lemme 2.3. (Massani, 1960).

Soit $\{X_t\}$ un processus stationnaire minimal et $\{X'_t\}$ son dual. Alors on a :

1. $Cov(X'_t, X_s) = \delta_{t,s}$,
2. $X'_t = \epsilon_t - \sum_{i=1}^{\infty} a_i \epsilon_{t+i}$.

Ce résultat donne l'expression de la fonction d'autocovariance γ'_k du processus dual $\{X'_t\}$ en fonction de celle de $\{X_t\}$. Dans ce qui suit, nous donnons des rapports plus spécifiques qui mettent en évidence le lien entre les représentations AR et MA de $\{X_t\}$ et celle de $\{X'_t\}$

Théorème 2.4. (Pourahmadi, 1989).

Les paramètres de la représentation autorégressive et moyenne mobile de $\{X_t\}$ et $\{X'_t\}$ vérifient

$$b'_0 = 1, \quad b'_k = -a_k, \quad (k = 1, 2, \dots), \quad (2.23)$$

$$a'_0 = 1, \quad a'_k = -b_k, \quad (k = 1, 2, \dots). \quad (2.24)$$

Ce théorème montre qu'il y a, en effet, une dualité structurelle entre la représentation AR de $\{X_t\}$ et la représentation MA de $\{X'_t\}$ et vice versa. Par conséquent, la tâche difficile d'estimer les paramètres de la représentation MA de $\{X_t\}$ peut être réduite à la tâche plus facile d'estimation des paramètres de la représentation AR de $\{X'_t\}$. Cette relation importante entre $\{X_t\}$ et $\{X'_t\}$ a été utilisée à maintes reprises dans l'analyse des séries chronologiques, en l'occurrence les problèmes d'identification et d'estimation (Pierce, 1970; Cleveland, 1972; Chatfield, 1979; Battaglia, 1988).

Les résultats précédents portant sur la dualité entre $\{X_t\}$ et $\{X'_t\}$ peuvent s'obtenir en imposant des conditions plus fortes que la minimalité du processus $\{X_t\}$. En effet, Pourahmadi 1985, utilise le concept de l'angle ρ_X entre les sous espaces passé-présent et futur pour étudier la représentation AR d'un processus stationnaire et montre que $\rho_X < 1$ est une condition suffisante de la convergence de la série (2.6) du prédicteur \hat{X}_k . En utilisant ce résultat et les notations précédentes, une relation de dualité entre $AR(\infty)$ et $MA(\infty)$ est établie par le théorème suivant :

Théorème 2.5. (Pourahmadi, 1985).

Soit $\{X_t\}$ un processus stationnaire avec $\rho_X < 1$. Alors

$$X_t = \epsilon_t + \sum_{k=1}^{\infty} b_k \epsilon_{t-k}$$

si et seulement si

$$X'_t = - \sum_{k=1}^{\infty} b_k X'_{t-k} + \epsilon'_t.$$

2.7 Interpolation suboptimale

Soit \mathcal{H} un espace de Hilbert, \mathcal{H}_1 et \mathcal{H}_2 deux sous espaces de \mathcal{H} et $\mathcal{H}_{1,2}$ le sous espace engendré par $\mathcal{H}_1 \cup \mathcal{H}_2$.

Le calcul de $PX = P_{\mathcal{H}}X$, quand $P_1X = P_{\mathcal{H}_1}X$ et $P_2X = P_{\mathcal{H}_2}X$ sont connus, nécessite des opérations complexes. En effet, Aronszajn (1950) démontre que

$$\begin{aligned} PX &= (I - P_2)(I - P_1P_2)^{-1}P_1X + (I - P_1)(I - P_1P_2)^{-1}P_2X \\ &= \sum_{k=1}^{\infty} [P_1(P_2P_1)^{k-1} + (P_2P_1)^{k-1} - (P_2P_1)^k - (P_1P_2)^k] X. \end{aligned} \quad (2.25)$$

Cette série converge en norme dans \mathcal{H} et l'angle entre \mathcal{H}_1 et \mathcal{H}_2 est positif, c'est à dire

$$\cos \phi = \sup_{h_i \in \mathcal{H}_i, \|h_i\|=1, i=1,2} |(h_1, h_2)| < 1.$$

Néanmoins, malgré la complexité de (2.25), l'idée d'alterner les projections donne lieu à beaucoup de résultats algorithmiques en statistique. Les plus importants sont l'algorithme de la factorisation d'une densité spectrale d'un processus, voir Massani et Weiner 1958, l'algorithme d'interpolation d'un processus stationnaire, voir Salehi 1974 et l'algorithme ACE proposé par Brieman et Friedman 1985. Pour d'autres applications voir Gaffke et Mathar 1989.

Dans cette partie, en utilisant (2.25), nous présentons une nouvelle approche pour résoudre le problème d'interpolation. Nous nous restreignons au problème d'interpolation semi-fini, lequel revient à trouver la meilleure interpolation linéaire de X_0 ainsi que l'erreur d'interpolation, basé sur $\{X_t, t \leq m, t \neq 0\}$, $0 < m < \infty$. Notons qu'avec $\mathcal{H}_1 = \overline{\text{sp}}\{X_t, t \leq -1\}$ et $\mathcal{H}_2 = \overline{\text{sp}}\{X_t, t \geq 1\}$, il est facile d'obtenir $P_1 X_0$ et $P_2 X_0$, mais le calcul de la projection $P X_0$ est plutôt compliquée, voir Nakazi 1980, Miamee et Pourahmadi 1988. Ainsi, en appliquant (2.25), la série résultante peut être simplifiée pour obtenir l'expression de $P X_0$ et celle de l'erreur $\|X_0 - P X_0\|^2$. Un examen minutieux et prudent de la formule suggère une nouvelle interpolation géométrique en transformant les deux sous espaces \mathcal{H}_1 et \mathcal{H}_2 en deux sous espaces orthogonaux.

En utilisant les mêmes notations que précédemment, dans ce qui suit, nous présentons une méthode d'interpolation suboptimale de valeurs manquantes d'une série chronologique. Cette méthode d'estimation peut se décomposer en deux problèmes partiels : on cherche à résoudre les problèmes de projection sur \mathcal{H}_1 et \mathcal{H}_2 séparément, le problème initial, ainsi affaibli, se ramène au problème de la construction de la meilleure combinaison linéaire des deux solutions obtenues.

Soit $\hat{X}_{0,m} = P_2 X_0$ le prédicteur basé sur les observations futures $\{X_t; 1 \leq t \leq m\}$. Définissons $\tilde{X}_{0,m} = \alpha_m \hat{X}_0 + \beta_m \hat{X}_{0,m}$ avec α_m et β_m à choisir de telle sorte à minimiser $\text{Var}(\hat{X}_0 - \tilde{X}_{0,m})$. Il est évident que α_m et β_m vérifient

$$\begin{pmatrix} \text{Var}(\hat{X}_0) & \text{Cov}(\hat{X}_0, \hat{X}_{0,m}) \\ & \text{Var}(\hat{X}_{0,m}) \end{pmatrix} \begin{pmatrix} \alpha_m \\ \beta_m \end{pmatrix} = \begin{pmatrix} \text{Cov}(X_0, \hat{X}_{0,m}) \\ \text{Cov}(X_0, \hat{X}_{0,m}) \end{pmatrix} \quad (2.26)$$

Comme

$$\begin{aligned} \text{Cov}(X_0, \hat{X}_{0,m}) &= \text{Var}(\hat{X}_0) = \gamma_0 - \sigma^2 \\ \text{Cov}(X_0, \tilde{X}_{0,m}) &= \text{Var}(\hat{X}_{0,m}) = \gamma_0 - \sigma_m^2, \end{aligned}$$

en définissant $\rho_m = \text{Corr}(\hat{X}_0, \hat{X}_{0,m})$ comme coefficient de corrélation entre les deux prédicteurs de X_0 et $r_m = \sqrt{\text{Var}(\hat{X}_0 / \tilde{X}_{0,m})}$, il s'en suit que la solution de (2.26) est

$$\alpha_m = \frac{1 - r_m \rho_m}{1 - \rho_m^2}, \quad \beta_m = \frac{1 - r_m^{-1} \rho_m}{1 - \rho_m^2} \quad (2.27)$$

avec

$$Var(X_0 - \tilde{X}_{0,m}) = \gamma_0 - \frac{1}{1 - \rho_m^2} [2\gamma_0 - \sigma^2 - \sigma_m^2 - 2Cov(\hat{X}_0, \hat{X}_{0,m})]. \quad (2.28)$$

Pour m assez grand, les formules (2.27) et (2.28) deviennent

$$\rho = \lim_{m \rightarrow \infty} \rho_m = 1 - \left(\frac{\gamma_0}{\sigma^2} \right) \sum_{k=1}^{\infty} a_k b_k, \quad (2.29)$$

donc

$$\lim_{m \rightarrow \infty} \alpha_m = \lim_{m \rightarrow \infty} \beta_m = (1 + \rho)^{-1} \quad (2.30)$$

et

$$Var(X_0 - \tilde{X}_{0,\infty}) = \gamma_0 - \frac{2\sigma^2}{1 - \rho^2} \left(\frac{\gamma_0}{\sigma^2} - 1 - \rho \right). \quad (2.31)$$

La formule (2.31) permet de comparer la performance de l'interpolation suboptimale avec la solution optimale obtenue dans le théorème 2.1 et donne une condition nécessaire et suffisante en fonction des paramètres a_k , b_k et ρ pour que la solution suboptimale soit optimale.

Théorème 2.6. (Pourahmadi, 1991).

Soit $\{X_t\}$ un processus stationnaire inversible (minimal) de représentation (2.1). Alors, la solution suboptimale $\tilde{X}_{0,\infty}$ est optimale si et seulement si

$$\sum_{k=0}^{\infty} b_k^2 = \frac{1 - \rho^2}{1 + \rho^2} \left[\frac{2}{1 - \rho} - \left(\sum_{k=0}^{\infty} a_k^2 \right)^{-1} \right]. \quad (2.32)$$

Exemple 5. Soit un ARMA(1,1) d'équation

$$X_t - \phi X_{t-1} - \theta \epsilon_{t-1} = \epsilon_t \quad \text{avec } |\phi| < 1, |\theta| < 1 \quad \text{et } \phi \neq \theta.$$

Notons que

$$\begin{aligned} b_0 &= 1, & b_j &= (\phi - \theta)\phi^{j-1}, & \sum_{j=1}^{\infty} b_j^2 &= 1 + \frac{(\theta - \phi)^2}{1 - \phi^2}, \\ a_0 &= 1, & a_j &= (\phi - \theta)\phi^{j-1}, & \sum_{j=1}^{\infty} a_j^2 &= 1 + \frac{(\theta - \phi)^2}{1 - \theta^2}, \\ \sum_{j=1}^{\infty} a_j b_j &= \frac{(\phi - \theta)^2}{1 - \phi\theta}, & \rho &= \frac{\phi^2 - \phi\theta}{1 - \theta\phi}. \end{aligned}$$

Pour $\phi = 0$ ou $\theta = 0$, l'équation (2.32) est vérifiée, par conséquent, pour un AR(1) ou MA(1), toute solution suboptimale est optimale.

2.8 Proximité de Pitman

Une approche alternative à la théorie de la décision standard a été développée par Pitman (1937). Afin de comparer deux estimateurs δ_1 et δ_2 du paramètre inconnu θ , il a proposé de comparer les distributions de leurs distances (ou proximité) à θ , soit,

$$P \{ \|\delta_1(x) - \theta\| < \|\delta_2(x) - \theta\| \}. \quad (2.33)$$

Si cette probabilité est plus grande que 0.5, δ_1 domine δ_2 au sens du Pitman, avec le message explicite que δ_1 devrait alors être préféré à δ_2 .

L'avantage de ce critère est de se passer de la condition d'existence des moments, condition souvent lourde ou difficile à vérifier.

Ce critère a été largement utilisé dans les deux dernières décennies, pour comparer de nombreux estimateurs, pour exemple, nous citons, Mason et al (1990), Peddada et Khattree (1991), Fountain et Keating (1994), Liqing Yan (2007), Sen et al (1990), Bose (1998). Saleh et Shen (1991) ont généralisé ce critère en introduisant le GPC. Certains de ces auteurs ont utilisé la notion de biais par rapport à la médiane pour caractériser la classe des estimateurs optimaux au sens de Pitman. Quoique formellement semblable à la domination stochastique, ce critère, dit de proximité, présente des défauts majeurs, Certains articles étudient les propriétés de la proximité de Pitman et mettent en évidence son caractère intrinsèque, puisqu'elle fait intervenir la distribution complète de $\|\delta_1(x) - \theta\|$ (par opposition à l'évaluation réductrice à travers une fonction de coût, quadratique par exemple). A l'opposé, Robert et al (1993b) exposent les défauts fondamentaux de ce critère. Nous présentons deux points caractéristiques.

Sans conteste, le point le plus criticable et le plus critiqué de la proximité de Pitman est la non transitivité. De fait, ce critère ne fournit pas le moyen de déterminer un estimateur optimal ou même de comparer des estimateurs entre eux. Pitman (1937) avait déjà remarqué cette difficulté, mais certains partisans de ce critère (voir notamment Blynth 1972) affirment de manière paradoxale que cette propriété est un avantage supplémentaire puisqu'elle reflète la complexité du monde, il peut effectivement arriver qu'un ordre de préférence raisonnable ne soit pas toujours transitif. Mais le besoin aigu de réduire une telle complexité mis à part, notons que la proximité est mise en avant comme un critère de comparaison, une alternative aux fonction de coûts usuelles lorsque il y a non transitivité, l'ordre déduit de ce critère n'est pas absolu, puisque, comme le montre beaucoup d'exemples, il y a toujours une possibilité d'obtenir un cycle de préférence. Dans de tels cas, ce critère ne peut pas fournir d'estimateur optimal.

Bien entendu, la non transitivité de ce critère l'empêche d'être équivalent à une fonction de coût, à ce titre, il ne peut relever de la théorie de la décision. Pour la même raison, il ne peut pas être équivalent à la domination stochastique. En fait, Blyt et Pathak (1985) fournissent un exemple où ces deux critères produisent des ordres opposés. Il est même impossible de définir un estimateur de Bayes (de décision) pour le critère de Pitman (bien qu'un estimateur à posteriori de Pitman puisse exister, voir Bose (1992) et Ghost et al (1993).

Dans ce qui suit, nous allons appliquer l'idée de Pitman pour comparer deux estimateurs de données manquantes d'une série chronologique. Nous soulignons que, dans la théorie de l'estimation des paramètres, la quantité à estimer est inconnue mais fixe, par contre, dans le cas des données manquantes la quantité à estimer est aléatoire.

Définition 2.1. Soit \tilde{X}_k et \hat{X}_k deux estimateurs d'une donnée manquante d'une série chronologique. Alors, \tilde{X}_k domine \hat{X}_k au sens de Pitman si et seulement si

$$P \left\{ |X_k - \tilde{X}_k| < |X_k - \hat{X}_k| \right\} > \frac{1}{2}. \quad (2.34)$$

Théorème 2.7. (Hamaz et Ibazizen, 2009).

Soit $\{X_t\}$ un processus autorégressif d'ordre 1, stationnaire, d'équation $X_t = \phi X_{t-1} + \epsilon_t$ avec $\epsilon_t \sim \mathcal{N}(0,1)$ et supposons que X_k est manquante. Alors, la valeurs interpolé $\tilde{X}_k = \frac{\phi}{1 + \phi^2}(X_{k+1} + X_{k-1})$ domine la valeur prédite $\hat{X}_k = \phi X_k$ au sens de Pitman.

Démonstration. Soit $p = P \left\{ |X_k - \tilde{X}_k| < |X_k - \hat{X}_k| \right\}$

on a

$$\begin{aligned} p &= P \left\{ \left| X_k - \frac{\phi}{1 + \phi^2}(X_{k-1} + X_{k+1}) \right| < |X_k - \phi X_{k-1}| \right\} \\ &= P \left\{ \frac{1}{1 + \phi^2} |X_k - \phi X_{k-1} + \phi^2 X_k - \phi X_{k+1}| < |X_k - \phi X_{k-1}| \right\} \\ &= P \left\{ \frac{1}{1 + \phi^2} |\epsilon_k - \phi \epsilon_{k-1}| < |\epsilon_k| \right\} \\ &= P \left\{ \left| 1 - \phi \frac{\epsilon_{k+1}}{\epsilon_k} \right| < 1 + \phi^2 \right\} \\ &= P \left\{ -2 - \phi^2 < -\phi \frac{\epsilon_{k+1}}{\epsilon_k} < \phi^2 \right\} \end{aligned}$$

Les variables ϵ_k et ϵ_{k+1} sont indépendantes de loi $\mathcal{N}(0,1)$, ce qui entraîne, $\frac{\epsilon_{k+1}}{\epsilon_k}$ suit une loi de Cauchy(0,1). Nous distinguons deux cas:

Cas 1. $\phi > 0$

$$\begin{aligned} p &= P\left\{\frac{-2-\phi^2}{\phi} < -\frac{\epsilon_{k+1}}{\epsilon_k} < \phi\right\} \\ &= P\left\{-\phi < \frac{\epsilon_{k+1}}{\epsilon_k} < \frac{2+\phi^2}{\phi}\right\} \\ &= \frac{1}{\pi} \left(\arctan\left\{\frac{2+\phi^2}{\phi}\right\} - \arctan\{-\phi\} \right) \\ &> \frac{1}{\pi} \left(\arctan\left\{\frac{1}{\phi}\right\} + \arctan\{\phi\} \right) \\ &\geq \frac{1}{2}. \end{aligned}$$

Cas 2. $\phi < 0$

Il suffit de poser $\psi = -\phi$.

□

Chapitre 3

Prédiction à passé incomplet

3.1 Introduction

L'estimation des paramètres d'un processus ARMA stationnaire dans le cas de valeurs manquantes a été intensivement relaté dans la littérature, beaucoup de ces recherches sont dues à Parzen (1983) et Brockwell et Davis (1991), alors que la prédiction à passé incomplet n'a émergé que ces dernières années.

Cette partie du travail se distingue fondamentalement du premier chapitre, mais n'en est pas moins son complément naturel. Nous avons jusqu'ici traité le problème de prédiction en se basant sur un passé complet. Dans ce chapitre, nous donnons une formule récurrente pour le calcul des prédicteurs à passé incomplet en résolvant simultanément le problème des données manquantes. Notons que ce dernier problème a été traité indépendamment du premier par plusieurs auteurs et la solution proposée exige des techniques avancées d'inversion de matrices (voir chapitre 2). L'idée est de subdiviser les sous espaces passé et futur d'une manière analogue à celle de l'algorithme des innovation dans Brockwell et Davis (1992). Cette idée peut être expliquée et mieux appréciée en essayant de prédire X_2 en se basant sur tout le passé $\{X_1, X_{-1}, \dots\}$ à l'exception de X_0 où X_1 n'est pas orthogonal à \mathcal{H}_{-1} . Pour $k \geq 0$, soit \hat{X}_k la projection orthogonale de X_k sur \mathcal{H}_{-1} , alors $X_1 - \hat{X}_1$ et \mathcal{H}_{-1} sont orthogonaux et $\mathcal{H}_1^* = \mathcal{H}_{-1} \oplus \overline{\text{sp}}\{X_1 - \hat{X}_1\}$. Sous cette dernière forme, le calcul des projections de X_0 et X_1 sur H_1^* devient plus simple.

Cette même technique est utilisée quand plusieurs valeurs sont manquantes. En effet, soit, $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ les variables observées après l'instant $t = 0$, les estimateurs des valeurs manquantes et les prévisions futures s'obtiennent par les projections futures sur le sous espace $\mathcal{H}_{-1} \oplus \{X_{t_1} \oplus \hat{X}_{t_1}, X_{t_2} \oplus \hat{X}_{t_2}, \dots, X_{t_n} \oplus \hat{X}_{t_n}\}$. L'algorithme proposé est récursif et offre une méthode satisfaisante pour le calcul des prédicteurs et des estimateurs des valeurs

manquantes. Cet algorithme est général et peut être appliqué même dans le cas multivarié. Ce chapitre est organisé comme suit : la section 2 est une généralisation de l'algorithme d'innovation. Une formule explicite de la variance de l'erreur de prédiction est donnée dans la section 3. Dans la section 4, Nous nous intéressons à la représentation autorégressive des prédicteurs à passé incomplet. La dernière section de ce chapitre porte sur l'influence des positions des données manquantes sur la variance de l'erreur de prédiction.

3.2 Généralisation de l'algorithme d'innovation

Soit \mathcal{H} l'espace de Hilbert des variables aléatoires centrées de variances finies et $\{Y_t; t \geq 1\}$ une série chronologique centrée de fonction d'autocovariance $c(i, j) = \text{cov}(X_i, X_j)$. Pour $n \geq 0$, soit $\mathcal{H}_n = \overline{\text{sp}}\{\mathcal{H}, Y_1, \dots, Y_n\}$ l'espace de toutes les combinaisons linéaires d'éléments de $\mathcal{H}, Y_n, \dots, Y_1$, et \hat{Y}_{n+1} la projection orthogonale de Y_{n+1} sur \mathcal{H}_n avec $\nu_n = \text{var}(Y_{n+1} - \hat{Y}_{n+1})$ la variance de l'erreur de prédiction. Comme $Y_1 - \hat{Y}_1, \dots, Y_n - \hat{Y}_n$ sont orthogonales et orthogonales à \mathcal{H} alors

$$\mathcal{H}_n = \mathcal{H} \oplus \text{sp}\{Y_1 - \hat{Y}_1\} \oplus \dots \oplus \text{sp}\{Y_n - \hat{Y}_n\} \quad (3.1)$$

et donc

$$\hat{Y}_{n+1} = \tilde{Y}_{n+1} + \sum_{j=1}^n \theta_{nj}(Y_{n+1-j} - \hat{Y}_{n+1-j}), \quad (3.2)$$

où \tilde{X} est la projection de X sur \mathcal{H} .

Dans ce qui suit, nous donnons un schéma récursif pour calculer $\{\theta_{nj}, j = 1, \dots, n; n = 1, 2, \dots\}$, qui se ramène à l'algorithme d'innovation en posant $\mathcal{H} = \{0\}$.

Pour $n \geq 1$, les coefficients de \hat{Y}_{n+1} ainsi que les erreurs quadratiques moyennes sont donnés par :

$$\begin{aligned} \nu_0 &= \text{var}(Y_1 - \hat{Y}_1) \\ \theta_{n, n-k} &= \nu_k^{-1} [c(n+1, k+1) - d(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} \nu_j], \quad k = 0, 1, \dots, n-1, \\ \nu_n &= c(n+1, n+1) - d(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n, n-j} \nu_j \end{aligned} \quad (3.3)$$

où $d(n+1, k+1) = \text{cov}(Y_{n+1}, \hat{Y}_{k+1})$. Les coefficients de (3.3) s'obtiennent dans l'ordre suivant : $\nu_0, \theta_{11}, \nu_1, \theta_{22}, \theta_{21}, \nu_2, \dots$

En utilisant (3.2) et l'orthogonalité des $Y_j - \hat{Y}_j, 1 \leq j \leq n$, on obtient

$$\theta_{n, n-k} = \frac{\text{cov}(Y_{n+1}, Y_{k+1} - \hat{Y}_{k+1})}{\text{var}(Y_{k+1} - \hat{Y}_{k+1})}. \quad (3.4)$$

En pratique, pour une série chronologique centrée $\{X_t\}$, \mathcal{H} est considéré habituellement comme segment d'observations passées, à savoir $\mathcal{H} = \overline{\text{sp}}\{X_{-N}, \dots, X_{-1}\}$ N fini ou infini. Supposons que $\{X_t\}$ est stationnaire, de moyenne nulle, de fonction d'autocovariance $\gamma_k = \text{cov}(X_{t+k}, X_t)$ et de représentations MA et AR

$$\begin{aligned} X_t &= \epsilon_t + b_1 \epsilon_{t-1} + \dots \\ X_t &= \epsilon_t + a_1 X_{t-1} + a_2 X_{t-2} + \dots \end{aligned} \quad (3.5)$$

avec $\sigma^2 = \text{var}(\epsilon_t)$, $\{b_k\}$ et $\{a_k\}$ sont les paramètres des représentations MA et AR respectivement. Ces paramètres vérifient les relations de récurrence suivantes

$$b_l = \sum_{k=0}^{l-1} b_k a_{l-k}, \quad l = 1, 2, \dots, b_0 = 1. \quad (3.6)$$

Pour $k \geq 0$, le meilleur prédicteur linéaire de X_k basé sur le passé $\{X_t; t \leq -1\}$ est la projection orthogonale de X_k sur $\overline{\text{sp}}\{X_t; t \leq -1\}$:

$$\hat{X}_k = \sum_{i=k+1}^{\infty} b_i \epsilon_{k-i}, \quad e_k = X_k - \hat{X}_k = \sum_{i=0}^k b_i \epsilon_{k-i} \quad (3.7)$$

$$c(k, l) = \text{cov}(e_k, e_l) = \sigma^2 \sum_{i=0}^{\min(k, l)} b_i b_{k-i}, \quad k, l \geq 0. \quad (3.8)$$

Notons que \mathcal{H} est orthogonal à $\{e_k, k \geq 1\}$ qui est un processus centré non stationnaire. Il convient donc de l'orthogonaliser.

Soit $K = \{t_1, \dots, t_m\}$ le sous ensemble des indices des variables observées. En se plaçant dans le cas $Y_t = e_t, t \geq 0$ et en utilisant (3.2) et (3.8) on obtient

$$\begin{aligned} \hat{e}_0 &= 0 & \nu_0 &= \text{var}(e_0 - \hat{e}_0) = \sigma^2, & e_0 - \hat{e}_0 &= \epsilon_0, \\ \theta_{11} &= c(1, 0) \nu_0^{-1} = b_1, & \nu_1 &= \sigma^2, & e_1 - \hat{e}_1 &= \epsilon_1, \\ \theta_{22} &= c(0, 2) \nu_0^{-1} = b_2, \theta_{21} = b_1, & \nu_2 &= \sigma^2, & e_2 - \hat{e}_2 &= \epsilon_2. \end{aligned}$$

Par déduction, pour tout $n \geq 1, j = 0, 1, \dots, n - 1$, nous avons

$$\theta_{n,n-j} = b_{n-j}, \quad \nu_0 = \sigma^2, \quad e_n - \hat{e}_n = \epsilon_n. \quad (3.9)$$

Pour simplifier les notations et sans perte de généralités, indexons la série chronologique de telle sorte à avoir X_0 première observation manquante et soit $K = \{t_i, 0 < t_1 < t_2 < \dots < t_m \leq \infty\}$ l'ensemble des indices des variables observées après X_0 . Considérons la série chronologique $\{Y_t; 1 \leq t \leq m\}$ avec Y_t le prédicteur de l'erreur $e_{t_i}, t_i \in K$. Les éléments de la matrice de covariance de $\{Y_t\}$ sont donnés par $\text{cov}(e_{t_i}, e_{t_j})$ dans (3.8). En utilisant (3.2)-(3.3), on obtient $\{X_{t_i} - \hat{X}_{t_i}, t_i \in K\} = \{Y_1, \dots, Y_m\}$ avec $\tilde{Y}_k = 0$ et $d(k, l) = 0$ pour tout $k, l \geq 0$.

Afin de calculer le prédicteur basé sur le passé incomplet, pour $0 \leq n \notin K$, notons \hat{X}'_n la projection orthogonale de X_n sur \mathcal{H}_{t_k} où t_k est le plus grand entier tel que $t_k \leq n - 1$. En utilisant (3.1), on obtient

$$\begin{aligned} \hat{X}'_n &= \hat{X}_n + \sum_{t_j \leq n-1} \beta_j (Y_j - \hat{Y}_j), & \beta_j &= \frac{\text{cov}(X_n, Y_j - \hat{Y}_j)}{\nu_j}, \\ X_n - \hat{X}'_n &= e_n - \sum_{t_j \leq n-1} \beta_j (Y_j - \hat{Y}_j), \\ \text{var}(X_n - \hat{X}'_n) &= \text{var}(e_n) + \sum_{t_j \leq n-1} \beta_j^2 \nu_j, \end{aligned} \quad (3.10)$$

où β_j et ν_j se calculent en utilisant les relations de récurrences données par (3.3).

Remarque 3.1. Sans recourir à la représentation espace d'états, l'algorithme d'innovation généralisé est très utile pour calculer les projections sur les espaces de dimensions finies ou infinies. Il se réduit à l'algorithme du filtre de Kalman ou à la procédure de Gram-Schmidt lorsque $\mathcal{H} = \{0\}$. Cet algorithme possède un avantage de taille puisqu'il permet d'expliciter les variances des prédicteurs et des interpolateurs d'un processus stationnaire plus général qu'un ARMA fini. L'outil clé pour calculer ces projections est basé sur la décomposition de Wold des processus stationnaires. Lorsque les covariances sont inconnues, l'algorithme proposé peut être mis en œuvre en utilisant des estimations de covariances comme dans Brockwell et Davis (1991, Chs. 5, 8).

Exemple 6. Pour une prédiction et une interpolation basées sur $\{X_m, \dots, X_1, X_{-1}, X_{-2}, \dots\}$. Nous avons besoin d'orthogonaliser $\{e_1, \dots, e_m\}$. En utilisant l'algorithme d'innovation généralisé on obtient :

$$\hat{e}_1 = 0, \quad e_1 - \hat{e}_1 = \epsilon_1 + b_1 \epsilon_0, \quad \nu_0 = \text{var}(e_1) = \sigma^2(1 + b_1^2),$$

$$\theta_{11} = c(1,2), \quad \nu_0^{-1} = (b_1 + b_1 b_2)(1 + b_1^2), \quad \nu_1 = \sigma^2 \frac{1 + b_1^2 + (b_2 - b_1^2)}{1 + b_1^2}$$

En utilisant la relation entre les coefficients des représentations MA et AR, on obtient $\nu_1 = \frac{1 + a_1^2 + a_2^2}{1 + a_1^2} \sigma^2$. Les autres paramètres θ_{nj} et ν_m se calculent identiquement,

$$\nu_m = \sigma^2 \frac{1 + a_1^2 + \dots + a_m^2 + a_{m+1}^2}{1 + a_1^2 + \dots + a_m^2} = \text{var}(X_{m+1} - \hat{X}'_{m+1}). \quad (3.11)$$

La croissance de la variance de l'erreur de prédiction X_{m+1} est due à l'observation manquante X_0

$$I_m = \text{var}(X_{m+1} - \hat{X}'_{m+1}) - \text{var}(X_{m+1} - \hat{X}_{m+1}) = \sigma^2 a_{m+1}^2 \left(\sum_{k=0}^m a_k^2 \right)^{-1}, \quad a_0 = 1.$$

Cette quantité est nulle, par exemple, pour un AR(p) et $m \geq p$. La variance de l'erreur d'interpolation de X_0 est

$$\text{var}(X_0 - \hat{X}'_0) = \sigma^2 \left(\sum_{k=1}^m a_k^2 \right)^{-1}. \quad (3.12)$$

3.3 Erreur de prévision à passé incomplet

Dans ce qui suit, nous supposons que les données $X_{-n_1}, \dots, X_{-n_N}$ sont manquantes, où N est un entier positif tel que $0 < n_1 < \dots < n_N$, et nous considérons le problème de prédire X_0 en se basant sur le passé incomplet

$$\mathcal{H}'_{-1} = \overline{\text{sp}}\{X_t; t \leq -1, t \neq -n_1, \dots, -n_N\}.$$

Soient $M = \{n_1, n_2, \dots, n_N\}$ l'ensemble des indices des valeurs manquantes et $P'_{-1}Y$ est la projection orthogonale de Y sur \mathcal{H}'_{-1} .

Dans le cas où $n_1 = 1, \dots, n_N = N$, X'_0 est le prédicteur d'horizon $N + 1$ qui se calcule facilement en utilisant la décomposition de Wold (Voir Chapitre 1). Quand les indices des valeurs manquantes sont quelconques, Chang et Pouramadi (1997) ont proposé un algorithme pour le calcul de \hat{X}'_0 , cette procédure utilise essentiellement la généralisation de l'algorithme d'innovation exposé dans la section précédente. L'expression résultante de cette méthode dépend seulement des paramètres de la représentation moyenne mobile (MA) obtenue par la décomposition de Wold et de la variance des innovations. La complexité

algorithmique de cette méthode est liée à l'entier N qui est le nombre d'observations manquantes.

En utilisant le résultat de Grenander et Rosenblatt (1954), une expression de $\text{var}(X_0 - \hat{X}'_0)$ qui utilise principalement les paramètres de la représentation AR de $\{X_t\}$ est donnée dans Pourahmadi (1994). Dans ce qui suit, nous donnons une expression explicite de $X_0 - \hat{X}'_0$. Cette expression permet d'obtenir la variance des erreurs et permet de déduire une représentation autorégressive du prédicteur \hat{X}'_0 . Le calcul de \hat{X}'_0 exige l'inversion d'une matrice dont la taille dépend du nombre d'observations manquantes N , mais indépendantes du rang de ces indices, les éléments de cette matrice dépendent seulement de la représentation autorégressive (AR) de $\{X_t\}$. Nous caractérisons également les processus pour lesquels la perte d'observations dans le passé n'effectent pas la prédiction, i.e. $\hat{X}_0 = \hat{X}'_0$.

Théorème 3.1. (Bondon, 2002).

Soit $\{X_t\}$ un processus non déterministe stationnaire admettant une représentation autorégressive $AR(\infty)$ de paramètres $\{a_k\}$ donnée par

$$X_k = \sum_{i=1}^{\infty} a_i X_{k-i} + \epsilon_k \quad (3.13)$$

Alors

$$X_0 - \hat{X}'_0 = -\sum_{p=0}^N \psi_p \sum_{j=0}^{n_p} a_{n_p-j} \epsilon_{-j}, \quad (3.14)$$

où les coefficients (ψ_p) satisfont les équations matricielles suivantes

$$U(\psi_0, \psi_1, \dots, \psi_N)' = (1, 0, \dots, 0)', \quad (3.15)$$

U est une $(N+1) \times (N+1)$ matrice non singulière d'éléments

$$U_{p,q} = \sum_{j=0}^{n_p \wedge n_q} a_{n_p-j} a_{n_q-j}, \quad p, q = 0, \dots, N. \quad (3.16)$$

La variance de l'erreur de prédiction est

$$\text{var}(X_0 - \hat{X}'_0) = \sigma^2 \psi_0. \quad (3.17)$$

Démonstration. Montrons d'abord que $X_0 - \hat{X}'_0$ est orthogonal à \mathcal{H}'_{-1} . Soit $k \in \mathbb{N}$, nous

avons

$$\begin{aligned}
 \langle X_0 - \hat{X}'_0, X_{-k} \rangle &= - \sum_{p=0}^N \psi_p \sum_{j=0}^{n_p} a_{n_p-j} \langle \epsilon_{-j}, X_{-k} \rangle \quad \text{De (3.14),} \\
 &= -\sigma^2 \sum_{p=0}^N \psi_p \sum_{j=k}^{n_p} a_{n_p-j} c_{j-k} \quad \text{De (3.15)} \quad (3.18) \\
 &= \sigma^2 \sum_{p=0}^N \psi_p \delta_{n_p-k}.
 \end{aligned}$$

Si $k \notin M$, la dernière égalité de (3.18) est nulle. Par conséquent, $X_0 - \hat{X}'_0 \perp \mathcal{H}'_{-1}$. Maintenant, vérifions que $\hat{X}'_0 \in \mathcal{H}'_{-1}$. Pour $j \in \{0, 1, \dots, n_N\}$, on a

$$\begin{aligned}
 X_{-j} &= \sum_{i=0}^{n_N-j} c_i \epsilon_{-j-i} + \sum_{i=n_N-j}^{\infty} c_i \epsilon_{-j-i} + V_{-j} \\
 &= \sum_{i=j}^{n_N} c_{i-j} \epsilon_{-i} + U_{-j}
 \end{aligned} \quad (3.19)$$

où $U_{-j} \in \mathcal{H}_{-n_N-1}$. Soient X , ϵ et U , les vecteurs de composantes X_{-j}, ϵ_{-j} et U_{-j} pour $j = 0, 1, \dots, n_N$. Il résulte de (3.19) que $X = C\epsilon + U$ où C est une $(n_N+1) \times (n_N+1)$ matrice triangulaire sup d'éléments $C_{p,q} = c_{q-p}$ pour $p, q = 0, \dots, n_N$. Pour $c_0 = 1$, la matrice C est non singulière et il s'en suit que $C^{-1} = -A$ où A est une matrice triangulaire sup d'éléments $A_{p,q} = a_{q-p}$ pour $p = 0, 1, \dots, n_N$.

Soit $T = AU$ et notons par T_{-j} les composantes de T . Nous avons

$$\epsilon_{-j} = - \sum_{i=j}^{n_N} a_{i-j} X_{-i} + T_{-j}, \quad j = 0, \dots, n_N \quad (3.20)$$

En insérant (3.20) dans (3.14), on obtient $\hat{X}'_0 = R_0 + S_0$ où

$$R_0 = X_0 - \sum_{p=0}^N \psi_p \sum_{j=0}^{n_p} a_{n_p-j} \sum_{i=j}^{n_N} a_{i-j} X_{-i} \quad (3.21)$$

$$S_0 = \sum_{p=0}^N \psi_p \sum_{j=0}^{n_p} a_{n_p-j} T_{-j} \quad (3.22)$$

Soit $i \in M$, i.e. $i = n_q$ où $q \in \{0, 1, \dots, N\}$ et soit $\alpha_{n_q} = 0$. On a $R_0 \in \text{sp}\{X_{-k}; k \in \{0, \dots, n_N\} \setminus M\} \subset \mathcal{H}'_{-1}$.

Comme chaque $U_{-j} \in \mathcal{H}_{-n_N-1}$ et $T = AU$, chaque $T_{-j} \in \mathcal{H}_{-n_N-1}$, il en résulte de (3.22)

que $S_0 \in \mathcal{H}'_{-1}$. D'où $\hat{X}'_0 = R_0 + S_0 \in \mathcal{H}'_{-1}$.

Montrons maintenant que la matrice U est définie positive et donc a fortiori U est non singulière. Soit $\alpha \in \mathbb{R}^{N+1}$, on déduit de (3.16) que

$$\alpha'U\alpha = \sum_{j=0}^{n_N} \left(\sum_{p=0}^N \alpha_p a_{n_p-j} \right)^2.$$

Donc, $\alpha'U\alpha = 0$ est équivalent à $\sum_{p=0}^N \alpha_p a_{n_p-j} = 0$ pour $j = 0, \dots, n_N$. En prenant successivement $j = n_N, n_{N-1}, \dots, n_0$ et en utilisant le fait que $a_0 = -1$, on obtient que $\alpha_N = \alpha_{N-1} = \dots = \alpha_0 = 0$. Finalement, l'expression de $\text{var}(X_0 - \hat{X}'_0)$ est tirée de (3.18) avec $k = 0$:

$$\text{var}(X_0 - \hat{X}'_0) = \langle X_0 - \hat{X}'_0, X_0 \rangle = \sigma^2 \sum_{p=0}^N \psi_p \delta_{n_p} = \sigma^2 \psi_0.$$

□

Remarque 3.2.

- a. La formule (3.17) s'obtient aussi en utilisant Pouramadi (1994, théorème 1). Si $\hat{X}'_0 = X_0$, alors $X_0 \in \mathcal{H}'_{-1} \subset \mathcal{H}_{-1}$, et donc $\{X_k\}$ est déterministe. Inversement, si $\{X_t\}$ est déterministe, alors $X_0 \in \mathcal{H}_{-n_{N-1}} \subset \mathcal{H}'_{-1}$, et donc $\hat{X}'_0 = X_0$. Par conséquent, $\text{var}(X_0 - \hat{X}'_0) > 0$ si et seulement si $\sigma^2 > 0$.
- b. Supposons que $n_1 = 1, \dots, n_N = N$, alors \hat{X}'_0 est le prédicteur d'horizon h de X_0 . Il s'en suit de (3.16) que $U = AA'$, et comme $C = -A^{-1}$, (3.16) est équivalente à

$$A(\psi_0, \dots, \psi_N)' = -(c_0, c_1, \dots, c_N)'. \tag{3.23}$$

On déduit de (3.14) et (3.23) que

$$X_0 - \hat{X}'_0 = -\sum_{j=0}^N \left(\sum_{p=0}^N \psi_p a_{p-j} \right) \epsilon_{-j} = \sum_{j=0}^N c_j \epsilon_{-j},$$

qui est une formule connue (voir Brockwell and Davis 1991, p. 189).

Comme le processus d'innovation $\{\epsilon_k\}$ n'est pas directement observable, la formule (3.14) ne peut pas être utilisée pour le calcul de \hat{X}'_0 à moins qu'on puisse exprimer ϵ_k en fonction des observations X_{k-i} pour $i \geq 0$. Ce qui est équivalent à trouver une

représentation autorégressive convergente en moyenne quadratique de \hat{X}'_0 dans l'espace temps. Le théorème suivant établit cette relation.

Théorème 3.2. (*Bondon, 2002*).

Soit $\{X_t\}$ un processus purement non déterministe, stationnaire de représentation AR de paramètres $\{a_k\}$. Le prédicteur \hat{X}'_0 admet une représentation autorégressive si et seulement si $\{X_k\}$ admet une représentation autorégressive. Dans ce cas, la représentation AR de \hat{X}'_k est unique et est donnée par

$$\hat{X}'_0 = \sum_{k \in \mathbb{N} \setminus M} h_k X_{-k}, \quad (3.24)$$

où

$$h_k = \delta_k - \sum_{p=0}^N \psi_p \sum_{j=0}^{n_p \wedge k} a_{n_p-j} a_{k-j}, \quad k \in \mathbb{N}. \quad (3.25)$$

Les coefficients (ψ_p) sont définis dans le théorème précédent.

Démonstration. Premièrement, notons que la représentation AR de \hat{X}'_0 existe, et elle est unique car $\{X_t\}$ est non déterministe et $\sum_{k=0}^{\infty} g_k X_{-k} = 0$, alors $g_k = 0$ pour tout $k \geq 0$.

Maintenant, si \hat{X}'_0 admet une représentation AR, alors $\hat{X}'_0 = \sum_{i=1}^{\infty} a_i X_{-i}$. La stationnarité

nous assure que $\hat{X}'_k = \sum_{i=1}^{\infty} a_i X_{k-i}$ pour tout $k \in \mathbb{Z}$.

Inversement, si X_k admet une représentation autorégressive, en remplaçant ϵ_{-j} dans (3.14) par son expression déduite de (3.13), nous obtenons

$$\hat{X}'_0 = X_0 - \sum_{k=0}^{\infty} \left(\sum_{p=0}^N \psi_p \sum_{j=0}^{n_p \wedge k} a_{n_p-j} a_{k-j} \right) X_{-k} = \sum_{k=0}^{\infty} h_k X_{-k},$$

où les coefficients (h_k) sont définis par (3.25). De (3.15) et (3.16), on obtient $h_k = 0$ pour tout $k \in M$, ce qui entraîne (3.24). \square

Remarque 3.3.

a. Sous les deux conditions, qui imposent à la densité spectrale f de $\{X_t\}$ d'être bornée presque partout et à f^{-1} d'être intégrable, Masani (1960, théorème 5.2) démontre que le prédicteur d'horizon $N+1$ de X_0 admet la représentation autorégressive (3.24) avec

$$h_k = \begin{cases} 0 & \text{si } 0 \leq k \leq N \\ - \sum_{j=N+1}^k c_j a_{k-j} & \text{si } k \geq N+1. \end{cases} \quad (3.26)$$

Bloomfield (1985, théorème 1) montre que si \hat{X}_0 admet une représentation AR, alors quelque soit $N > 0$, le prédicteur linéaire de X_0 en fonction des observations $\{X_k; k < -N\}$ admet aussi une représentation AR que l'on obtient au moyen de relations données par exemple dans Bondon 2002.

Cheng et Pourahmadi (1997) ont proposé une méthode récursive pour calculer le prédicteur linéaire \hat{X}'_0 de X_0 en fonction de toutes les observations du passé à l'exception d'un nombre fini d'entre elles dont les indices temporels sont quelconques. La complexité de leur algorithme dépend des indices des observations manquantes et leur méthode ne fournit pas la représentation AR de \hat{X}'_0 .

b. La représentation (3.24) a été démontrée dans Bondon (2000, Corollaire1) en imposant deux conditions différentes :

1. f est bornée presque sûrement et f^{-1} est intégrable.
2. L'angle entre le sous espace futur $\overline{\text{sp}}\{X_i, i \geq 1\}$ et le sous espace passé-présent \mathcal{H}_0 est positif.

c. D'après (3.25), $h_0 = 0$ et pour tout $k \geq 1$, nous avons

$$h_k = -\sum_{j=0}^{n_N} \alpha_j a_{k-j}, \quad (3.27)$$

où $\alpha_j = \sum_{p=0}^N \psi_p a_{n_p-j}$. Par conséquent, $h_k \leq \left(\sum_{j=0}^{n_N} \alpha_j^2 \right) \left(\sum_{j=0}^{n_N} a_{k-j}^2 \right)$ pour tout $k \in \mathbb{N}$, et si $\{X_k\}$ est minimal, nous avons

$$\sum_{k=0}^{\infty} h_k^2 \leq (n_N + 1) \left(\sum_{j=0}^{n_N} \alpha_j^2 \right) \sum_{k=0}^{\infty} a_k^2 < \infty.$$

(e.) Si $\sum_{k=0}^{\infty} |a_k| < \infty$, alors la série de (3.13) est convergente dans L_2 , voir Brockwell et Davis (1991, Proposition 3.1.1). D'où \hat{X}'_0 admet une représentation autorégressive donnée par (3.24), et on déduit de (3.27) que

$$\sum_{k=0}^{\infty} |h_k| \leq \left(\sum_{j=0}^{n_N} |\alpha_j| \right) \sum_{k=0}^{\infty} |a_k| < \infty.$$

De (3.17), la croissance de la variance de l'erreur de prédiction de X_0 est due aux données manquantes $X_{-n_1}, \dots, X_{-n_N}$, elle est égale à $\sigma^2(\psi_0 - 1)$.

Dans le théorème qui suit, les processus pour lesquels les données manquantes n'influent pas sur les prédicteurs sont caractérisés.

Théorème 3.3. (Bondon, 2002).

Soit $\{X_k\}$ un processus non déterministe stationnaire de représentation AR de paramètres (a_k) . Alors $\hat{X}'_0 = \hat{X}_0$ si et seulement si $a_{n_i} = 0$ pour $i = 1, \dots, N$.

Démonstration. Il résulte de (3.14) que $\hat{X}'_0 = \hat{X}_0$ si et seulement si

$$-\epsilon_0 = \left(\sum_{p=0}^N \psi_p a_{n_p} \right) \epsilon_0 + \sum_{j=1}^{n_N} (\psi_p a_{n_p-j}) \epsilon_{-j}$$

ce qui est équivalent

$$\sum_{p=0}^N \psi_p a_{n_p} = -1, \quad (3.28)$$

et

$$\sum_{p=1}^N \psi_p a_{n_p-j} = 0, \quad j = 1, \dots, n_N. \quad (3.29)$$

En prenant successivement $j = n_N, n_{N-1}, \dots, n_1$ dans (3.29) et en utilisant le fait que $a_0 = -1$, on déduit que (3.28) et (3.29) sont équivalentes à $(\psi_0, \psi_2, \dots, \psi_N) = (1, 0, \dots, 0)$. Comme la première colonne de U est $(1, -a_{n_1}, \dots, a_{n_N})'$, alors $(1, 0, \dots, 0)$ est solution de (3.15) si et seulement si $a_{n_i} = 0$ pour $i = 1, \dots, N$. \square

Remarque 3.4.

a. Sous l'hypothèse de la représentation AR donnée par (3.13), le théorème précédent se démontre facilement. En effet, si $\hat{X}'_0 = \hat{X}_0$, on déduit de (3.24) et de (3.13) que

$$\sum_{k \in \mathbb{N} \setminus M} h_k X_{-k}, \text{ ce qui entraîne } a_k = 0 \text{ pour tout } k \in M \setminus \{0\}.$$

Inversement, si $a_k = 0$ pour tout $k \in M \setminus \{0\}$, $\hat{X}'_0 = \sum_{k \in \mathbb{N} \setminus M} a_k X_{-k}$ et donc, $\hat{X}_0 \in \mathcal{H}'_{-1}$

et $X_0 - \hat{X}_0 \perp \mathcal{H}_{-1} \supset \mathcal{H}'_{-1}$. Par conséquent, $\hat{X}_0 = \hat{X}'_0$.

b. Soit $\mathcal{H}_{1, n_1-1} = \overline{\text{sp}}\{X_{-k}, 1 \leq k \leq n_1 - 1\}$ et \hat{X}_{0, n_1} la projection orthogonale de X_0 sur \mathcal{H}_{n_1-1} . Comme $K \subset \mathcal{H}'_{-1}$, nous avons $\text{var}(X_0 - \hat{X}'_0) \leq \text{var}(X_0 - \hat{X}_{0, n_1})$, et du fait que, $\hat{X}_0 - \hat{X}'_0 \in \mathcal{H}_{-1}$, $X_0 - \hat{X}_0 \perp \hat{X}_0 - \hat{X}'_0$ alors $\text{var}(X_0 - \hat{X}'_0) = \sigma^2 + \text{var}(\hat{X}_0 - \hat{X}'_0)$. D'où, $\text{var}(X_0 - \hat{X}'_0) \leq \text{var}(X_0 - \hat{X}_{0, n_1}) - \sigma^2$. Quand $n_1 \rightarrow \infty$, \hat{X}_{0, n_1} converge dans L_2 , et donc \hat{X}'_0 converge vers X_0 pour tout processus stationnaire.

Exemple 7. Supposons que seule la valeur X_{-m} , $m > 0$ est manquante. On déduit de (3.16) que

$$U = \begin{pmatrix} 1 & -a_m \\ -a_m & S_m \end{pmatrix}$$

où $S_m = \sum_{i=0}^m a_i^2$, et donc la solution de (3.15) est $(\psi_0, \psi_1) = S_{m-1}^{-1}(S_m, a_m)$. De (3.14) et (3.17), on obtient,

$$\begin{aligned} X_0 - \hat{X}_0 &= \epsilon_0 - \frac{a_m}{S_{m-1}} \sum_{j=1}^m a_{m-j} \epsilon_{-j}, \\ \text{var}(X_0 - \hat{X}'_0) &= \sigma^2 \left(1 + \frac{a_m^2}{S_{m-1}} \right). \end{aligned} \quad (3.30)$$

Si $\{X_k\}$ est purement non déterministe de représentation AR donnée par (3.13), on déduit du théorème précédent que \hat{X}'_0 admet une représentation AR donnée par (3.24) où

$$h_k = a_k - \frac{a_m}{S_{m-1}} \sum_{j=1}^{m \wedge k} a_{m-j} a_{k-j}, \quad k \geq 1.$$

Exemple 8. Supposons que $\{X_t\}$ est un processus autorégressif d'ordre r causal, d'équation

$$X_k = \sum_{i=1}^r a_i X_{k-i} + \epsilon_k, \quad 1 \leq r < \infty. \quad (3.31)$$

Alors

- (a) \hat{X}'_0 admet la représentation AR donnée par (3.14) avec $h_k = 0$ pour tout $k > n_N + r$,
- (b) $\hat{X}_0 = \hat{X}'_0 = \sum_{i=1}^r a_i X_{-i}$ si et seulement si $a_k = 0$ pour tout $k \in \{n_i, i = 1, \dots, N \setminus n_i \leq r\}$.

Comme $\{X_t\}$ est causal, $\{\epsilon_k\}$ est l'innovation du processus $\{X_t\}$ et (3.31) est la représentation AR de $\{X_k\}$. Par conséquent, (a) résulte du théorème (3.2) et (b) du théorème (3.3).

Remarque 3.5.

- a. Si $\{X_t\}$ est un AR(r) causal, alors \hat{X}'_0 est le meilleur prédicteur linéaire de X_0 basé sur le passé fini $\{X_{-k}, k \in \{1, \dots, n_N + r\} \setminus M\}$. Cette propriété généralise le résultat connu lorsque le passé est complet.
- b. Posons $r = 1$ dans (3.31), et soit $a = a_1$ avec $|a| < 1$. Si $n_1 > 1$, on a $\hat{X}'_0 = \hat{X}_0 = aX_{-1}$. Si $n_k = n$ pour $1 \leq k < j$ et $n_j > j$ pour tout j , nous avons $a^j X_{-j} \in \mathcal{H}'_{-1}$ et $X_0 - a^j X_{-j} = \sum_{k=0}^{j-1} a^k \epsilon_{-k} \perp \mathcal{H}_{-j} \supset \mathcal{H}'_{-1}$. $\hat{X}'_0 = a^j X_{-j}$ et $\text{var}(X_0 - \hat{X}'_0) = \sigma(1 - a^{2j})(1 - a^2)^{-1}$.

3.4 Représentation AR du prédicteur à passé infini incomplet

Dans cette section, la représentation autorégressive du meilleur prédicteur linéaire en moyenne quadratique d'une série chronologique stationnaire basé sur toutes les observations du passé à l'exception d'un nombre fini d'entre elles dont les indices temporels sont quelconques est établie. Ce résultat est obtenu sous les hypothèses suffisantes classiques d'existence d'une représentation autorégressive pour le prédicteur basé sur tout le passé.

Soit $(X_k)_{k \in \mathbb{Z}}$ une série chronologique centrée stationnaire au second ordre et purement non déterministe de densité spectrale f . On note \hat{X}_0 le meilleur prédicteur linéaire en moyenne quadratique de X_0 en fonction du passé $\{X_k, k \leq -1\}$.

Pour être calculable en pratique, \hat{X}_0 doit avoir une représentation autorégressive (AR) convergente en moyenne quadratique en fonction des $X_k, k \leq -1$. Deux conditions bien connues garantissent l'existence (et l'unicité) d'une telle représentation. Il s'agit de ($f \in L_\infty$ et $f^{-1} \in L_1$) et de $\theta > 0$ où θ est l'angle entre le sous espace "passé et présent" et le sous espace "futur" de (X_k) . Aucune de ces conditions n'implique l'autre.

Bloomfield (1985) a montré que si \hat{X}_0 admet une représentation AR alors quelque soit $N > 0$, le prédicteur linéaire de X_0 en fonction des observations $X_k; k < -N$ admet aussi une représentation autorégressive que l'on obtient au moyen de relations données par exemple dans Akutowicz (1957).

Cheng et Pourahmadi (1997) ont proposé une méthode récursive pour calculer le prédicteur linéaire \hat{X}'_0 de X_0 en fonction de toutes les observations du passé à l'exception d'un nombre fini d'entre elles dont les indices temporels sont quelconques. La complexité de leur algorithme dépend des indices des observations manquantes et leur méthode ne fournit pas la représentation AR de \hat{X}'_0 .

Dans la suite, $\overline{\text{sp}}\{X_k; k \in \mathbb{Z}\}$ est le domaine temporel de $\{X_t\}$ et $L_2(f)$ est son domaine fréquentiel. L'espace $L_2(f)$ est muni du produit scalaire

$$\langle g, h \rangle = \int_{-\pi}^{\pi} g(\lambda) \overline{h(\lambda)} d\lambda,$$

où $d\lambda$ est la mesure de Lebesgue normalisée sur $]-\pi, \pi]$. L'application $\tau : X_k \rightarrow e^{-ik\lambda}$ définit un isomorphe entre $\overline{\text{sp}}\{X_k, k \in \mathbb{Z}\}$ et $L_2(f)$. On désigne par L_2^{0+}, L_2^+ et L_2^{0-} les sous-espaces de L_2 des fonctions g dont les k èmes coefficients de Fourier $g_k = \int_{-\pi}^{\pi} g(\lambda) e^{-ik\lambda} d\lambda$ s'annulent respectivement pour $k < 0$, $k \leq 0$ et $k > 0$; on note $[g]_+$ et $[g]_{0-}$ les fonctions de L_2^+ et L_2^{0-} dont les k ème coefficient de Fourier sont g_k pour $k > 0$ et $k \leq 0$. Enfin, pour

tous entiers k et n , $k \wedge n$ désigne le minimum de k et n , $\delta_k = 0$ pour $k \neq 0$ et $\delta_0 = 1$.

Soit

$$\mathcal{M} = \{n_0, n_1, \dots, n_N\}, \quad 0 = n_0 < n_1 < \dots < n_N,$$

un ensemble fini d'entiers et \hat{X}'_0 la projection de X_0 sur $\overline{\text{sp}}\{X_{-k}, k \in \mathbb{N} \setminus \mathcal{M}\}$. Comme τ est une isométrie, $\tau(\hat{X}'_0)$ est la projection de la fonction constante 1 sur le sous espace \mathcal{H} de $L_2(f)$ défini par :

$$\mathcal{H} = \overline{\text{sp}}\{e^{-ik\lambda}; k \in \mathbb{N} \setminus \mathcal{M}\}. \quad (3.32)$$

Théorème 3.4. (Bondon, 2000).

Soit $\{X_k\}$ une série chronologique centrée stationnaire du second ordre purement non déterministe de densité spectrale f . Soient θ l'angle entre le sous-espace "futur et présent" et le sous-espace "futur" de (X_k) et (a_k) les paramètres AR de (X_k) avec $a_0 = 1$. Si $f \in L_\infty$ et $f^{-1} \in L_1$, ou si $\theta > 0$, la projection h de 1 sur le sous-espace \mathcal{H} de $L_2(f)$ défini par (3.32) est donnée par

$$h = 1 - \bar{\phi}[\phi\psi]_{0-}, \quad (3.33)$$

où

$$\phi(\lambda) = \sum_{k=0}^{\infty} a_k e^{ik\lambda}, \quad (3.34)$$

$$\psi(\lambda) = \sum_{k=0}^N \psi_k e^{-in_k\lambda} \quad (3.35)$$

et les coefficients (ψ_k) satisfont l'équation matricielle

$$U(\psi_0, \psi_1, \dots, \psi_N) = (1, 0, \dots, 0)', \quad (3.36)$$

U étant une matrice inversible de dimension $N + 1$ d'éléments

$$U_{p,q} = \sum_{j=0}^{n_p \wedge n_q} a_{n_p-j} a_{n_q-j}, p, q = 0, \dots, N. \quad (3.37)$$

De plus, $\|1 - h\|^2 = \sigma^2 \psi_0$, où

$$\sigma^2 = \exp \left[\int_{-\pi}^{\pi} \log f(\lambda) d\lambda \right]. \quad (3.38)$$

Remarque 3.6. Sous les hypothèses du théorème 3.4, on a $\text{var}(X_0 - \hat{X}'_0) = \|1 - h\|^2 = \sigma^2 \psi_0$. Une expression équivalente de $\text{var}(X_0 - \hat{X}'_0)$ a été obtenue par Grenander et Rosenblatt (1954).

Corollaire 3.1. (Bondon, 2000).

Sous les hypothèses du théorème 3.4, \hat{X}'_0 admet la représentation AR

$$\hat{X}'_0 = \sum_{k \in \mathbb{N} \setminus \mathcal{M}} h_k X_{-k} \quad (3.39)$$

où

$$h_k = \delta_k - \sum_{q=0}^N \psi_q \sum_{j=0}^{n_q \wedge k} a_{n_q-j} a_{k-j}. \quad (3.40)$$

Exemple 9. Lorsque $\mathcal{M} = \{0, 1, \dots, N\}$, \hat{X}'_0 est la prédiction à $(N+1)$ pas de X_0 et il est facile de déduire de (3.40) que $h_k = \delta_k - \sum_{j=0}^{N \wedge k} c_j a_{k-j}$, ce qui est équivalent à la relation donnée par exemple dans Akutowicz (1957).

Exemple 10. Soit $\{X_t\}$ un processus AR causal d'ordre r d'équation

$$\sum_{j=0}^r a_j X_{k-j} = \epsilon_k, \quad a_0 = 1, \quad 1 \leq r < \infty.$$

La densité spectrale f de (X_k) vérifie : $f \in L_\infty$ et $f^{-1} \in L_\infty$. Comme $\{X_t\}$ est causal, $\{\epsilon_t\}$ est le processus d'innovation de $\{X_t\}$ et a_0, \dots, a_r sont les paramètres AR de $\{X_t\}$. On déduit alors de (3.40) que $h_k = 0$ pour $k > n_N + r$ et il découle de (3.39) que

$$\hat{X}'_0 = \sum_{k=0, k \notin \mathcal{M}}^{n_N+r} h_k X_{-k}. \quad (3.41)$$

Lorsque le passé est complet, $\mathcal{M} = \{0\}$, (3.41) donne le résultat connu $\hat{X}'_0 = -\sum_{k=1}^r a_k X_{-k}$.

3.5 Influence de valeurs manquantes sur la prédiction de séries chronologiques

Dans ce qui suit, nous donnons une estimation de l'influence des valeurs manquantes sur la variance de l'erreur de la prédiction. Nous mettons l'accent sur le rôle fondamental des coefficients de la représentation autorégressive et de la position des données manquantes. Pour quantifier l'impact causé par les valeurs manquantes $X_{-t_1}, \dots, X_{-t_N}$ dans la prédiction linéaire de X_0 , Il est important de remarquer que $P'_{-1}X_0 \in \mathcal{H}'_{-1} \subset \mathcal{H}_{-1}$, et $\epsilon_0 \perp \mathcal{H}_{-1}$. D'où, $P_{-1}X_0 - P'_{-1}X_0 \perp \epsilon_0$, et nous avons

$$\|X_0 - P'_{-1}X_0\|^2 = \|\epsilon_0 + P_{-1}X_0 - P'_{-1}X_0\|^2 = \sigma_\epsilon^2 + \|P_{-1}X_0 - P'_{-1}X_0\|^2.$$

Théorème 3.5. (Bondon, 2005).

Pour un processus non déterministe stationnaire $\{X_t\}$, nous avons

$$\sigma_\eta \max_{i \in M} |a_i| \leq \|P_{-1}X_0 - P'_{-1}X_0\| \leq \sigma_\epsilon \sum_{i \in M} \phi_i |a_i| \leq \sigma_X \sum_{i \in M} |a_i| \quad (3.42)$$

où $\phi_i = \left(\sum_{j=0}^{t_N-i} c_j^2 \right)^{1/2}$.

Démonstration. Nous avons

$$\begin{aligned} X_0 - \sum_{i=1}^{t_N} a_i X_{-i} &= - \sum_{i=0}^{t_N} a_i X_{-i} = - \sum_{i=0}^{t_N} a_i \sum_{c_j} \epsilon_{-i-j} - \sum_{i=0}^{t_N} a_i V_{-i} \\ &= - \sum_{k=0}^{\infty} s_k \epsilon_{-k} - \sum_{a_i} V_{-i}, \end{aligned}$$

où

$$U_0 = - \sum_{k=t_N+1}^{\infty} s_k \epsilon_{-k} - \sum_{i=0}^{t_N} a_i V_{-i} \in \mathcal{H}_{-t_N-1}.$$

Par conséquent,

$$\begin{aligned} P_{-1}X_0 &= \sum_{i=1}^{t_N} a_i X_{-i} + U_0 \\ P'_{-1}X_0 &= \sum_{i=1}^{t_N} a_i X_{-i} + \sum_{i \in M} a_i P'_{-1}X_{-i} + U_0, \\ P_{-1}X_0 - P'_{-1}X_0 &= \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}). \end{aligned}$$

Comme $\mathcal{H}_{-t_N-1} \subset \mathcal{H}'_{-1}$, nous avons

$$\begin{aligned} \left\| \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}) \right\| &\leq \sum_{i \in M} |a_i| \|X_{-i} - P_{-1}X_{-i}\| \\ &\leq \sum_{i \in M} |a_i| \|X_{-i} - P_{-t_N-1}X_{-i}\|. \end{aligned} \quad (3.43)$$

Comme pour tout $t \in \mathbb{Z}$, on a

$$\begin{aligned} P_{-t_N-1}X_t &= \sum_{i=t_N+t+1}^{\infty} c_i \epsilon_{t-i} + V_t, \\ X_t - P_{-t_N-1}X_t &= \sum_{i=0}^{t_N+t} c_i \epsilon_{t-i}, \end{aligned}$$

il vient $\|X_{-i} - P_{-t_{N-1}}X_{-i}\| = \sigma_\epsilon \varphi_i$, ce qui donne la majoration en question.

Comme $\varphi_i \leq \left(\sum_{j=0}^{\infty} c_j^2\right)^{1/2}$ et $\sigma_X = \sigma_\epsilon^2 \sum_{j=0}^{\infty} c_j^2 + \|V_t\|^2$, la seconde majoration s'en suit immédiatement.

Soit $(i, j) \in M^2$. Nous avons

$$\langle X_{-i}, \eta_{-j} \rangle = \begin{cases} \sigma_\eta^2 & \text{si } i = j \\ 0 & \text{sinon.} \end{cases}$$

De plus, comme $\mathcal{H}'_{-1} \subset I_{-j}$ et $\eta_{-j} \perp I_{-j}$, nous avons $\eta_{-j} \perp \mathcal{H}'_{-1}$. Par conséquent,

$$\left\langle \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}), \eta_{-j} \right\rangle = a_j \sigma_\eta^2,$$

en utilisant l'inégalité de Schwarz, on obtient

$$|a_i| \sigma_\eta \leq \left\| \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}) \right\|,$$

d'où la minoration.

□

Pour un processus stationnaire minimal, la minoration montre que la dégradation de la prédiction est due aux observations manquantes qui décroît avec $\max_{i \in M} |a_i|$. Dans le cas particulier, où $M = \{m\}, m > 0$, L'inégalité (3.42) donne

$$\frac{\sigma_\epsilon |a_m|}{\left(\sum_{i=0}^{\infty} a_i^2\right)^{1/2}} \leq \|P_{-1}X_0 - P'_{-1}X_0\| \leq \sigma_\epsilon |a_m|, \quad (3.44)$$

et il résulte de Pourahmadi et Soofi (2000, thm 3.1) que

$$\|P_{-1}X_0 - P'_{-1}X_0\| = \frac{\sigma_\epsilon |a_m|}{\left(\sum_{i=0}^{\infty} a_i^2\right)^{1/2}}.$$

Plus m est grand, meilleure est la précision dans (3.42).

Le théorème qui suit est une conséquence imédiate du théorème 2.2 de Bondon (2000).

Théorème 3.6. (Bondon, 2005).

Soit $\{X_t\}$ un processus stationnaire non déterministe. Alors

- (a) $P'_{-1}X_0 = P_{-1}X_0$ si et seulement si $a_i = 0$ pour tout $i \in M$,
- (b) $P'_{-1}X_0 \rightarrow P_{-1}X_0$ dans L_2 quand $t_1 \rightarrow \infty$.

Pour un processus minimal, le théorème (3.6) découle de l'inégalité (3.42).

En effet, (a) est immédiate et pour démontrer (b), notons que la minimalité est équivalente à $\sum_{i=0}^{\infty} a_i^2 < \infty$, ce qui entraîne que $a_i \rightarrow 0$ quand $i \rightarrow \infty$. Par conséquent, $a_i \rightarrow 0$ pour tout i dans M quand $t_1 \rightarrow \infty$, la minoration de (3.42) implique (b).

Le théorème (3.6) donne le degré de convergence de $P'_{-1}X_0$ vers $P_{-1}X_0$ quand $t_1 \rightarrow \infty$.

Dans le théorème qui suit un résultat asymptotique plus précis est présenté.

Théorème 3.7. (*Bondon, 2005*).

Soit $\{X_t\}$ un processus stationnaire non déterministe. Alors, quand $t_1 \rightarrow \infty$

(a) $\|P_{-1}X_0 - P'_{-1}X_0\| = O(\alpha^{t_1})$ si $|a_i| = O(\alpha^i)$ quand $i \rightarrow +\infty$ pour tout $\alpha \in (0,1)$,

(b) $\|P_{-1}X_0 - P'_{-1}X_0\| \asymp t_1^\alpha$ si $a_i \sim ci^\alpha$ quand $i \rightarrow \infty$ avec $c > 0$ et $\alpha < -1/2$.

Démonstration. (a) découle de $\sum_{i \in M} |a_i| = O(\alpha^{t_1})$ et de la minoration du théorème 3.5.

Comme $\alpha < -1/2$ et $|a_i| \sim ci^\alpha$ quand $i \rightarrow \infty$ alors $\sum_{i=0}^{\infty} a_i^2 < \infty$, par conséquent, $\{X_t\}$ est minimal. On déduit de (3.42), pour tout $\alpha \in (0,1)$

$$0 < \sigma_\epsilon c(1 - \epsilon) \leq \frac{\|P_{-1}X_0 - P'_{-1}X_0\|}{t_1^\alpha} \leq \sigma_X c(1 + \alpha)N < \infty,$$

pour t_1 suffisamment grand, ce qui donne (b). □

Les processus ARMA ont un développement autorégressif vérifiant $|a_j| = O(\alpha^j)$, et les processus FARIMA un développement vérifiant $|a_j| \sim Cj^\alpha$. La perte d'un nombre fini d'observations du passé dégrade donc plus l'erreur quadratique de prévision pour les processus à long mémoire que pour les processus à courte mémoire.

Conclusion

Le traitement des séries chronologiques avec observations manquantes est un problème concret et toujours embarrassant lorsqu'il s'agit de données réelles. En effet, dans les applications, on est très souvent en présence d'observations pour lesquelles on ne dispose pas de l'ensemble des valeurs des variables descriptives, et ceci se produit pour de nombreuses raisons : erreurs de saisie, rubriques non renseignées dans des enquêtes, valeurs aberrantes qu'on préfère supprimer, données recueillies difficilement, statistiques officielles non disponibles, etc. La plupart des logiciels statistiques (comme SAS par exemple) suppriment purement et simplement les observations incomplètes. Toutefois, cela peut supprimer tout intérêt à l'étude si le nombre de données restantes est trop faible.

Cette thèse est une rétrospective des diverses techniques d'estimation des données manquantes et de prédiction de séries chronologiques à passé complet et incomplet.

En s'appuyant sur le théorème (2.2), l'algorithme d'espérance maximisation (EM) est une méthode itérative efficace pour calculer l'estimation de vraisemblance maximale en présence de données manquantes. Le but est d'estimer les paramètres du modèle pour lesquelles les données observées sont les plus vraisemblables tout en tenant compte de l'existence de données manquantes. Dans la pratique, ce résultat est un peu plus difficile à mettre en œuvre car il n'est pas toujours aisé d'estimer les paramètres du modèle avec les données manquantes.

Sans recourir à la représentation espace d'états, l'algorithme d'innovation généralisé, présenté dans le chapitre 3, est très utile pour calculer les projections sur les espaces de dimensions finies ou infinies. Cet algorithme possède un avantage de taille puisqu'il permet d'explicitier les variances des prédicteurs et des interpolateurs d'un processus stationnaire plus général qu'un ARMA fini. L'outil clé pour calculer ces projections est basé sur la décomposition de Wold des processus stationnaires. Lorsque les covariances sont inconnues, l'algorithme proposé peut être mis en œuvre en utilisant des estimations de covariances.

Au delà des problèmes liés à la prévision elle-même, nous nous sommes intéressés à la caractérisation des processus p -stationnaires. Plus précisément, à la représentation autorégressive et moyenne mobile de tels processus.

Enfin, ce travail ouvre le chemin à plusieurs axes de recherches et d'investigation et les extensions sont nombreuses. Parmi les éclairages possibles, on peut citer :

- Généralisation des algorithmes et relations de récurrences donnés au chapitre 1 aux processus p -stationnaire.
- Étude comparative des méthodes paramétriques d'estimation de données manquantes.

Bibliographie

Abraham, B. (1981). *Missing observations in time series*, Communication in Statistics : Theory and Methods, **10**, 1643-1653.

Akutowic, E.J. (1957). *On a explicit formula in linear least square predicion*, Math. Scand., **5**, 261-266.

Bentarzi, M. and Hallin, M. (1994). *On the invertibility of periodic moving-average models*, Journal of Time Series Analysis, **15**, 263-268.

Bhansli, R.J. (1974). *Asymptotic properties of the Weinner-Kolmogorov predictor*, I.J. Royal Statist. Soc. Ser. B, **36**, 61-73.

Bhansli, R.J. (1978). *Linear prediction by autoregressive model fitting in the time time domain*, Ann. Statisti., **6**, 224-231.

Bloomfield, R. (1972). *On the error of prediction of a time series*, Biometrika, **59**, 501-507.

Bloomfield, P. (1985). *On series representations for linear predictions*, Ann. Probab, **13** (1), 226-233.

Bondon, P. (2002). *Prediction with incomplete past of a stationnary process*, Stochastic Processes and their Applications, **98** (1), 67-76.

Bondon, P. (2005). *Influence of missing values on the prediction of a stationary time series*, Journal of Time Series Analysis, **26** (4), 519-525.

Bondon, P. (2001). *Recursive relations for multistep prediction of stationary time series*, Journal of Time Series Analysis, **22** (3), 339-410.

- Box, G. E. P and Jenkins, G.M. (1976). *Time Series Analysis, Forecasting and Control (2nd edn)*, Holden-Day San Francisco, CA.
- Brockwell, P.J. and Davis R.A. (1991). *Time Series: Theory and Methods (2nd eds)*, New York: Springer Verlag.
- Brubacher, S.R. and Wilson, G.T. (1976). *Interpolating time series with application to the estimation of holiday effect on electricity demand*, Journal of Applied Statistics, **25**, (2), 107-116.
- Cambanis, S. and Hardin, C.d. (1988). *Innovation and Wold decompositions of stable processes*, Prob. Th. Rel. Fields, **79**, 1-27.
- Cambanis, S. and Zakeri, I.F. (1996). *Forward and reversed time prediction of autoregressive sequences*, Journal of Applied Probability, **33**, (3), 1053-1060.
- Chatefield, C. (1980). *Calculating interval forecast*, J.Bus.Econm. Statist., **11**, 121-144.
- Cheng, R. and Pourahmadi, M. (1997). *Prediction with incomplete past and interpolation of missing values*, Statistics and Probability Letters, **33**, 346-346.
- Cheng, R., Miamee, A.G. and Pourahmadi, M. (2000). *Regularity and minimality of infinite variance processes*, Journal of Theoretical Probability, **13**, 1115-1122.
- Cline, D.B.H. and Brockwell, P.J. (1985). *Linear prediction of ARMA processes with infinite variance*, Stochastic Proc, Appl., **19**, 281-296.
- Damselth, E. (1980). *Interpolating missing values in a time series*, Scandinavian Journal of Statistics, **7**, 33-39.
- Doob, J.L. (1953). *Stochastic Processes*. Wiley, New York.
- Dunsmuir, w. (1983). *Large sample proprieties of estimation in time series observed at unequally spaced times*, Lecture Notes in Statistics, **25**, 58-77.
- Ferreiro, O. (1987). *Methodologies for the estimation of missing observations in time series*, Statistics and Probability Letters, **5**, 65-69.
- Ghosh, M. and Sen, P.K. (1989). *Median unbiasedness and Pitman closeness*, J. Amer. Statist. Assoc., **84**, 1089-1091.

Gladyshev, E.G. (1961). *Periodically Correlated Random Sequences*, Soviet Mathematics **2**, 385-388.

Grenander, U. and Rosenblatt M. (1954). *An extension of a theorem of G. Szego and its application to the study of stochastic processes*, Trans. Amer. Math. Soc., **76**, 112-126.

Hamaz, A. and Ibazizen, M. (2009). *Comparison of two estimation methods of missing values using Pitman-Closeness Criterion*, Communication in Statistics: Theory and Methods, **38**, 2210-2214.

Hannan, E.J. and Kanter, M. (1977). *Autoregressive process with infinite variance*, J. Appl. Probab., **14**, 411-415.

Jones, R.H. (1980). *Maximum likelihood fitting of ARMA models in time series with missing observations*, Technometrics, **22**, 389-395.

Jones, R.H. and Brelsford, W.M. (1967). *Time series with periodic structure*, Biometrika, **54**, 403-408.

Keating, J.P. and Masson, R.L. (1985a). *Pitman measure of closeness*, Sankhya. Ser. B., **47**, 22-32.

Keating, J.P. and Masson, R.L. (1985b). *Practical relevance of an alternative criterion in estimation*, Amer. Statist., **39**, 203-205.

Kohn, R. and Pierse, R.G. (1984). *Estimating missing observations in economic time series*, J. Amer. Statist. Assoc., **81**, 751-761.

Ljung, G.M. and Box, G.E.P. (1979). *The likelihood function of stationary autoregressive-moving average models*, Biometrika, **66**, 256-270.

Ljung, G.M. (1982). *The likelihood function for a stationary autoregressive-moving average processes with missing data*, Biometrika, **69**, 265-280.

Ljung, G.M. (1989). *A note on the estimation of missing values of missing values in time series*, Communications in Statistics: Simulation and computation, **18** (2), 459-462.

Lund, R. and Basawa, I.V. (2000). *Recursive prediction and likelihood evaluation for periodic ARMA models*, Journal of Time Series Analysis, **21** (1), 75-93.

Masani P. (1960). *The prediction theory of multivariate stochastic processes*. Acta Math., **104**, 141-162.

Mcleod, A.I. (1993). *Parcimony, model adequacy and periodic correlation in time series forecasting*, Int. Stat. Rev. **61**, 387-393.

Miamme, A.G. and Salehi, H. (1980). *On an explicit representation of the linear predictor of a weakly stationary stochastic sequence*. Tech. Report, Dept. of Statistics and Probability.

Miamme, A.G. and Salehi, H. (1983). *On an explicit representation of the linear predictor of a weakly stationary sequence*. Boletín de la Sociedad Matemática Mexicana, **44**, 1023-1030.

Mohanty, R. and Pourahmadi, M. (1996). *Estimation of the generalised error variance of a multiple time series*. J. Amer. Statist. Assoc. **91**, 294-299.

Parzen, E. (1974). *Some recent advances in time series modelling*, I.E.E.E. Trans. Automatic Control, Ac-19, 723-729.

Parzen, E. and Pagano, M. (1979). *An approach to modeling seasonally stationary time series*, J. Economet. **9**, 137-153.

Parzen, E. (1983). *Proc. of time series analysis of irregularly observed data*, Lecture Notes in Statistics, **25**, Springer, New York.

Parzen, E. (1984). *Time series ARMA model identification by estimating information*, Proceeding of the 15th Annual Symposium on the Interface of Computer Science and Statistics. Amsterdam: North Holland.

Pourahmadi, M. (1989). *Estimation and interpolation of missing values of stationary time series*, Journal of Time Series Analysis, **10** (2), 149-169.

Pourahmadi, M. (1992). *Alternating projections and interpolation of stationary Processes*, J. Applied. Proba., **29**, 921-931.

Pourahmadi, M. (1992b). *Fundamental roles of the idea of regression and Wold decomposition in time series*. New direction in time series analysis. Part I, 287-314, IMA Vol, Math. Appl., **45**, Springer, New York.

Pourahmadi, M. (1994). *Two prediction problems and extension of a theorem of Szego*, Iranian Math. Soc., **2**, 1-12.

Pourahmadi, M. and Soofi, E.S. (2000). *Prediction variance and information worth of observations in time series*, Journal of Time Series Analysis, **21** (4), 413-434.

Pourahmadi, M. (2001). *Foundation of Time Series Analysis and Prediction Theory*, New York: Wiley.

Priestly, M.B. (1981). *Spectral Analysis and Time Series*, Academic Press, London.

Roy, S.S. and Chakraborty, S. (2006). *Prediction problem related to the first order autoregressive processes in the presence of outliers*, Applic. Math., **33**, 265-274.

Shaman P. (1982). *The mean square of one-step-ahead prediction for an autoregressive process when the coefficients are estimated with missing data*, Time Series Analysis : Theory and Practice, **1**, 53-63.

Shumway, R.H. and Stoffer, D.S. (2000). *Time Series Analysis and its Applications*, New York: Springer Verlag.

Sen, P.K. (1990). *On the Pitman closeness of some sequential estimators*, Sequential Analysis, **9**, 383-400.

Salehi, H. (1974). *On the bilateral linear predictor for minimal stationary stochastic processes*, Siam J. Appl. Math., **26**, 502-507.

Vacchia, A.V. (1985). *Periodic autoregressive-moving average (PARMA) modeling with applications to water resources*, Water Resources Bulletin, **21**, 721-730.

Wenzel, T. (2002). *Pitman-Closeness as a measure to evaluate the quality of forecasts*, Communication in Statistics : Theory and Methods, **31**, 535-550.

Wiener, N. and Massani, P.R. (1958). *The prediction theory of multivariate stationary processes*, Acta. Math., **99**, 119-137.

