

Séries temporelles : Modélisation ARIMA et modélisation espace-état

8 septembre 2004

Enseignant : Florin Avram

Objectif : L'interpolation –prévision "ponctuelle, déterministe" – et la régression : démarche "statistique", qui va au delà de l'interpolation en analysant les résidus et en produisant des intervalles de confiance, sont parmi les méthodes les plus importantes dans les mathématiques et statistiques appliquées. On les utilise par exemple pour la prédiction des phénomènes spatio-temporaux en géostatistique, économétrie, météorologie, sciences environnementales, ..., etc. Nous allons aborder ces thèmes dans le contexte des séries temporelles uni-dimensionnelles, en commençant par la modélisation statistique la plus simple : ARIMA, et en suivant avec la modélisation d'espace-état.

Contenu :

1. Introduction
2. Inspection graphique des séries temporelles. Lissage et filtres.
3. Modélisation statistique. Modèles graphiques et d'espace-état.
4. Modélisation ARIMA. Prévision et tests pour les résidus.
5. Le filtre du Kalman.

Compétences acquises : Les étudiants apprendront à utiliser des diverses méthodes de filtrage et prévision des séries temporelles, notamment par la modélisation ARMA, et à tester les résidus pour évaluer la fiabilité des modèles choisies.

Matériels :

1. Notes de cours/TD, qui utilisent parties des notes de M. Lavielle (Université Paris-Sud) et A. Korabinski (Heriot-Watt) sur les séries temporelles (toutes les coquilles sont de ma responsabilité).
2. Notes WEB : A. Charpentier, M. Kratz, J-M. Dufour (en Français) et R.H. Smith, R. Weber (** En Anglais), etc
 - <http://www.crest.htfr/pageperso/lfa/charpent/charpent.htm#TS>
 - <http://www.math-info.univ-paris5.fr/kratz/cours.html>
 - <http://www.statslab.cam.ac.uk/rrw1/timeseries/index.html>
3. A. C. Harvey, Time Series Models.
4. J. Durbin and S.J. Koopman, Time series analysis by state space methods.
5. C. Gouriéroux et A. Monfort, Cours de séries temporelles.

Table des matières

- 1 **Introduction : Méthodes déterministes (interpolation exacte et approximative, extrapolation) et statistiques**

2

2	Les séries temporelles	5
2.1	Les composantes d'une chronique	6
2.2	Quelques types de décomposition	7
2.2.1	le modèle additif	7
2.2.2	le modèle multiplicatif	7
2.2.3	les modèles mixtes	7
2.3	Lissage et filtres	8
2.3.1	L'inversion des filtres	8
2.4	Discussion sur la manipulation formelle des fonctions dans l'opérateur B	9
2.5	Exercices : TD 1	9
3	Quelques rappels de probabilité et des statistiques	11
3.1	Analyse statistique univariée	11
3.2	Analyse statistique bivariée	11
3.3	Distribution gaussienne multivariée	12
4	Modélisation stochastique de séries temporelles	14
4.1	Quelques exemples des processus stochastiques	14
4.1.1	Le bruit blanc	14
4.1.2	Les processus stationnaires	15
4.1.3	Les processus linéaires et moyennes mobiles MA(q)	15
5	Les modèles autorégressifs	18
5.1	Exercices : TD 2	19
5	Les modèles ARMA(p,q)	21
5.1	Causalité et inversibilité des modèles ARMA(p,q)	21
5.2	Équations Yule-Walker pour les covariances/corrélations des modèles autorégressifs AR(p) causales	23
5.3	Contrôle continu : TD3	25
5.4	Équations de Yule-Walker pour les covariances/corrélations des processus ARMA(p,q) causales	26
6	La prévision linéaire	28
6.1	La prévision des processus stationnaires en représentation autorégressive AR(p)	29
7	Les modèles ARIMA(p,d,q)	31
7.1	Prévision linéaire des modèles autorégressifs ARIMA(p,d,0)	31
7.2	Exercices : TD 4	32
7.3	Prévision linéaire des modèles à bruit moyenne mobile	32
7.4	Exercices : TD 5	34
7.5	Le filtre de Kalman	35
8	Examen d'entraînement 2	38

1 Introduction : Methodes déterministes (interpolation exacte et approximative, extrapolation) et statistiques

On considère un processus stochastique (i.e. une famille des variables aléatoires) $U(x)$, où $x \in \mathbb{R}^d$ sera d'habitude une variable spatiale, comme en géostatistique, ou temporelle $x \in \mathbb{N}, \mathbb{Z}$ ou \mathbb{R}_+ , comme pour les séries chronologiques.

On se propose d'estimer la valeur de la variable $U(x)$ en un point x quelconque connaissant les valeurs $U(x_i)$ aux points de mesure x_i , pour $i = 1, \dots, N$. Le but principal est donc la prédiction des valeurs inobservables (comme les valeurs futures des séries temporelles, ou moins accessibles

physiquement, couteuses, etc), a partir des valeurs connues. On veut à la fois : a) enlever du bruit éventuel et b) "extrapoler" du connu au inconnu.

Domaines d'application :

- Prospection et exploitation pétrolières et minières
- Traitement du signal
- Imagerie médicale
- Océanographie, météorologie, hydrogéologie, environnement, ...
- Séries temporelles, appliquées en économie, finances, météo, médecine, ...

Définition 1.1 Une série chronologique (ou temporelle) est une succession d'observations au cours du temps : $\{U_t : t = 1, 2, \dots, n, \dots\} = (U_1, U_2, \dots, U_n, \dots)$

Par rapport aux autres types de données statistiques, la particularité des séries chronologiques tient à la présence d'une relation d'antériorité qui ordonne l'ensemble des informations. Les dates d'observations sont souvent équidistantes les unes des autres : on a des séries mensuelles, trimestrielles, etc, dans quel cas on peut les indexer par $t \in \mathbb{N}$. **Exemples :** a) Nombre des moutons par année en Angleterre, entre 1867 et 2003. b) Nombre de voyageurs par mois (SNCF) entre 1990 et 2003. c) Nombre de voitures vendues par un garage, par trimestre entre 1995 et 1999.

d) **Taux de mortalité, per age, entre 55 et 104** (le premier exemple d'utilisation de splines, par Whittaker (1923).

Les séries temporelles vont fournir notre principal exemple pour examiner l'application des diverses méthodes étudiées.

Méthodes : En statistique, il y a deux familles principales de méthodes :

1. les méthodes classiques, appelées "paramétriques" utilisent d'abord une interpolation déterministe simple (comme la regression lineaire), et ensuite verifient si "le choix du modèle" est bon (par des methodes parfois assez compliquées)
2. les méthodes appelées "non paramétriques", qui commencent par une interpolation déterministe sophistiquée, comme splines, noyaux RBF, ondelettes, pour l'aspect spatial (en adoptant souvent une démarche différente, comme le filtre de Kalman, pour l'aspect temporel). Pour ces méthodes, "la verification du choix du modèle" est plus compliquée, mais ce desavantage est compensé par le fait que la prédiction déterministe est meilleure.

Il existent donc plusieurs démarches : 1) **déterministes**, qui aboutissent dans une prédiction ponctuelle et 2) **stochastiques**, qui fournissent aussi des intervalles de confiance.

1. Prédiction des modèles stochastiques :
 - Modèles Gaussiens, démarches parametrique et non-parametrique.
 - Modèles non-gaussiens : modèle lineaire généralisé, etc.
2. Interpolation déterministe :
 - interpolation exacte : polynomiale, polynomiale par morceaux, par "splines", etc.
 - interpolation approximative : polynôme de moindre carrées, "splines ajustées/de moindre carrées", etc

L'interpolation déterministe la plus simple est celle par polynômes, mais elle est rarement satisfaisante (voir exemple de Runge). Dans le cas des series composées des morceaux à allure différente, la première idée serait d'utiliser interpolation polynomiale par morceaux. Mais, une meilleure approche est de demander aussi **lissité** dans les points de contact, ce qui nous ramène à splines, et, finalement, dans le cas d'observations incertaines, à **l'interpolation déterministe approximative par splines ajustées**, qui cherche une approximation $\xi(x)$ minimisant l'objectif :

$$\min_{\xi(x) \in C^2} \sum_i (u(x_i) - \xi(x_i))^2 + \lambda \int_{\mathbb{R}} (\xi''(x))^m dx \tag{1}$$

Cet objectif essaie de trouver un équilibre entre deux buts opposés : 1) celui d'un bon ajustement de données, obtenu en minimisant la somme des écarts au carré et 2) la "lissité", obtenu en

minimisant la dérivé d'ordre m au carré (on voit par exemple que avec $m = 2$ ce terme est minimisé par la droite, qui est "très lisse"). Ou, en discrétisant sur les entiers, on minimisera :

$$\min_{\xi(x) \in C^2} \sum_i (u(x_i) - \xi(x_i))^2 + \lambda \sum_i (\Delta^m \xi(x_i))^2$$

où Δ^m est la discrétisation de la dérivé d'ordre m . Cet objectif a été pour la première fois proposé par Whittaker (1923). Quand $m = 2$ et donc $\Delta^2 \xi(x_i) = \xi(x_{i+2}) - 2\xi(x_{i+1}) + \xi(x_i)$ est la discrétisation de la seconde dérivé, on arrive au "filtre de Hodrick et Prescott", utilisé dans l'économie pour prédire la tendance du marché.

Pour mieux comprendre l'objectif de minimisation (1) quand $m = 2$, nous examinons maintenant les cas limite $\lambda \rightarrow \infty$ et $\lambda \rightarrow 0$. Dans ces deux cas, le problème devient une minimisation "du objectif moins important, avec l'objectif plus important imposé d'être 0", i.e.

1. Dans la limite $\lambda \rightarrow \infty$, la deuxième contrainte est imposé. Donc, le problème devient :

$$\min_{\xi(x) \in C^2} \sum_i (u(x_i) - \xi(x_i))^2$$

$$\int_{\mathbb{R}} (\xi''(x))^2 dx = 0$$

La solution est la droite des moindres carrés.

2. Dans la limite $\lambda \rightarrow 0$, la première contrainte est imposé. Donc, le problème devient :

$$\min_{\xi(x) \in C^2} \int_{\mathbb{R}} (\xi''(x))^2 dx$$

$$\sum_i (u(x_i) - \xi(x_i))^2 = 0$$

Il s'avère que la minimisation de $\int_{\mathbb{R}} (\xi''(x))^2 dx$ est exactement le principe qui détermine le positionnement d'une tige mince, flexible, obligée à passer par des points donnés u_i . Donc, dans ce cas, la solution est une "tige d'interpolation exacte", dont le nom mathématique est spline naturelle d'interpolation.

Notes : 1) Le cas général est une pondération de ces deux cas extrêmes par un **facteur ad-hoc** λ . Avec λ faible on est proche de l'interpolation spline, et avec λ grand on est proche de la droite des moindres carrés.

2) Il reste encore la question du choix du paramètre λ . A priori, sans ajouter un nouveau critère d'optimisation, qui spécifie "la lissité désirée", chaque λ est légitime. Pour une première solution pour le choix d'un critère supplémentaire d'optimisation, on pourrait ajouter au objectif une fonction donnée convexe, comme $(\lambda - a)^2$, et minimiser aussi par rapport à λ , ce qui remplacerait le problème du choix de λ par celui d'un a ... Les logiciels proposent plusieurs possibilités plus sophistiquées et qui marchent assez bien, mais un choix définitif n'a pas encore émergé.

Les méthodes déterministes, qui se veulent indépendantes des caractéristiques statistiques sont plus "robustes" et recommandées quand il n'y a pas assez de données. Pour quelques exemples encore plus simples, on pourrait utiliser pour le "debruitage" des images une moyenne simple des valeurs voisines, et pour la prédiction des séries temporelles, des moyennes simples des valeurs passées; souvent, l'information disponible ne permet de faire rien de mieux.

Par contre, nous allons nous occuper ici aussi parfois des cas plus rares quand l'information suffit pour une modélisation statistique permettant de classer le processus/série observés, et d'arriver ainsi à une inférence spécifique au problème, i.e. à une analyse des résidus et aux intervalles de confiance.

2 Les séries temporelles

Une règle générale en statistique descriptive consiste à commencer par regarder ses données, avant d'effectuer le moindre calcul. Ainsi, la figure 1 montre différentes séries chronologiques, qui méritent quelques commentaires.

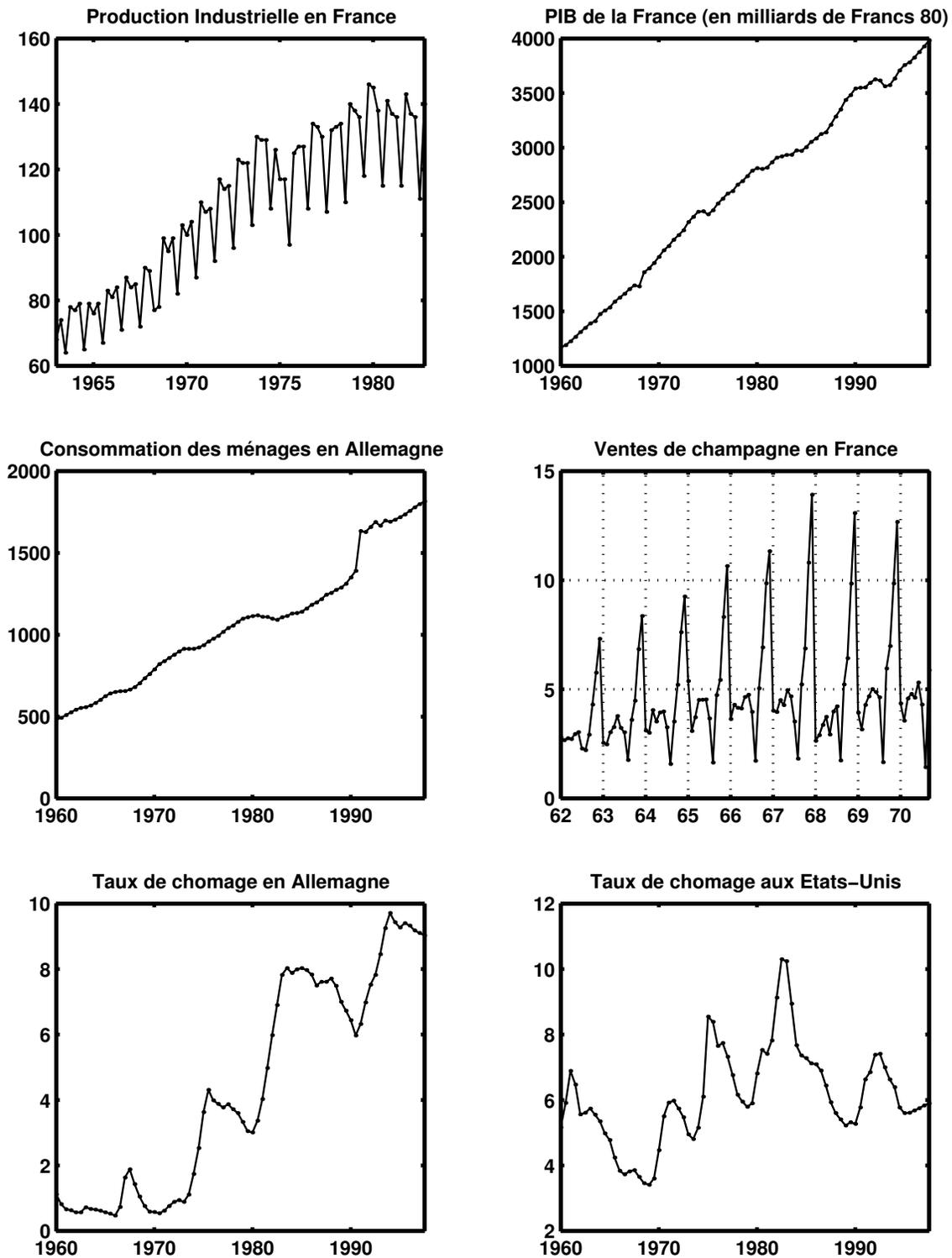


FIG. 1 – Quelques exemples de séries chronologiques

- La consommation des ménages en Allemagne et le Produit Intérieur Brut en France semblent avoir augmenté régulièrement.

- Le taux de chômage en Allemagne semble avoir globalement augmenter depuis 1960, mais avec une alternance de baisses et de hausses soudaines. Le taux de chômage des Etats-Unis ne semble pas évoluer globalement, mais présente également cette alternance de baisses et de hausses.
- Les ventes de champagnes, tout comme la production industrielle semblent exhiber un caractère périodique (ventes importantes de champagne en fin d’année, baisse de la production industrielle en été, ...).
- D’autre part, les variations de ces 2 séries (indice de production industrielle et ventes de champagne) ont une amplitude qui semble augmenter au cours du temps.
- Toutes ces séries ont un aspect irrégulier. Ces fluctuations irrégulières ont parfois une amplitude anormalement élevée (PIB et production industrielle en France au second trimestre 1968, consommation en Allemagne en 1991).

Cette liste de remarques n’est bien sûr pas exhaustive. Elles traduisent simplement quelques comportements que l’on retrouve sur la plupart des séries chronologiques. Puisque notre ambition est de décrire et d’analyser ce genre de chroniques, il nous faut donc proposer des modèles qui intègrent les différentes caractéristiques que nous venons de relever.

2.1 Les composantes d’une chronique

Dans un premier temps, l’examen graphique de la série étudiée ($y_i, 1 \leq i \leq n$) permet de dégager, lorsqu’on envisage une période de temps suffisamment longue, un certain nombre de composantes fondamentales de l’évolution de la grandeur étudiée.

Il faut alors analyser ces composantes, en les dissociant les unes des autres, c’est-à-dire que l’on considère cette série comme résultant de la combinaison de différentes séries, tel que chacune d’elles ait une évolution simple.

1. *La tendance* ($f_i, 1 \leq i \leq n$) représente l’évolution à long terme de la grandeur étudiée, et traduit l’aspect général de la série.
2. *Le cycle* (ou cycle conjoncturel) ($c_i, 1 \leq i \leq n$) regroupe les variations autour de la tendance avec des alternances de phases d’expansion et de recession. Ces phases durent généralement plusieurs années, mais n’ont pas de durée fixe. Sans informations spécifiques, il est généralement très difficile de dissocier la tendance du cycle. Dans le cadre de ce cours, la composante appelée tendance regroupera donc la tendance et le cycle.
3. *Les variations saisonnières* ($s_i, 1 \leq i \leq n$) sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz, ...), ou encore par des activités économiques et sociales (fêtes, vacances, solde, etc). Elles sont de nature périodique, c’est-à-dire qu’il existe un entier p , appelé *période*, tel que $s_i = s_{i+p}$ pour tout $i \geq 1$. Cette composante est donc entièrement déterminée par ses p premières valeurs s_1, s_2, \dots, s_p .
4. *Les fluctuations irrégulières* ($e_i, 1 \leq i \leq n$) ont souvent un effet de faible intensité et de courte durée. Ils sont de nature aléatoire (ce qui signifie ici, dans un cadre purement descriptif, qu’ils ne sont pas expliqués).
5. *Les variations accidentelles* sont des valeurs isolées anormalement élevées ou faibles. Ces variations brusques de la série sont généralement explicables (Mai 68, réunification de l’Allemagne, tempête, ...). La plupart du temps, ces accidents sont intégrés dans la série des fluctuations irrégulières.

En résumé, nous considérerons une série chronologique comme issue de la composition de 3 composantes :

- ($f_i, 1 \leq i \leq n$) la tendance (intégrant éventuellement un cycle),
- ($s_j, 1 \leq j \leq p$) les coefficients saisonniers,
- ($e_i, 1 \leq i \leq n$) les fluctuations irrégulières (intégrant éventuellement des accidents).

2.2 Quelques types de décomposition

Après avoir détecté graphiquement quelles sont les composantes présentes, il faut proposer un modèle :

2.2.1 le modèle additif

$$y_i = f_i + s_i + e_i, \quad 1 \leq i \leq n. \quad (2)$$

Pour bien séparer la tendance de la composante saisonnière, et pour des raisons d'unicité dans la décomposition proposée, on impose que la somme des facteurs saisonniers soit nulle :

$$\sum_{j=1}^p s_j = 0.$$

Dans ce modèle, on suppose que les amplitudes des variations saisonnières et des fluctuations résiduelles ne dépendent pas de la tendance.

Imaginons que nous étudions la série des températures moyennes relevées chaque mois en un même site, depuis janvier 1990. Que peut-on dire des composantes présentes ?

- la série (f_i) représente la tendance générale (réchauffement ? cycle ?).
- Les données étant mensuelles, la période est de un an, et donc $p = 12$.
- Des valeurs $s_1 = -10$ et $s_6 = +8$ signifient que le mois de janvier est plus froid de 10° par rapport à l'ensemble de l'année, alors que juin est plus chaud de 8° .
- Une fluctuation irrégulière $e_{14} = -2$ signifie qu'il a fait 2° de moins que prévu pour un mois de février, en 1991 (c'est-à-dire ce que nous laissent prévoir la tendance et l'effet saisonnier pour février 1991).

2.2.2 le modèle multiplicatif

$$y_i = f_i(1 + s_i)(1 + e_i), \quad 1 \leq i \leq n. \quad (3)$$

Là encore, on impose que la somme des facteurs saisonniers soit nulle : $\sum_{j=1}^p s_j = 0$.

Dans ce modèle, on considère maintenant que les amplitudes des fluctuations dépendent du niveau. Considérons le nombre d'entrées quotidiennes dans un cinéma. Des valeurs $s_4 = -0.5$ et $s_6 = +0.8$ signifient ici que la fréquentation de cette salle diminue de 50% le jeudi et augmente de 80% le samedi (par rapport à l'ensemble de la semaine). Une valeur $e_9 = +0.2$ signifie que le nombre d'entrée du deuxième mardi a été de 20% supérieur au chiffre attendu pour ce jour là.

Remarque : Le modèle multiplicatif est généralement utilisé pour des données de type économique.

2.2.3 les modèles mixtes

Il s'agit là des modèles où addition et multiplication sont utilisées. On peut supposer, par exemple, que la composante saisonnière agit de façon multiplicative, alors que les fluctuations irrégulières sont additives :

$$y_i = f_i(1 + s_i) + e_i, \quad 1 \leq i \leq n. \quad (4)$$

(toutes les autres combinaisons sont également possibles ...).

La modélisation stochastique des séries temporelles commence en observant leur graphique et en cherchant une décomposition additive ou multiplicative. Nous étudierons en suite le modèle additif (le modèle multiplicatif revient à un modèle additif pour le log des données).

2.3 Lissage et filtres

Définition 2.1 Soit S l'espace des séries stationnaires à moyenne 0, ou l'espace des séries déterministes avec norme L_2 fini. Soit ψ une suite tq $\sum \psi_i^2 < \infty$. Un opérateur de la forme $F_\psi : S \rightarrow S$ défini par :

$$Y_t = F_\psi(\epsilon_t) = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

est appelé opérateur de convolution ou **filtre**, et noté par $\psi(B)$.

C'est facile de voir que les opérations des filtrages commutent, i.e. que

$$F_{\psi_1} F_{\psi_2} = F_{\psi_2} F_{\psi_1} = F_{\psi_1 * \psi_2}$$

En effet, l'ensemble des filtres est isomorphe au ensemble des fonctions complexes $\psi(z)$, l'isomorphisme étant "la transformé z " des suites. Cet isomorphisme explique quelques manipulations formelles avec les filtres, mais pas celle liés à l'inversion.

Une question de portée pratique pour un filtre c'est la détermination de son noyau et son espace invariant.

Théorème 2.1 a) Un filtre conserve les polynômes de degré $\leq p$ ssi 1 est une racine d'ordre au moins p de l'équation $\psi(z) = 1$. a) Un filtre enlève les composantes périodiques d'ordre p ssi $\psi(z)$ est divisible par $1 + z + \dots + z^{p-1}$ (donc si $\psi(z) = 0$, pour toutes les racine d'ordre p de l'unité, sauf $z = 1$).

Plus tard, nous discuterons des processus Y_t "ARMA(p,q)", qui satisfont des équations de la forme :

$$\varphi(B)Y_t = \theta(B)\epsilon_t$$

Exercice 2.1 Trouvez un filtre $f(B)$ qui conserve les polynômes de degré ≤ 1 , et qui enlève les composantes périodiques d'ordre 4, et déduisez que pour une série ayant une composante périodique d'ordre 4 et une tendance linéaire m_t , la tendance est donné par $m_t = \mathbb{E}Y + f(B)Y_t$.

2.3.1 L'inversion des filtres

Une question importante pour ce modèle sera le calcul des filtres inverses.

Exemple 2.1 Le processus $AR(1)$ et le filtre inverse de $1 - \varphi B$

Le résultat suivant sera utile pour les questions d'inversion des ARMA(p,q).

Lemma 2.1 a) Pour un polynôme $\varphi(z) = \prod_{i=1}^p (1 - z/\lambda_i)$ qui a toutes ses racines λ_i à l'extérieur du cercle unitaire $|z| \leq 1$, $\frac{1}{\varphi(z)}$ a un développement en série de Taylor

$$\frac{1}{\varphi(z)} = \sum_{n=0}^{\infty} \psi_n z^n$$

qui est convergente à l'intérieur du cercle unitaire $|z| = 1$. Dans le cas le plus simple des racines λ_i distinctes, on a

$$\psi_n = \sum_{i=1}^p \frac{K_i}{\lambda_i^{n+1}} \quad (5)$$

où $K_i = -\frac{1}{\varphi'(\lambda_i)}$. (Dans le cas des racines confondues, on a des formules similaires qui utilisent dérivées de degré supérieur).

b) Pour un polynôme $\varphi(z) = \prod_{i=1}^p (1 - z/\lambda_i)$ qui a toutes ses racines λ_i à l'intérieur du cercle unitaire $|z| \leq 1$, $\frac{1}{\varphi(z)}$ a un développement en série de Laurent

$$\frac{1}{\varphi(z)} = \sum_{n=-1}^{-\infty} \psi_n z^n$$

qui est convergente sur le cercle unitaire $|z| = 1$. Dans le cas le plus simple des racines λ_i distinctes, on a

$$\psi_n = - \sum_{i=1}^p K_i \lambda_i^{n+1} \quad (6)$$

où $K_i = -\frac{1}{\varphi'(\lambda_i)}$ c) Dans le cas mixte avec racines à l'intérieur et aussi à l'extérieur du cercle unitaire on a un mélange des formules ci-dessus.

Note : Les calculs formels avec des polynômes $f(B)$ peuvent-être justifiés en observant que les matrices correspondants sont des des matrices Toeplitz, et qu'il y a un isomorphisme entre l'anneau des matrices Toeplitz est celui des polynômes $f(z)$ (voir dernier sous-chapitre).

Mais, ces approches ne résolvent pas de suite le problème fondamental ici, qui est de choisir entre plusieurs définitions possible pour le développement de l'inverse d'un polynome $\frac{1}{\varphi(z)}$ en série des puissances (car le choix dépend du positionnement du point d'intérêt z par rapport aux racines).

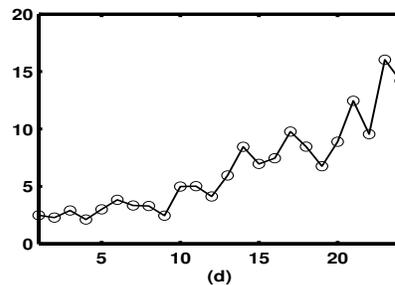
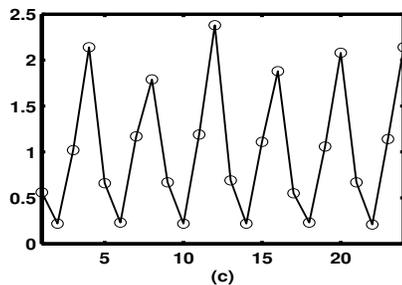
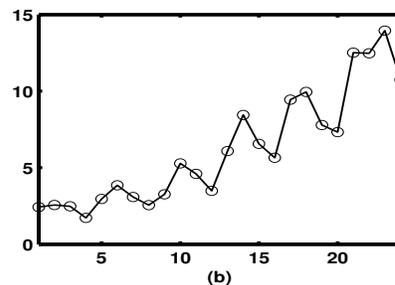
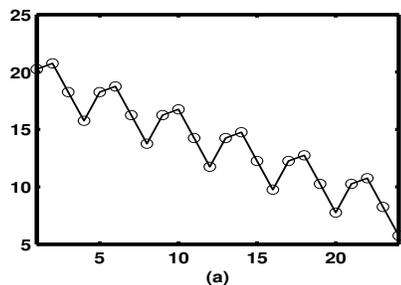
2.4 Discussion sur la manipulation formelle des fonctions dans l'operateur B

On peut approcher de manière rigoureuse les manipulations formelles comme l'inversion du polynôme $\varphi(B)$ par plusieurs démarches :

1. **Les fonctions génératrices.** Cette approche associe à chaque suite ψ_n avec $n \in \mathbb{N}$, $-n \in \mathbb{N}$ ou $n \in \mathbb{Z}$ la fonction $\tilde{\psi}(z) = \sum_n \psi_n z^n$. Dans le premier cas appelé série de Taylor, la série est convergente dans l'intérieur d'un certain "cercle de convergence", dans le deuxième cas, la série est convergente dans l'exterieur d'un certain "cercle de divergence" et dans le troisième cas, appelé série de Laurent, la série est convergente dans l'intérieur d'un certain "anneau de convergence" (le role joué par la convergence dans les calculs n'est pas crucial ; on peut utiliser parfois même des séries divergentes partout, en les définissant comme objets isomorphes à un certain anneau algebrique).
2. **Les matrices Toeplitz.** On s'aperçoit que les operateurs correspondant à des polynômes en B sont représenté par des matrices Toeplitz ; on peut démontrer que il y a un isomorphisme entre l'anneau des matrices Toeplitz est celui des fonctions génératrices, donc les deux approches sont équivalents.

2.5 Exercices : TD 1

1. Pour les différentes séries proposées dans la Figure 1, déterminer les composantes présentes et proposer un modèle
2. Pour chacune des quatre séries suivantes,



- écrire le modèle qui vous semble convenir, en précisant les composantes présentes et le type du modèle,
- si une composante saisonnière est présente, que vaut la période? Si une tendance est présente, comment peut-elle être modélisée?

3. On considère la série suivante

t_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y_i	7.5	4.4	3.3	7.6	3.9	2.4	6.9	4.5	2.7	8.2	4.1	3.0	7.5	3.5	2.8

- Représenter graphiquement cette série.
 - Quel modèle proposeriez-vous pour cette série (justifier)?
 - Calculer les facteurs saisonniers ($s_j, 1 \leq j \leq p$) ainsi que leur somme $\sum_{j=1}^p s_j$.
 - En notant ($e_i, 1 \leq i \leq n$) la série des fluctuations irrégulières, calculer e_1, e_2 et e_3 .
4. On considère un modèle simple où la tendance est une constante ($f(t) = a$).
- On considère tout d'abord le modèle sans composante saisonnière. Comment choisir a si le modèle est additif? que peut-on alors dire sur les fluctuations irrégulières? que se passe-t-il si le modèle est multiplicatif?
 - On considère maintenant qu'une composante saisonnière ($s_j, 1 \leq j \leq p$) est présente. On suppose que le nombre d'observations n est un nombre entier L de périodes : $n = Lp$. Comment choisir a et (s_j) si le modèle est additif? que peut-on alors dire sur les fluctuations irrégulières? que se passe-t-il si le modèle est multiplicatif?
 - Reprendre la question b) lorsque le nombre d'observations n'est pas un nombre entier de périodes : $n = Lp + m$.

3 Quelques rappels de probabilité et des statistiques

3.1 Analyse statistique univariée

Définition 3.1 Variable aléatoire : Une variable X est dite aléatoire si elle peut prendre un ensemble de plusieurs valeurs possibles :

- discret : $X = x_1, x_2, \dots, x_n$
- ou continu : X appartient à un intervalle (a, b)

Chaque X est caractérisée par une distribution de probabilité, spécifié par une probabilité cumulée ou "cdf" $F(x) = \mathbb{P}\{X \leq x\}$, et aussi par :

- la "masse de probabilité" : $\mathbb{P}\{X = x_i\} = p_i$ dans le cas discret, ou
- la densité de probabilité ou "pdf" $f(x) = \frac{dF(x)}{dx}$ dans le cas continu.

La loi de probabilité peut être représentée par ses moments, en particulier :

- la moyenne ou espérance mathématique :

$$m_X = \mathbb{E}X = \int x dF(x)$$

- la variance ou espérance des carrés des écarts à la moyenne :

$$\sigma_X^2 = \text{Var } X = \mathbb{E}(X - m_X)^2 = \int (x - m_X)^2 dF(x)$$

ou σ_X est l'écart type ou paramètre de dispersion, exprimé dans les mêmes unités que X . Les intervalles de la forme $[m_X - z_\alpha \sigma_X, m_X + z_\alpha \sigma_X]$, appelés **intervalles de confiance** remplacent la "prévision ponctuelle" de X , qui est m_X . Ici, z_α sont choisies en fonction du "niveau de confiance" demandé α et de la distribution du "residu normalisé" $\frac{X - m_X}{\sigma_X}$ (par exemple, pour une confiance de 95% et résidus Gaussiens, $z_\alpha = 2$).

3.2 Analyse statistique bivariée

La codispersion de deux variables aléatoires X et Y est caractérisée par la covariance ou espérance du produit des écarts à leur moyenne respectives m_X et m_Y :

$$\sigma_{XY} = \text{Cov}(X, Y)$$

Cette quantité peut être normalisée : le coefficient de corrélation est le rapport - sans dimension - de la covariance aux produit des deux écarts-types :

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

On montre que le coefficient de corrélation est nécessairement compris entre -1 et +1. Il indique le degré de dépendance linéaire entre les deux variables : - s'il est proche de ± 1 , il indique l'existence d'une relation linéaire entre variables ;

- s'il est proche de 0, il indique que les deux variables ne sont pas linéairement dépendantes.

L'indépendance complète entre les variables correspond à une hypothèse plus forte, soit en terme de probabilités : $f_{XY}(x, y) = f_X(x)f_Y(y)$. Elle implique que la covariance et le coefficient de corrélation correspondants sont nuls - l'inverse n'étant pas nécessairement vérifiée.

Les géostatisticiens utilisent également le **variogramme**

$$\gamma_{XY} = \frac{\text{Var}(X - Y)}{2}$$

d'habitude pour des variables avec la même espérance et variance σ^2 , dans quel cas on trouve :

$$\gamma_{XY} = \sigma^2 - \text{Cov}(X, Y)$$

De lors, ce coefficient et parfaitement équivalent à la covariance. Cependant, la variogramme a une intepretation intuitive de "mesure de dissimilitude" qui la rend préféré dans la géostatistique (en plus, le fait que dans ce domaine les covariances sont presque toujours positives afaibli l'intérêt descriptive de la covariance).

3.3 Distribution gaussienne multivariée

Définition 3.2 Un vecteur colonne $\mathbf{X} = (X_1, \dots, X_I)$ a une distribution gaussienne multivariée $N(\boldsymbol{\mu}, \mathbf{C})$, où $\boldsymbol{\mu}, \mathbf{C}$ denotent un vecteur ligne et une matrice positive, ssi sa densité est de la forme

$$f_{\mathbf{X}}(\mathbf{x}) = \sqrt{\frac{\det(\mathbf{Q})}{(2\pi)^I}} e^{-(\mathbf{X}-\boldsymbol{\mu})^{(t)} \mathbf{Q} (\mathbf{X}-\boldsymbol{\mu})/2}$$

où $\mathbf{Q} = \mathbf{C}^{-1}$ est appelée la **matrice de précision**.

L'espérance et la matrice de covariance de la distribution gaussienne multivariée sont

$$\mathbb{E}\mathbf{X} = \boldsymbol{\mu} \quad \text{et} \quad \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^{(t)}] = \mathbf{C}$$

On voit que la distribution gaussienne multivariée est uniquement caractérisé par son espérance et matrice de covariance (ou de précision). Les distributions conditionnelles de la gaussienne multivariée sont aussi gaussiennes.

Exercice 3.1 Soit (Y, Z) une partition des coordonnées d'un vecteur \mathbf{X} à distribution gaussienne $N(\mathbf{0}, \mathbf{C})$, et soit \hat{Y} la distribution de Y , en sachant Z .

a) Montrez que cette distribution est aussi gaussienne, avec matrice de précision \mathbf{Q}_{YY} et espérance $\hat{Y} = \mathbb{E}[Y|Z]$:

$$\hat{Y} = -(\mathbf{Q}_{YY})^{-1} \mathbf{Q}_{YZ} Z$$

b) Finalement, déduisez les formules :

$$\hat{Y} = \mathbf{C}_{YZ} (\mathbf{C}_{ZZ})^{-1} Z \quad \text{et} \quad \hat{\mathbf{C}}_{YY} = \mathbf{C}_{YY} - \mathbf{C}_{YZ} (\mathbf{C}_{ZZ})^{-1} \mathbf{C}_{ZY}$$

où $\hat{\mathbf{C}}_{YY}$ représente la covariance conditionnée de Y .

En utilisant ces formules d'abord quand $Y = X_i$ est une seule composante (dans quel cas on dénote les autres composantes par $Z = X_{-i}$), et ensuite quand $Y = (X_i, X_j)$ est un couple (dans quel cas on dénote les autres composantes par $Z = X_{-(i,j)}$), on arrive à :

Exercice 3.2

$$\mathbb{E}[X_i|X_{-i}] = -\frac{1}{Q_{i,i}} \sum_{j \neq i} Q_{i,j} x_j \quad \text{et} \quad \text{Cor}[(X_i, X_j)|X_{-(i,j)}] = -\frac{Q_{i,j}}{\sqrt{Q_{i,i} Q_{j,j}}}$$

L'extension au cas à moyennes nonnules, mais **connues** et immédiate; on arrive ainsi à la fameuse formule de régression qui prédit une variable Y par une fonction lineaire d'autres variables "explicatives" $Z_i, i = 1, \dots, I$:

$$\boxed{\hat{Y} = \mu_y + \sum_{i=1}^I a_i (Z_i - \mu_i)} \quad (7)$$

Le vecteur colonne $\mathbf{a} = (a_i, i = 1, \dots, I) = -(\mathbf{Q}_{YY})^{-1} \mathbf{Q}_{YZ}$ vérifie le système d'équations "normales" :

$$\boxed{\mathbf{C}_{Z,Z} \mathbf{a} = \mathbf{C}_{Z,Y}} \quad (8)$$

(où $\mathbf{C}_{i,j} = \text{Cov}(Z_i, Z_j)$ et $\mathbf{C}_{Z,Y} = (\text{Cov}(Z_i, Y), i = 1, \dots, I)$). On note que :

- les covariances entre les variables explicatives interviennent dans le premier membre
- les covariances entre les variables explicatives et la variable estimée interviennent dans le second membre.

La variance d'estimation correspondante est :

$$\sigma_E^2 = \mathbb{E}[(Y - \hat{Y})^2] = \sigma_Y^2 - \sum_{i=1}^N a_i \text{Cov}(Y, Z_i) = \sigma_Y^2 - \mathbf{C}_{Y,Z} \mathbf{a} \quad (9)$$

Elle permet de voir la contribution de chacune des variables dans la réduction de la variance (dans le cas univarié, ça donne : $\sigma_E^2 = \sigma_Y^2(1 - \rho^2)$).

Notes : 1) Le système normal (8) est très général ; en particulier, il généralise la formule de prévision univariée $\hat{Y} = \mu_Y + a(Z - \mu_Z)$, où $a = \frac{\text{Cov}(Y,Z)}{\text{Cov}(Z,Z)} = \frac{\sigma_{YZ}}{\sqrt{\sigma_Z \sigma_Y}}$ et le système de Yule-Walker pour les séries temporelles autoregressives.

2) On arrive au même système pour des modèles non gaussiennes, si on demande de trouver le "BLUE" : l'estimateur linéaire non biaisé de variance minimale (voir ci-dessous).

4 Modélisation stochastique de séries temporelles

La modélisation stochastique distingue en général entre trois types des processus :

1. observations Y_t
2. variables "latentes" inobservables directement X_t , soumis parfois aux diverses contraintes, et qui doivent être estimées.
3. des bruits inobservables ϵ_t

Un exemple de modélisation très générale est fournie par les **modèles espace-état** :

$$X_{t+1} = A_t X_t + b_t + \epsilon_X(t) \text{ équation d'évolution} \quad (10)$$

$$Y_t = C_t X_t + d_t + \epsilon_Y(t) \text{ équation d'observation} \quad (11)$$

Exemple 4.1 Le modèle additif

Idée : On cherche une décomposition de la forme :

$$Y_t = m_t + \epsilon_t \quad \text{où :}$$

- m_t représente la "tendance", qui est intuitivement un "mouvement lisse à long terme".
- $\epsilon_t = Y_t - m_t$, les "résidus" qui restent après qu'on enlève la partie structurée m_t , représentent des "irrégularités/fluctuations imprévisibles", qui au début semblent inutilisables (à ignorer).

Donc, ce modèle contient seulement l'équation d'observation. m_t peut-être traité aussi comme une variable latente, et avec l'ajout des informations sur son évolution, une deuxième équation d'évolution pourrait-être ajoutée.

Une telle décomposition est plus utile si on peut assurer que les résidus aient une structure statistique simple comme "bruit blanc Gaussien" ou "bruit blanc de second ordre", moyennes mobiles, etc.

4.1 Quelques exemples des processus stochastiques

4.1.1 Le bruit blanc

Définition 4.1 Un processus $\epsilon_t, t \in \mathbb{N}$ ou $t \in \mathbb{Z}$ est appelé **bruit blanc Gaussien** si les variables ϵ_t sont **i.i.d.** (indépendants et identiquement distribués) Gaussiennes et à espérance $\mathbb{E}\epsilon_t = 0$.

Notes : 1) Cette structure stochastique est la plus désirable possible, car la distribution Gaussienne possède plusieurs propriétés importantes, comme celle d'être invariante par rapport aux rotations (ce qui est évidemment une réquise pour un bruit aleatoire).

2) Un tel processus a une covariance

$$\gamma(s, t) = \mathbb{E}[\epsilon_s \epsilon_t] = 0, \forall s \neq t \quad \text{et donc le coefficient de corrélation} \quad (12)$$

$$\rho(s, t) = \frac{\gamma(s, t)}{\sigma_s \sigma_t} = \delta_{s,t} \quad (13)$$

où $\delta_{s,t}$ est le symbole du Kronecker).

3) Comme les tests d'indépendance et Gaussianité demandent beaucoup de données, qui ne sont pas toujours disponibles, il faut faire parfois avec un objet moins structuré : le "bruit blanc de second ordre" défini par les deux dernières formules équivalentes (12), (13) :

Définition 4.2 Un processus $\epsilon_t, t \in \mathbb{N}$ ou $t \in \mathbb{Z}$ est appelé **bruit blanc de second ordre** s'il a la moyenne 0, la variance constante $\mathbb{E}\epsilon_t^2 = \sigma^2$ et une covariance $\gamma(s, t) = \mathbb{E}[\epsilon_s \epsilon_t] = 0, \forall s \neq t$ (et donc les coefficients de corrélation $\rho(s, t) = \delta_{s,t}$).

La définition 4.1 la plus forte est celle requise pour des tests distributionnels des résidus comme Fisher, Student, etc ; la définition 4.2 peut servir pour une inspection graphique préliminaire (examination des corrélations des résidus), avant le test final du modèle 4.1.

L'intention est donc que la partie bruit blanc ne contient "aucune information", mais ce concept d'information n'est pas uniquement défini ; en plus, même la requise plus modeste du bruit noncorréolé n'est pas toujours testable (dans l'absence des données suffisantes), ce qui ramène au besoin de travailler parfois avec du bruit corréolé (voir le prochain sous-chapitre).

4.1.2 Les processus stationnaires

Quand on rencontre des résidus à des corrélations nonnules ("non-blancs"), le niveau de complexité suivant qu'on essaie d'adopter est celui d'un processus stationnaire, appelé aussi "bruit corréolé/coloré".

Définition 4.3 Soit X un processus aléatoire indexé par $T = \mathbb{N}$ ou \mathbb{Z} . On dit que X est **stationnaire strict** si pour toute famille finie d'instantants $t_1 \dots t_r \in T$ et tout entier s , les lois jointes de $(X_{t_1} \dots X_{t_r})$ et de $(X_{t_1+s} \dots X_{t_r+s})$ sont les mêmes.

Définition 4.4 Soit X un processus aléatoire indexé par $T = \mathbb{N}$ ou \mathbb{Z} . On dit que X est **stationnaire à l'ordre 2** si la moyenne $m(t)$ et la covariance $\Gamma(s, t)$ sont invariantes par translation dans le temps, i.e. si la moyenne est constante :

$$\mathbb{E}X_t = m_t = m, \forall t$$

et si la covariance/corrélation dépend seulement de l'écart de temps $k = t - s$, i.e. il existe une fonction d'une variable $\gamma(k)$, paire, telle que :

$$\text{Cov}(X_t, X_s) = \Gamma(t, s) = \gamma(t - s) = \gamma(k), \forall k = -2, -1, 0, 1, 2, 3, \dots$$

Un exemple important des processus stationnaires sont les "processus linéaires" $Y_t = \sum_i \psi_i \epsilon_{t-i}$ avec ϵ_t bruit blanc et leur cas particulier avec un nombre fini des coefficients ψ_i nonnuls, les "moyennes mobiles".

Les suites de nombres qui peuvent-être covariances sont uniquement caractérisés par leur transformé Fourier.

Théorème 4.1 Une suite paire $\gamma_k \in L_2$ peuvent-être les covariances d'une série stationnaire ssi la transformé Fourier -appelée aussi **densité spectrale**

$$f(w) = \gamma_0 + 2 \sum_1^{\infty} \gamma_k \cos(wk)$$

est nonnegative pour chaque w .

4.1.3 Les processus linéaires et moyennes mobiles MA(q)

Définition 4.5 Un processus Y_t sera appelé linéaire en ϵ_t s'il peut être représenté dans la forme :

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i} \tag{14}$$

où $\sum \psi_i^2 < \infty$ et ϵ_t est un bruit blanc.

Théorème 4.2 *Un processus linéaire*

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i}$$

où $\sum \psi_i^2 < \infty$ est : a) bien défini dans L_2 sense

b) à variance constante stationnaire $\text{Var } Y_t = \sigma_\epsilon^2 \sum_{i=-\infty}^{\infty} \psi_i^2$

c) à autocovariance donnée par :

$$\gamma(t, t+k) = \sigma_\epsilon^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+k} \quad (15)$$

d) stationnaire à ordre deux.

Démonstration : En considérant $\text{Var}(Y_t)$, on voit que la condition est nécessaire. En suite, on voit qu'elle suffit, car $\text{Cov}(Y_t, Y_{t+k})$ est bien définie et ne dépend pas de t . En effet, on utilise l'inégalité de Cauchy-Schwartz (qui est équivalente à $|\rho_k| \leq 1$).

Définition 4.6 On appelle processus $MA(q)$ un processus $Z_t, t \in \mathbb{Z}$ vérifiant une relation de récurrence :

$$Z_t = \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (16)$$

où $\theta_0 = 1$ et ϵ_t est un bruit blanc de variance σ^2 .

Il est convenable d'introduire un opérateur de retard B défini sur l'ensemble des suites par

$$BY_t = Y_{t-1}, B\epsilon_t = \epsilon_{t-1} \quad \text{et donc } B^i Y_t = Y_{t-i}, B^i \epsilon_t = \epsilon_{t-i}, \text{ et donc } \sum_{i=1}^q \theta_i \epsilon_{t-i} = \theta(B)\epsilon_t$$

où $\theta(B)$ dénote le polynôme $\sum_{i=1}^q \theta_i B^i$.

La notation des polynômes de retard ramène (16) à la forme :

$$Z_t = \theta(B)\epsilon_t$$

EXERCICE :

1. Calculer la fonction d'autocovariance $\gamma(s, t)$ d'un processus $MA(1)$.
2. Calculer la fonction d'autocovariance d'un processus $MA(q)$

Evidemment, du point de vue pratique (pour la prédiction), on ne s'intéresse que dans le cas –qui sera appelé **causal**– quand la représentation n'utilise pas "le bruit du futur" :

Définition 4.7 Un processus stationnaire Y_t s'appelle **causal** s'il peut être représenté dans la forme :

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (17)$$

où $\sum \psi_i^2 < \infty$

Viceversa,

Définition 4.8 Une représentation

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

d'un processus stationnaire Y_t s'appelle **inversible** si on peut aussi représenter le bruit comme :

$$\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i} \quad (18)$$

où $\sum \pi_i^2 < \infty$

Remarques :

1. La plupart des séries ne sont pas stationnaires, mais on essaie quand-même de se ramener à ce cas par des transformations (logarithmes, Box-Cox, etc).
2. Pour un processus du second ordre, la stationnarité stricte implique la stationnarité au sens large (à l'ordre 2). La réciproque est fautive. Une suite Y de v.a. indépendantes de même moyenne et même variance est toujours stationnaire à l'ordre 2 ; mais si les Y_n n'ont pas tous la même loi, Y n'est pas stationnaire au sens strict.
3. La stationnarité à l'ordre 2 est bien plus facile à étudier et vérifier que la stationnarité stricte. Son importance pratique tient surtout aux problèmes de prédiction ou de régression. En effet, on se limite souvent à des critères de moindres carrés pour avoir des estimateurs calculables. Cela signifie alors utiliser des prédicteurs linéaires optimaux dont le calcul ne fait pas intervenir dans sa totalité la structure probabiliste du processus X observé, mais seulement la géométrie (angles et longueurs) de la suite (X_k) considérée comme suite de vecteurs dans l'espace de Hilbert $L^2(\Omega, P)$. Or, cette géométrie ne dépend que des moments d'ordre 2 de X ; la notion naturelle de stationnarité est donc l'invariance de ces moments d'ordre 2 par translation dans le temps.
4. Théoriquement, l'estimation des caractéristiques statistiques observables une seule fois (le cas souvent dans les séries temporelles et la géostatistique !) est faisable seulement si on a la chance d'avoir à faire avec un processus stationnaire.

Avantages de la modélisation stochastique :

1. Produit des intervalles de confiance (au lieu des prévisions ponctuelles), qui expriment l'incertitude.
2. Permet, spécialement dans le cas où il y a beaucoup d'observations, une examination des hypothèses du modèle (par exemple par l'examination des résidus), qui peut suggérer finalement une amélioration du modèle.
3. La modélisation déterministe nécessite des choix arbitraires d'une forme paramétrique d'interpolation et d'un objectif d'optimisation (qui déterminera les paramètres du modèle). Par contre, la modélisation stochastique lie ces deux choix à l'espérance et covariance (ou plus généralement vraisemblance) du modèle $Y(x)$.

5 Les modèles autorégressifs

La prédiction d'une série est particulièrement simple quand elle peut-être "bien approximée" par un modèle "autorégressif" paramétrique :

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots)$$

avec des fonctions f tel que $f(x_1, x_2) = a + b_1x_1 + b_2x_2$, $f(x_1, x_2) = ae^{b_1x_1 + b_2x_2}$, etc. Dans ce cas, la régression classique nous donne les coefficients a, b_1, b_2 , et aussi des tests (Fisher, etc) pour décider si le modèle est une "bonne approximation".

Nous allons considérer ici les modèles autorégressifs linéaires :

Définition 5.1 Un processus stationnaire $Y_t, t \in \mathbb{Z}$ sera appelé **processus autorégressif d'ordre p : AR(p)** s'il existe un bruit blanc ϵ_t (de variance constante σ^2) et des réels $\varphi_i, i = 1, \dots, p$ tel que une relation de récurrence :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \epsilon_t, \forall t \in \mathbb{Z} \quad (19)$$

est vérifié.

La notation des polynômes de retard ramène (19) à la forme :

$$\varphi(B)Y_t = \epsilon_t$$

Définition 5.2 Le polynôme

$$\varphi(B) = 1 - \sum_{i=1}^p \varphi_i B^i$$

sera appelé **polynôme caractéristique**, ou **symbole** du modèle (19).

Rq : Les processus autorégressifs sont définies par une équation, qui à priori, peut ne pas avoir des solutions.

Exemple : processus de Markov AR(1). Montrez que l'équation :

$$Y_t = \phi Y_{t-1} + \epsilon_t \quad (20)$$

1. a une solution stationnaire unique si $|\phi| < 1$, qui depend seulement du **bruit présent et passé**.

Indication : vous pouvez le faire en calculant la solution (i) par des substitutions répétées ou (ii) en utilisant des operateurs, en posant $Y_t = (1 - \phi B)^{-1} \epsilon_t$, et en developpant la fraction comme une série de puissances en B . En suite, calculez les covariances, pour montrer la stationnarité.

2. pour $|\phi| > 1$, l'équation : (20) a aussi une solution stationnaire unique, qui depend seulement du bruit **futur**.
3. pour le cas $|\phi| = 1$, l'équation : (20) (appelée *marche aléatoire*) n'a pas de solution stationnaire.

Le bruit blanc de léquation autorégressive dans le premier cas a la propriété tr'es convenable de coincider avec l'erreur de prédiction par rapport au passé $Y_t - \hat{Y}_t = Y_t - \mathbb{E}[Y_t / Y_{t-1}, Y_{t-2}, \dots]$:

Définition 5.3 Si pour une suite stationnaire Y_t la différence $\epsilon_t = Y_t - \hat{Y}_t$ satisfait : a) ϵ_t est une fonction de $(Y_t, Y_{t-1}, Y_{t-2}, \dots)$ et

b)

$$\mathbb{E}[\epsilon_t / [Y_{t-1}, Y_{t-2}, \dots]] = 0 \quad (21)$$

alors elle sera appelée **bruit d'innovation** (par rapport au passé).

Rq : On montre facilement que pour un bruit d'innovation ϵ_t , les innovations satisfont $\mathbb{E}\epsilon_t = 0$ et $\mathbb{E}\epsilon_t\epsilon_{t-k} = 0$, pour $k \neq 0$. Donc, les innovations constituent un bruit blanc de deuxième ordre.

Nous allons démontrer plus tard que pour un modèle (19) pour le quel le polynôme caractéristique a seulement des racines plus grandes en valeur absolue que 1, on a :

$$\hat{Y}_t := \mathbb{E}[Y_t / \{Y_{t-1}, Y_{t-2}, \dots\}] = \sum_{i=1}^p \varphi_i Y_{t-i}$$

et le bruit est un bruit d'innovation.

Rémarque : Le modèle AR(p) sous la condition des racines plus grandes en valeur absolue que 1 fournit un exemple de modèle additif à bruit blanc d'innovation, satisfaisant (21).

Dans le prochain chapitre on verra un deuxième exemple, obtenu en admettant des modèles autoregressifs AR(p) avec du bruit **non-blanc**, de type moyenne mobile MA(q), appelées modèles ARMA(p, q) (toujours sous la même condition sur les racines).

Définition 5.4 On appellera "filtrage adapté" ou "information" pour la suite $Y_t, t \in \text{Nout} \in \mathbb{Z}$ une suite des sous-algèbres $F_t, t \in \text{Nout} \in \mathbb{Z}$ tel que

$$F_{t-1} \subset F_t \subset F_{t+1} \subset \dots \quad \text{et tel que} \\ Y_t \in F_t$$

i.e. F_t est croissante et Y_t est complètement déterminé en sachant F_t (ou encore "mesurable par rapport à F_t ").

L'information la plus simple est "autorégressive", donc la seule chose connue sont les valeurs passées du Y_t , i.e. $F_t = \{Y_t, Y_{t-1}, Y_{t-2}, \dots\}$. On vérifie facilement dans cet exemple que l'information est croissante et qu'elle contient Y_t .

Définition 5.5 Pour une suite et un filtrage adapté donné $(Y_t, F_t), t \in \text{Nout} \in \mathbb{Z}$, avec $F_{t-1} \subset F_t \subset F_{t+1} \subset \dots$, la décomposition adapté de Y_t est :

$$Y_t = m_t + \epsilon_t \quad \text{où :} \tag{22}$$

$$m_t = \hat{U}_t := \mathbb{E}[Y_t / F_{t-1}] \tag{23}$$

est la "prévision" de Y_t en sachant F_{t-1} . La différence $\epsilon_t = Y_t - \hat{Y}_t$ est appelée **innovation**.

Dans le cas autorégressif par exemple, la prévision est

$$m_t = \mathbb{E}[Y_t / Y_{t-1}, Y_{t-2}, \dots] = f(Y_{t-1}, Y_{t-2}, \dots)$$

Rémarque : L'information du passé F_t contient parfois, comme dans la régression classique, un ensemble des variables explicatives. Comme le nombre des variables explicatives dépend du choix de l'utilisateur, nous voyons que le concept d'information n'a pas uniquement définie.

5.1 Exercices : TD 2

1. Trouvez l'élément suivant des séries :

$$2, 6, 12, 20, 30, 42, .. \quad \text{et} \quad 4, 10, 20, 36, 62, 104, \dots$$

2. *Stationarité des processus*

En considérant les caractéristiques statistiques de second ordre (moyennes, variances, covariances), investiguer si les processus suivants sont stationnaires (de second ordre). On convient de noter par $\{Z_t\}$ le bruit blanc.

$$(a) \quad Y_t = Z_t + \theta_1 Z_{t-1} \tag{MA(1)}$$

$$(b) \quad Y_t = Y_{t-1} + \varphi + Z_t \quad (\varphi \neq 0) \tag{marche aléatoire avec tendance}$$

3. *Une question d'unicité - est ce que deux processus distincts peuvent avoir la même FAC (fonction d'autocovariance) ?*

Soient $\{u_t, t \in \mathbb{Z}\}$ et $\{v_t, t \in \mathbb{Z}\}$ deux bruit blancs de variances respectives σ^2 et $\theta^2\sigma^2$, o $0 < |\theta| < 1$. On considère alors les processus aléatoires $\{X_t, t \in \mathbb{Z}\}$ et $\{Y_t, t \in \mathbb{Z}\}$ tels que :

$$X_t = u_t + \theta u_{t-1} \\ Y_t = v_t + \frac{1}{\theta} v_{t-1}$$

Montrer que $\{X_t, t \in \mathbb{Z}\}$ et $\{Y_t, t \in \mathbb{Z}\}$ ont la même fonction d'autocovariance.

4. Une question d'inversibilité - est ce qu'un processus à représentation MA noninversible peut aussi avoir une autre représentation inversible ? Soit $\{U_t, t \in \mathbb{Z}\}$ le processus aléatoire défini par l'équation

$$U_t = u_t + \frac{1}{\theta}u_{t-1};$$

où u_t est bruit blanc. Montrer que cette représentation du processus U_t n'est pas inversible.

On pose

$$w_t = \sum_{j=0}^{+\infty} \theta^j U_{t-j}$$

- (a) Montrer que $\{w_t, t \in \mathbb{Z}\}$ est un bruit blanc dont on précisera la variance en fonction de σ^2 et θ
- (b) Montrer que $U_t = w_t + \theta w_{t-1}$ et que $\{U_t, t \in \mathbb{Z}\}$ est alors inversible.
5. (*) Restrictions sur les valeurs des coefficients d'autocorrélation pour les processus MA, et l'inversibilité
- a) Pour le processus MA(1) trouvez les valeurs maximales et minimales de ρ_1 et les valeurs de θ pour les quelles ces valeurs sont atteintes.
- b) Même question pour ρ_1 et ρ_2 pour le processus MA(2) (on trouvera θ_1 et θ_2).
- c) Investiguez sous quelles conditions le processus MA(1) est inversible.
6. a) Déterminez la corrélogramme des processus suivants :
- (i) le processus MA(2) $Y_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$
- (ii) le processus MA(3) $Y_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3}$
- b) Calculez et tracez la corrélogramme pour les cas :
- (i) MA(2) : $\theta_1 = 0.8, \theta_2 = 0.5$
- (ii) MA(3) : $\theta_1 = 0.8, \theta_2 = -0.4, \theta_3 = -0.3$
- c) (*) Investiguez si ces deux processus sont inversibles.
7. Calculer la fonction d'autocovariance du processus :

$$Y_n = \begin{pmatrix} a_0 \varepsilon_n + a_1 \varepsilon_{n-1} \\ b_1 \varepsilon_{n-1} + b_2 \varepsilon_{n-2} \end{pmatrix}$$

2) Quand ε_t est du bruit blanc, l'utilisation finale dans la prédiction est seulement à travers sa variance : $\sigma^2 = \mathbb{E}\varepsilon_t^2$; par contre, celle du bruit coloré se fait aussi à travers les covariances, comme en regression. Pour ça, il convient parfois de décomposer encore le bruit coloré comme $\varepsilon_t = Z_t + \tilde{\varepsilon}_t$ où $\tilde{\varepsilon}_t$ est du bruit blanc "supplémentaire" inobservable (provenant par exemple de l'imprecision des observations). Les deux composantes Z_t et $\tilde{\varepsilon}_t$ sont stochastiques, mais dans des sens différents :

- Z_t est un "bruit corréolé/coloré" stationnaire, qui reflète la structure locale du champs (les "influences locales entre les voisins"). En moyennant sur des zones locales, on arrive à estimer et à utiliser ensuite sa covariance dans la prévision.
- $\tilde{\varepsilon}_t$ est du "bruit blanc" totalement indéchiffrable, sauf sa variance ponctuelle σ^2 , qui détériore la prévision.

Quand-même, du point de vue des formules mathématiquement, les deux termes agissent toujours ensemble, i.e. à travers leur somme Y_t . Une des premières applications, qui a rendu Wiener célèbre, a été de décider à partir du quel point on peut être sur que des observations radar contiennent un signal ("cible"). La solution a relevé l'importance du "quotient signal/bruit". Comme dans la géostatistique la présence du signal est sûre, et l'introduction du bruit est facile, on suppose parfois que $\tilde{\varepsilon}_t = 0$.

5 Les modèles ARMA(p,q)

Il y a deux types principaux de modèle additif :

1. à bruit blanc, par exemple le modèle autoregressif AR(p), et
2. à bruit stationnaire non-blanc par exemple le modèle ARMA(p,q).

Définition 5.1 On appelle processus ARMA(p,q) un processus stationnaire $Y_t, t \in \mathbb{Z}$ vérifiant une relation de récurrence :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (24)$$

où les φ_i, θ_i sont des réels et ϵ_t est un bruit blanc de variance σ^2 .

La notation des polynômes de retard ramène (24) à la forme :

$$\varphi(B)Y_t = \theta(B)\epsilon_t$$

Il s'avère que les modèles ARMA(p,q) qui nous intéressent ont toujours aussi des représentations AR(p), MA(q), mais avec $p = \infty, q = \infty$.

Par exemple, rappelons (exercice 2 d) du TD antérieur) qu'un processus AR(1) a aussi une représentation causale MA(∞) ssi $|\varphi| < 1$ (obtenue : a) en résolvant la récurrence ou b) par l'inversion formelle du polynôme $\varphi(B) = 1 - \varphi B$. Donc, on a une représentation causale MA(∞) (en termes du bruit passé) du processus AR(1) ssi le **polynôme caractéristique** $\varphi(z) = 1 - \varphi z$ a sa racine à l'extérieur du cercle unitaire $|z| \leq 1$.

Nous verrons dessous que cette propriété se généralise : **les processus ARMA(p,q) avec des polynômes caractéristiques $\varphi(B), \theta(B)$ à racines dehors le cercle unitaire ont deux autres représentations équivalentes :**

1. MA(∞), de Y_t en termes de ϵ_t (appelée aussi représentation lineaire), et
2. AR(∞), de ϵ_t en termes de Y_t

Ces représentations peuvent être obtenues par des **inversions formelles** de l'équation (24), suivies par un développement de la fraction correspondante dans une série des puissances :

$$Y_t = \frac{\theta(B)}{\varphi(B)} \epsilon_t = \left(\sum_{i=0}^{\infty} \psi_i B^i \right) \epsilon_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad \epsilon_t = \frac{\varphi(B)}{\theta(B)} Y_t = \left(\sum_{i=0}^{\infty} \pi_i B^i \right) Y_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$$

Exemple 5.1 MA(1) Trouver la représentation AR(∞) (i.e. $\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$) du processus MA(1)

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

5.1 Causalité et inversibilité des modèles ARMA(p,q)

Nous avons vu qu'il y a un problème (non-causalité) avec les modèles AR(p) qui ont des racines de $\varphi(z)$ à l'intérieur du cercle unitaire, causés par le choix du développement formel du $\varphi(B)^{-1}$. Par exemple, pour le modèle AR(1) avec polynôme $\varphi(z) = 1 - z\varphi$ et racine $\lambda = \varphi^{-1}$, nous avons un développement

$$\frac{1}{1 - z\varphi} = \sum_{n=0}^{\infty} \varphi^n z^n \quad \text{si } |\lambda| > 1, \text{ à l'intérieur du cercle } |z| \leq \lambda, \text{ mais}$$

$$\frac{1}{1 - z\varphi} = - \sum_{n=-\infty}^{-1} \varphi^n z^n \quad \text{si } |\lambda| < 1, \text{ à l'extérieur du cercle unitaire, } |z| \geq \lambda$$

Ces problèmes disparaissent pour les modèles ARMA(p,q) qui ont toutes les racines de $\varphi(z)$ et $\theta(z)$ à l'extérieur du cercle unitaire :

Théorème 5.1 a) Un processus ARMA(p,q) avec toutes les racines du polynôme caractéristique $\varphi(z)$ à l'extérieur du cercle unitaire est causal, i.e. il peut être représenté sous la forme : $Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ où $\sum \psi_i^2 < \infty$ et donc Y_t appartient au espace linéaire engendré par le passé du bruit

$$Y_t \in sp\{\epsilon_{t-i}, i = 0, 1, \dots\}$$

Les coefficients ψ_i sont dans ce cas les coefficients de la série Taylor de $\psi(z) = \frac{\theta(z)}{\varphi(z)}$

b) Un processus ARMA(p,q) avec les racines du polynôme caractéristique $\theta(z)$ à l'extérieur du cercle unitaire est inversible, i.e. le bruit peut être représenté sous la forme : $\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ où $\sum \pi_i^2 < \infty$ et donc ϵ_t appartient au espace linéaire engendré par le passé du signal observé

$$\epsilon_t \in sp\{Y_{t-i}, i = 0, 1, \dots\}$$

Les coefficients π_i sont dans ce cas les coefficients de la série Taylor de $\pi(z) = \frac{\varphi(z)}{\theta(z)}$

En conclusion, comme du point de vue pratique les développements Laurent sont inacceptables pour la prévision (parce-qu'elles impliquent les valeurs futures, imprévisibles du bruit), on acceptera désormais seulement les modèles ARMA(p,q) avec toutes les racines de $\varphi(z)$ et $\theta(z)$ à l'extérieur du cercle unitaire, qui sont causales et inversibles.

Rémarque : Dans le cas le plus simple avec les racines λ_i de l'équation $\varphi(z) = 0$ distinctes, on peut calculer les coefficients ψ_n en commençant par un développement en fractions simples $\pi(z) = \frac{\varphi(z)}{\theta(z)} = \sum_i K_i \frac{1}{1-z/\lambda_i}$ où λ_i sont les racines du $\theta(z)$ et $K_i = -\frac{\theta(\lambda_i)}{\varphi'(\lambda_i)}$, en arrivant à : $\psi_n = \sum_{i=1}^p \frac{K_i}{\lambda_i^{n+1}}$. Des formules pareilles existent pour π_n .

Corollaire 5.1 Pour un processus ARMA(p,q) avec toutes les racines des polynômes caractéristiques $\varphi(z), \theta(z)$ à l'extérieur du cercle unitaire, les espaces linéaires engendrés par le bruit et le passé du signal coïncident :

$$sp\{Y_{t-i}, i = 0, 1, \dots\} = sp\{\epsilon_{t-i}, i = 0, 1, \dots\}$$

et

$$\boxed{\mathbb{E}Y_t \epsilon_{t+k} = 0, \forall k \geq 1}$$

Rémarque : Ce corollaire permettra un développement immédiate d'une approche de prévision (=régression) par projection dans l'espace de Hilbert engendré par le passé.

Théorème 5.2 a) Pour un processus ARMA(p,q) $\varphi(B)Y_T = \theta(B)\epsilon_t$ avec toutes les racines du polynôme caractéristique $\varphi(z)$ à l'extérieur du cercle unitaire, les coefficients $\psi_i = \sigma^{-2} \mathbb{E}Y_t \epsilon_{t-i}$ de la représentation causale $Y_t = \sum \psi_i \epsilon_{t-i}$ satisfont la récurrence

$$\begin{aligned} \psi_0 &= 1, \quad \psi_k = \theta_k + \sum_{i=1}^{\min[k,p]} \varphi(i)\psi(k-i), \quad 1 \leq k \leq q \\ \psi_k &= \sum_{i=1}^{\min[k,p]} \varphi(i)\psi(k-i), \quad k > q \end{aligned}$$

Note : La dernière équation est appelée **récession de Yule-Walker**

b) Pour un processus ARMA(p,q) $\varphi(B)Y_T = \theta(B)\epsilon_t$ avec toutes les racines du polynôme caractéristique $\theta(z)$ à l'extérieur du cercle unitaire, les coefficients π_i de la représentation inverse $\epsilon_t = \sum \pi_i Y_{t-i}$ satisfont la récurrence

$$\begin{aligned} \pi_0 &= 1, \quad \pi_k = -(\varphi_k + \sum_{i=1}^{\min[k,q]} \theta(i)\pi(k-i)), \quad 1 \leq k \leq p \\ \pi_k &= \sum_{i=1}^{\min[k,q]} \theta(i)\pi(k-i), \quad k > p \end{aligned}$$

Note : Les représentations inverse/causale permettent d'appliquer aux processus ARMA(p,q) les méthodes adaptés aux modèles AR(∞)/MA(∞).

Ainsi, $\psi_1 = \varphi_1 + \theta_1$, $\psi_2 = \varphi_2 + \varphi_1^2 + \theta_1\varphi_1 + \theta_2$, $\psi_3 = \varphi_3 + 2\varphi_1\varphi_2 + \varphi_3 + (\varphi_2 + \varphi_1^2)\theta_1 + \varphi_1\theta_2 + \theta_3, \dots$

Exemple 5.2 ARMA(1,1)

1. Trouver la représentation $MA(\infty)$ d'un processus ARMA(1,1) causale

$$Y_t = \varphi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

2. Trouver la représentation AR(∞) d'un processus invertible ARMA(1,1).

Exemple 5.3 Le domaine (φ_1, φ_2) de causalité d'un processus AR(2)

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \epsilon_t$$

est beaucoup plus compliquée que pour le AR(1). On obtient le triangle situé en dessous de $\varphi_2 + \varphi_1 < 1$, $\varphi_2 - \varphi_1 < 1$ (obtenues en jettant dehors les paires qui ont des racines réelles dedans, parce que $\varphi(1) < 0 / \varphi(-1) < 0$), et dessus $\varphi_2 > -1$ (obtenue en jettant dehors les paires qui ont des racines complexes dedans parce que $|z_i|^2 = |z_1 z_2| = |\frac{c}{a}| = \frac{1}{-\varphi_2} < 1$) Les racines sont réelles/complexes dessus/dessous la parabole $\varphi_2 = \frac{-\varphi_1^2}{4}$. **Note :** L'analyse relève deux autres régions qui pourraient "diminuer" le triangle : c'est le cas d'une parabole convexe avec deux racines réelles positives en $(0, 1)$ et son symétrique. Mais ces cas sont impossibles : par exemple, le premier demande $-1 < \varphi_2 < 0, 0 < z^* = \frac{-\varphi_1}{2\varphi_2} < 1$, qui n'intersecte pas la "partie réelle" du triangle.

La corrélogramme des processus autorégressifs AR(p) et ARMA(p,q) n'est pas aussi facile à calculer que celle des processus MA(q). Pour cela, c'est intéressant de comparer deux solutions :

1. Solution indirecte, en représentant d'abord le processus AR(p) comme un processus $MA(\infty)$ en inversant le filtre $\varphi(B)$; en suite on utilise la formule (15) de la corrélogramme pour les processus $MA(\infty)$.
2. Solution directe, en utilisant les équations de Yule-Walker.

5.2 Équations Yule-Walker pour les covariances/corrélations des modèles autorégressifs AR(p) causales

Exemple 5.4 AR(1) : Calcul de la corrélogramme par :

1. Les équations Yule-Walker.
2. La représentation $MA(\infty)$.

Tracez la corrélogramme $\{\rho_k\}$ pour

(a) $\varphi = 0.5$, (b) $\varphi = 0$, and (c) $\varphi = -0.5$.

Les covariances et les corrélations d'un processus AR(p) sont liées aux coefficients $\varphi = (\varphi_1, \dots, \varphi_p)$ par les équations de Yule-Walker. Ces équations s'obtiennent des "équations normales" en régression, ou, directement :

1. Pour les covariances, en multipliant la formule autoregressive de Y_{t+k} par Y_t et en prenant l'espérance, on obtient :

$$\begin{aligned} \gamma_k &= \sum_{i=1}^p \varphi_i \gamma_{k-i} && \text{pour } k \geq 1 && (25) \\ \gamma_0 &= \sum_{i=1}^p \varphi_i \gamma_i + \mathbb{E}Y_t \epsilon_t = \sum_{i=1}^p \varphi_i \gamma_i + \sigma^2 && \text{pour } k = 0 \end{aligned}$$

2. Pour trouver les corrélations, on remarque d'abord, en divisant par γ_0 , qu'elles satisfont aussi la récurrence (25) :

$$\rho_k = \sum_{i=1}^p \varphi_i \rho_{k-i} \quad \text{pour } k \geq 1 \quad (26)$$

Alors, il suffit de déterminer les premières p corrélations, ce qu'on fait en appliquant la récurrence (26) pour $k = 1, \dots, p$, en tenant compte aussi de la symétrie de γ_k . En général, pour le processus $AR(p)$ on arrive ainsi au **système Yule-Walker** pour les premières p corrélations $\boldsymbol{\rho} = (\rho(1), \dots, \rho(p))$:

$$R\boldsymbol{\rho} = \boldsymbol{\rho} \quad (27)$$

où R est la matrice Toeplitz symétrique :

$$\begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \dots & \dots & \dots & \dots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix}$$

Les équations (27) permettent de calculer d'abord les premières p corrélations ; en suite, on peut trouver aussi les autres corrélations en utilisant la récurrence.

Note : En prenant γ_0 comme facteur commun dans la deuxième équation en (25), on trouve

$$\gamma_0 = \frac{\sigma^2}{1 - \sum_i \varphi_i \rho_i},$$

i.e. γ_0 en fonction des corrélations ρ_i . Ça permet en suite d'obtenir les covariances, en partant des corrélations.

Exemple 5.5 $AR(2)$:

Vérifiez si les processus $AR(2)$:

a) $Y_t = -0.5Y_{t-1} + 0.14Y_{t-2} + \epsilon_t$ et

b) $Y_t = -0.6Y_{t-2} + \epsilon_t$ sont stationnaires causals. Montrez, en partant directement du système de Yule-Walker que leur corrélogrammes sont :

a) $\rho_k = \frac{17}{129}(0.2)^k + \frac{112}{129}(-0.7)^k, \quad k = 0, 1, 2, \dots$ et

b) $\rho_k = \frac{1}{2}i^k(0.6)^{k/2}\{1 + (-1)^k\} = (0.6)^{k/2} \cos(k\pi/2), \quad k = 0, 1, 2, \dots$

Exemple 5.6 Tracez les corrélogrammes pour les processus $AR(2)$ avec :

(i) $\varphi_1 = 0.2, \quad \varphi_2 = 0.35$ et (ii) $\varphi_1 = -0.8, \quad \varphi_2 = -0.16$

Théorème 5.3 (*) **Formule générale des corrélations pour $AR(2)$:** Pour le processus $AR(2)$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \epsilon_t$$

le système de Yule-Walker (27) donne :

$$\rho_1 = \frac{\varphi_1}{1 - \varphi_2}, \quad \rho_2 = \varphi_2 + \frac{\varphi_1^2}{1 - \varphi_2}$$

(ce qui implique la restriction $\rho_1^2 < \frac{1}{2(1+\rho_2)}$).

1. Si les racines λ_1, λ_2 de $\varphi(z) = 0$ sont distinctes, on obtient

$$\rho_k = \frac{\rho_1 - \lambda_2^{-1}}{\lambda_1^{-1} - \lambda_2^{-1}} \lambda_1^{-k} + \frac{\lambda_1^{-1} - \rho_1}{\lambda_1^{-1} - \lambda_2^{-1}} \lambda_2^{-k}$$

Aussi, en utilisant $\varphi_1 = \lambda_1^{-1} + \lambda_2^{-1}, \varphi_2 = -\lambda_1^{-1} \lambda_2^{-1}$, on arrive à

$$\rho_k = \frac{(1 - \lambda_2^2)\lambda_1^{k+1} - (1 - \lambda_1^2)\lambda_2^{k+1}}{(\lambda_1 - \lambda_2)(1 + \lambda_1\lambda_2)}, \quad k \geq 0.$$

termes de racines seulement.

2. Dans le cas de racines confondues $\lambda_i = \lambda$, on obtient :

$$\rho_k = \left\{ 1 + \left(\frac{1 - \lambda^2}{1 + \lambda^2} \right) k \right\} \lambda^k, \quad k \geq 0.$$

5.3 Contrôle continu : TD3

1. Trouvez les représentations $MA(\infty)$ et $AR(\infty)$ des processus $ARMA(2,1)$ définies par :

a) $Y_t - \frac{1}{2}Y_{t-1} - \frac{3}{16}Y_{t-2} = \epsilon_t + 1.25\epsilon_{t-1}$ b) $(1 - B + \frac{B^2}{4})Y_t = (1 + B)\epsilon_t$

2. Obtenez, en partant directement du système de Yule-Walker, les premières cinq corrélations pour un processus $AR(2)$ avec : a) $\phi_1 = 0.6, \phi_2 = -0.2$ b) $\phi_1 = -0.6, \phi_2 = 0.2$ Calculez aussi la variance $\gamma(0)$. Tracez les corrélations.

3. a) Vérifiez si le processus $AR(2)$ $Y_t = -0.3Y_{t-1} + 0.10Y_{t-2} + \epsilon_t$ est stationnaire causal. Calculez son corrélogramme, en partant directement du système de Yule-Walker, et tracez le.

Même questions pour le processus $AR(2)$ $Y_t = -Y_{t-1} - 0.34Y_{t-2} + \epsilon_t$.

4. Soit Y_t un processus $ARMA(1,1)$ vérifiant l'équation $Y_t - 0.5Y_{t-1} = \epsilon_t + 0.4\epsilon_{t-1}$ avec ϵ_t un bruit blanc.

(a) Précisez si le processus est stationnaire, causal et inversible, et calculez sa fonction d'autocovariance.

(b) Trouvez les coefficients ψ_j de sa représentation comme processus $MA(\infty)$.

Mêmes questions pour $Y_t = -0.5Y_{t-1} + \epsilon_t - 0.8\epsilon_{t-1}$, $Y_t = 0.7Y_{t-1} - 0.1Y_{t-2} + \epsilon_t$.

5. (*) *Le triangle d'inversibilité pour les processus $MA(2)$*

Rappel : Un processus MA est inversible si les racines du polynôme $\theta(B)$ sont toutes en module supérieures à un.

(a) Trouvez les inégalités qui définissent la région pour laquelle un processus $MA(1)$ est inversible.

(b) Trouvez les inégalités (il y en a trois) qui définissent la région (triangulaire) du plan (θ_1, θ_2) pour laquelle un processus $MA(2)$ est inversible. Tracez la région sur un graphe.

6. (*) *Restrictions sur les valeurs des coefficients d'autocorrélation pour les processus MA*

Pour le processus $MA(2)$, trouvez, en partant de l'exercice antérieur, un domaine S contenant toutes les valeurs possibles de ρ_1, ρ_2 tel que le processus soit inversible, et les valeurs de θ_1, θ_2 pour lesquelles les valeurs sur la frontière de S sont atteintes.

Solutions :

5) Pour simplifier, nous prenons le polynôme $\tilde{\theta}(z) = z^2 + \theta_1 z + \theta_2$. La condition d'avoir racines de module moins que 1 est différente pour le cas des racines complexes et celui des racines réelles.

(a) racines complexes : $|z_i|^2 = |z_1 z_2| = \left| \frac{c}{a} \right| = |\theta_2| < 1$.

(b) racines réelles : $\tilde{\theta}(1) > 0, \tilde{\theta}(-1) > 0$

Les racines sont réelles/complexes dessous/dessus la parabole $\theta_2 = \frac{\theta_1^2}{4}$.

En conclusion, la région de inversibilité dans le domaine (θ_1, θ_2) :

$$\begin{aligned}\theta_2 &> -\theta_1 - 1 \\ \theta_2 &> \theta_1 - 1 \\ \theta_2 &< 1\end{aligned}$$

est le triangle situé dessus les deux lignes $\theta_2 + \theta_1 = -1, \theta_2 = \theta_1 - 1$ et dessous la ligne $\theta_2 < 1$.

6) Transformant les équations antérieures, on trouve :

$$\begin{aligned}\rho_2 &\geq \rho_1 - .5 \\ \rho_2 &\geq -\rho_1 - .5 \\ \rho_1^2 &= 4\rho_2(1 - 2\rho_2) \text{ dessous}\end{aligned}$$

5.4 Équations de Yule-Walker pour les covariances/corrélations des processus ARMA(p,q) causales

On obtient les mêmes équations de Yule-Walker $\gamma_k = \sum_{i=1}^p \varphi_i \gamma_{k-i}$ pour $k \geq N = q + 1$. Par contre, les premières équations pour $k < N$ deviennent plus compliquées. Soit $\gamma_k^{(Y,\epsilon)} = \mathbb{E}[Y_t \epsilon_{t-k}]$ les corrélations de la série avec le bruit, données par $\psi_k \sigma^2$ si $k \geq 0$ et par 0 outrement (par la causalité). On obtient :

$$\gamma_k = \sum_{i=1}^p \varphi_i \gamma_{k-i} + \sum_{0 \leq j \leq q} \theta_j \gamma_{k-j}^{(Y,\epsilon)} = \sum_{i=1}^p \varphi_i \gamma_{k-i} + \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{k-j} \quad (28)$$

Pour appliquer la recursion, il faut obtenir les coefficients $\psi_j, j = 1, \dots, p$ et aussi les p valeurs initiales $\gamma_q, \dots, \gamma_{q-p+1}$, qu'on trouvera en utilisant les équations (28) et la symétrie de γ_k .

Rappel : Pour un processus ARMA(p,q) causal, les coefficients $\psi_i = \sigma^{-2} \mathbb{E}Y_t \epsilon_{t-i}$ de la représentation causale $Y_t = \sum \psi_i \epsilon_{t-i}$ satisfont la récurrence

$$\begin{aligned}\psi_0 &= 1, \quad \psi_k = \theta_k + \sum_{i=1}^{\min[k,p]} \varphi(i) \psi(k-i), \quad 1 \leq k \leq q \\ \psi_k &= \sum_{i=1}^{\min[k,p]} \varphi(i) \psi(k-i), \quad k > q\end{aligned}$$

Exemple 5.7 ARMA(1,1)

1. Montrez que $\rho_1 = \varphi + \frac{\theta \sigma^2}{\gamma_0}, \gamma_0 = \frac{\sigma^2(1+\theta^2+2\theta\varphi)}{1-\varphi^2}$, et $\{\rho_k\}$ est :

$$\rho_k = \frac{(\varphi + \theta)(1 + \varphi\theta)}{1 + 2\varphi\theta + \theta^2} \varphi^{k-1}, k \geq 1.$$

2. Tracez la corrélogramme pour le cas : $\varphi = \pm 0.7, \theta = \pm 0.5$.

En général, nous trouvons :

Théorème 5.4 (*) Les premières $p + 1$ covariances s'obtiennent du système à $p + 1$ équations et $p + 1$ inconnues :

$$\Gamma \begin{pmatrix} 1 \\ -\varphi_1 \\ -\varphi_2 \\ \dots \\ -\varphi_p \end{pmatrix} = \sigma^2 \Psi \begin{pmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_q \end{pmatrix} \quad (29)$$

ou $\mathbf{\Gamma}$ est la matrice des covariances

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-1) \\ \dots & \dots & \dots & \dots \\ \gamma(p) & \gamma(p-1) & \dots & \gamma(0) \end{pmatrix}$$

$\mathbf{\Psi}$ est la matrice des dimensions $(p+1) \times (q+1)$:

$$\mathbf{\Psi} = \begin{pmatrix} \psi(0) & \psi(1) & \dots & \psi(q) \\ 0 & \psi(0) & \dots & \psi(p-1) \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \psi(0) \end{pmatrix}$$

et $\psi_i = \sigma^{-2} \mathbb{E}Y_t \epsilon_{t-i}$, $i = 0, \dots, q$ sont calculés par la récurrence

$$\psi_0 = 1, \quad \psi_k = \theta_k + \sum_{i=1}^{\min[k,p]} \varphi(i) \psi(k-i), \quad 1 \leq k \leq q$$

6 La prévision linéaire

On se propose de donner au temps t une prévision $\hat{X}_t(k)$ de la valeur X_{t+k} d'un processus. Donc

1. t est le temps de prévision
2. $k > 0$ et l'écart de prévision : "lead time"
3. $t + k$ est le temps à prédire.
4. $\hat{X}_t(k)$ est la prévision
5. $e_t(k) = X_{t+k} - \hat{X}_t(k)$ seront les erreurs de prévision.

Comme les processus ARMA(p,q) satisfont des contraintes linéaires, il est naturel (même avec des bruits non-Gaussiens) de chercher une **prévision linéaire** $X_t(k)$ par une combinaison linéaire de valeurs passées ou du bruit blanc, à variance minimale, c'est à dire,

$$X_t(k) = \sum_{i=0}^{\infty} \pi_{t,k}(i) X_{t-i} \text{ ou}$$

$$X_t(k) = \sum_{i=0}^{\infty} a_{t,k}(i) \epsilon_{t-i}, k = 1, 2, \dots$$

La prévision linéaire à variance minimale des processus ARMA(p,q) coïncide avec l'espérance conditionnelle

$$\hat{X}_t(k) = \mathbb{E}[X_{t+k} | F_t]$$

où $F_t = \{X_t, X_{t-1}, \dots, X_0, X_{-1}, \dots\}$, et donc on peut profiter du fait que l'opérateur d'espérance conditionnelle $\mathbb{E}[\cdot | F_t]$ est linéaire. Deux notations pour cette opérateur apparaissent dans la littérature :

$$\hat{X}_t(k) = \hat{X}(t+k|t) = \mathbb{E}[X_{t+k} | F_t]$$

Toutes les trois représentations $MA(\infty)$, $AR(\infty)$ et $ARMA(p, q)$ nous aideront dans la prévision.

Théorème 6.1 (*) Soit $e_t(k) = X_{t+k} - X_t(k)$ l'erreur de prévision de la prédiction linéaire optimale $X_t(k) = \sum_{i=0}^{\infty} a_{t,k}(i) \epsilon_{t-i}$ d'un processus stationnaire causal X_t , et soit $V(k) = \mathbb{E}[X_{t+k} - X_t(k)]^2$ sa variance. Alors, les coefficients de prévision ont donné par

$$a_{t,k}(i) = \psi_{k+i}, i = 0, 1, \dots$$

et donc

$$X_t(k) = \sum_{i=0}^{\infty} \psi_{k+i} \epsilon_{t-i}, \text{ l'erreur est}$$

$$e_t(k) = \sum_{i=0}^{k-1} \psi_i \epsilon_{t+k-i}$$

et la variance de ce prédicteur est :

$$V(k) = \sigma^2 \sum_{i=0}^{k-1} \psi_i^2$$

Démonstration : Utilisant le développement linéaire en bruit de X_t , i.e. $X_t = \psi(B)\epsilon_t$ avec $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$ et ψ_i sont les coefficients de la représentation causale. On trouve que les coefficients $a_k(i)$ qui minimisent la variance de l'erreur : $V(k) = \mathbb{E}[X_{t+k} - X_t(k)]^2$ sont $a_{t,k}(i) = \psi_{k+i}$ et l'erreur de prévision peut s'exprimer comme $R_t(k) = \sum_{i=0}^{k-1} \psi_i \epsilon_{t+k-i}$.

Notes :

1. Ce résultat fournit des **intervalles de confiance pour la prévision**.
2. Pour $k=1$, $e_t(1) = \epsilon_t(1)$ et $V(1) = \sigma^2$

3. Ce résultat suggère que pour un processus non stationnaire, la variance de l'erreur de prévision converge vers infini :

$$\lim_{k \rightarrow \infty} V(k) = \infty$$

La représentation autoregressive $AR(\infty)$ $\epsilon_t = X_t + \sum_{i=1}^{\infty} \pi_i X_{t-i}$ nous fournit directement une formule explicite :

$$\hat{X}(t, t-1) = \sum_{i=1}^{\infty} \pi_i X_{t-i} = (1 - \pi(B))X_t$$

Donc, le problème de prévision 1 pas en avant se réduit formellement à travers la décomposition

$$X_t = \hat{X}(t, t-1) + \epsilon_t = (1 - \pi(B))X_t + \pi(B)X_t$$

au calcul des deux filtres $1 - \pi(B)$ et $\pi(B)$. Dans la prochaine section, nous verrons que des formules plus explicites sont aussi disponibles quand la représentation autorégressive $AR(p)$ a un ordre p fini.

6.1 La prévision des processus stationnaires en représentation autorégressive $AR(p)$

Nous considérons maintenant la prévision des processus stationnaires dans la forme $AR(p)$, en permettant $p = \infty$, et en utilisant la notation φ_i au lieu de π_i . La méthode sera "d'appliquer l'opérateur chapeau" dans l'équation définissant le modèle.

Théorème 6.2 *Pour un modèle $AR(p)$ $\varphi(B)X_t = \epsilon_t$ tel que le symbole $\varphi(z)$ ne s'annule pas dans le cercle unitaire, le bruit ϵ_t est un bruit d'innovation, i.e. $\mathbb{E}\epsilon_{s+k}X_s = 0$ si $k > 0$, et :*

$$\hat{X}_t := \mathbb{E}[X_t / \{X_{t-1}, X_{t-2}, \dots\}] = \sum_{i=1}^p \varphi_i X_{t-i}$$

Démonstration immédiate, par la causalité.

Exemple 6.1 *La prévision linéaire $X_t(k)$ pour un processus $AR(1)$ à moyenne 0 satisfait la récursion Yule Walker*

$$X_t(k) = \varphi X_t(k-1)$$

et donc est simplement

$$X_t(k) = X_t \varphi^k$$

Pour un processus $AR(1)$ à moyenne μ elle est

$$X_t(k) - \mu = (X_t - \mu) \varphi^k$$

Exemple 6.2 *Soit*

$$\varphi(B)X_t = (1 - \lambda_1 B)(1 - \lambda_2 B)X_t = \epsilon_t$$

un processus $AR(2)$, avec λ_1, λ_2 étant les inverses des racines de $\varphi(z) = 0$. Montrez que les prévisions Box-Jenkins $X_t(k)$ au temps t satisfont la récursion :

$$\varphi(B)X_t(k) = X_t(k) - (\lambda_1 + \lambda_2)X_t(k-1) + \lambda_1 \lambda_2 X_t(k-2) = 0 \quad (30)$$

et donc les prévisions sont de la forme :

$$X_t(k) = A_1(t)\lambda_1^k + A_2(t)\lambda_2^k \quad (31)$$

En termes des deux dernières valeurs observées X_t et X_{t-1} , les prévisions sont données par :

$$X_t(k) = \frac{\lambda_1^{k+1} - \lambda_1^{k+1}}{\lambda_1 - \lambda_2} X_t - \frac{\lambda_1 \lambda_2 (\lambda_1^k - \lambda_1^k)}{\lambda_1 - \lambda_2} X_{t-1} \quad (32)$$

Supposons maintenant que p est fini et considérons le problème de déterminer la projection $\hat{X}_t(1) = \sum_{i=1}^l \pi_{t,l}(i) X_{t+1-i}$ dans l'espace $F_{t,k} = \{X_t, X_{t-1}, \dots, X_{t-l+1}\}$, où $l \geq p$. Soit $\boldsymbol{\pi}^{(t,l)} = (\pi_{t,l}(i), i = 1, \dots, l)$ le vecteur des coefficients de la régression. Par le théorème 6.2, il est nécessaire que $\boldsymbol{\pi}^{(t,l)} = (\varphi_1, \dots, \varphi_p, 0, \dots, 0)$.

Note : Comme nous sommes ici dans le cadre d'une régression classique avec du bruit indépendant, il est nécessaire que les coefficients $\boldsymbol{\pi}^{(t,l)}$ satisfont

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(l-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(l-2) \\ \dots & \dots & \dots & \dots \\ \gamma(l-1) & \gamma(l-2) & \dots & \gamma(0) \end{pmatrix} \boldsymbol{\pi}^{(t,l)} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \dots \\ \gamma(l) \end{pmatrix}$$

ou encore (en divisant par $\gamma(0)$),

$$\begin{pmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p-2) \\ \dots & \dots & \dots & \dots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{pmatrix} \boldsymbol{\pi}^{(t,l)} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(l) \end{pmatrix}$$

Pour $l = p$, on retrouve ainsi le système Yule-Walker, et pour $l > p$ il n'est pas difficile de voir que les vecteurs des coefficients φ_i étendu par zéros satisfont aussi des systèmes Yule-Walker.

Considérons maintenant le problème de déterminer si $p < \infty$; autrement dit, est-ce que les systèmes Yule-Walker ont toujours comme dernière composante un 0, à partir d'un point p ?

Définition 6.1 La suite $\varphi_n^{(n)}$ des dernières composantes des systèmes Yule-Walker d'ordre n est appelée la suite des **corrélations partielles**.

Par le théorème 6.2

Théorème 6.3 Une série stationnaire est $AR(p)$ avec $p < \infty$ ssi toutes les corrélations partielles sont 0 à partir du point $p + 1$.

Le vrai ordre p du modèle est inconnu (et potentiellement infini). Pour le déterminer, on peut commencer en supposant $p \geq 1$; on calcule $\hat{\varphi}_1$ en supposant $p = 1$:

$$\hat{\varphi}_1^{(1)} = \rho_1$$

Si ce coefficient est "statistiquement" 0, ça sugere qu'on a à faire peut-etre avec du bruit blanc. On continue de toute façon en investiguant $p \geq 2$; On calcule $\hat{\varphi}_1, \hat{\varphi}_2$ en supposant $p = 2$; on trouve alors :

$$\hat{\varphi}_2^{(2)} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (33)$$

Si le deuxième coefficient $\hat{\varphi}_2^{(2)}$ est "statistiquement" 0, on soupçonne qu'on a à faire avec un modèle

$AR(1)$. On passe à investiguer $p \geq 3$, en calculant le coefficient $\hat{\varphi}_3^{(3)} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}}$ qui est

l'estimation de φ_3 obtenue du système d'ordre 3, puis à $p \geq 4, \dots$

Donc pratiquement, on calcule un après l'autre quelques corrélations partielles, en essayant de voir s'il sont "statistiquement" 0. Si oui, à partir d'un point $p + 1$, on conclut qu'on a à faire avec un modèle $AR(p)$. Si non (et si les corrélations ne deviennent 0 non plus), on passe à chercher un modèle mixte $ARMA(p, q)$.

Pour les modèles $ARMA(p, q)$, au lieu d'utiliser les représentations $MA(\infty), AR(\infty)$, il est plus simple d'utiliser directement la définition du modèle. On vera ça dans le cadre plus général des processus **non stationnaires** $ARIMA(p, d, q)$.

7 Les modèles ARIMA(p,d,q)

Définition 7.1 On appelle processus ARIMA(p,d,q) un processus X_t pour le quel le processus différencié d'ordre d , $Y_t = (1 - B)^d X_t, t \in \mathbb{Z}$ vérifie une relation de récurrence ARMA(p,q) :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (34)$$

où les φ_i, θ_i sont des réels et ϵ_t est un bruit blanc de variance σ^2 .

La notation des polynômes de retard ramène (34) à la forme :

$$\boxed{\varphi(B)Y_t = \varphi(B)(1 - B)^d X_t = \theta(B)\epsilon_t}$$

Soit X_t un processus ARIMA(p,d,q)

$$\phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t$$

où ϵ_t est "bruit blanc" (i.e. un processus stationnaire à corrélation 0) et $\phi(B), \theta(B)$ sont des polynômes dans l'opérateur de retard B à ordres p, q et avec coefficient libre 1.

7.1 Prédiction linéaire des modèles autorégressifs ARIMA(p,d,0)

Exemple 7.1 La prédiction linéaire $X_t(k)$ pour le processus ARIMA(0,1,0) à moyenne μ satisfait la récursion Yule Walker

$$X_t(k) = X_t(k - 1)$$

et est donc constante

$$X_t(k) = X_t$$

(c'est un cas particulier de la formule dans l'exercice sur la prédiction AR(1)).

Exemple 7.2 Déduisez la formule de prédiction Box-Jenkins pour un processus ARIMA(1,1,0) Calculez la limite $\lim_{k \rightarrow \infty} X_t(k)$ pour un processus ARIMA(1,1,0)

En conclusion, nous voyons que le "type" de la fonction de prédiction $X_t(k)$ dans le cas des bruits indépendants (sans partie MA) est déterminé complètement par la fonction $\varphi(z)$, et on verra que ça reste vrai pour les processus ARIMA(p,d,q), pour $k > q$.

Proposition 1 La fonction de prédiction "eventuelle" de Box-Jenkins pour les processus ARIMA(p,d,q) est un élément de l'espace linéaire des solutions de la récursion $\varphi(B)X_t(k)$, pour $k > q$.

Par exemple, pour les processus ARIMA(0,d,q) la fonction de prédiction "eventuelle" est un polynôme d'ordre $d - 1$.

Exemple 7.3 On considère un processus $\{X_t\}$ pour lequel la série différenciée deux fois est un bruit blanc, c'est à dire $\{X_t\}$ est un processus ARIMA(0,2,0). Montrez que la fonction de prédiction Box-Jenkins est donnée par

$$X_t(k) = X_t + k(X_t - X_{t-1}), k \geq 0.$$

donc les prévisions se trouvent sur la droite qui passe par les deux derniers points.

Définition 7.2 Les dernières $p + d$ valeurs $(X_t(q), X_t(q - 1), \dots, X_t(q - d - p + 1))$ qui précèdent $X_t(q)$ (donc avant le point où la récursion de Yule Walker devient valable) s'appellent les **valeurs pivots**.

Il suit clairement que :

Proposition 2 La prédiction "eventuelle" de Box-Jenkins pour les processus ARIMA(p,d,q) est la fonction dans l'espace linéaire des solutions de la récursion $\varphi(B)X_t(k)$ qui passe par les valeurs pivots.

Corollaire 7.1 La prédiction linéaire $X_t(k)$ pour le processus ARIMA(0,d,0) est donnée par le polynôme d'ordre $d - 1$ qui passe par les d derniers points.

7.2 Exercices : TD 4

1. Trouvez les coefficients d'autocorrélation et tracez la corrélogramme pour le processus ARMA(1,2)
 $Y_t = 0.6Y_{t-1} + \epsilon_t - 0.3\epsilon_{t-1} - 0.1\epsilon_{t-2}$
2. Prédiction du modèle ARIMA(2,1,0)

$$(1 - \phi_1 B)(1 - \phi_2 B)(1 - B)X_t = \epsilon_t \quad \text{with } -1 < \phi_1 < \phi_2 < 1.$$

(a) Vérifiez que :

- (i) $X_t(k) = (1 + \phi_1 + \phi_2)X_t(k-1) - (\phi_1 + \phi_2 + \phi_1\phi_2)X_t(k-2) + \phi_1\phi_2X_t(k-3)$, $k \geq 1$
- (ii) $X_t(k) = A_t + B_t\phi_1^k + (X_t - A_t - B_t)\phi_2^k$, $k = 0, 1, 2, \dots$ pour certaines A_t et B_t qui dépendent seulement du t .

(b) Trouvez les poids des valeurs pivots X_t, X_{t-1}, X_{t-2} dans la prédiction Box-Jenkins de X_t

(c) Trouvez la limite $\lim_{k \rightarrow \infty} X_t(k)$

7.3 Prédiction linéaire des modèles à bruit moyenne mobile

Pour la prédiction linéaire $X_t(k)$ des processus ARIMA(p,d,q), on aura toujours besoin d'une estimation de $\epsilon_{t-1}, \epsilon_{t-2}, \dots$, ou au moins de $\epsilon_{-1}, \epsilon_{-2}, \dots$, i.e. du "bruit inobservable passé" du modèle. On peut aussi recourir à la représentation AR(∞), dans quel cas on aura besoin de X_{-1}, X_{-2}, \dots , qui sont aussi inobservables. Dans tout cas, le résultat final demandera une approximation des valeurs précédant le début d'observations 0; l'approximation la plus simple dans l'absence des moyennes est $\epsilon_k = Y_k = 0$ pour $k < 0$.

Exemple 7.4 Pour le processus MA(1), on vérifie facilement que $X_t(k) = 0$ pour $k = 2, 3, \dots$, (pour une généralisation, voir le théorème 6.1). Pour $k = 1$, la définition :

$$X_{t+1} = \epsilon_{t+1} + \theta\epsilon_t$$

donne :

$$X_t(1) = \theta\epsilon_t$$

Pour se débarrasser de ϵ_t , on peut utiliser la représentation :

$$\epsilon_t = \sum_{i=0}^{\infty} (-1)^i \theta^i X_{t-i} = X_t + \sum_{i=1}^{\infty} (-1)^i \theta^i X_{t-i}$$

Donc, $X_{t+1} = \epsilon_{t+1} + \sum_{i=1}^{\infty} (-1)^{i-1} \theta^i X_{t+1-i}$ et

$$X_t(1) = \hat{X}_t = \sum_{i=0}^{\infty} (-1)^i \theta^{i+1} X_{t-i}$$

$$X_t(k) = 0, k = 2, 3, \dots$$

Il est naturel de supposer que l'information est finie, i.e. $F_t = \{X_t, X_{t-1}, \dots, X_1\}$. La formule dans ce cas, obtenue par récursion, est :

$$\hat{X}_t = \mathbb{E}[X_{t+1}|F_t] = \sum_{i=0}^{t-1} (-1)^i \theta^{i+1} X_{t-i} - (-\theta)^{t+1} \epsilon_0$$

Comme ϵ_0 n'est pas connue, en pratique on utilise : $\hat{X}_t(1) = \sum_{i=0}^t (-1)^i \theta^{i+1} X_{t-i}$. Si $\theta < 1$ et t est large, la différence sera négligeable.

Donc, cet exemple montre déjà qu'une estimation du "bruit inobservable" $\epsilon_t, \dots, \epsilon_1, \epsilon_0, \dots$ est incontournable pour les modèles ARMA avec $q \geq 1$.

Théorème 7.1 Dans le cas d'un modèle ARIMA(p, d, q), la meilleure prévision linéaire au temps t est :

$$\hat{X}_t(k) = \mathbb{E}[X_{t+k}|F_t] = \sum_{i=1}^p \tilde{\varphi}_i \hat{X}_t(k-i) + \sum_{i=k}^q \theta_i \hat{\epsilon}_{t+k-i}$$

où les $\tilde{\varphi}_i$ sont les coefficients du polynôme $\varphi(B)(1-B)^d$ (dans le cas d'un modèle ARMA(p, q) $\tilde{\varphi}_i = \varphi_i$).

Pour $k > q$, cette formule est exactement la récurrence homogène Yule-Walker $\varphi(B)\hat{X}_t(k) =$, et donc la prévision sera donnée par la solution de cette équation qui passe par les $p + d$ points pivots.

Les inconnues $\hat{\epsilon}_{t-i}, i \geq 0$ peuvent être enlevés en utilisant la représentation inverse "π" du bruit en fonction de la série, ou en utilisant $\hat{\epsilon}_t = Y_t - \hat{Y}_{t-1}(1)$ (les dernières se calculent récursivement). Une estimation arbitraire de $\epsilon_0, \epsilon_{-1}, \dots$ sera nécessaire.

Exercice 7.1 On considère le processus ARMA(1,1) à moyenne 0 $(1 - \varphi B)Y_t = (1 + \theta B)\epsilon_t$ où $-1 < \varphi < 1$ et $-1 < \theta < 1$.

- Montrez que la fonction de prévision Box-Jenkins est donnée par $Y_t(k) = Y_t(1)\varphi^{k-1}, k \geq 1$, et que

$$\begin{aligned} Y_t(1) &= \varphi Y_t + \theta \epsilon_t \\ &= (\varphi + \theta)Y_t - \theta Y_{t-1}(1) \\ &= (\varphi + \theta)\{Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} + \dots\} \end{aligned}$$

Est ce que ces résultats restent vrais si $\varphi = 1$, donc pour ARIMA(0,1,1) ?

- On utilise ce modèle pour ajuster une série et on obtient comme estimations des paramètres $\varphi = 0.8, \theta = 0.3$ et $\mu = ?$. Les dix dernières valeurs disponibles sont :

$t :$	51	52	53	54	55	56	57	58	59	60
$y_t :$	2.98	4.10	6.10	9.36	8.57	8.82	7.31	7.19	2.36	0.40

Donnez les prévisions des trois valeurs suivantes de la série. Quelle parmi les trois formules pour $Y_t(1)$ ci-dessus paraît la plus convenable à appliquer ?

Exercice 7.2 Le processus ARIMA(0,1,1) (appelé aussi IMA(1,1)) est défini par :

$$(1 - B)Y_t = (1 + \theta B)\epsilon_t$$

Si $\theta < 1$, les coefficients de la représentation du bruit sont :

$$\pi_i = (1 + \theta)(-\theta)^{i-1}, i \geq 1,$$

(à vérifier).

- Montrez qu'on peut le représenter :

$$Y_t = \epsilon_t + (1 + \theta) \sum_{k=1}^{t-1} \epsilon_{t-k} + \theta \epsilon_0$$

- Montrez que $\bar{Y}_t = (1 + \theta)Y_t - \theta \bar{Y}_{t-1}$.

Note : La dernière formule est appelée lissage exponentiel, au moins quand $\theta \in (-1, 0)$ et donc $\alpha = 1 + \theta \in (0, 1)$. La formule donne une moyenne pondérée : $\bar{Y}_t = \alpha Y_t + (1 - \alpha)\bar{Y}_{t-1}$
 α s'appelle constante de lissage.

Rémarques : 1) Plus α est petit, plus la nouvelle série est lisse et les valeurs passées ont un plus grand poids dans la prévision. Quand α est proche de 1 les valeurs les plus récentes ont le poids le plus important.

2) On peut voir la prévision Box-Jenkins comme une généralisation du lissage exponentiel, en utilisant des paramètres estimés à partir des données (au-lieu de ad-hoc).

7.4 Exercices : TD 5

1. Prévisions sous le modèle ARIMA(1,1,1)

Considérons le processus ARIMA(1,1,1) $(1 - \varphi B)(1 - B)Y_t = (1 + \theta B)\epsilon_t$, avec $-1 < \varphi < 1$ et $-1 < \theta < 1$.

- (a) Montrez que $Y_t(1) = (1 + \varphi)Y_t - \varphi Y_{t-1} + \theta \epsilon_t$ et $Y_t(k) = (1 + \varphi)Y_t(k-1) - \varphi Y_t(k-2)$, $k \geq 2$.
- (b) Montrez que $Y_t(k) = A_t + B_t \varphi^k$ pour $k \geq 0$, et trouvez des expressions pour A_t et B_t en termes de $Y_t, Y_{t-1}, \epsilon_t, \varphi$ et θ , en utilisant $Y_t(0)[= Y_t]$ et $Y_t(1)$ du (a) ci-dessus. Montrez que :

$$Y_t(k) = Y_t + \varphi \frac{(1 - \varphi^k)}{1 - \varphi} (Y_t - Y_{t-1}) + \theta \frac{(1 - \varphi^k)}{1 - \varphi} \epsilon_t, \quad k \geq 0.$$

Trouvez la limite $\lim_{k \rightarrow \infty} Y_t(k)$

- (c) Montrez que $Y_t(1) = -\theta Y_{t-1}(1) + (1 + \varphi + \theta)Y_t - \varphi Y_{t-1}$ et $Y_t(k) = Y_{t-1}(k+1) + \psi_k \epsilon_t$.
- (d) Montrez que $Y_t(k)$ peut s'exprimer en fonction seulement des valeurs passées de la série. [Indication : utilisez les π pour vous débarrasser de ϵ_t]
- (e) En utilisant le modèle $(1 - 0.6B)(1 - B)Y_t = (1 + 0.3B)\epsilon_t$ obtenez les prévisions des trois termes suivants de la série :

$t :$	1	2	3	4	5	6	7	8	9	10
$y_t :$	14.8	12.4	9.4	7.7	7.3	9.0	10.5	11.2	10.4	11.6
$t :$	11	12	13	14	15	16	17	18	19	20
$y_t :$	12.1	11.6	9.9	8.1	6.6	5.4	4.2	5.3	6.8	9.2

2. Considérons le processus ARIMA(1,1,2) :

$$(1 - \alpha B)(1 - B)Y_t = (1 + \theta_1 B + \theta_2 B^2)\epsilon_t$$

où $-1 < \alpha < 1$. Soit $Y_t(k)$ la prévision de Y_{t+k} au temps t .

- (a) Montrez que $Y_t(1) = (1 + \alpha)Y_t - \alpha Y_{t-1} + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}$ et trouvez les expressions correspondantes pour $Y_t(2)$ et $Y_t(k)$ pour $k \geq 3$
- (b) Montrez que la fonction de prévision peut s'exprimer sous la forme $Y_t(k) = a_t + b_t \alpha^k$, $k \geq 1$, et donnez la formule de a_t, b_t comme fonctions de $Y_t, Y_{t-1}, \epsilon_t, \epsilon_{t-1}$.
- (c) Montrez que $Y_t(k)$ peut s'exprimer en fonction seulement des valeurs passées de la série.
- (d) Un statisticien a utilisé le modèle ARIMA (1,1,2) décrit ci-dessus pour une série (dénommée prix) qui exprime le prix d'une action à la bourse pour 100 jours consécutifs. En sachant que $Y_{98}(1) = 686,996$ et $Y_{99}(1) = 659,416$ et $\sigma_\epsilon = 2$, calculer les prévisions $Y_{101|100}, Y_{102|100}$ de Y_{101} et Y_{102} et donner les 95% intervalles de confiance associés avec ces prévisions.

3. **Projet** : Il serait intéressant de déterminer analytiquement "la tendance asymptotique", i.e. le polynôme d'ordre $d - 1$ vers le quel les prévisions converge asymptotiquement pour les processus ARIMA(p,d,q).

Considerons par exemple ARIMA(p,2,0) ; ce modèle inclue une tendance lineaire, pour la quelle le premier candidat est la droite par les deux dernières points pivots (comme dans le cas $p = 0$, quand les pivots coincident avec les valeurs X_t, X_{t-1}). En général, les prévisions doivent encore converger asymptotiquement vers une droite. Pour $p = 0$, on commence deja exactement sur la "droite de tendance" (due à l'absence d'autres racines dans la partie autorégressive) ; mais, pour $p \geq 1$, nous serons obligé de tenir compte d'autres valeurs pivots et donc de X_{t-2}, X_{t-3}, \dots . A priori donc, les p points qui precedent les 2 dernières point auront aussi une influence sur la "droite de tendance".

7.5 Le filtre de Kalman

1. (a) Déduisez les formules des coefficients de corrélation ρ_1, ρ_2 pour le processus $MA(1)$.
 (b) Trouvez les valeurs maximales et minimales de ρ_1 et les valeurs de θ pour les quelles ces valeurs sont atteintes.
 (c) Déduisez les formules des coefficients de corrélation ρ_1, ρ_2 pour le processus $AR(2)$.
2. On considère le processus aléatoire $AR(2)$ suivant :

$$X_t = 10 + (-.3) X_{t-1} + .01 X_{t-2} + \epsilon_t$$

ou ϵ_t est $BB(0, \sigma^2 = 1)$.

- (a) Calculer l'espérance de X_t , en supposant que le processus est stationnaire.
 - (b) Est-ce que le processus est stationnaire causal ?
 - (c) Donner les équations de Yule-Walker du processus, calculer la variance, ainsi que les 3 premières valeurs des autocorrélations.
 - (d) Pour un modèle non-stationnaire, comment on pourrait détecter la manque de stationnarité à partir des expressions numériques fournies par les équations de Yule-Walker pour les autocorrélations ?
 - (e) Calculer les 3 premières autocorrélations partielles.
3. Soit le processus :

$$(1 - B + .25B^2)Y_t = (1 + \theta B)Z_t$$

- (a) Est-ce que ce processus est stationnaire causal ? Si oui, trouvez la fonction d'autocorrélation de Y_t .
 - (b) Est-ce que ce processus est inversible ? Calculez les premiers cinq coefficients de la "représentation π " du bruit Z_t en termes de la série. De quel problème on s'aperçoit si le processus Y_t n'est pas inversible ?
 - (c) Donnez, si possible, une méthode de prévision pour un processus (i) causal et noninversible (ii) noncausal et inversible.
 - (d) Donnez la prévision un pas en avant $\hat{Y}_t(1)$ en fonction des valeurs passées Y_t, Y_{t-1}, \dots
4. Considérons le processus : $ARIMA(1,1,2)$:

$$(1 - \alpha B)(1 - B)Y_t = (1 + \theta_1 B + \theta_2 B^2)Z_t$$

où $-1 < \alpha < 1$. Soit $Y_t(k)$ la prévison de Y_{t+k} au temps t .

- (a) Montrez que $Y_t(1) = (1 + \alpha)Y_t - \alpha Y_{t-1} + \theta_1 Z_t + \theta_2 Z_{t-1}$ et trouvez une expression correspondante pour $Y_t(2)$. Donnez une formule de récurrence pour $Y_t(k)$ pour $k \geq 3$.
- (b) Montrez que la fonction de prévision $Y_t(k)$ peut s'exprimer sous la forme $Y_t(k) = a_t + b_t \alpha^k$, et donnez les formules de a_t, b_t comme fonctions de $Y_t, Y_{t-1}, Z_t, Z_{t-1}$.

Solutions :

1. (c) Le processus peut s'écrire $Y_t = (1 + \tau_1 B + \tau_2 B^2)Z_t = (1 + .4B)^2 Z_t$. Il est inversible car la racine $-5/2$ est à l'extérieur du disque unité. Par identification des coefficients on trouve que $\pi_1 = \theta_1, \pi_2 = \theta_2 - \theta_1^2, \pi_3 = \theta_1^3 - 2\theta_2\theta_1, \pi_4 = -\theta_1^4 + 3\theta_2\theta_1^2 - \theta_2^2, \dots$ et alors

$$\hat{Y}_t(1) = Y_t + \sum_{i=1} \pi_i Y_{t-i}$$

avec $\pi_1 = .8, \dots$

5. (a) Ce processus AR(2) pas centré peut s'écrire $(1 - .2B - .35B^2)X_t = (1 - .7B)(1 + .5B)X_t = 40 + \epsilon_t$ En prenant l'espérance on trouve $\mathbb{E}(X_t) = \frac{40}{1 - .2 - .35} = \frac{40}{.45}$.

(b) Le processus est stationnaire causal car les racines du polynôme sont à l'extérieur du disque unité.

(c) Soit $Y_t = X_t - \mathbb{E}X_t$. Alors Y_t est un processus AR(2) centré satisfaisant $(1 - .2B - .35B^2)Y_t = (1 - .7B)(1 + .5B)X_t = \epsilon_t$

La fonction d'autocovariance de Y_t (qui est la même que celle de X_t) est obtenue de l'équation : $\mathbb{E}(Y_t Y_{t-h}) = 0.2\mathbb{E}(Y_{t-1} Y_{t-h}) + 0.35\mathbb{E}(Y_{t-2} Y_{t-h}) + \mathbb{E}(\epsilon_t Y_{t-h})$, ce qui donne le système de Yule Walker :

$$\rho_1 = 0.2 + 0.35\rho_1$$

$$\rho_2 = 0.2\rho_1 + 0.35$$

La première équation donne $\rho_1 = \frac{2}{6.5} = .31$, et la seconde donne $\rho_2 = 0.2\rho_1 + .35 = .41$. Finalement, $\gamma_0 = \frac{1}{1 - \sum \varphi_i \rho_i} = 2.52$.

(d) Les autocorrélations partielles $\hat{\rho}_i, i = 1, 2, 3$ se calculent à l'aide des déterminants des matrices d'autocorrélations et sont $.31, .35, \approx 0$. La troisième autocorrélation est en effet une erreur d'arrondissement, car le modèle AR(p) a les autocorrélations partielles nulles au delà du rang p.

8 Examen d'entraînement 2

2. Considérons le filtre $P(B) = \frac{1}{3}(2 + B + B^2 - B^3)$.
- (a) Montrez qu'il "enlève" les composantes saisonnières de période 3, i.e. qu'il transforme chaque fonction de période 3 dans une fonction constante.
 - (b) Trouvez l'ordre de la tendance polynômiale maximale conservée (laissée invariante) par ce filtre.
3. Préviation du modèle $ARIMA(1,2,0)$ $(1 - \varphi B)(1 - B)^2 X_t = \epsilon_t$.
- (a) Trouvez la forme de la prévision Box-Jenkins $X_t(k)$ de X_t et les poids des points pivotales X_t, X_{t-1}, X_{t-2} .
 - (b) Trouvez la limite $\lim_{k \rightarrow \infty} X_t(k)$
4. Soit le processus :

$$(1 - .8B + .16B^2)Y_t = (1 + \theta B)Z_t$$

- (a) Est-ce que ce processus est stationnaire causal? Si oui, obtenez la "représentation ψ " du Y_t par rapport au bruit Z_t et trouvez la fonction d'autocorrélation de Y_t .
 - (b) Sous quelles conditions est ce processus inversible? Obtenez la "représentation π " du bruit Z_t en termes de la série. De quel problème on s'aperçoit si le processus n'est pas inversible?
 - (c) Donnez la prévision k pas en avant $\hat{Y}_t(k)$, en utilisant les valeurs Y_t, Y_{t-1}, Z_t .
5. Quelle est la limite $\lim_{k \rightarrow \infty} X_t(k)$ de la prévision linéaire d'un processus ARMA (p,q)?