

Séries temporelles : régression, modélisation ARIMA(p,d,q), modélisation espace-état et splines

2 novembre 2005

Enseignant : Florin Avram

Objectif : L'interpolation : prévision "ponctuelle, déterministe", et la régression : démarche "statistique" plus complexe, 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.

Le premier dilemme dans les séries temporelles et la statistique spatiale est le choix entre modèles stochastiques et déterministes (qui peuvent être vues en effet comme cas particuliers simples des premiers). Le deuxième dilemme est le choix entre modélisation globale (régression, modélisation ARMA) et modélisation locale, par exemple par splines (qui change en effet de modèle quand cela semble opportune).

Nous allons aborder ces thèmes dans le contexte des séries temporelles uni-dimensionnelles, en commençant par l'approche d'interpolation déterministe la plus simple : inspection graphique, lissage par filtres, identification paramétrique de la tendance et prédiction des moindres carrés. En suite, nous examinons l'approche itérative statistique qui consiste à raffiner des modèles ARIMA ou des modèles d'espace-état, choisies conformément aux tests pour les résidus. Eventuellement, la démarche stochastique pourra aussi être mise en oeuvre en partant d'une l'interpolation déterministe plus sophistiqué, par **splines**.

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.

Volume horaire :

- 12 heures de cours : 1 heure et demi Jeudi 8 à 13 :40, S06, pour 8 semaines, et qui se transformeront en suite en TD pour la neuvième et dixième semaines.
- 18 heures de TD : 1 heure et demi Vendredi 08 :00, S23, pour 10 semaines (et complètes par trois heures pour les deux dernières semaines, ayant place dans la salle. de cours, Jeudi).

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 RH. 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. Gourieroux et A. Monfort, Cours de séries temporelles.

Table des matières

1	Introduction	3
2	Premier abord aux séries temporelles/chroniques	4
2.1	Les composantes d'une chronique	4
2.2	Quelques types de décomposition	6
2.2.1	le modèle additif	6
2.2.2	le modèle multiplicatif(*)	7
2.2.3	les modèles mixtes(*)	7
3	Filtres/moyennes mobiles	7
3.1	Fonctions génératrices	8
3.2	Filtres de lissage	9
3.3	Filtres qui enlèvent les composantes saisonnières	11
3.4	Exercices : TD 1	11
4	Modélisation stochastique des séries temporelles	14
4.1	Introduction	14
4.2	Processus stochastiques stationnaires	15
4.3	Exemples des processus stationnaires	16
4.3.1	Le bruit blanc	16
4.3.2	Les processus linéaires et les moyennes mobiles MA(q)	17
4.3.3	Les modèles autorégressifs AR(p)	18
4.3.4	Les modèles ARMA(p,q)	20
4.3.5	Les modèles ARIMA(p,d,q)	21
4.4	(*)L'inversion des series des puissances et des filtres $\varphi(B)$	21
4.4.1	Causalité des modèles AR(p)	22
4.4.2	Inversibilité des processus MA(q)	23
4.4.3	Causalité et inversibilité des modèles ARMA(p,q)	23
4.5	Exercices : TD 2	25
4.6	TP	27
4.7	(*) La positivité : caractérisation des suites de covariance	28
5	La prévision linéaire	29
5.1	La prévision des processus stationnaires AR(p)	30
5.2	Bruit d'innovation	30
5.3	Prévision linéaire des modèles autorégressifs ARIMA(p,d,0)	32
5.4	Prévision linéaire des modèles ARIMA(p,d,q)	33
5.5	La détermination de l'ordre d'un modèle autorégressif	34
5.6	Exercices : TD 3	35
6	L'estimation des modèles ARIMA(p,d,q)	36
6.1	Équations Yule-Walker pour les covariances/corrélations des modèles autorégressifs AR(p)	36
6.2	Équations de Yule-Walker pour les covariances/corrélations des processus ARMA(p,q)	38
6.3	Intervales de confiance pour la prédiction	38
6.4	Le lissage exponentiel (*)	39
6.4.1	Le lissage exponentiel simple	39
6.4.2	Le lissage exponentiel de Holt	41
6.5	Le lissage exponentiel de Winters	42
6.5.1	modèle de Winters additif	42
6.5.2	modèle de Winters multiplicatif	42
6.6	TP	43

6.7	Modèles d'espace-état	44
6.8	Contrôle continu en séries temporelles	44
7	Examens d'entraînement	46
7.1	Examen d'entraînement 1	46
7.2	Examen d'entraînement 2	50
7.3	Examen d'entraînement 3	52
8	Sommaire des définitions et résultats dans les séries temporelles	53
8.1	Filtres	53
8.2	Causalité et inversibilité des modèles ARMA(p,q)	54
8.3	Équations Yule Walker	54
8.4	Prévision linéaire	55
9	Appendice : Rappels des statistique et probabilités multivariées	55
9.1	Analyse statistique univariée	55
9.2	Rudiments de statistique bivariée	56
10	La régression lineaire	56
10.1	Méthode des moindres carrés et méthode des 2 points	56
10.2	le coefficient de corrélation linéaire ne suffit pas pour mesurer la qualité de l'ajustement	58
10.3	Les tendances polynomiales	59
10.3.1	la courbe des moindres carrés	59
10.3.2	le choix du degré du polynome	60
10.4	La méthode du changement de variable	62
10.5	Courbes de tendance qui ne se reduisent pas à une régression lineaire	62
10.6	La fonction d'autocorrélation empirique	63
10.7	(*) Autocorrélation d'une chronique possédant une composante périodique	64
10.8	(*) Autocorrélation d'une chronique présentant une tendance	65
10.9	(*) Autocorrélation d'une série de fluctuations irrégulières	66

1 Introduction

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 (c'est le premier exemple d'utilisation de splines, par Whittaker (1923)).

Les séries temporelles sont le plus simple exemple d'une thématique plus large : l'estimation et prévision des processus stochastique, i.e. des familles des variables aléatoires $U(x)$. Pour les séries temporelles/chronologiques, on s'intéresse en $x \in \mathbb{N}, \mathbb{Z}$ ou \mathbb{R}_+ , pendant que dans la statistique spatiale, (par exemple en géostatistique) on s'intéresse dans le cas $x \in \mathbb{Z}^d$ ou $x \in \mathbb{R}^d$.

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 donnés x_i , pour $i = 1, \dots, N$. Le but principal est le choix d'un modèle ("estimation") raisonnable, qui permettra à partir des valeurs connues la prédiction des valeurs inobservables (comme les valeurs futures des séries temporelles, ou moins accesibles physiquement, couteuses, etc). On veut à la fois : a) enlever du bruit d'observation eventuel 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, ...

2 Premier abord aux séries temporelles/chroniques

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.

- 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 augmenté depuis 1960, mais avec une alternance de baisses et de hausses soudaines. Le taux de chômage des États-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.

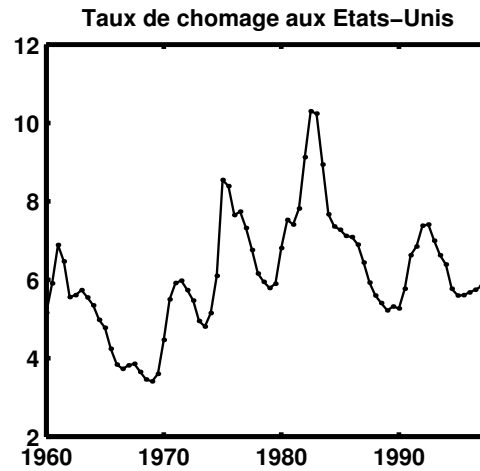
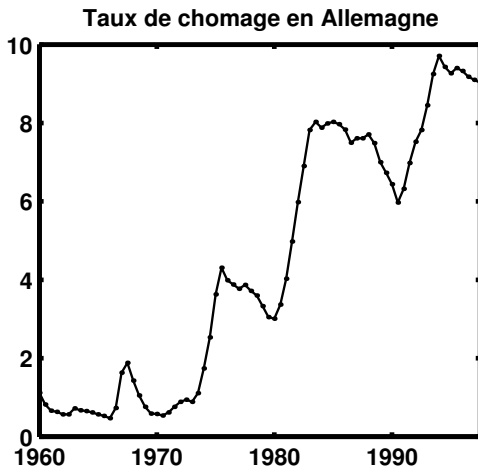
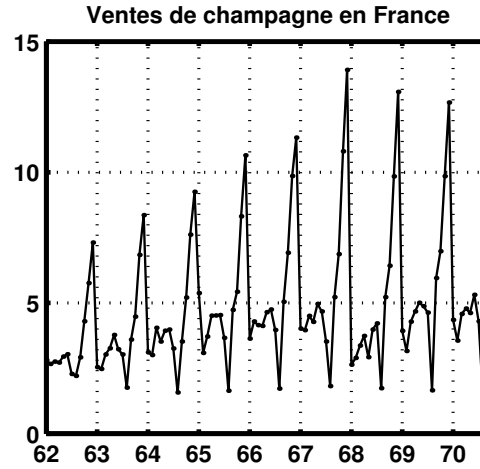
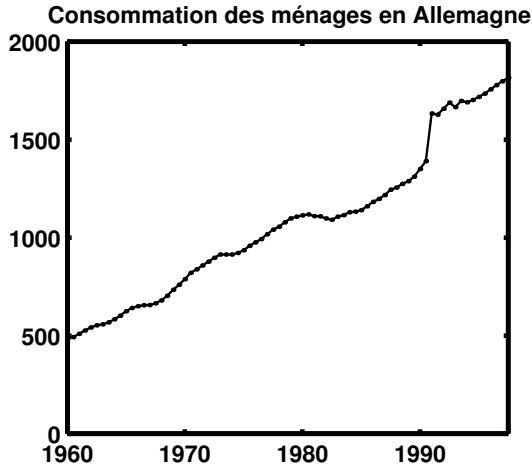
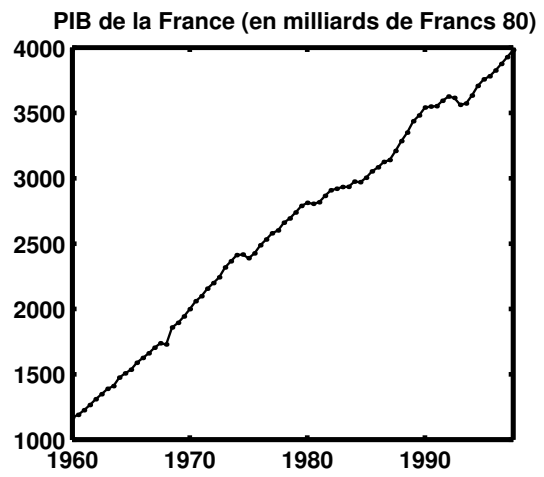
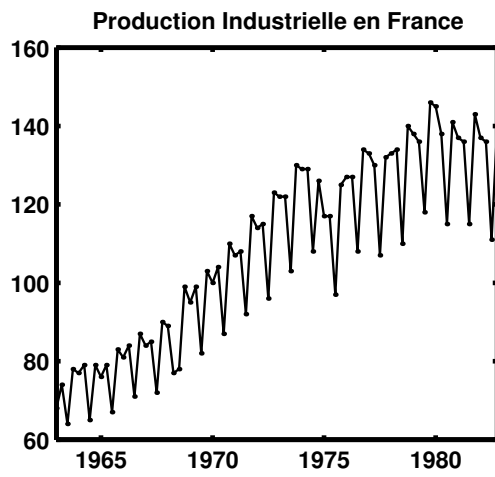


FIG. 1 – Quelques exemples de séries chronologiques

Il faut alors analyser ces composantes, en les dissociant les unes des autres, c'est-à-dire en considérant une série comme résultant de la combinaison de différentes composantes, 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. C'est une fonction monotone, souvent polynomiale.
2. *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 acti-

tivités économiques et sociales (fêtes, vacances, solde, etc).

Mathématiquement, ce sont des fonctions périodiques, c'est-à-dire qu'il existe un entier p , appelé *période*, tel que $s_i = s_{i+p}$ pour tout $i \geq 1$. Au premier abord, cette composante est entièrement déterminée par ses p premières valeurs s_1, s_2, \dots, s_p . Mais on rencontre souvent aussi des phénomènes pour les quelles la période peut elle même varier. On parle alors de

3. *Cycles* ($c_i, 1 \leq i \leq n$), qui regroupent des variations à période moins précise autour de la tendance, par exemple les phases économiques 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 pour la plupart du temps aussi les cycles.
4. *Les fluctuations irrégulières/résidues/bruit* ($e_i, 1 \leq i \leq n$) sont des variations de faible intensité et de courte durée, et de nature aléatoire (ce qui signifie ici, dans un cadre purement descriptif, qu'elles ne sont pas complètement expliquables). En effet, elles ne sont pas clairement apercevables dans les graphiques, à cause de leur faible intensité par rapport aux autres composantes. Elles apparaissent clairement seulement après "l'enlèvement du signal"; la question qui se posera alors sera : est-ce qu'ils contiennent encore du signal, ou est-ce que c'est vraiment du "bruit" ?
5. *Les variations accidentelles/observations aberrantes* sont des valeurs isolées anormalement élevées ou faibles de courte durée. 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 bruits (les fluctuations irrégulières).
6. *Points de changement* Ce sont des points où la série change complètement d'allure, par exemple de tendance. Ils sont normalement explicables, et imposent une analyse séparée de la série, par morceaux.

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).

Exemple : Trouvez l'élément suivant de la série y_t ci-dessous, une équation de récurrence pour

$$y_t = \{1, 3, 7, 13, 21, 31, \dots\}$$

Obtenez une formule analytique pour y_t , en utilisant :

a) la théorie des équations de récurrence à coefficients constants. R : $n^2 + n + 1$

b) la méthode des fonctions génératrices, décomposition en fractions partielles et l'expansion en série des puissances :

$$\frac{1}{(a-z)^k} = \sum_{n=0}^{\infty} C_{k-1}^{n+k-1} \frac{z^n}{a^{n+1}}, \quad |z| \leq a$$

$$R : a(z) = \frac{1+z^2}{(1-z)^3} = \frac{2}{(1-z)^3} - \frac{2}{(1-z)^2} + \frac{1}{(1-z)}$$

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

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

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 :

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

Exemple : Imaginons que nous étudions la série des températures moyennes relevées chaque mois en un même site, depuis janvier 1990, et que la tendance (plutôt faible) a une allure linéaire. Le modèle additif est :

$$y_i = a + bi + \sum_{k=1}^{11} s_k 1_{i \cong k \pmod{12}} - \left(\sum_{k=1}^{11} s_k \right) 1_{i \cong 0 \pmod{12}} + e_i$$

Les coefficients a, b, s_1, \dots, s_{11} et les résidus peuvent être déterminés en minimisant la somme des carrés des résidus $\sum_i e_i^2$, i.e. par régression.

Que peut-on dire des composantes présentes dans cet exemple ?

- 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 laissions prévoir la tendance et l'effet saisonnier pour février 1991).

2.2.2 le modèle multiplicatif(*)

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

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 :

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

(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).

Une fois un modèle est obtenue, il peut être utilisé pour la prédiction des valeurs futurs.

3 Filtres/moyennes mobiles

Souvent il semble une bonne idée de baser les prédictions sur l'information locale fournie par les voisins, ce qui suggère de construire des "moyennes mobiles".

Définition 3.1 La série Y_t s'appelle une **moyenne mobile** de X_t ou **filtre** si

$$Y_t = \sum_{i=-k_1}^{k_2} \theta_i X_{t-i} \quad (4)$$

où $k_1, k_2 \geq 0$. L'ordre du filtre est $k = k_1 + k_2 + 1$.

Le cas des θ_i à somme égale à 1 s'appelle **lissage**, celui avec des θ_i égaux s'appelle **moyenne arithmétique**, et le cas d'une moyenne arithmétique avec $k_1 = k_2 = q$ sera appelé **moyenne arithmétique symétrique ou centré**.

Exemples : filtres arithmétiques, causaux, ...

Il est convenable d'introduire un **opérateur de retard** B (ou encore de "rétro-décalage") défini sur l'ensemble des suites par

$$BX_t = X_{t-1} \quad \text{donc } B^i X_t = X_{t-i}, \quad \text{et finalement } \sum_{i=-k_1}^{k_2} \theta_i X_{t-i} = \theta(B)X_t$$

où $\theta(B)$ dénote le polynôme $\sum_{i=-k_1}^{k_2} \theta_i B^i$. La notation des polynômes de retard ramène (4) à la forme :

$$Y_t = \theta(B)X_t$$

et les équations de récurrence $\sum_{i=-k_1}^{k_2} \theta_i X_{t-i} = 0$ à la forme :

$$\theta(B)X_t = 0$$

Exemples et applications dans la théorie des récurrences à coefficients constants ...

$\theta(B)$ est appelé **opérateur de différences** (sur l'espace des séries). La série de Laurent associée $\theta(z) = \sum_{i=-k_1}^{k_2} \theta_i z^i$ sera appelée le symbole de l'opérateur.

Définition 3.2 Le symbole d'un filtre $\theta(B)$ est la fonction $\theta(z) : \mathbb{C}^- \rightarrow \mathbb{C}$.

Nous travaillerons surtout avec des **filtres causaux**

$$\theta(B) = \sum_{i=0}^k \theta_i B^i$$

où k peut être aussi ∞ . Dans ce cas, les coefficients du filtre seront notés surtout par ψ_i , et le symbole par $\psi(z)$.

3.1 Fonctions génératrices

Un fait très important est que la fonction génératrice

$$y(z) = \sum_{t=0}^{\infty} Y_t z^t$$

d'une série définie par un filtre causal $Y_t = \psi(B)X_t$ est essentiellement le produit de $x(z) = \sum_{t=0}^{\infty} X_t z^t$, la fonction génératrice de X_t et du symbole $\psi(z)$. Plus précisément, notons par $\psi_{\leq m}(z) =: \sum_{i=0}^m \psi_i z^i$ la troncature de n'importe quelle série des puissances aux premiers termes.

Théorème 3.1 Pour chaque filtre causal d'ordre fini, on a :

$$y(z) - y_{\leq k-1}(z) = \psi(z)x(z) - \sum_{i=0}^{k-1} (X_i z^i) \psi_{\leq k-1-i}(z)$$

où k est l'ordre du filtre.

Démonstration : Nous allons vérifier "formellement" ¹ le cas particulier $k = 2$ des filtres quadratiques $\psi_0 + \psi_1 B + \psi_2 B^2$, quand ce théorème devient :

$$y(z) - Y_0 - Y_1 z = \psi(z)x(z) - X_0(\psi_0 + \psi_1 z) - (X_1 z) \psi_0$$

Rémarques : 1) Ce théorème est un exemple de la fameuse méthode des fonctions génératrices de Laplace, qui transforme les récurrences en des équations algébriques pour les fonctions génératrices !
2) Pour les séries Y_t, X_t doublement infinies, i.e. avec $t \in \mathbb{Z}$, le résultat est plus simple

$$\boxed{y(z) = \psi(z)x(z)} \quad (5)$$

car on peut remplacer 0, 1 par un point arbitraire de départ $-k - 1, -k$ et ensuite on fait k tendre vers ∞ .

3) Pour les séries Y_t, X_t doublement infinies, on peut inverser formellement cette relation, obtenant X_t à partir de Y_t :

$$x(z) = \frac{y(z)}{\psi(z)}$$

Mais, le travail avec les séries Y_t, X_t doublement infinies contient des "pièges" qu'on discutera plus tard. De l'autre coté, travailler avec des séries indicées par $t \in \mathbb{N}$ nous force à définir l'égalité d'une serie comme égalité des composantes, à partir d'un certain point, i.e.

$$A = B \iff \exists K \in \mathbb{N} \text{ tel que } A_n = B_n, \forall n \geq K$$

C'est facile de vérifier que les opérations de filtrage commutent :

Théorème 3.2 Soit $\theta_1(B), \psi_2(B)$ deux filtres et posons $\psi(B) = \psi_1(B)\psi_2(B)$. Alors :

$$\psi_1(B)\psi_2(B)X_t = \psi_2(B)\psi_1(B)X_t = \psi(B)X_t$$

Rémarque : L'approche de décomposition additive demande de rompre une serie donnée $Y_t = m_t + \epsilon_t$ dans un "signal" m_t et du bruit ϵ_t . Il est naturel d'essayer de depister le signal par un filtre de lissage causal $l(B)$, donc $m_t = l(B)Y_t$, tq ce filtre "détruit le bruit mais laisse passer le signal". Il en suit que le bruit aussi est donné par un filtre causal

$$\epsilon_t = Y_t - m_t = (I - l(B))Y_t := \pi(B)Y_t$$

où $\pi(B)$ "détruit le signal mais laisse passer le bruit".

On peut voir donc l'analyse des séries temporelles comme la recherche pour un filtre qui transformera notre serie en bruit. Pour cela, il va être important de définir des tests pour decider quand une serie est un bruit. Nous allons examiner plus tard des statistiques comme les correlations, correlations partielles, le periodogramme, etc., issues de la modelisation probabiliste des séries temporelles.

Pour l'instant, les prochains paragraphes nous donnent quelques outils pour juger le comportement des filtres appliqués au séries déterministes.

3.2 Filtres de lissage

Un filtre de lissage (à $\sum_i \theta_i = 1$)

$$Y_t = \sum_{i=1}^k \theta_i X_{t-i} := \hat{X}_t$$

¹en ignorant la convergence des séries; cet aspect n'est pas foncier, car on peut justifier algébriquement même des manipulations avec séries à rayon de convergence 0.

peut être utilisé pour la prédiction de X_t . Remarquez que le fait que la prédiction est "non-biaisée pour les séries stationnaires", i.e. :

$$\mathbb{E}\hat{X}_t = \mathbb{E}\sum_{i=1}^k \theta_i X_{t-i} = \left(\sum_{i=1}^k \theta_i\right)\mathbb{E}X_1$$

est assuré par la condition $\sum_{i=1}^k \theta_i = 1$.

Cette condition assure aussi qu'une série égale à $\mathbf{1}$ sera "prédite" exactement, i.e. $\theta(B)\mathbf{1} = \mathbf{1}$, et en fait chaque série constante $X_t = k$ sera prédite exactement :

$$\theta(B)k = k \cdot (\theta(B)\mathbf{1}) = k \cdot 1 = k$$

La vérification est très facile pour ça, remarquons que

Il est possible en fait, en choisissant les coefficients θ_i d'un filtre, d'assurer qu'il laisse invariantes toutes les séries polynomiales p_t d'un degré donné.

Exercice 3.1 a) Montrez qu'une moyenne arithmétique symétrique d'ordre $2q + 1 = 3$, donné par

$$\theta(B) = \frac{1}{3}(1 + B + B^{-1})$$

conserve (laisse invariantes) les tendances linéaires $p_t = a + bt$. b) Généraliser pour q quelconque.

Nous verrons maintenant un résultat désirable de l'application des filtres de lissage : la réduction de la variance des observations.

Exercice 3.2 Montrez qu'une moyenne arithmétique symétrique d'ordre $2q + 1$ diminue la variance σ^2 d'un bruit blanc (=série i.i.d. de moyenne 0) par $2q + 1$.

En conclusion, si la série observée est de la forme

$$X_t = p_t + \epsilon_t$$

où $p_t = a + bt$ est une tendance linéaire, que l'opération de prendre une moyenne arithmétique symétrique d'ordre q n'affecte pas la tendance, i.e. $\theta(B)p_t = p_t$, mais a un effet de diminution du bruit stochastique ϵ_t , ramenant à :

$$\hat{X}_t = \theta(B)(p_t + \epsilon_t) = p_t + (\theta(B)\epsilon_t) = p_t + \frac{\epsilon_{t+q} + \dots + \epsilon_t + \dots + \epsilon_{t-q}}{2q + 1} := p_t + \epsilon'_t$$

avec un nouveau bruit $e'_t = \frac{\epsilon_{t+q} + \dots + \epsilon_t + \dots + \epsilon_{t-q}}{2q+1}$ de variance inférieure à celle de e_t .

Donc, si on constate une tendance linéaire dans le comportement d'une chronique dans un voisinage, on peut estimer la tendance dans ce voisinage en prenant des moyennes mobiles arithmétiques symétriques, car ça va réduire (atténuer) le bruit et mettre en évidence la tendance linéaire. L'effet du lissage augmente en augmentant q .

Exercice 3.3 Montrez que la droite obtenue en lissant $2q + 2$ observations avec des moyennes mobiles arithmétiques symétriques d'ordre $2q + 1$ est :

$$y - \frac{\sum_{i=1}^{2q+1} X_i}{2q + 1} = (x - (q + 1)) \frac{X_{2q+2} - X_1}{2q + 1}$$

Le théorème suivant nous donne un critère pour identifier le degré maximal des polynômes laissés invariants par un filtre $\theta(B)$; autrement dit, de déterminer le degré maximal des polynômes inclus dans l'espace invariant des séries Z_t satisfaisant $\theta(B)Z_t = Z_t$:

Théorème 3.3 L'espace invariant d'un filtre contient les polynômes de degré $\leq p$ ssi 1 est une racine d'ordre au moins $p + 1$ de l'équation $\theta(z) = 1$, i.e. $\theta(1) = 1, \theta'(1) = 0, \theta''(1) = 0, \theta^{(p)}(1) = 0$.

Exercice 3.4 Demontrez le théorème pour $p = 0, 1$

Outre l'exploration de l'espace invariant d'un filtre, une autre question importante est celle de l'exploration du noyau, i.e. l'espace des séries Z_t satisfaisant $\theta(B)Z_t = 0$. Cette question a une portée pratique pour l'enlèvement de composantes saisonnières (et leur détermination).

3.3 Filtrés qui enlèvent les composantes saisonnières

Définition 3.3 a) Une série s_t sera appelée périodique de période p ssi

$$s_{t+p} = s_t \iff (1 - B^p)s_t = 0, \quad \forall t \quad (6)$$

b) Une série s_t sera appelée saisonnière de période p ssi

$$\sum_{i=1}^p s_{t+i} = 0 \iff \left(\sum_{i=0}^{p-1} B^i \right) s_t = 0, \quad \forall t \quad (7)$$

Exercice 3.5 Montrez qu'un filtre $\theta(z)$ qui est divisible par $1 + z + \dots + z^{p-1}$, i.e. de la forme $\theta(z) = (1 + z + \dots + z^{p-1})\theta_1(z)$, "enlève" les composantes saisonnières de période p , i.e. :

$$\theta(B)s(t) = 0 \quad \forall t$$

pour chaque série s_t satisfaisant (7).

En effet, la réciproque est aussi vraie (admis) :

Théorème 3.4 Un filtre $\theta(B)$ annule (ou enlève) les composantes saisonnières d'ordre p ssi son symbole $\theta(z)$ est divisible par $1 + z + \dots + z^{p-1}$ (donc si $\theta(z) = 0$, pour toutes les racine d'ordre p de l'unité, sauf $z = 1$).

Exemples : Pour enlever les composantes saisonnières d'ordre 4, on peut utiliser donc la moyenne mobile arithmétique d'ordre 4, pour une périodicité mensuelle on peut utiliser la moyenne mobile arithmétique d'ordre 12, etc... En général, en utilisant un filtre arithmétique d'ordre p on peut enlever la partie saisonnière de cet ordre, pour mieux décèler ensuite la tendance.

Alternativement, après le choix d'une forme appropriée pour la tendance et une période spécifique, selon le graphe, on peut déterminer au même temps les coefficients de la tendance et de la partie périodique par une régression linéaire.

Exercice 3.6 Montrez que le filtre $\frac{1}{9}(-B^2 + 4B + 3 + 4B^{-1} - B^{-2})$ laisse invariants les polynômes de troisième degré, et enlève les composantes saisonnières d'ordre 3.

3.4 Exercices : TD 1

1. Trouvez l'élément suivant des séries y_t ci-dessous, ainsi que des équations de récurrences qu'elles satisfont et leurs solutions analytiques :

$$2, 6, 12, 20, 30, 42, \dots \quad (8)$$

$$4, 10, 20, 36, 62, 104, \dots \quad (9)$$

$$3, 2, 1, 6, 3, 2, 1, \dots \quad (10)$$

$$0, -1, -2, 3, 0, -1, -2, \dots \quad (11)$$

Indication. a), b) Calculez les séries différenciées : $z_t = \Delta y_t = y_t - y_{t-1}$. La deuxième série admet deux continuations naturelles (au moins).

2. Une série vérifie la récurrence $y_t - 2y_{t-1} + y_{t-2} = (-1)^{t-1}$, $t \geq 2$, $y_0 = 1, y_1 = 3$ Obtenez une formule analytique pour y_t , en utilisant : la méthode des fonctions génératrices, décomposition en fractions partielles et l'expansion en série des puissances :

$$\frac{1}{(a-z)^k} = \sum_{n=0}^{\infty} C_{k-1}^{n+k-1} \frac{z^n}{a^{n+1}}, \quad |z| \leq a$$

$$R : a(z) = \frac{1+2z}{(1+z)(1-z)^2} = \frac{3}{2(1-z)^2} - \frac{1/4}{(1-z)} - \frac{1/4}{(1+z)}, \quad a(n) = (5 - (-1)^n + 6n)/4$$

3. (a) Montrez que le filtre $P(B) = \frac{1}{3}(2+B+B^2-B^3)$ "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.

Sol : Ordre 1.

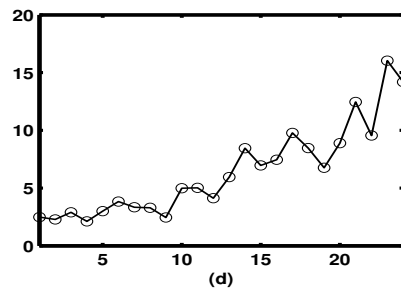
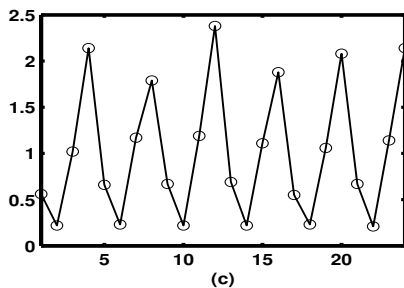
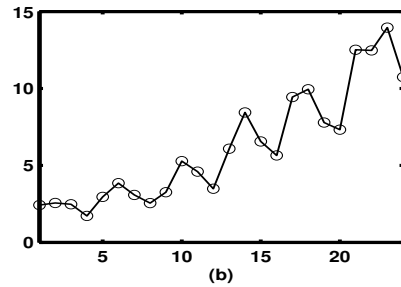
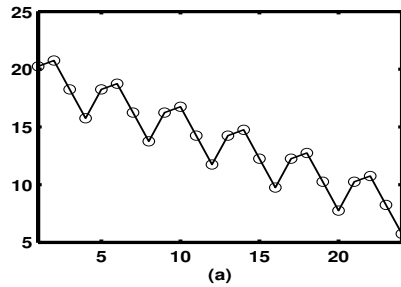
4. Trouver un filtre $1 + \alpha B + \beta B^2 + \gamma B^3$ qui laisse passer une tendance affine sans distortion et élimine les périodicités d'ordre 2. Indication : Trouver un système des $2 + 1 = 3$ équations et résoudre.
 5. Trouvez un filtre $f(B)$ qui conserve les polynômes de degré ≤ 1 , et qui enlève les composantes saisonnières 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 donnée par $m_t = f(B)Y_t$.

Sol : $\frac{1+B+B^2+B^3}{4} \frac{5-3B}{2}$

6. a) Montrez qu'une série saisonnière est périodique, et que chaque série périodique p_t est la somme d'une série saisonnière et d'une série constante.
 b) Trouvez une base de l'espace vectorielle des séries périodiques d'ordre p .
 c) Trouvez une base de l'espace vectorielle des séries saisonnières d'ordre p , et ensuite une base des séries périodiques qui la contient.
 7. On considère la série suivante :

t_i	1	2	3	4	5	6	7	8	9	10
y_i	58	40	31	15	18	15	9	9	10	8

- a) Représenter graphiquement cette série.
 b) On se propose d'ajuster une tendance f de la forme $f(t) = \frac{1}{a + bt}$. Justifier ce choix.
 c) Déterminer les coefficients a et b , en utilisant un changement de variable approprié :
 - par la méthode des 2 points (en choisissant judicieusement les 2 points)
 - par la régression linéaire.
 d) représenter les 2 tendances ainsi obtenues sur le graphique précédent et comparer les résultats. Est-ce que les résidus ont une allure irrégulière ?
 8. Pour chacune des quatre séries suivantes,



- a) écrire le modèle qui vous semble convenir, en précisant le type du modèle (par "défaut additif"), la tendance et la période
 b) Exprimez le modèle choisi sous la forme d'une équation vectorielle **linéaire** dans les paramètres inconnus, et donnez la formule de la régression qui permet à déterminer ces paramètres.

9. 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

- a) Représenter graphiquement cette série.
 - b) Quel modèle proposeriez-vous pour cette série (justifier) ?
 - c) Calculer les facteurs saisonniers $(s_j, 1 \leq j \leq p)$ ainsi que leur moyenne $p^{-1} \sum_{j=1}^p s_j$, en supposant une tendance constante $m_t = a$.
 - d) En notant $(e_i, 1 \leq i \leq n)$ la série des fluctuations irrégulières, calculer e_1, e_2 et e_3 .
 - e) Proposer une méthode pour l'estimation des paramètres, en supposant cette fois une tendance lineaire $m_t = at + b$. Implementez le calcul en utilisant un logiciel. Proposez un teste pour choisir entre les deux modèles.
10. On considère un modèle simple où la tendance est une constante ($f(t) = a$).
- 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 ?
 - b) 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 ?
 - c)* Reprendre la question b) lorsque le nombre d'observations n'est pas un nombre entier de périodes : $n = Lp + m$.
11. On considère une série $(y_i, 1 \leq i \leq n)$ périodique, de période p . On suppose que le nombre d'observations n est un multiple de p : $n = Lp$. Montrer alors que les corrélations suivantes sont :

$$\rho(p) = \frac{L-1}{L} \quad ; \quad \rho(2p) = \frac{L-2}{L} \quad ; \quad \dots \quad ; \quad \rho(jp) = \frac{L-j}{L} \dots$$

4 Modélisation stochastique des séries temporelles

4.1 Introduction

Rappelons le modèle additif sans saisonnalité, qui cherche une décomposition de la forme :

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

- m_t représente la "tendance" (intuitivement un "mouvement lisse à long terme"), qui sera la composante la plus importante dans la prévision.
- $\epsilon_t = Y_t - m_t$ sont les "résidus" qui restent après qu'on enlève la partie structurée m_t . Elles représentent des "irrégularités/fluctuations imprévisibles", qui au début semblent inutilisables (à ignorer) pour la prévision (c'est correct du point de vue de la prévision ponctuelle, mais elles nous serviront quand-même dans le calcul des intervalles de confiance).

On s'arrangera toujours tel que les résidus ont la moyenne 0, mais ça n'est pas suffisant pour qu'ils soient un bruit totalement sans structure="bruit blanc" (et s'il y a encore une partie structurée, elle devrait être incluse en m_t).

Le "bruit blanc" est notre premier exemple d'un processus stochastique : une formalisation du concept de séries temporelles, ayant des propriétés bien définies (voir prochaine chapitre). Inspirés par les propriétés de ce processus, on proposera des tests statistiques correspondant à ce modèle, qui nous permettront de décider si ϵ_t ont les propriétés de manque de structure désirées.

Pour tendance, plusieurs modèles se sont avérés utiles :

1. régression sur des prédicteurs exogènes ("covariates"), implémenté en logiciels comme R par "formules" :

$$m_t \sim X_t^{(1)} + X_t^{(2)} + \dots$$

2. modèles de superposition des chocs extérieurs/moyennes mobiles/FIR **inobservables** ϵ_t :

$$m_t = \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

3. modèles autoregressifs :

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

Dans le manque des prédicteurs exogènes, il est assez naturel d'adopter une modélisation autoregressive pour la tendance. Sous certaines conditions de régularité, ça ramènera à des prévisions autoregressives un pas en avant :

$$\hat{Y}_t = f(Y_{t-1}, Y_{t-2}, \dots)^2$$

Le modèle le plus simple est le processus AR(1) :

$$Y_t = \phi Y_{t-1} + b + \epsilon_t$$

Ce modèle est recommandable si on envisage une prévision

$$\hat{Y}_t = \phi Y_{t-1} + b \iff (\hat{Y}_t - a) = \phi(Y_{t-1} - a)$$

où $b = a(1 - \phi)$.

On vérifie que si la moyenne de Y_t est 0 on a $a = b = 0$; pour simplifier, on supposera normalement qu'on a déjà enlevé la moyenne de Y_t .

²La modélisation autoregressive permettra aussi des prédictions k pas en avant :

$$\hat{Y}_{t+k} = f_k(Y_t, Y_{t-1}, Y_{t-2}, \dots), \quad k = 1, 2, \dots$$

Les valeurs (positives) de k correspondent au futur et doivent être extrapolées/prévues. La fonction de prévision f_k représente une projection de Y_{t+k} sur l'espace engendré par $Y_t, Y_{t-1}, Y_{t-2}, \dots$. Plusieurs : choix sont possibles par exemple extrapolation à partir d'un ajustement/interpolation polynomiale ou par splines.

Pour utiliser ce modèle, on estime le paramètre ϕ par une régression linéaire des points

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

Le fait d'avoir enlevé la moyenne ramène à une droite passant par l'origine $y = \phi x$.
En suite, on utilise la valeur trouvée pour résoudre l'équation. On trouve

$$Y_t = \sum_{i=0}^{t-1} \phi^i \epsilon_{t-i} + \phi^t Y_0$$

et examiner solution stationnaire unique ssi : $|\phi| < 1$.

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

b) Montrez que l'équation : (20) a une solution stationnaire unique, qui dépend seulement du **bruit futur** ssi : $|\phi| > 1$.

En conclusion

1. pour $|\phi| < 1$, l'équation : (20) a une solution stationnaire unique causale, qui dépend seulement du **bruit passé**.
2. pour $|\phi| > 1$, l'équation : (20) a une solution stationnaire unique, qui dépend seulement du **bruit futur**.

Pour tester la validité des modèles, proposés, il faut d'abord préciser rigoureusement les propriétés désirées des résidus ou des chocs en tant que processus stochastiques ; en plus, les modèles seront utiles seulement si on peut vérifier leur "stationnarité", i.e une certaine uniformité de structure par rapport au temps. Ça nous ramène à considérer les processus stochastiques stationnaires, les distributions jointes de quelles ne varient pas avec le temps.

4.2 Processus stochastiques stationnaires

Définition 4.1 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'instant $t_1 \dots t_r \in T$ 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 4.2 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 :

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

Comme la plupart de séries n'est observable qu'une seule fois, l'utilité du concept de distributions et covariances théoriques n'est pas évidente pour les applications. Par contre, on peut toujours calculer des distributions et covariances empiriques, et sous l'hypothèse de stationnarité, les moyennes empiriques convergent vers les théoriques.

Théorème 4.1 Pour un processus stationnaire, les covariances empiriques

$$\gamma_n(k) = (n - k)^{-1} \sum_{t=1}^{n-k} (X_t - m_n)(X_{t+k} - m_n)$$

estimées à partir de n observations convergent vers les covariances théoriques quand $n \rightarrow \infty$.

D'ici l'importance du concept de stationarité, qui justifie l'estimation des modèles statistiques observables une seule fois (le cas souvent dans les séries temporelles et la géostatistique !) : ceci est faisable ssi on a la chance d'avoir à faire avec un processus stationnaire.

Remarques :

1. La plupart des séries ne sont pas stationnaires, mais on peut essayer 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 fausse. 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.

4.3 Exemples des processus stationnaires

L'idéalisation probabiliste de la propriété que les résidus sont "complètement irréguliers", ne retenant aucune structure, est le "bruit blanc" stationnaire.

Un deuxième exemple important des processus stationnaires "non-blancs" sont les "processus linéaires" $MA(\infty) Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ avec ϵ_t bruit blanc et leur cas particulier avec un nombre fini des coefficients ψ_i nonnuls, les "**moyennes mobiles**" $MA(q)$.

Le troisième exemple étudié seront les "**processus autorégressifs**" $AR(\infty) \epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ avec ϵ_t bruit blanc et leur cas particulier avec un nombre fini des coefficients π_i nonnuls, les processus $AR(p)$.

4.3.1 Le bruit blanc

L'exemple le plus simple de modèle stochastique est le bruit blanc discret, la structure "revée" des résidus qui restent après qu'on enlève la tendance/moyenne d'un processus.

Définition 4.3 *Un processus $\epsilon_t, t \in \mathcal{T}$, où \mathcal{T} est un ensemble dénombrable quelconque, est appelé **bruit blanc stationnaire** si les variables ϵ_t sont **i.i.d.** (indépendants et identiquement distribués) à espérance $\mathbb{E}\epsilon_t = 0$. Il sera appelé **bruit blanc Gaussien** si la distribution de chaque v.a. ϵ_t est Gaussiennes.*

Un bruit blanc a la 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).

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

Définition 4.4 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 $\rho(s, t) = \delta(s - t)$).

Notes :

1. Le bruit blanc Gaussien est une structure probabiliste très naturelle, 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. Le bruit blanc stationnaire est une idéalisation du processus des residus de la regression lineaire, qu'on aimerait "independents". Mais, comme l'indépendance est un concept probabiliste, et les residus sont le résultat determinist d'une regression appliqué a une serie observée une seule fois, il est difficile de la verifier rigoureusement. Parmi les tests possibles, mentionnent celui de "turning points", qui demande de verifier que la frequence de ces points est environ 4/6, et le teste qui verifie si la somme des correlations empiriques est proche de 0. Si ces deux testes sont positives, on sait au moins "qu'on ne peut pas repousser l'hypothèse de l'indépendance". Il y aussi des tests distributionnels des résidus comme Fisher, Student, qui testent la Gaussianité.
3. Quand les tests des données rejettent l'hypothèse du bruit blanc, i.e. quand on a du **bruit corrélé**, la regression classique doit etre remplace par une analyse plus fine, appelée krigeage en geostatistique.

4.3.2 Les processus linéaires et les 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 :

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i} \text{ avec } \sum \psi_i^2 < \infty \quad (14)$$

où ϵ_t est un bruit blanc.

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 4.6 Un processus linéaire Y_t s'appelle **causal** s'il peut être représenté dans la forme :

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (15)$$

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

Définition 4.7 On appelle processus MA(q) un processus lineaire $Z_t, t \in \mathbb{Z}$ vérifiant une relation :

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

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

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

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

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 (i.e. $\text{Var } Y_t < \infty$),
b) à variance constante stationnaire $\text{Var } Y_t = \sigma_\epsilon^2 \sum_{i=-\infty}^{\infty} \psi_i^2$
c) à autocovariance donnée par :

$$\gamma(t, t+k) = \sigma_\epsilon^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+k} < \infty \quad (17)$$

d) stationnaire à ordre deux.

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

Exercice 4.1 1. Calculer la fonction d'autocovariance $\gamma(k)$ d'un processus MA(1).
2. Calculer la fonction de covariance $\gamma(k)$ d'un processus MA(q)

Le fait que les fonctions de covariance et corrélation $\gamma(k)$, respectivement $\rho(k)$ d'un processus MA(q) s'annulent pour $k > q$ permet de reconnaître des séries qui peuvent être modélisées comme MA(q). Plus précisément, pour accepter l'hypothèse qu'une série est MA(q) pour un q donné, on vérifie que toutes les corrélations pour $k > q$ satisfont

$$|\rho_n(k)| \leq z_\alpha \sigma_q$$

où

$$\sigma_q^2 = \frac{1 + 2(\hat{\rho}(1)^2 + \hat{\rho}(2)^2 + \dots + \hat{\rho}(q)^2)}{n}$$

(formule de Bartlett) et z_α , la "fractile" d'ordre α de la distribution Gaussienne, dépend du niveau de confiance α désiré (par exemple, $z_{.95} = 2$). Donc, si toutes les corrélations pour $k > q$ sont à l'intérieur de cette bande de confiance, on accepte l'hypothèse que la série est MA(q).

4.3.3 Les modèles autorégressifs AR(p)

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

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + \epsilon_t \quad (18)$$

Dans ce cas il s'avère typiquement que la formule de prévision ponctuelle pour Y_t un pas en avant est simplement :

$$\hat{Y}_t = f(Y_{t-1}, Y_{t-2}, \dots)$$

Nous allons considérer ici surtout des modèles autorégressifs linéaires (où f est une fonction linéaire) AR(p) :

Définition 4.8 Un processus stationnaire $Y_t, t \in \mathbb{Z}$ sera appelé **processus autorégressif linéaire d'ordre p** : **AR(p)** s'il existe un bruit blanc ϵ_t et des réels $\varphi_i, i = 1, \dots, p$ tels qu'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 vérifiée.

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

$$\varphi(B)Y_t = \epsilon_t \quad \text{où } \varphi(B) = 1 - \sum_{i=1}^p \varphi_i B^i$$

Définition 4.9 *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 définis par une équation, qui à priori, peut ne pas avoir des solutions ; comme "solution" de l'équation (19) nous aimerions avoir une représentation du processus Y_t par rapport au processus ϵ_t .

Nous verrons maintenant que le processus de Markov AR(1) $Y_t = \phi Y_{t-1} + \epsilon_t$ a une représentation causale ssi $|\phi| \neq 1$ et cette représentation est causale ssi $|\phi| < 1$.

Exercice 4.2 processus AR(1). a) Montrez que l'équation :

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{20}$$

a une solution stationnaire unique, qui dépend seulement du **bruit présent et passé**, étant donc **causale**, ssi : $|\phi| < 1$.

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

b) Montrez que l'équation : (20) a une solution stationnaire unique, qui dépend seulement du **bruit futur** ssi : $|\phi| > 1$.

En conclusion

1. pour $|\phi| < 1$, l'équation : (20) a une solution stationnaire unique causale, qui dépend seulement du bruit **passé**. On vérifie alors que ϵ_t est un bruit d'innovation.
2. pour $|\phi| > 1$, l'équation : (20) a une solution stationnaire unique, qui dépend seulement du bruit **futur**. On vérifie alors que ϵ_t n'est pas un bruit d'innovation.
3. pour le cas $|\phi| = 1$, l'équation : (20) (appelée *marche aléatoire*) n'a pas de solution stationnaire. Par contre, les **increments** $Y_t - Y_{t-1} = \epsilon_t$ sont stationnaires ; cette situation plus compliquée sera analysé dans le chapitre sur les processus ARIMA(p,d,q).

Exercice 4.3 Montrez que si un processus $AR(2)$ $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$ a une représentation stationnaire causale

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

alors les coefficients ψ_i satisfont la **recursion Yule-Walker**

$$\begin{aligned} \psi_t &= \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2}, \quad t \geq 2 \\ \psi(0) &= 1, \quad \psi(1) = \phi_1 \end{aligned}$$

Exercice 4.4 Montrez que la formule des coefficients ψ_j de la représentation $MA(\infty)$ d'un processus $AR(2)$ $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$, en fonction des racines de "l'équation caractéristique" de la récurrence Yule-Walker

$$0 = \lambda^2 - \phi_1 \lambda - \phi_2 = \lambda^2 \varphi(\lambda^{-1})$$

est :

$$\psi(k) = \begin{cases} \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} & \text{if } \lambda_1 \neq \lambda_2 \\ (k+1)\lambda_k & \text{if } \lambda_1 = \lambda_2 = \lambda \end{cases}$$

Montrez que si l'équation caractéristique a ses racines λ_1, λ_2 dedans le cercle unitaire $|\lambda| < 1$ (et donc le "symbole" $1 - \phi_1 z - \phi_2 z^2$ a ses racines z_1, z_2 dehors le cercle unitaire $|z| < 1$), alors la condition

$$\sum_i \psi_i^2 \leq \infty,$$

qui assure la convergence des représentations $MA(\infty)$, est satisfaite.

Donnez les coefficients ψ_j dans les cas particuliers :

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

Indication : La solution générale de la récurrence d'ordre 2 $\psi(k) = \varphi_1 \psi(k-1) + \varphi_2 \psi(k-2)$, pour valeurs initiales arbitraires $\psi(0) = x_0, \psi(1) = x_1$ est :

- avec des racines distinctes λ_1, λ_2 :

$$\psi(k) = \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} x_1 - \frac{(\lambda_1^k \lambda_2 - \lambda_2^k \lambda_1)}{\lambda_1 - \lambda_2} x_0$$

- avec des racines confondues $\lambda_1 = \lambda_2 = \lambda$:

$$\psi(k) = x_0 \lambda^k + (x_1 - \lambda x_0) k \lambda^{k-1}$$

En conclusion, les processus $AR(1)$ et $AR(2)$ ont une représentation $MA(\infty)$ ssi l'équation $0 = \varphi(z)$ a ses racines z_1, z_2 dehors le cercle unitaire $|z| \leq 1$. Il s'avère que cette situation est typique pour tous les modèles $AR(p)$.

4.3.4 Les modèles $ARMA(p,q)$

Définition 4.10 On appelle processus $ARMA(p,q)$ un processus stationnaire $Y_t, t \in \mathbb{Z}$ vérifiant une relation de récurrence :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}, \quad \forall t \in \mathbb{Z} \quad (21)$$

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 (21) à la forme :

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

Nous verrons dessous que **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 causale), et
2. AR(∞), de ϵ_t en termes de Y_t (appelée aussi représentation inverse).

Ces représentations peuvent être obtenues par des **inversions formelles** de l'équation (21), suivies par un développement de la fraction correspondante dans une série des puissances :

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

On peut aussi formuler ces représentations en utilisant les fonctions génératrices des séries intervenantes. Comme vue au (5), la représentation lineaire $Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ est équivalente à :

$$Y(z) = \psi(z)\epsilon(z)$$

où $Y(z) = \sum_{i=-\infty}^{\infty} Y_t z^t$, $\epsilon(z) = \sum_{i=-\infty}^{\infty} \epsilon_t z^t$ sont les fonctions génératrices doublement infinies, et $\psi(z)$ est le symbole ou **fonction de transfert**.

L'hypothèse ARMA(p,q) est alors équivalente à la rationalité de la fonction de transfert :

$$\psi(z) = \frac{\theta(z)}{\varphi(z)}$$

et la modélisation ARMA peut être vue aussi comme une approximation Padé.

Le besoin de travailler avec des représentations causales et inversibles (voir dessous) nous forcent à accepter seulement des fonctions de transfert $\psi(z)$ qui n'ont ni des racines ni des pôles dans l'intérieur du cercle unitaire $|z| < 1$.

Par exemple, rappelons 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 $|z| \leq 1$.

Exemple 4.1 ARMA(1,1) Trouver la représentation AR(∞) (i.e. $\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$) du processus ARMA(1)

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

4.3.5 Les modèles ARIMA(p,d,q)

Définition 4.11 On appelle processus ARIMA(p,d,q) un processus X_t pour lequel 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) :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (22)$$

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 (22) 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.

4.4 (*)L'inversion des series des puissances et des filtres $\varphi(B)$

Le résultat suivant est utile pour l'inversion des modèles AR(p), et aussi des ARMA(p,q), qui sont des processus Y_t satisfaisant des équations de la forme : $\varphi(B)Y_t = \theta(B)\epsilon_t$.

Théorème 4.3 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

$$\psi_n = \sum_{i=1}^p \frac{K_i}{\lambda_i^{n+1}} \quad (23)$$

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

$$\psi_n = - \sum_{i=1}^p K_i \lambda_i^{n+1} \quad (24)$$

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.

Ce resultat justifie des manipulations formelles analogues qu'on fait avec des fonctions dans l'operateur B ³.

³On 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 puissances/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, mais a des expressions differentes dans l'intérieur des "anneaux de convergence" qui evitent les singularités. 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 sur les suites correspondant à des polynômes en B sont representé par des matrices Toeplitz ; on peut démontrer que il y a un isomorphisme entre l'anneau des matrices Toeplitz est celui des fonctions génératrices. Cet isomorphisme explique l'équivalence des deux approches. Formellement, la conclusion est que l'operateur B doit-être traité comme le scalaire $z = 1$ (qui est sa valeur propre), et donc "l'expansion correcte" pour les inversions $\frac{1}{\varphi(z)}$ en série des puissances dépendront du positionnement du point $z = 1$ par rapport aux racines.

4.4.1 Causalité des modèles AR(p)

Rappelons qu'il y a un problème (non-causalité) avec le modèle AR(1) quand la racine $\lambda = \varphi^{-1}$ de son polynôme $\varphi(z) = 1 - \varphi z$ est à l'intérieur du cercle unitaire. Ce problème est lié à l'existence des plusieurs développement possibles pour la fonction $\varphi(z)^{-1} = \frac{1}{1-z\varphi}$:

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

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

On a la même situation de pour tous les modèles AR(p)

Théorème 4.4 a) *Un processus AR(p) est causal, i.e. il peut être représenté sous la forme : $Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ où $\sum \psi_i^2 < \infty$ ssi toutes les racines de son polynôme caractéristique $\varphi(z)$ sont à l'extérieur du cercle unitaire. Les coefficients ψ_i sont dans ce cas les coefficients de la série Taylor de $\pi(z) = \frac{1}{\varphi(z)}$*

4.4.2 Inversibilité des processus MA(q)

Définition 4.12 *Une représentation causale*

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

d'un processus stationnaire Y_t s'appelle **inversible** si on peut aussi représenter le bruit par une représentation causale :

$$\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i} \quad (25)$$

où $\sum \pi_i^2 < \infty$

Exemple 4.2 *Le processus MA(1) $Y_t = \epsilon_t + \theta \epsilon_{t-1}$ est inversible ssi $\theta < 1$. En effet, comme dans la résolution de la recursion AR(1), on voit que :*

$$\epsilon_t = Y_t - \theta Y_{t-1} + \dots (-\theta)^{t-1} Y_1 + (-\theta)^t \epsilon_0$$

Pour $\theta < 1$, ça converge vers $\sum_{i=0}^{\infty} \pi_i Y_{t-i}$, où $\pi_i = (-\theta)^i$.

Théorème 4.5 *Un processus MA(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 $\sum |\pi_i| < \infty$. Les coefficients π_i sont dans ce cas les coefficients de la série Taylor de $\pi(z) = \frac{1}{\theta(z)}$ ⁴*

Remarque 4.1 *Donc, ϵ_t appartient à l'espace linéaire engendré par le passé du signal observé*

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

et les espaces engendrés par $\{Y_{t-i}, i = 0, 1, \dots\}$ et $\{\epsilon_{t-i}, i = 0, 1, \dots\}$ coïncident.

⁴La démonstration est basée sur un théorème concernant le développement en série des puissances des inverses des polynômes complexes –voir la prochaine section– et finalement sur le fait que l'ensemble des filtres est isomorphe à l'ensemble des fonctions complexes $\psi(z)$, l'isomorphisme étant "la transformée z " des suites. Cet isomorphisme explique quelques manipulations formelles avec les filtres (mais pas celles liés à l'inversion).

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

Les problèmes de non-causalité et non-inversibilité des modèles ARMA(p,q) disparaissent quand toutes les racines de $\varphi(z)$ et $\theta(z)$ sont à l'extérieur du cercle unitaire :

Théorème 4.6 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 $\sum |\psi_i| < \infty$ 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 $\sum |\pi_i| < \infty$ 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 $\pi(z) = \frac{\varphi(z)}{\theta(z)}$

Corollaire 4.1 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 immediate d'une approche de prévision (=régression) par projection dans l'espace de Hilbert engendré par le passé.

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), nous allons considere désormais surtout 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, et on s'appuyera sur le corollaire ci-dessus.

Exercice 4.5 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.

1. Précisez si le processus est stationnaire, causal et inversible, et calculez sa fonction d'autocovariance.
2. Trouvez les coefficients ψ_j de sa représentation comme processus MA(∞) et les coefficients π_j de sa représentation comme processus AR(∞) et précisez si ces représentations sont convergentes.

Mêmes questions pour le processus ARMA(2,1) défini par :

$$Y_t - 0.7Y_{t-1} + 0.1Y_{t-2} = \epsilon_t + 2\epsilon_{t-1}$$

Rémarque : Dans le cas le plus simple avec les racines λ_i de l'équation $\varphi(z) = 0$ distinctes, on obtient facilement des formules generales pour 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 donc $K_i = -\frac{\theta(\lambda_i)}{\varphi'(\lambda_i)}$. On arrive à : $\psi_n = \sum_{i=1}^p \frac{K_i}{\lambda_i^{n+1}}$. Des formules pareilles existent pour π_n , et dans le cas des racines non-distinctes.

Dans le cas des racines non-distinctes et complexes, il est preferable d'aborder le developpement Taylor $\frac{\theta(z)}{\varphi(z)} = \psi(z)$ directement, en obtenant des équations de recurrence pour ψ_k , à partir des coefficients du developpement $\varphi(z) \psi(z) = \theta(z)$. Cette méthode, est en effet applicable toujours :

Exercice 4.6 ARMA(2,1)

1. Trouver par la méthode directe la représentation MA(∞) 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).

Cet exercice peut être généralisé :

Théorème 4.7 (*) 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

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

Rappel : La dernière équation est appelée **recursion de Yule-Walker**.

Notes : a) Rémarquez que pour le processus ARMA(p,q), la récurrence pour $k > q$ est exactement comme pour le processus AR(p) ; mais, les premières équations obtenues en comparant les coefficients des puissances en $\boxed{\psi(B)\varphi(B) = \theta(B)}$ pour $k \leq q$ changent par l'ajout de θ_k de côté droite.

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

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

Exemple 4.3 Ainsi, $\psi_1 = \varphi_1 + \theta_1, \psi_2 = \varphi_2 + \varphi_1^2 + \theta_1\varphi_1 + \theta_2, \psi_3 = \varphi_3 + 2\varphi_1\varphi_2 + \varphi_1^3 + (\varphi_2 + \varphi_1^2)\theta_1 + \varphi_1\theta_2 + \theta_3, \dots$

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

4.5 Exercices : TD 2

1. Calculer la fonction d'autocovariance du processus à valeurs en \mathbb{R}^2 :

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

où ϵ_n est un bruit blanc standard.

2. Restrictions sur les valeurs des coefficients d'**autocorrélation** pour les processus MA.
Trouvez, pour le processus MA(1), les valeurs maximales et minimales de la corrélation ρ_1 et les valeurs de θ pour les quelles ces valeurs sont atteintes.
3. 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 correlogramme pour les cas :
- (i) $MA(2)$: $\theta_1 = -5/6, \theta_2 = 1/6$
 - (ii) $MA(2)$: $\theta_1 = 0.8, \theta_2 = 0.5$
 - (iii) $MA(3)$: $\theta_1 = 0.8, \theta_2 = -0.4, \theta_3 = -0.3$
4. Investiguez si les deux processus $MA(2)$ sont inversibles, en examinant la recurrence obtenue par la méthode des coefficients indéterminés.
5. 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.
 - (b) Trouvez les coefficients ψ_j de sa représentation comme processus $MA(\infty)$ et les coefficients π_j de sa représentation comme processus $AR(\infty)$ et précisez si ces représentations sont convergentes.
6. Mêmes questions pour les processus $ARMA(2,1)$ et $ARMA(2,2)$ 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) $(1 - B + \frac{B^2}{4})Y_t = (1 + B + B^2)\epsilon_t$
7. Soit le processus :

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

- (a) Est-ce que ce processus est stationnaire causal? Si oui, obtenez la "représentation ψ " du Y_t par rapport au bruit ϵ_t .
 - (b) Sous quelles conditions est ce processus inversible? Obtenez la "représentation π " du bruit ϵ_t en termes de la série. De quel problème on s'aperçoit si le processus n'est pas inversible?
8. a) 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. Indiquez le domaine des racines réelles et celui des racines complexes. *Indication : Les conditions pour avoir racines de module plus grand que 1 sont différentes pour le cas des racines complexes et celui des racines réelles, et pour un polynôme $\theta(z) = 1 + \theta_1 z + \theta_2 z^2$, la condition pour racines réelles de module plus grand que 1 sont plus compliquées que les conditions (équivalentes) que le polynôme "reciproque" $\tilde{\theta}(z) = z^2 \theta(1/z) = z^2 + \theta_1 z + \theta_2$ ait des racines réelles z_i de module plus petit que 1. Pour ce dernier polynôme, les conditions sont :*
- (a) racines complexes : $|z_i|^2 = |z_1 z_2| = |\frac{c}{a}| = |\theta_2| < 1$.
 - (b) racines réelles : $\tilde{\theta}(1) = 1 + \theta_1 + \theta_2 > 0$, $\tilde{\theta}(-1) = 1 - \theta_1 + \theta_2 > 0$
- b) (*) Pour le processus $MA(2)$, trouvez un domaine S contenant toutes les valeurs possibles des coefficients d'autocorrélation ρ_1, ρ_2 tel que le processus soit inversible, et les valeurs de θ_1, θ_2 pour les quelles les valeurs sur la frontière de S sont atteintes.
9. (*) Trouver le domaine de causalité dans le plan (φ_1, φ_2) d'un processus $AR(2)$.
10. 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.
11. a) Vérifiez si le processus $AR(2)$ $Y_t = -0.3Y_{t-1} + 0.10Y_{t-2} + \epsilon_t$ est stationnaire causal. Calculez son correlogramme, en partant directement du système de Yule-Walker, et tracez le.
- b) Mêmes questions pour le processus $AR(2)$ $Y_t = -Y_{t-1} - 0.34Y_{t-2} + \epsilon_t$.
12. Calculez la fonction d'autocovariance et la fonction d'autocorrélation des processus dans les exercices antérieurs.

13. 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 :

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

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

14. (*) 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

$$U_t = \epsilon_t + \frac{1}{\theta} \epsilon_{t-1} \text{ où } |\theta| < 1$$

et ϵ_t est bruit blanc.

- (a) Montrer que cette représentation du processus U_t n'est pas inversible.
 (b) On pose maintenant $w_t = \sum_{j=0}^{+\infty} \theta^j U_{t-j}$. Montrer que $\{w_t, t \in \mathbb{Z}\}$ est un bruit blanc dont on précisera la variance en fonction de σ^2 et θ
 (c) Montrer que $U_t = w_t + \theta w_{t-1}$ et que cette représentation de $\{U_t, t \in \mathbb{Z}\}$ est inversible.

Solutions :

- 4) 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 $\theta_2 < 1$. Les racines sont réelles/complexes dessous/dessus la parabole $\theta_2 = \frac{\theta_1^2}{4}$.

b) Pour passer de (θ_1, θ_2) à (ρ_1, ρ_2) on utilise

$$\rho_1 = \frac{\theta_1(1 + \theta_2)}{1 + \theta_1^2 + \theta_2^2} \quad \rho_2 = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

Transformant les équations antérieures, on trouve :

$\theta_2 = 1$ implique $\rho_1 = \frac{2\theta_1}{2 + \theta_1^2}$, $\rho_2 = \frac{1}{2 + \theta_1^2}$, $\theta_1 = \frac{\rho_1}{2\rho_2}$, $\rho_2(2 + \frac{\rho_1^2}{4\rho_2^2}) = 2\rho_2 + \frac{\rho_1^2}{4\rho_2} = 1$ et donc $\rho_1^2 = 4\rho_2(1 - 2\rho_2)$. Finalement, on trouve

$$\begin{aligned} \text{dessous} \rho_1 &= 2\sqrt{\rho_2(1 - 2\rho_2)} \\ \rho_2 + 1/2 &\geq \rho_1 \\ \rho_2 + 1/2 &\geq -\rho_1 \end{aligned}$$

où les dernières deux inegalités viennent de l'inegalité entre les moyennes arithmétique et géométrique de $(1 + \theta_2), \theta_1$.

- 5) Le domaine de causalité d'un processus $AR(2)$

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

(beaucoup plus compliquée que pour le $AR(1)$), obtenu comme le domaine d'inversibilité du processus $MA(2)$, est le triangle situé en dessous de $\varphi_2 + \varphi_1 < 1$, $\varphi_2 - \varphi_1 < 1$ et dessus $\varphi_2 = -1$.

4.6 TP

1. Effectuez une analyse **Box-Jenkins** d'un jeu de données : par exemple "WWWusage" en R (qui représente le nombre d'utilisateurs connectés à un serveur Internet chaque minute), en utilisant les commandes "acf(x)", "pacf(x)" (ou "eda.ts(x)", si disponible) "arima()", "pnorm()" et "eda.ts(x\$res)". L'analyse devrait aboutir dans un modèle avec résidus tel qu'au plus 1 sur 20 des coefficients acf et pacf sortent de l'intervalle de confiance autour de 0, et avec p-valeurs des coefficients inférieures à .05 (rappel : p-val $\approx \mathbb{P}\{|t_{val}| \geq 2\}$, où val sont les valeurs "standardisées", i.e. divisées par l'erreur standard (s.e.)).
2. Répétez, avec un jeu de données de votre choix (à trouver à partir de la liste "data(package=NULL)" o'ù "data(package="ts)").
3. Ecrivez des programmes qui **simulent** (sans utiliser la commande arima.sim) des processus :
 - a) MA(2) avec $\theta_0 = 1$, les autres coefficients à choisir, et à bruit petit, b) AR(2) à bruit petit, c) ARMA(2,2) (en passant la sortie du premier program au deuxième), et d) ARIMA(2,1,2) (en appliquant cumsum à la sortie du program antérieur).

Enoncez les théorèmes satisfaits par l'acf et le pacf des premiers deux cas, et vérifiez ensuite que vos programmes produisent des résultats adéquats.

Pour le troisième et quatrième cas, **estimez** le modèle par la commande arima, avec les ordres simulés, et aussi avec des ordres plus grands. Est-ce que la commande retrouve les coefficients que vous aviez choisi quand le bruit est très petit (en supposant que l'analyse est bonne, et donc que l'acf et pacf des résidus indiquent un bruit blanc) ?

4. Interpretation de l'acf

- (a) Soit x un vecteur de nombres consecutifs. Simulez une série a) linéaire $y = ax$
 - b) quadratique $y = ax^2 + bx$
 - c) périodique $y = \sin(ax)$
 - d) "presque périodique" $y = \sin(ax) + \sin(bx)$en donnant deux exemples de a, b pour chaque problème (donc 12 exemples). Obtenez l'acf de toutes les séries et indiquez vos observations sur la dépendance de a, b .
- (b) Pour les séries avec acf non zero, indiquez quels filtres/transformations pourront nous amener aux résidus bruit blanc.
- (c) Démontrez sur trois des exercices antérieurs l'effet sur l'acf de l'ajout du bruit blanc d'écart type $\sigma = R/4$, $\sigma = R$ et $\sigma = 3R$, où $R = \max y_i - \min y_i$ est "l'écart du signal déterministe".
Inclure au moins un exemple qui a besoin du filtrage, et étudier encore une fois l'effet du même filtrage.

4.7 (*) La positivité : caractérisation des suites de covariance

Les suites de nombres qui peuvent être covariances sont uniquement caractérisées par leur transformée Fourier.

Théorème 4.8 (*Bochner) Une suite paire $\gamma_k \in L_2$ peut représenter les covariances d'une série stationnaire ssi la transformé Fourier -apellée aussi **densité spectrale**

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

est nonnegative pour chaque w .

Il y a une caractérisation équivalente en termes des matrices de covariance. Soit X un processus p -dimensionnel, stationnaire à l'ordre 2, supposé centré. On note γ la fonction d'autocovariance de X

- $\gamma(k) = E(X_n \overline{X_{n-k}^t})$ est une matrice carrée d'ordre p .
- $\gamma(k) = \overline{\gamma(-k)^t}$. En particulier, la matrice de variance-covariance du processus X est une matrice hermitienne (symétrique dans le cas réel) puisque $\gamma(0) = \overline{\gamma^t(0)}$.
- Dans le cas $p = 1$, $|\gamma(k)| \leq \gamma(0)$.
- $(\gamma(k))_{k \in \mathbb{Z}}$ est une famille de type positif, c'est-à-dire que pour tout $A_1 \dots A_k$ de \mathbb{C}^p et tout $n_1 \dots n_k$

$$\sum_{i=1}^k \sum_{j=1}^k A_i^t \gamma(n_i - n_j) \overline{A_j} \geq 0.$$

Preuve Soit $W = \sum_{i=1}^k A_i^t X_{n_i}$. On a alors

$$\begin{aligned} \text{Var}(W) &= E\left[\sum_{i=1}^k A_i^t X_{n_i} \overline{\sum_{j=1}^k A_j^t X_{n_j}}\right] \\ &= E\left[\sum_{i=1}^k \sum_{j=1}^k A_i^t X_{n_i} \overline{X_{n_j}^t A_j}\right] \\ &= \sum_{i=1}^k \sum_{j=1}^k A_i^t \gamma(n_i - n_j) \overline{A_j} \geq 0. \square \end{aligned}$$

- Dans le cas univarié, la matrice d'autocovariance est une matrice de Toeplitz

$$C_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(n-1) \\ \gamma(-1) & \gamma(0) & \gamma(1) & \dots & \gamma(n-2) \\ \gamma(-2) & \gamma(-1) & \gamma(0) & \dots & \gamma(n-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(-n+1) & \gamma(-n+2) & \gamma(-n+3) & \dots & \gamma(0) \end{pmatrix}$$

Cette matrice carrée d'ordre n est positive, en effet

$$a^t C_n \bar{a} = \sum_{i=1}^n \sum_{j=1}^n a_i \gamma(i-j) \overline{a_j} \geq 0.$$

Au lieu de considérer la fonction d'autocovariance, on peut choisir d'utiliser la fonction d'autocorrélation

$$\rho(k) = (\rho_{i,j}(k)) = \left(\frac{\gamma_{i,j}(k)}{\sqrt{\gamma_{i,i}(0) \gamma_{j,j}(0)}} \right)$$

Cela revient à considérer non pas le processus X mais un processus $Y = (Y_n)$ dont la coordonnée i définie par $Y_n^{(i)} = X_n^{(i)} / \sqrt{\gamma_{i,i}(0)}$ est de variance 1.

5 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. $k > 0$ est l'écart de prévision
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 ARIMA(p,d,q) satisfont des contraintes lineaire, il est naturel 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,

$$\begin{aligned}\hat{X}_t(k) &= \sum_{i=0}^{\infty} \pi_{t,k}(i) X_{t-i} \text{ ou} \\ \hat{X}_t(k) &= \sum_{i=0}^{\infty} a_{t,k}(i) \epsilon_{t-i}, k = 1, 2, \dots\end{aligned}$$

Théorème 5.1 *La prévision linéaire à variance minimale des processus ARMA(p,q) avec du bruit blanc Gaussien coincide avec l'espérance conditionnelle*

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

où $F_t = \{X_t, X_{t-1}, \dots, X_0, X_{-1}, \dots\}$.

Donc, nous allons pouvoir profiter du fait que l'operateur d'espérance conditionnelle $\hat{X}_t(k) = \hat{X}(t+k|t) = \mathbb{E}[X_{t+k} | F_t]$ (toutes ces notations sont rencontrées dans la littérature) est lineaire.

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

5.1 La prévision des processus stationnaires AR(p)

Nous considerons maintenant plus en detail 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 , même quand $p = \infty$. La méthode sera "d'appliquer l'operateur chapeau" dans l'équation lineaire définissant le modèle.

5.2 Bruit d'innovation

Définition 5.1 *Si un processus Y_t défini par une autorégression (18) $Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + \epsilon_t$ est tel que la différence $\epsilon_t = Y_t - f(Y_{t-1}, Y_{t-2}, \dots) = Y_t - \hat{Y}_t$ est indépendante du passé $\mathcal{F}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$, il sera appelée **bruit d'innovation** (par rapport au passé).*

Dans ce cas,

$$\mathbb{E}[\epsilon_t | Y_{t-1}, Y_{t-2}, \dots] = 0 \quad (26)$$

et ϵ_t est noncorrélé avec Y_{t-1}, Y_{t-2}, \dots .

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

Le bruit blanc d'une équation autorégressive causale a la propriété très convenable de coincider avec l'erreur de prédiction par rapport au passé $Y_t - \hat{Y}_t = Y_t - \mathbb{E}[Y_t | Y_{t-1}, Y_{t-2}, \dots]$:

Nous verrons plus tard que le bruit d'un modèle (19) est d'innovation ssi le modèle est causal, et ssi le polynôme caractéristique a seulement des racines plus grandes en valeur absolue que 1. Dans ce cas, on a :

$$\hat{Y}_{t|t-1} = \mathbb{E}[Y_t | \{Y_{t-1}, Y_{t-2}, \dots\}] = f(Y_{t-1}, Y_{t-2}, \dots) = \sum_{i=1}^p \varphi_i Y_{t-i}$$

et donc la prédiction devient très aisée, car on applique simplement la relation d'autorégression, **en oubliant le bruit**.

Théorème 5.2 *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. $\mathbb{E}[\epsilon_{t+k} | \mathcal{F}_t] = 0$ si $k > 0$, et les prévisions satisfont la récurrence Yule-Walker :

$$\hat{X}_t(k) := \mathbb{E}[X_{t+k} | \{X_t, X_{t-1}, \dots\}] = \sum_{i=1}^p \varphi_i \hat{X}_t(k-i)$$

En particulier,

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

Démonstration immédiate, par la causalité et donc l'indépendance de ϵ_{t+k} de \mathcal{F}_t .

Exemple 5.1 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 connue μ elle est

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

Théorème 5.3 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 :

$$\varphi(B)X_t(k) = X_t(k) - \varphi_1 X_t(k-1) - \varphi_2 X_t(k-2) = X_t(k) - (\lambda_1 + \lambda_2)X_t(k-1) + \lambda_1 \lambda_2 X_t(k-2) = 0 \quad (27)$$

et donc les prévisions sont de la forme :

$$\boxed{X_t(k) = A_1(t)\lambda_1^k + A_2(t)\lambda_2^k} \quad (28)$$

En termes des deux dernières valeurs observées X_t et X_{t-1} , quand $\lambda_1 \neq \lambda_2$, les prévisions sont données par :

$$X_t(k) = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} X_t - \frac{(\lambda_1^{k+1}\lambda_2 - \lambda_2^{k+1}\lambda_1)}{\lambda_1 - \lambda_2} X_{t-1} \quad (29)$$

Démonstration : Les prévisions $\hat{X}_t(k)$ satisfont l'équation Yule Walker

$$\hat{X}_t(k) = \varphi_1 \hat{X}_t(k-1) + \varphi_2 \hat{X}_t(k-2)$$

avec valeurs initiales $\hat{X}_t(0) = X_t, \hat{X}_t(-1) = X_{t-1}$.

La solution générale de la récurrence d'ordre 2 $\psi(k) = \varphi_1 \psi(k-1) + \varphi_2 \psi(k-2)$, pour valeurs initiales arbitraires $\psi(0) = x_0, \psi(1) = x_1$ est :

- avec des racines distinctes λ_1, λ_2 :

$$\psi(k) = \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} x_1 - \frac{(\lambda_1^k \lambda_2 - \lambda_2^k \lambda_1)}{\lambda_1 - \lambda_2} x_0$$

- avec des racines confondues $\lambda_1 = \lambda_2 = \lambda$:

$$\psi(k) = x_0 \lambda^k + (x_1 - \lambda x_0) k \lambda^{k-1}$$

ce qui rend le résultat.

Notea : a) Dans le cas causal, $\lambda_i < 1$, et donc la **prévision éventuelle**

$$\lim_{k \rightarrow \infty} X_t(k) = 0$$

est toujours 0.

b) Il est facile d'étendre cette approche pour tous les processus autorégressifs ARIMA(p,d,0) à ordres p, d finis, et d'obtenir des formules explicites de prévision en termes de racines de l'équation $\varphi(z) = 1$.

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

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

(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 $\pi(B)$) et des formules pareilles pour des écarts plus grands que 1.

Exemple 5.2 La prévision des processus stationnaires ARMA(p,1)

Nous considérons maintenant plus en détail 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 , même quand $p = \infty$. La méthode sera "d'appliquer l'opérateur chapeau" dans l'équation linéaire définissant le modèle.

Théorème 5.4 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. $\mathbb{E}[\epsilon_{t+k} | \mathcal{F}_t] = 0$ si $k > 0$, et les prévisions satisfont la récurrence Yule-Walker :

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

En particulier,

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

Démonstration immédiate, par la causalité et donc l'indépendance de ϵ_{t+k} de \mathcal{F}_t .

5.3 Prévision linéaire des modèles autorégressifs ARIMA(p,d,0)

Exemple 5.3 La prévision 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 prévision AR(1)).

Exemple 5.4 Déduisez la formule de prévision 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 prévision $X_t(k)$ 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 prévision "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 prévision "eventuelle" est un polynôme d'ordre $d - 1$.

Exemple 5.5 On considère un processus $\{X_t\}$ pour lequel la série différencié deux fois est un bruit blanc, c'est à dire $\{X_t\}$ est un processus ARIMA(0,2,0). Montrez que la fonction de prévision 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 dernières points.

Définition 5.2 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 reccursion de Yule Waker devient valable) s'appellent les **valeurs pivots**.

Il suit clairement que :

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

Corollaire 5.1 La prévision 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 dernières points.

5.4 Prévision linéaire des modèles ARIMA(p,d,q)

Pour la prévision 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 resultat final demandera une approximation des valeurs precedant le debut d'observations 0 ; l'approximation la plus simple dans l'absence des moyennes est $\epsilon_k = Y_k = 0$ pour $k < 0$.

Exemple 5.6 Pour le processus MA(1), on verifie facilement que $X_t(k) = 0$ pour $k = 2, 3, \dots$, (pour une généralisation, voir le théorème 6.3). Pour $k = 1$, la définition :

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

donne :

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

Pour se debarasser 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 reccursion, 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 negligeeable.

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 5.5 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 5.1 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 5.2 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).

5.5 La détermination de l'ordre d'un modèle autorégressif

Supposons d'abord 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 5.4, 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 5.3 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 5.4

Théorème 5.6 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-être 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 :

$$\hat{\varphi}_2^{(2)} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (30)$$

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 $ARMA(p, q)$.

5.6 Exercices : TD 3

1. Soit le processus :

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

- 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 .
 - 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?
 - Donnez la prévision k pas en avant $\hat{Y}_t(k)$, en utilisant les valeurs Y_t, Y_{t-1}, Z_t .
2. Trouvez les coefficients d'autocorrélation et tracez la corrélogramme pour le processus ARMA(1,2) $Y_t = 0.6Y_{t-1} + \epsilon_t - 0.3\epsilon_{t-1} - 0.1\epsilon_{t-2}$
3. Prévision du modèle ARIMA(2,1,0)

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

- Verifiez que :
 - $X_t(k) = (1 + \phi_1 + \phi_2)X_t(k-1) - (\phi_1 + \phi_2 + \phi_1\phi_2)X_t(k-2) + \phi_1\phi_2X_t(k-3)$, $k \geq 1$
 - $X_t(k) = A_t + B_t\phi_1^k + (X_t - A_t - B_t)\phi_2^k$, $k = 0, 1, 2, \dots$ pour certaines A_t et B_t qui dépendent seulement du t .
 - Trouvez les poids des valeurs pivots X_t, X_{t-1}, X_{t-2} dans la prévision Box-Jenkins de X_t
 - Trouvez la limite $\lim_{k \rightarrow \infty} X_t(k)$
4. 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 $-1 < \theta < 1$.

- Montrez que $Y_t(1) = (1 + \varphi)Y_t - \varphi Y_{t-1} + \theta\epsilon_t$ et $Y_t(k) = (1 + \varphi)Y_t(k-1) - \varphi Y_t(k-2)$, $k \geq 2$.
- 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 $\lim_{k \rightarrow \infty} Y_t(k)$

- Montrez que $Y_t(1) = -\theta Y_{t-1}(1) + (1 + \varphi + \theta)Y_t - \varphi Y_{t-1}$ et $Y_t(k) = Y_{t-1}(k+1) + \psi_k \epsilon_t$.
- 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]
- 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

5. 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 .

- 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 $k \geq 3$
- 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 $Y_t, Y_{t-1}, \epsilon_t, \epsilon_{t-1}$.
- Montrez que $Y_t(k)$ peut s'exprimer en fonction seulement des valeurs passées de la série.
- 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.

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

Considérons par exemple ARIMA(p,2,0) ; ce modèle inclue une tendance linéaire, pour la quelle le premier candidat est la droite par les deux dernières points pivots (comme dans le cas $p = 0$, quand les pivots coïncident avec les valeurs X_t, X_{t-1}). En général, les prévisions doivent encore converger asymptotiquement vers une droite. Pour $p = 0$, on commence déjà 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 précèdent les 2 dernières point auront aussi une influence sur la "droite de tendance".

6 L'estimation des modèles ARIMA(p,d,q)

Les deux méthodes principales pour l'estimation des paramètres sont la méthode des moments et la maximisation de la vraisemblance. La première méthode s'appuie sur les formules théoriques des moments, en l'occurrence les corrélations.

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

Pour les modèles AR(p) causaux, la régression classique nous donne les coefficients, et aussi des tests (Fisher, etc) pour décider si on a enlevé toute la structure des résidus.

La corrélogramme des processus autorégressifs AR(p) 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(∞) en inversant le filtre $\varphi(B)$; en suite on utilise la formule (17) de la corrélogramme pour les processus MA(∞).
2. Solution directe, en utilisant les équations de Yule-Walker.

Exemple 6.1 AR(1) : Calcul de la corrélogramme par :

1. Les équations Yule-Walker.
2. La représentation MA(∞).

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 :

$$\begin{aligned} \gamma_k &= \sum_{i=1}^p \varphi_i \gamma_{k-i} \quad \text{pour } k \geq 1 \\ \gamma_0 &= \sum_{i=1}^p \varphi_i \gamma_i + \mathbb{E}Y_t \epsilon_t = \sum_{i=1}^p \varphi_i \gamma_i + \sigma^2 \quad \text{pour } k = 0 \end{aligned} \quad (31)$$

2. Pour trouver les corrélations, on remarque d'abord, en divisant par γ_0 , qu'elles satisfont aussi la récurrence (31) :

$$\rho_k = \sum_{i=1}^p \varphi_i \rho_{k-i} \quad \text{pour } k \geq 1 \quad (32)$$

Alors, il suffit de déterminer les premières p corrélations, ce qu'on fait en appliquant la récurrence (32) 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 $\rho = (\rho(1), \dots, \rho(p))$:

$$\boxed{\mathbf{R} \varphi = \rho} \quad (33)$$

où \mathbf{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 (33) permettent de calculer les coefficients φ à partir des corrélations et viceversa (on calcule d'abord les premières p corrélations ; en suite, on trouve aussi les autres corrélations en utilisant la récurrence).

Note : En prenant γ_0 comme facteur commun dans la deuxième équation en (31), 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 6.2 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 causaux. 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 6.3 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 6.1 (*) **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 (33) 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 $0 = \lambda^2 - \phi_1 \lambda - \phi_2 = \lambda^2 \varphi(\lambda^{-1})$ (qui sont dedans le cercle unitaire, par la causalité) sont distinctes, on obtient

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

Finalement, en utilisant $\varphi_1 = \lambda_1 + \lambda_2, \varphi_2 = -\lambda_1 \lambda_2$, 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 :

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

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

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 autrement (par la causalité). On obtient :

$$\gamma_k = \sum_{i=1}^p \varphi_i \gamma_{k-i} + \sum_{0 \leq j \leq q} \theta_j \gamma_{k-j}^{(Y, \epsilon)} = \sum_{i=1}^p \varphi_i \gamma_{k-i} + \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{k-j} \quad (34)$$

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 (34) et la symmetrie 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 recurrence

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

Exemple 6.4 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 :

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

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

En général, nous trouvons :

Théorème 6.2 (*) Les premières $p + 1$ covariances s'obtiennent du système à $p + 1$ équations et $p + 1$ inconnues :

$$\mathbf{\Gamma} \begin{pmatrix} 1 \\ -\varphi_1 \\ -\varphi_2 \\ \dots \\ -\varphi_p \end{pmatrix} = \sigma^2 \mathbf{\Psi} \begin{pmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_q \end{pmatrix} \quad (35)$$

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

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

Ψ est la matrice des dimensions $(p+1) \times (q+1)$:

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

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

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

6.3 Intervalles de confiance pour la prédiction

Théorème 6.3 (*) 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

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

et donc

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

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

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 :

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

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 verra ça dans le cadre plus général des processus **non stationnaires** ARIMA(p,d,q).

6.4 Le lissage exponentiel (*)

Les méthodes de lissage exponentiel sont des méthodes de prévision très simples à mettre en oeuvre. On considère qu'une série (y_i) est observée entre des instants t_1 et t_n . On cherche alors à prédire y_{n+1} .

Les trois variantes proposées ici sont adaptées à trois situations différentes :

- le lissage exponentiel simple : pas de tendance, pas de saisonnalité,
- le lissage exponentiel de Holt : tendance, pas de saisonnalité,
- le lissage exponentiel de Winters : tendance et saisonnalité.

Pour simplifier les notations, on considèrera ici que $t_i = i$ pour tout $i \geq 1$.

6.4.1 Le lissage exponentiel simple

Cette méthode ne doit être employée qu'en présence d'une série de fluctuations irrégulières, c'est-à-dire une série qui ne présente ni tendance ni composante saisonnière.

Pour prédire y_{n+1} , l'idée est particulièrement simple : on combine la dernière observation y_n avec la dernière prévision effectuée. Notons \hat{y}_{n+1} la prévision de y_{n+1} . On pose alors :

$$\hat{y}_{n+1} = \alpha y_n + (1 - \alpha)\hat{y}_n,$$

où α , appelé *la constante de lissage*, est le plus souvent compris entre 0 et 1.

Remarques :

1. Notons $e_i = y_i - \hat{y}_i$ l'erreur de prévision au temps t_i . On peut alors exprimer \hat{y}_i comme suit :

$$\begin{aligned}\hat{y}_{i+1} &= \alpha(y_i - \hat{y}_i) + \hat{y}_i \\ &= \hat{y}_i + \alpha e_i.\end{aligned}$$

La nouvelle prévision apparaît donc comme l'ancienne prévision corrigée. L'ampleur de la correction est proportionnelle à la dernière erreur de prévision. C'est ce qu'on appelle la présentation de *mise à jour par l'erreur* du lissage exponentiel simple.

2. Il est possible d'exprimer la nouvelle prévision \hat{y}_{n+1} en fonction des données passées. En effet, on peut successivement remplacer $\hat{y}_n, \hat{y}_{n-1}, \dots$ par leur expression :

$$\begin{aligned}\hat{y}_{n+1} &= \alpha y_n + (1 - \alpha)(\alpha y_{n-1} + (1 - \alpha)\hat{y}_{n-1}) \\ &= \alpha y_n + \alpha(1 - \alpha)y_{n-1} + (1 - \alpha)^2(\alpha y_{n-2} + (1 - \alpha)\hat{y}_{n-2}) \\ &= \sum_{j=0}^{n-1} \alpha(1 - \alpha)^j y_{n-j} + (1 - \alpha)^n \hat{y}_1.\end{aligned}$$

On peut donc interpréter le lissage exponentiel simple comme une moyenne mobile. Si $\hat{y}_1 = y_1$, les coefficients de cette moyenne mobile sont $[\alpha, (1 - \alpha)\alpha, (1 - \alpha)^2\alpha, \dots, (1 - \alpha)^{n-2}\alpha, (1 - \alpha)^{n-1}]$. On vérifie bien que la somme de ces coefficients vaut 1. Cela explique le terme de *lissage*. D'autre part, ces coefficients décroissent de façon *exponentielle*.

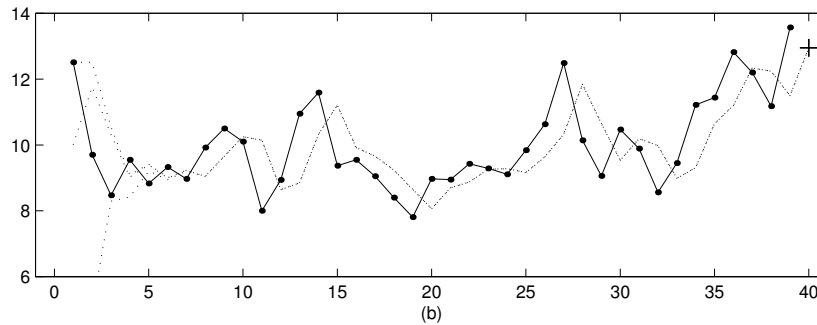
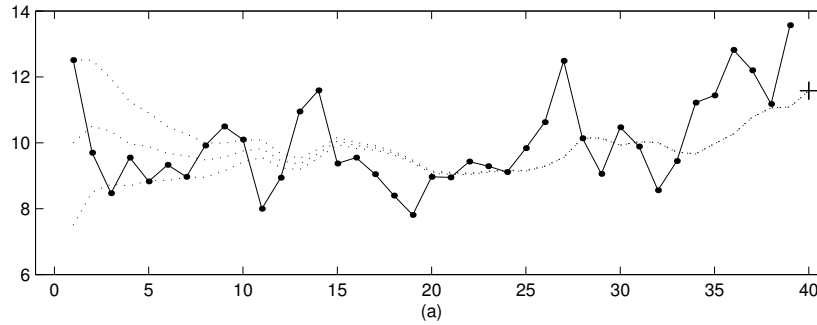
Cette méthode nécessite bien sûr d'initialiser la série des prévisions en choisissant une prévision initiale \hat{y}_1 . On peut par exemple :

- choisir $\hat{y}_1 = y_1$,
- choisir $\hat{y}_1 = \bar{y}$.

Quand au choix de α , remarquons les deux cas particuliers suivants :

- si $\alpha = 0$, alors $\hat{y}_{n+1} = \hat{y}_1$: toutes les prévisions sont identiques,
- si $\alpha = 1$, alors $\hat{y}_{n+1} = y_n$: la prévision est égale à la dernière valeur observée.

La Figure 2 nous montre les séries de prévisions obtenues pour une même série de longueur $n = 39$, avec la constante de lissage égale à 0.2 et à 0.7, ainsi qu'avec différentes initialisations ($\hat{y}_1 = y_1, \hat{y}_1 = \bar{y}$ et $\hat{y}_1 = 2\bar{y} - y_1$).



Le choix de l'initialisation importe peu, puisque cette initialisation est rapidement "oubliée". Cet oubli est d'autant plus rapide que la constante est proche de 1. Par contre, le choix de α est clairement important. Les séries de prévisions obtenues avec $\alpha = 0.2$ et $\alpha = 0.7$ sont très différentes.

FIG. 2 – *Lissage exponentiel simple. Différentes prévisions obtenues avec différentes constantes de lissage et différentes initialisations, (a) $\alpha = 0.2$, (b) $\alpha = 0.7$. La prévision \hat{y}_{40} est indiquée par une croix.*

En particulier, la figure montre bien que la valeur de \hat{y}_{40} dépend du choix de α ($\hat{y}_{40} = 11.58$ si $\alpha = 0.2$ et $\hat{y}_{40} = 12.95$ si $\alpha = 0.7$).

Une façon de choisir le coefficient de lissage consiste à déterminer la valeur de α qui minimise les écarts entre la série observée et la série des prédictions. On peut en fait ne considérer que les écarts obtenus sur la deuxième moitié de la série, afin de ne pas tenir compte de l'initialisation. On cherche alors α qui minimise

$$S(\alpha) = \sum_{i=[n/2]+1}^n (y_i - \hat{y}_i)^2.$$

Dans l'exemple proposé Figure 2, $S(\alpha)$ est minimum pour $\alpha = 0.61$, et ce, quelque soit l'initialisation choisie pour \hat{y}_1 . Donc, puisque la série des prévisions s'ajuste le mieux à la série des observations pour cette valeur de la constante de lissage, on peut espérer que la prévision \hat{y}_{40} obtenue avec $\alpha = 0.61$ soit meilleure que celle obtenue avec une autre valeur de α .

La Figure 3 nous présente les prévisions obtenue avec $\alpha = 0$, $\alpha = 1$ et $\alpha = 0.61$.

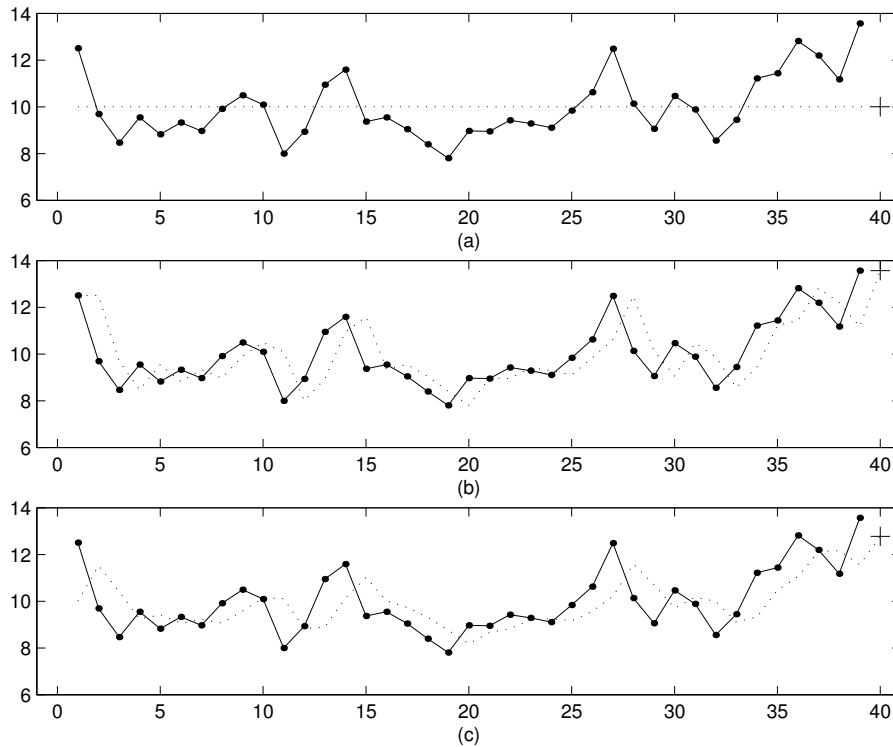


FIG. 3 – *La série des prévisions obtenues avec (a) $\alpha = 0$, (b) $\alpha = 1$, (c) $\alpha = 0.61$. La prévision \hat{y}_{40} est indiquée par une croix.*

Pour ces trois constantes de lissage, on obtient les résultats suivants :

α	$S(\alpha)$	\hat{y}_{40}
0	44.30	10.01
1	30.04	13.57
0.61	28.13	12.78

6.4.2 Le lissage exponentiel de Holt

La méthode de *lissage exponentiel de Holt*, encore appelée méthode de Holt-Winters avec tendance, mais sans saisonnier, utilise une fonction de prévision *localement linéaire* :

$$\hat{y}_{n+h} = a_n + h b_n$$

pour prédire y_{n+h} à partir des observations y_1, \dots, y_n .

Le niveau a_n et la pente b_n sont mis à jour au moyen des formules suivantes :

$$\begin{aligned} a_n &= \alpha y_n + (1 - \alpha) \hat{y}_n \\ b_n &= \gamma (a_n - a_{n-1}) + (1 - \gamma) b_{n-1}. \end{aligned}$$

La droite de tendance locale au temps t_{n-1} d'équation $f(t) = a_{n-1} + b_{n-1}(t - t_{n-1})$ passe par les points (t_{n-1}, a_{n-1}) et $(t_n, a_{n-1} + b_{n-1})$.

Pour obtenir a_n , la première des deux relations réalise une moyenne pondérée entre l'observation y_n et la prévision $a_{n-1} + b_{n-1}$ donnée par la droite.

Pour trouver la pente b_n , on effectue une moyenne pondérée entre la pente entre les points (t_{n-1}, a_{n-1}) et (t_n, a_n) , et la dernière prévision de la pente b_{n-1} .

Un choix raisonnable des prévisions initiales consiste à poser, par exemple,

$$a_1 = y_1, \quad b_1 = 0.$$

On peut, comme pour le lissage exponentiel simple, mettre en évidence la présentation de mise à jour par l'erreur. En effet, on montre facilement que, pour tout $i \geq 2$,

$$\begin{aligned} a_i &= a_{i-1} + b_{i-1} + \alpha e_i \\ b_i &= b_{i-1} + \alpha \gamma e_i \end{aligned}$$

6.5 Le lissage exponentiel de Winters

La méthode de *lissage exponentiel de Winters*, encore appelée méthode de Holt-Winters avec saisonnalité, généralise la méthode de Holt, en introduisant une composante saisonnière (s_i), de période p . Cette composante intervient de façon additive ou de façon multiplicative.

6.5.1 modèle de Winters additif

La composante est introduite de façon additive comme suit :

$$\hat{y}_{n+h} = a_n + h b_n + s_{n+h-p}$$

où les trois composantes sont réactualisées de la façon suivante :

$$\begin{aligned} a_n &= \alpha (y_n - s_{n-p}) + (1 - \alpha) (\hat{y}_n - s_{n-p}) \\ b_n &= \gamma (a_n - a_{n-1}) + (1 - \gamma) b_{n-1} \\ s_n &= \delta (y_n - a_n) + (1 - \delta) s_{n-p} \end{aligned}$$

On voit donc que les deux premières formules de mise à jour de la méthode de Holt ont été reprises, mais celle relative au niveau est appliquée sur la donnée corrigée des variations saisonnières $y_n - s_{n-p}$, avec l'estimation la plus récente du coefficient saisonnier correspondant.

La composante saisonnière s_n est une combinaison de la variation saisonnière $y_n - a_n$ et de l'ancienne prévision du coefficient saisonnier s_{n-p} .

On peut choisir les initialisations suivantes :

$$\begin{aligned} a_1 = a_2 = \dots = a_p &= \frac{1}{p} \sum_{i=1}^p y_i \\ T_1 &= 0 \\ s_i &= y_i - a_i \quad 1 \leq i \leq p. \end{aligned}$$

Une présentation de mise à jour par l'erreur peut être aisément établie :

$$a_i = a_{i-1} + b_{i-1} + \alpha e_i$$

$$b_i = b_{i-1} + \alpha \gamma e_i$$

$$s_i = s_{i-p} + \delta(1 - \alpha)e_i$$

6.5.2 modèle de Winters multiplicatif

La composante est introduite de façon multiplicative comme suit :

$$\hat{y}_{n+h} = (a_n + hb_n)(1 + s_{n+h-p})$$

où les trois composantes sont réactualisées de la façon suivante :

$$a_n = \alpha \frac{y_n}{1 + s_{n-p}} + (1 - \alpha)(a_{n-1} + b_{n-1})$$

$$b_n = \gamma(a_n - a_{n-1}) + (1 - \gamma)b_{n-1}$$

$$s_n = \delta \left(\frac{y_n}{a_n} - 1 \right) + (1 - \delta)s_{n-p}$$

6.6 TP

1. On considère les quatre couples de variables suivants,

X1	Y1	X2	Y2	X3	Y3	X4	Y4
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

- Pour chacun de ces quatre couples, calculer les moyennes et variances de X et Y , ainsi que la covariance et le coefficient de corrélation entre X et Y .
- En déduire les équations des quatre droites de régression de Y en X .
- Représenter graphiquement ces quatre nuages de points et les droites de régression.
- Commenter.

2. On considère la série des ventes trimestrielles de parapluies suivante :

Trimestre	I	II	III	IV
1995	30	15	5	30
1996	36	18	9	36
1997	45	15	10	60
1998	48	16	8	72
1999	54	18	9	45

Calculer la moyenne mobile d'ordre 4 et la représenter graphiquement avec la série des ventes. Commenter ce graphique.

3. En utilisant la méthode du lissage exponentiel simple, déterminer les prévisions $(\hat{y}_i, 1 \leq i \leq 10)$ obtenues pour la série (y_i) suivante, avec une constante de lissage $\alpha = 0, 0.1, 0.2, \dots, 1$

i	1	2	3	4	5	6	7	8	9	10
y_i	4	2	3	5	1	4	2	3	8	7

Calculer $S(\alpha)$ pour chacune de ces valeurs de α . Quelle prévision proposeriez vous pour y_{11} ?

2. Reprendre l'exercice 1. avec la série (y_i) suivante :

i	1	2	3	4	5	6	7	8	9	10
y_i	3	3	2	2	3	1	0	0	1	1

4. En utilisant la méthode de Holt (avec tendance, sans composante saisonnière), déterminer les prévisions $(\hat{y}_i, 1 \leq i \leq 10)$ obtenues pour la série (y_i) suivante, avec $\alpha = 0, 0.5, 1$ et $\gamma = 0, 0.5, 1$.

i	1	2	3	4	5	6	7	8	9	10
y_i	4	3	6	5	7	8	10	9	7	6

Calculer $S(\alpha, \gamma)$ pour chacune de ces valeurs de (α, γ) . Quelle prévision proposeriez vous pour y_{11} ?

5. On considère les ventes trimestrielles de parapluies d'un magasin (en milliers), entre 1998 et 2000 .

i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	30	17	9	36	44	28	21	50	61	33	30	82

- a) En utilisant la méthode de Winters (avec tendance et composante saisonnière), dans le cas additif et multiplicatif, déterminer les prévisions $(\hat{y}_i, 1 \leq i \leq 12)$ obtenues avec différentes valeurs (α, γ, δ) , choisies dans $\{0, 0.5, 1\}$ par exemple.
b) Calculer $S(\alpha, \gamma, \delta)$ pour chacune de ces valeurs de (α, γ, δ) et chacun des modèles.
c) Les valeurs "optimales" des coefficients sont $(0.2, 0.8, 0.4)$ pour un modèle additif et $(0.1, 0.9, 0.4)$ pour un modèle multiplicatif. Quel modèle vous semble le mieux convenir ?
d) Quelles prévisions proposeriez vous pour les ventes de 2001 ?
6. Appliquer la méthode de Winters pour prédire le nombre de tests VIH de novembre 1999 à octobre 2000 (les valeurs "optimales" des coefficients sont $(0.1, 0., 0.3)$ pour un modèle additif comme pour un modèle multiplicatif).

6.7 Modèles d'espace-état

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** :

$$X_{t+1} = A_t X_t + b_t + \epsilon_X(t) \quad \text{équation d'évolution} \quad (36)$$

$$Y_t = C_t X_t + d_t + \epsilon_Y(t) \quad \text{équation d'observation} \quad (37)$$

Exemple 6.5 Le modèle additif cherche une décomposition de la forme :

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

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 devient utile ssi on peut assurer que les résidus aient une structure statistique simple comme "bruit blanc Gaussien" ou "bruit blanc de second ordre", moyennes mobiles, etc.

6.8 Contrôle continu en séries temporelles

1. Déterminer une moyenne mobile causale $\theta(B) = \sum_{i=0}^q \theta_i B^i$ d'ordre q minimal, qui laisse passer une tendance quadratique sans distortion et qui enlève les composantes saisonnières d'ordre 3.
2. Soit $X_t = \sum_{i=0}^k \theta_i \epsilon_{t-i}$ un processus moyenne mobile, utilisé pour lisser un bruit blanc $\epsilon_t = BB(0, \sigma^2 = 1)$.
 - a) Quelle est la variance de X_t ?
 - b) Trouvez le filtre (i.e. les coefficients $(\theta_0, \dots, \theta_k)$) qui ramènent à un processus X_t à variance minimale, parmi toutes les moyennes mobiles d'ordre k qui laissent passer une tendance constante sans distortion.
3. On considère le processus aléatoire suivant :

$$X_t - .2 X_{t-1} - .35 X_{t-2} = 45 + \epsilon_t$$

ou ϵ_t est $BB(0, \sigma^2 = 1)$.

- (a) Calculer l'espérance de X_t , $m = \mathbb{E}X_t$, en supposant que le processus est stationnaire.
- (b) Quelle est l'équation de récurrence satisfaite par le processus $Y_t = X_t - m$?
- (c) Donner les équations de Yule-Walker pour les autocorrélations du processus Y_t , calculer les 3 premières autocorrélations, ainsi que la variance.
- (d) Est-ce que le processus Y_t est stationnaire causal ? Si oui, donnez une formule générale pour les coefficients ψ_k de sa représentation causale $Y_t = \sum_k \psi_k \epsilon_{t-k}$.

Solutions :

1. En demandant que $\theta(B) = \frac{1+B+B^2}{3}(a_0 + a_1B + a_2B^2)$ satisfait $\theta(1) = 1, \theta'(1) = 0, \theta''(1) = 0$, on trouve : $\theta(B) = 1/9(8 - 7B + 2B^2)(1 + B + B^2) = 8/9 + B/9 + B^2/3 - 5B^2/9 + 2B^4/9$.
2. a) $\text{Var } X_t = \sum_i \theta_i^2$. b) Elle est minimisé sous la contrainte $\sum_i \theta_i = 1$ par la moyenne arithmétique à coefficients égaux.
3. (a) Ce processus AR(2) pas centré peut s'écrire $(1 - .2B - .35B^2)X_t = (1 - .7B)(1 + .5B)X_t = 45 + \epsilon_t$ En prenant l'espérance on trouve $\mathbb{E}(X_t) = \frac{45}{1-.2-.35} = \frac{40}{.45} = 100$.
(b) Soit $Y_t = X_t - \mathbb{E}X_t$. Alors Y_t est un processus AR(2) centré satisfaisant $(1 - .2B - .35B^2)Y_t = (1 - .7B)(1 + .5B)X_t = \epsilon_t$
d) 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 l'équation Yule Walker

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2}$$

Pour $k = 1, 2$ on a le système de Yule Walker :

$$\begin{aligned}\rho_1 &= 0.2 + 0.35\rho_1 \\ \rho_2 &= 0.2\rho_1 + 0.35\end{aligned}$$

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) Le processus est stationnaire causal car les racines du polynôme sont à l'extérieur du disque unité. Les coefficients $\psi(k)$ satisfont aussi l'équation Yule Walker

$$\psi_k = \varphi_1 \psi_{k-1} + \varphi_2 \psi_{k-2}$$

avec valeurs initiales $\psi_0 = 1, \psi_1 = \varphi_1$.

La solution générale de cette récurrence d'ordre 2, pour valeurs initiales arbitraires x_0, x_1 et avec des racines distinctes λ_1, λ_2 , est :

$$\psi(k) = \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} x_1 - \frac{(\lambda_1^k \lambda_2 - \lambda_2^k \lambda_1)}{\lambda_1 - \lambda_2} x_0$$

et donc ici on obtient cette formule, avec $x_0 = 1, x_1 = \varphi_1$.

7 Examens d'entraînement

7.1 Examen d'entraînement 1

1. Soit Y_t un processus ARMA(1,1) vérifiant l'équation $Y_t = -0.5Y_{t-1} + \epsilon_t - 0.8\epsilon_{t-1}$, 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 $MA(\infty)$ et les coefficients π_j de sa représentation comme processus $AR(\infty)$ et précisez si ces représentations sont convergentes.
2. 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 prévision 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 $Y_t, Y_{t-1}, Z_t, Z_{t-1}$.

3. 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.

4. *Processus ARMA(2,1)* 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

5. 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 prévision 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 $Y_t, Y_{t-1}, Z_t, Z_{t-1}$.

6. On considère la série suivante :

t_i	1	2	3	4	5	6	7	8	9	10
y_i	8	4	13	15	18	15	19	19	20	28

- a) Représenter graphiquement cette série.
- b) On se propose d'ajuster une droite de régression

$$f(t) = at + b$$

de moindre carrées. Déterminer les coefficients a et b .

- c) Quel est le point le plus "éloigné" de $f(t)$?

- c) Calculez une droite de régression $f(t)$ en imposant $b \geq 6$.
7. On considère une série $(y_i, 1 \leq i \leq n)$ périodique, de période p . On suppose que le nombre d'observations n est un multiple de p : $n = Lp$. Calculez les corrélations empiriques :

$$\rho(p), ; \rho(2p), ; \dots ; \rho(jp) \dots$$

en utilisant la définition des corrélations empiriques :

$$\rho(k) = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

où $\bar{y} = (\sum_{i=1}^n y_i) / n$ est la moyenne de l'ensemble de la série $(y_i, 1 \leq i \leq n)$.

Commentez sur le résultat.

8. Quelle est la limite $\lim_{k \rightarrow \infty} X_t(k)$ de la prévision linéaire d'un processus ARMA (p,q) ?

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 $\pi_1 = .8, \dots$

Examen : 10 Janvier 2005, UPPA, Dépt. de Mathématiques
Module : Séries temporelles : TMUZ44U
Durée : deux heures

1. Déterminer une moyenne mobile causale $\theta(B) = \sum_{i=0}^q \theta_i B^i$ d'ordre q minimal, qui laisse passer une tendance quadratique sans distortion et qui enlève les composantes saisonnières d'ordre 4.
2. Soit $X_t = \sum_{i=0}^3 \theta_i \epsilon_{t-i}$ un processus moyenne mobile, utilisé pour lisser un bruit blanc $\epsilon_t = BB(0, \sigma^2 = 1)$.
- a) Quelle est la variance de X_t ?
- b) Trouvez le filtre (i.e. les coefficients $(\theta_0, \theta_1, \theta_2, \theta_3)$) qui ramènent à un processus X_t à variance minimale, parmi toutes les moyennes mobiles d'ordre 3 qui laissent passer une tendance affine $X_t = a + bt$ sans distortion.

3. Stationarité des processus Soit

$$X_{t+1} = \phi X_t + \epsilon_{t+1}, t = 0, 1, 2, \dots,$$

où ϵ_t est bruit blanc Gaussien de distribution $N(0, \sigma^2)$, un processus $AR(1)$ défini seulement à partir du temps $t = 0$. Supposons aussi que X_0 est Gaussien de distribution $N(0, \sigma_0^2)$. Déterminez la valeur de σ_0^2 qui assure que le processus X_t est stationnaire.

4. Considérez le modèle **ARIMA(1,1,0)**

$$(1 - \lambda B)(1 - B)X_t = \epsilon_t$$

où $|\lambda| < 1$ et ϵ_t est bruit blanc Gaussien de distribution $N(0, 1)$.

- (a) Donnez une formule de récurrence et ensuite une formule générale pour les coefficients ψ_k de sa représentation causale $X_t = \sum_k \psi_k \epsilon_{t-k}$.
- (b) Est-ce que ce processus est stationnaire causal ?
- (c) Donner une formule de récurrence et ensuite une formule générale pour les autocorrélations ρ_k du processus X_t . Est-ce que $\sum_k \rho_k < \infty$?
- (d) Donnez une formule de récurrence et ensuite une formule générale pour la prévision Box-Jenkins $X_t(k) = \mathbb{E}[X(t+k)|X_t]$ de $X(t+k)$, $k = 0, 1, 2, \dots$
- (e) Trouvez la limite $\lim_{k \rightarrow \infty} X_t(k)$.

Solutions :

1. En demandant que $\theta(B) = \frac{1+B+B^2+B^2}{3}(a_0 + a_1 B + a_2 B^2)$ satisfait $\theta(1) = 1, \theta'(1) = 0, \theta''(1) = 0$, on trouve : $\theta(B) = \frac{15-16B+5B^2}{4} \frac{1+B+B^2+B^3}{4} = \frac{15}{16} - \frac{1}{16}B + \frac{4}{16}B^2 - \frac{4}{16}B^3 - \frac{1}{16}B^4 + \frac{5}{16}B^5$.
2. a) $\text{Var } X_t = \sum_i \theta_i^2$. b) Elle est minimisé sous les contraintes $\sum_i \theta_i = 1, \sum_i i\theta_i = 0$ par $(7/10, 4/10, 1/10, -2/10)$.
3. $\sigma_0^2 = \frac{\sigma^2}{1-\varphi^2}$
4. (a) Les coefficients $\psi(k)$ satisfont l'équation Yule Walker

$$\psi_k = \varphi_1 \psi_{k-1} + \varphi_2 \psi_{k-2}$$

avec valeurs initiales $\psi_0 = 1, \psi_1 = \varphi_1$.

La solution générale de cette récurrence d'ordre 2, pour valeurs initiales arbitraires x_0, x_1 et avec des racines distinctes λ_1, λ_2 , est :

$$\psi(k) = \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} x_1 - \frac{(\lambda_1^k \lambda_2 - \lambda_2^k \lambda_1)}{\lambda_1 - \lambda_2} x_0$$

(et avec des racines non-distinctes $\lambda_1 = \lambda_2 = \lambda$ nous avons $\psi(k) = k\lambda^{k-1}x_1 + (1 - k\lambda^k)x_0$).

Ici, avec $\lambda_2 = 1$, on obtient par cette formule, (avec $x_0 = 1, x_1 = \varphi_1$) $\psi(k) = \frac{\lambda^{k+1}-1}{\lambda-1}$ (et avec des racines non-distinctes nous avons $\psi(k) = (k+1)\lambda^k$).

b) Le processus n'est pas stationnaire causal car la racines 1 du polynôme caractéristique n'est pas à l'extérieur du disque unité.

(c) Pour un processus AR stationnaire X_t , la fonction d'autocorrélation satisfait l'équation Yule Walker

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2}$$

Pour $k = 1$ ça donne :

$$\rho_1 = \frac{\varphi_1}{1 - \varphi_2} = \frac{1 + \lambda}{1 + \lambda} = 1$$

et ensuite $\rho_k = 1$. Le seul processus stationnaire Gaussien satisfaisant cette recurrence est $X_t = X_{t-1} = \dots$, quand $\sigma = 0$. Pour $\sigma \neq 0$, il n'existe pas de processus stationnaire satisfaisant notre recurrence.

d) Avec la même recurrence, on arrive à :

$$X_t(k) = \frac{\lambda^{k+1} - 1}{\lambda - 1} X_t + \frac{\lambda^{k+1} - \lambda}{\lambda - 1} X_{t-1}$$

7.2 Examen d'entraînement 2

1. 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'auto-corrélation 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

2. Soit Y_t un processus ARMA(2,1) vérifiant l'équation $Y_t = 0.7Y_{t-1} - 0.1Y_{t-2} + \epsilon_t - 2\epsilon_{t-1}$, avec ϵ_t un bruit blanc.

- Précisez si le processus est stationnaire, causal et inversible, et calculez sa fonction d'autocovariance.
- Trouvez les coefficients ψ_j de sa représentation comme processus MA(∞) et les coefficients π_j de sa représentation comme processus AR(∞) et précisez si ces représentations sont convergentes.

3. Prédiction du modèle ARIMA(1,2,0) $(1 - \varphi B)(1 - B)^2 X_t = \epsilon_t$.

- 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} .
- Trouvez la limite $\lim_{k \rightarrow \infty} X_t(k)$.
- Calculez les coefficients ψ_k de l'expansion causale de ce processus.

4. Calculer la fonction d'autocovariance $\Gamma_k = \mathbb{E}Y_{n+k}Y_n^t$ du processus à valeurs en \mathbb{R}^2 :

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

où ϵ_n est un bruit blanc standard.

5. Stationnarité des processus

Investiguer si les processus suivants sont stationnaires de second ordre, en calculant leurs caractéristiques statistiques de premier et second ordre (moyennes, variances, covariances). On convient de noter par $\{\epsilon_t\}$ le bruit blanc.

- $Y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$ (MA(1))
- $Y_t = Y_{t-1} + \mu + \epsilon_t$ (marche aléatoire avec tendance)
- $Y_t = \varphi Y_{t-1} + \mu + \epsilon_t$ ($\varphi \neq 1$) (AR(1) avec tendance)

6. Processus AR(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, $\sigma^2 = 1$).

- Calculer l'espérance de X_t , en supposant que le processus est stationnaire.
- Est-ce que le processus est stationnaire causal ?
- Donner la recursion de Yule-Walker pour les prévisions $\hat{X}(t+k|t)$ du processus, et résolvez la en termes des valeurs pivotales X_t, X_{t-1} .
Mêmes questions pour

$$(1 - B + .25B^2)X_t = \epsilon_t$$

7. Processus ARMA(2,1)

Soit le processus :

$$(1 - .6B + .08B^2)Y_t = (1 + \theta B)\epsilon_t$$

- Est-ce que ce processus est stationnaire causal ? Si oui, obtenez la "représentation ψ " du Y_t par rapport au bruit Z_t .
- 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 ?
- Donnez la prévision k pas en avant $\hat{Y}_t(k)$, en utilisant les valeurs Y_t, Y_{t-1}, ϵ_t .

8. Montrez que pour ARIMA(1,2,0),

$$\psi_k = \{\varphi^{k+2} + (1 - 2\varphi) + k(1 - \varphi)\}(1 - \varphi)^{-2}, k \geq 0$$

7.3 Examen d'entraînement 3

2. On considère le processus aléatoire suivant :

$$X_t = 20 - .4 X_{t-1} + .12 X_{t-2} + \epsilon_t$$

ou ϵ_t est $BB(0, \sigma^2 = 1)$.

- Calculer l'espérance de X_t , en supposant que le processus est stationnaire.
- Est-ce que le processus est stationnaire causal ?
- Donner les équations de Yule-Walker du processus, calculer la variance, ainsi que les 3 premières valeurs des autocorrélations.
- 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 ?
- Calculer les 3 premières autocorrélations partielles.

3. Soit le processus :

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

- Est-ce que ce processus est inversible ? Calculez la formule générale des coefficients $\pi(k)$ de la "représentation π " du bruit ϵ_t en termes de la série. Quels sont les premiers quatre coefficients De quel problème on s'aperçoit si le processus n'est pas inversible ?
- Donnez la prévision un pas en avant $\hat{Y}_t(1)$ qui utilise les trois valeurs $Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}$.
- Est-ce que ce processus est stationnaire causal ? Si oui, trouvez la fonction d'autocorrélation de Y_t .

4. Considérons le processus :

$$(1 - B)Y_t = (1 + \beta_1 B + \beta_2 B^2)\epsilon_t$$

Soit $Y_t(k)$ la prévision de Y_{t+k} au temps t .

- Trouvez des expressions pour $Y_t(1), Y_t(2)$, comme fonctions de $Y_t, Y_{t-1}, \epsilon_t, \epsilon_{t-1}$. Donnez une formule de récurrence pour $Y_t(k)$ pour $k \geq 3$, et écrivez ensuite $Y_t(k)$ comme une fonction de $Y_t, Y_{t-1}, \epsilon_t, \epsilon_{t-1}$.
- Trouvez des expressions pour $Y_t(1), Y_t(2)$, comme fonctions de $Y_t, Y_{t-1}, Y_{t-2}, \dots$ en utilisant la représentation π du bruit.

- Donnez les formules des coefficients de corrélation ρ_1, ρ_2 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.

Solutions :

- Le processus peut s'écrire $Y_t = (1 + \tau_1 B + \tau_2 B^2)\epsilon_t = (1 + .4B)^2\epsilon_t$. 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 $\pi_1 = .8, \dots$

3. (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 :

$$\begin{aligned}\rho_1 &= 0.2 + 0.35\rho_1 \\ \rho_2 &= 0.2\rho_1 + 0.35\end{aligned}$$

La première équation donne $\rho_1 = \frac{.2}{.65} = .31$, 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.

Soit $X_t = X_0 + \epsilon_1 + \epsilon_2 + \dots + \epsilon_n$ une marche aléatoire "infinitésimale" sur \mathbb{R} , i.e. une marche avec (ϵ_n) une suite de variables aléatoires réelles indépendantes de même loi $P[\epsilon_n = D] = p$ et $P[\epsilon_n = -D] = q = 1 - p$, ou $p, q = \frac{1}{2} \pm \frac{\mu}{2\sigma^2}D$, $D^2 = \sigma^2h$, et $h \rightarrow 0, n \rightarrow \infty$ et $nh = t$.

Montrez que alors l'espérance et la variance de X_t approchent l'espérance et la variance d'un mouvement Brownien \tilde{X}_t de tendance μ et variabilité σ , ($\tilde{X}_t = X_0 + \sigma W_t + \mu t$, où W_t est le mouvement Brownien standard.)

Pour tout x, l, u de la forme $kD, k \in \mathbb{Z}$, nous considérons le processus \tilde{X}_t conditionné sur $\tilde{X}_0 = x$ pour $x \in [l, u]$, jusqu'au temps d'arrêt τ quand le processus sort de l'intervalle $[l, u]$.

1. Obtenez l'équation de récurrence et les conditions frontière satisfaites par $p_x = \mathbb{P}_x\{\tilde{X}_\tau = K\}$, $t_x = \mathbb{E}_x\tau$, et $f_x = \mathbb{E}_x\tilde{X}_\tau$ et $d_x = \mathbb{E}_x e^{-r\tau}$. *Indication* : Approximez par la marche aléatoire "infinitésimale" X_t , conditionnez sur le premier pas ϵ_1 et utilisez les relations :

$$\mathbb{E}_x[g(X_\tau) | \epsilon_1 = \pm 1] = p_{x\pm 1}, \quad \mathbb{E}_x[\sum_0^{\tau-1} c(X_i) | \epsilon_1 = \pm 1] = c(x) + t_{x\pm 1}, \dots$$

2. Résolvez les équations de récurrence qui en résultent pour p_x, t_x, f_x et d_x dans le cas $p = q = 1/2$.

3. (a) Obtenez l'équation différentielle et les conditions frontière satisfaites par le coût total espéré

$$f_\sigma(x) = \mathbb{E}_x \int_0^\tau c(\tilde{X}_t) dt$$

(où \tilde{X}_t est un mouvement Brownien de tendance μ et variabilité σ). *Indication* : Prenez la limite d'une équation de récurrence semblable à celle obtenue dans le problème 1 (a).i.

(b) Résolvez l'équation dans le cas $l = 0, u = 1, \mu < 0$, pour la fonction de coût unitaire $c(x) = 1$. Est-ce que le résultat dépend de σ ? Donnez une interprétation géométrique du cas $\sigma = 0$.

(c) Résolvez l'équation dans le cas $l = 0, u = 1, \mu < 0$, pour la fonction de coût unitaire $c(x) = x$. Donnez une interprétation géométrique du cas $\sigma = 0$. Quelle est la limite $\lim_{\sigma \rightarrow 1} \frac{f_\sigma(x)}{f_0(x)}$?

8 Sommaire des définitions et résultats dans les séries temporelles

8.1 Filtres

Définition 8.1 a) Une série s_t sera appelée saisonnière de période p ssi

$$\sum_{i=1}^p s_{t+i} = 0 \iff \left(\sum_{i=0}^{p-1} B^i \right) s_t = 0, \quad \forall t \quad (38)$$

b) Une série s_t sera appelée périodique de période p ssi

$$s_{t+p} = s_t \iff (1 - B^p)s_t = 0, \quad \forall t \quad (39)$$

Théorème 8.1 Un filtre $\psi(B)$ annule (ou "enlève") les composantes saisonnières 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 $z = 1$).

Théorème 8.2 L'espace invariant d'un filtre contient les polynômes de degré $\leq p$ ssi 1 est une racine d'ordre au moins $p+1$ de l'équation $\psi(z) = 1$, i.e. $\psi(1) = 1, \psi'(1) = 0, \psi''(1) = 0, \psi^{(p)}(1) = 0$.

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

Théorème 8.3 a) Un processus ARMA(p) est causal(inversible), i.e. il peut être représenté sous la forme : $Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ où $\sum \psi_i^2 < \infty$

($\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ où $\sum \pi_i^2 < \infty$) ssi toutes les racines de son polynôme caractéristique $\varphi(z)$ ($\theta(z)$) sont à l'extérieur du cercle unitaire. Les coefficients ψ_i ($\pi(i)$) sont dans ce cas les coefficients de la série Taylor de $\psi(z) = \frac{\theta(z)}{\varphi(z)}$ ($\pi(z) = \frac{\varphi(z)}{\theta(z)}$)

Théorème 8.4 (*) 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 : Cette dernière équation est appelée **recursion 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

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

Exemple 8.1 Ainsi, $\psi_1 = \varphi_1 + \theta_1, \psi_2 = \varphi_2 + \varphi_1^2 + \theta_1\varphi_1 + \theta_2, \psi_3 = \varphi_3 + 2\varphi_1\varphi_2 + \varphi_1^3 + (\varphi_2 + \varphi_1^2)\theta_1 + \varphi_1\theta_2 + \theta_3, \dots$

8.3 Équations Yule Walker

1. Les premières p corrélations $\boldsymbol{\rho} = (\rho(1), \dots, \rho(p))$ peuvent être obtenues du système réduit à p équations et p inconnues :

$$\mathbf{R} \boldsymbol{\varphi} = \boldsymbol{\rho} \quad (40)$$

ou R est la matrice Toeplitz symétrique :

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

2. En suite, pour $k > p$ on utilise la récurrence : $\rho_k = \sum_{i=1}^p \varphi_i \rho_{k-i}$
3. La variance est $\gamma_0 = \frac{\sigma^2}{1 - \sum_{i=1}^p \varphi_i \rho_i}$ (et en suite, on obtient les covariances par $\gamma_k = \rho_k \gamma_0, k > 1$).
4. **Exemple : AR(2)** Pour le processus AR(2)

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

avec racines λ_1, λ_2 de $0 = \lambda^2 - \phi_1 \lambda - \phi_2 = \lambda^2 \varphi(\lambda^{-1})$ dedans le cercle unitaire (pour assurer la causalité), on obtient, en résolvant (40) :

$$\rho_1 = \frac{\varphi_1}{1 - \varphi_2} = \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 \lambda_2}, \quad \rho_2 = \varphi_2 + \frac{\varphi_1^2}{1 - \varphi_2}, \dots$$

(a) Si les racines λ_1, λ_2 sont distinctes, on obtient

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

Finalement, en utilisant $\varphi_1 = \lambda_1 + \lambda_2, \varphi_2 = -\lambda_1 \lambda_2$, on arrive à

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

en termes de racines seulement.

(b) Dans le cas de racines *confondues* $\lambda_i = \lambda$, on obtient :

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

8.4 Prévision linéaire

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

Dans le cas d'un modèle ARIMA(p,d,q), les $\tilde{\varphi}_i$ sont les coefficients du polynôme $\varphi(B)(1-B)^d$, et dans le cas d'un modèle ARMA(p,q) $\tilde{\varphi}_i = \varphi_i$. Les $\hat{\epsilon}_t$ peuvent être enlevés en utilisant la représentation "π" du bruit en fonction de la série.

9 Appendice : Rappels des statistique et probabilités multivariées

9.1 Analyse statistique univariée

Définition 9.1 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, $z_\alpha = 2$).

9.2 Rudiments de statistique bivariée

La codispersion de deux variables aléatoires X et Y est caractérisée par leur "covariance", i.e. par l'espérance du produit des écarts à leur moyennes respectives m_X et m_Y :

$$\sigma_{XY} = \text{Cov}(X, Y) = \mathbb{E}(X - m_X)(Y - m_Y)$$

En normalisant cette quantité on obtient le coefficient de corrélation : le rapport - sans dimension - de la covariance aux produit des deux écarts-types :

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

Exercice 9.1 Montrez que le coefficient de corrélation est nécessairement compris entre -1 et +1.

Le coefficient de corrélation indique le degré de dépendance linéaire entre 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 : $f_{XY}(x, y) = f_X(x)f_Y(y)$, ce qui implique que la covariance et le coefficient de corrélation correspondants sont nuls - la réciproque n'étant pas nécessairement vraie).

Les géostatisticiens utilisent également la **sémivariance**

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

d'habitude pour des variables avec la même variance.

Exercice 9.2 Montrez que si $\text{Var}(X, X) = \text{Var}(Y, Y) = \sigma^2$, la covariance et sémivariance satisfont :

$$\gamma_{XY} = \sigma^2 - \text{Cov}(X, Y) = \sigma^2(1 - \rho_{XY})$$

Dés lors, ces coefficients sont parfaitement équivalents. Cependant, la sémivariance a une interprétation intuitive de "mesure de dissimilitude" qui la rend préférée en géostatistique.

10 La régression lineaire

10.1 Méthode des moindres carrés et méthode des 2 points

On cherche ici à modéliser la tendance par une droite d'équation $m(t) = at + b$. Comment déterminer alors $\theta = (a, b)$?

La **droite des moindres carrés** est celle qui ajuste le mieux les observations, *au sens des moindres carrés*. Elle est obtenue en calculant $\theta = (a, b)$ qui minimise $\sum_{i=1}^n (y_i - m(t_i))^2$. Le calcul des coefficients a, b se fait par la méthode classique **déterministe** d'approximation par projection/régression, que nous rappelons maintenant.

Soit Y un vecteur qu'on veut approximer par un ensemble $X_i, i = 1, \dots, I$ des vecteurs "régresseurs" (plus accessibles, plus connus que Y ...). Alors, l'approximation de Y par une combinaison linéaire des $X_i, i = 1, \dots, I$ qui est la plus proche de Y **au sens des moindres carrés** est :

$$\hat{Y} = \sum_{i=1}^I a_i X_i$$

où le vecteur colonne $\mathbf{a} = (a_i, i = 1, \dots, I)$ vérifie le système d'équations "normales" :

$$\mathbf{C}_{X,X} \mathbf{a} = \mathbf{C}_{X,Y} \tag{41}$$

et où $\mathbf{C}_{X,X}$ est la **matrice Gramm** de produits scalaires des regresseurs et $\mathbf{C}_{X,Y} = (\text{Cov}(X_i, Y), i = 1, \dots, I)$. Pour le cas des deux régresseurs : X et $\mathbf{1}$, ça ramène aux formules de la régression univariée :

$$a = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}, \quad b = \bar{Y} - a \bar{X}$$

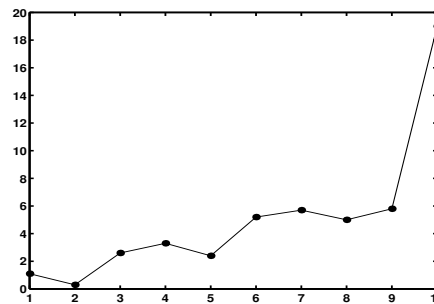
Exercice 10.1 Dans le cas des regresseurs $X_1 = \mathbf{i}, X_2 = \mathbf{1}$, on trouve à partir des équations normales (41) que $a = \frac{\text{Cov}(Y, \mathbf{i})}{\text{Var}(\mathbf{i})} = \frac{\sum_i (i-1/2)Y_i}{(I^3+2I)/12}$, $b = \bar{Y} - a \bar{\mathbf{i}} = \bar{Y} - a \frac{I}{2}$.

Méthode Alternative : La méthode des 2 points consiste à faire passer une tendance droite par deux points $((t_I, y_I), (t_{II}, y_{II}))$ choisis arbitrairement. On peut par exemple choisir de constituer 2 paquets d'observations, puis calculer les points moyens, ou encore les points médians de ces deux sous séries, puis faire passer la droite par ces 2 points. On pourrait aussi utiliser les quantiles d'ordre 1/3 et 2/3.

Exemple : Soit la chronique suivante,

t_i	1	2	3	4	5
y_i	1.1	0.3	2.6	3.3	2.4

t_i	6	7	8	9	10
y_i	5.2	5.7	5.0	5.8	19.0



On choisit de modéliser la tendance de cette chronique par une droite. Plutôt que de calculer immédiatement la droite des moindres-carrés, il faut prendre le temps de regarder un peu cette chronique et remarquer que l'observation y_{10} semble être un point aberrant. On peut donc utiliser la méthode des 2 points, et faire passer la droite par les points médians de coordonnées (3, 2.4) et (8, 5.7). Les résultats obtenus avec ces deux méthodes sont présentés Figure 4 :

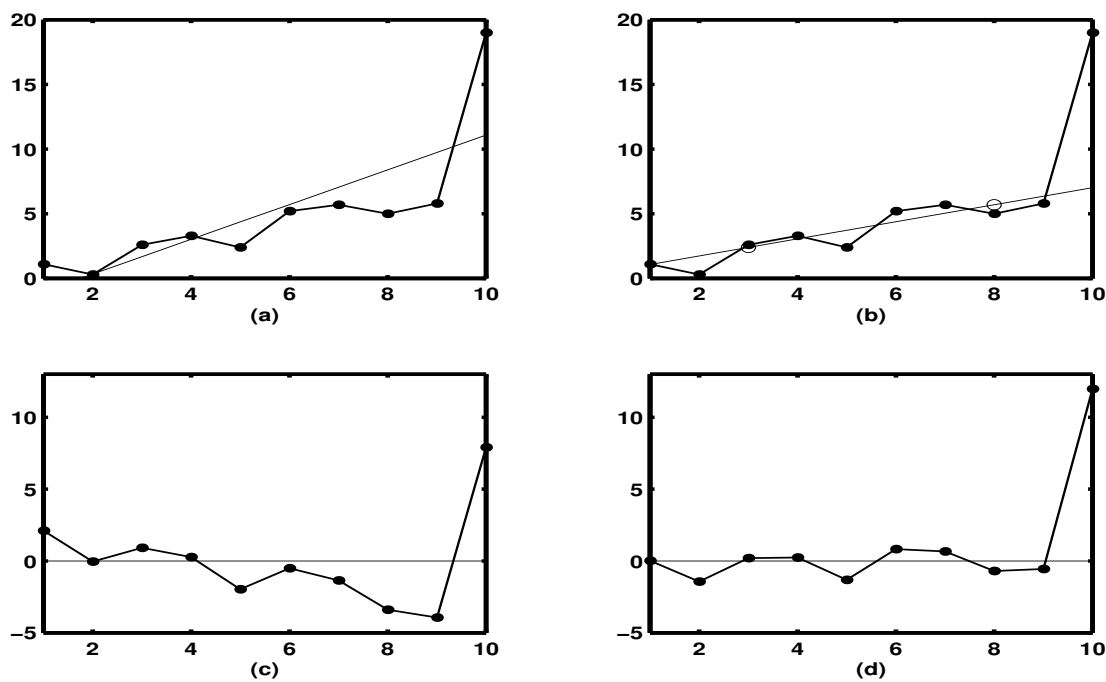


FIG. 4 – Ajustements linéaires obtenu par : (a) la méthode des moindres carrés, (b) la méthode des 2 points (indiqués par des cercles), (c) et (d) résidus associés.

La première droite (a) est meilleure au sens des moindres carrés : c’est elle qui passe le plus près de l’ensemble des points. Néanmoins, elle modélise mal la tendance de cette chronique. La seconde droite (b) au contraire passe par deux points représentatifs du nuage, sans tenir compte de la dernière observation. Cette droite fournit alors un très bon ajustement du reste de la chronique. L’examen des résidus conduit à la même conclusion : les résidus obtenus avec la droite des moindres carrés (c) ne fluctuent pas irrégulièrement autour de 0 (ils ont tendance à décroître jusqu’à t_9). Les résidus obtenus avec la seconde droite (d) fluctuent irrégulièrement autour de 0 et avec une plus faible amplitude jusqu’à l’instant t_9 . La grande amplitude du dernier résidu e_{10} nous laisse supposer qu’il s’agit là d’un “accident” dont il vaut mieux ne pas tenir compte.

10.2 le coefficient de corrélation linéaire ne suffit pas pour mesurer la qualité de l’ajustement

⚠ Attention : le coefficient de corrélation linéaire ne mesure pas la qualité de l’ajustement offert par un modèle linéaire. Il indique simplement si, en moyenne, les points de la série sont éloignés ou non de la droite des moindres carrés.

Considérons les deux séries y et z à titre d’exemple :

t_i	1	2	3	4	5	6	7	8	9	10
y_i	19.4	17.6	24.6	30.4	35.0	38.4	40.6	41.6	41.4	40.0
z_i	1.0	3.3	13.4	8.3	11.9	11.4	8.8	10.2	16.2	10.3

Le coefficient de corrélation linéaire entre t et la série y vaut 0.91 alors que celui entre t et z ne vaut que 0.65. Doit-on en conclure qu’un ajustement linéaire convient mieux à la première chronique qu’à la seconde ? NON, sûrement pas avant d’avoir regardé ces deux chroniques :

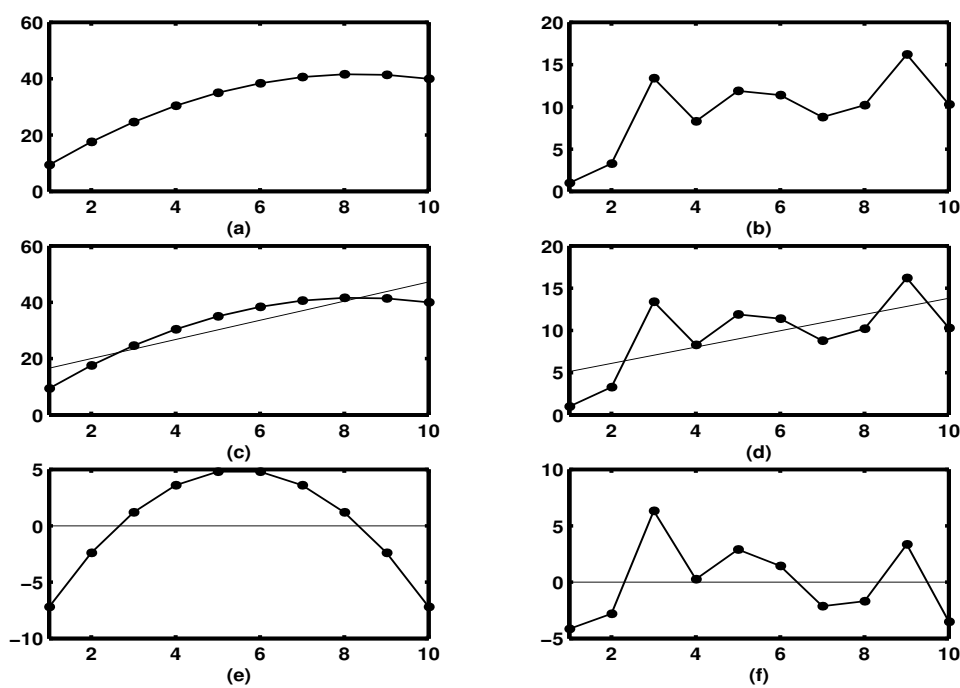


FIG. 5 – (a) et (b) les deux séries y et z , (c) et (d) les ajustements linéaires obtenu par la méthode des moindres carrés, (e) et (f) résidus associés.

Il est clair, au vu de ces graphiques, qu'une tendance linéaire ne convient pas du tout à la première chronique, malgré un coefficient de corrélation linéaire de 0.91! Au contraire, la tendance de la seconde chronique semble linéaire. La valeur plus faible du coefficient de corrélation linéaire provient simplement du fait que les résidus ont une amplitude assez importante. La tendance linéaire n'explique donc que 42% (0.65^2) de la variance de cette chronique.

10.3 Les tendances polynomiales

On considère maintenant que la tendance f est un polynôme de degré d

$$f(t) = a_0 + a_1t + a_2t^2 + \dots + a_dt^d$$

Ici, $\theta = (a_0, a_1, \dots, a_d)$.

Bien sûr, le cas $d = 0$ correspond à une tendance constante, et le cas $d = 1$ à une tendance linéaire. Si la tendance semble posséder des minima et/ou des maxima (points où la dérivée s'annule), un degré supérieur doit être choisi.

10.3.1 la courbe des moindres carrés

Ici, les coefficients de la courbe des moindres carrés peuvent être calculés explicitement. En effet, posons

$$T = \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^d \\ 1 & t_2 & t_2^2 & \dots & t_2^d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^d \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad F = \begin{pmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{pmatrix} \quad \theta = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix}$$

On a donc $F = T\theta$. La courbe des moindres carrés est obtenue en calculant la valeur de θ qui minimise $\|Y - F\|^2 = \|Y - T\theta\|^2$. La solution de ce problème de minimisation est bien connue :

$$\theta^{MC} = (T'T)^{-1}T'Y.$$

En particulier, on peut vérifier que si $d = 0$, alors $T'T = n$ et $T'Y = \sum y_i$. On obtient bien $a_0^{MC} = \bar{y}$.

Si $d = 1$, alors

$$T'T = \begin{pmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix} \quad T'Y = \begin{pmatrix} \sum y_i \\ \sum t_i y_i \end{pmatrix}$$

On retrouve alors les formules bien connues

$$\begin{pmatrix} a_0^{MC} \\ a_1^{MC} \end{pmatrix} = \begin{pmatrix} (\text{var}(t)\bar{y} - \text{cov}(t, y)\bar{t}) / \text{var}(t) \\ \text{cov}(t, y) / \text{var}(t) \end{pmatrix}$$

Pour des valeurs de d supérieures à 1, l'expression explicite de θ^{MC} est compliquée. En fait, on n'a absolument pas besoin de cette expression, puisque seul le calcul numérique de θ^{MC} est utile et que d'autre part, les formules données ci-dessus pour ce calcul sont très faciles à programmer (simple calcul matriciel).

10.3.2 le choix du degré du polynome

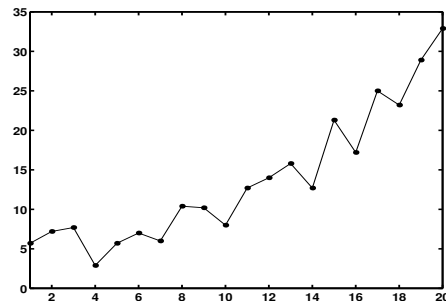
Un examen visuel de la série permet en général de se faire une idée du degré du polynome à utiliser. Comme dans tous les problèmes d'ajustement de modèle, le choix d'un modèle est basé sur l'examen des résidus. En effet, on aimerait

1. obtenir des résidus qui fluctuent autour de 0 avec une amplitude la plus faible possible.
2. utiliser un polynome de degré le plus faible possible.

Le premier critère sera d'autant plus satisfait que le degré du polynome sera élevé. En particulier, le choix extrême qui consiste à utiliser un polynome de degré $n - 1$ conduit à un ajustement parfait, puisque ce polynome passe par tous les points de la série. Ce choix est absurde dès que l'on sait qu'il y a des résidus : il ne faut pas chercher à les éliminer à tout prix . . . D'un autre côté, le second critère sera totalement satisfait avec un polynome de degré nul. Ce choix est également absurde dès que l'on pense qu'une tendance (non constante) est présente. Il y a donc un compromis à faire entre ces deux critères.

Exemple : Soit la chronique suivante,

t_i	1	2	3	4	5
y_i	5.7	7.2	7.7	2.9	5.7
t_i	6	7	8	9	10
y_i	7.0	6.0	10.4	10.2	8.0
t_i	11	12	13	14	15
y_i	12.7	14.0	15.8	12.7	21.3
t_i	16	17	18	19	20
y_i	17.2	25.0	23.2	28.9	32.9



Considérons les différents ajustements obtenus sur une même série, avec des polynomes de degré 0, 1, 2 et 3 (en utilisant la méthode des moindres carrés) :

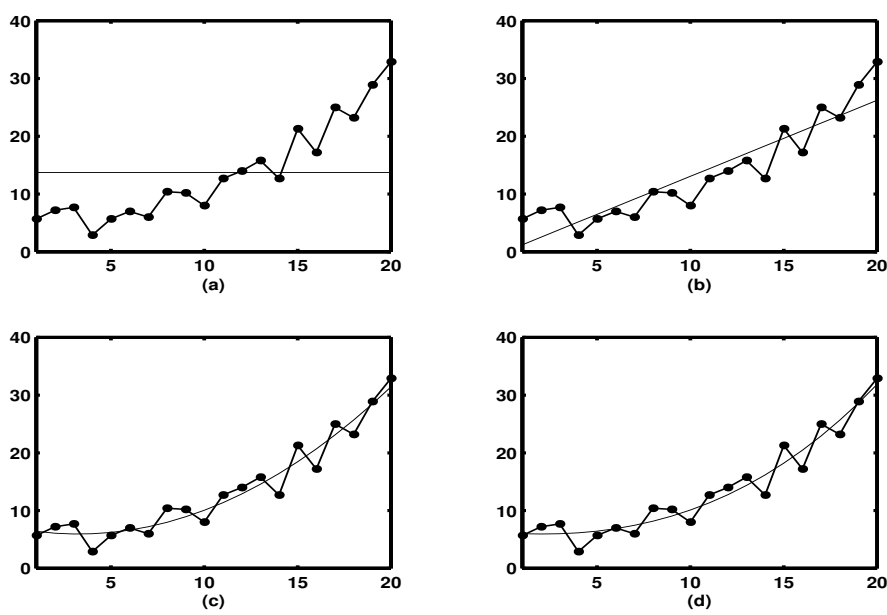


FIG. 6 – La série originale (y_i , $1 \leq i \leq n$) et les ajustements polynomiaux ($f(t_i)$, $1 \leq i \leq n$) obtenus par la méthode des moindres carrés avec différents degrés. (a) $f(t) = 13.73$, (b) $f(t) = -0.10 + 1.32t$, (c) $f(t) = 6.92 - 0.60t + 0.09t^2$, (d) $f(t) = 6.15 - 0.20t + 0.05t^2 + 0.0015t^3$.

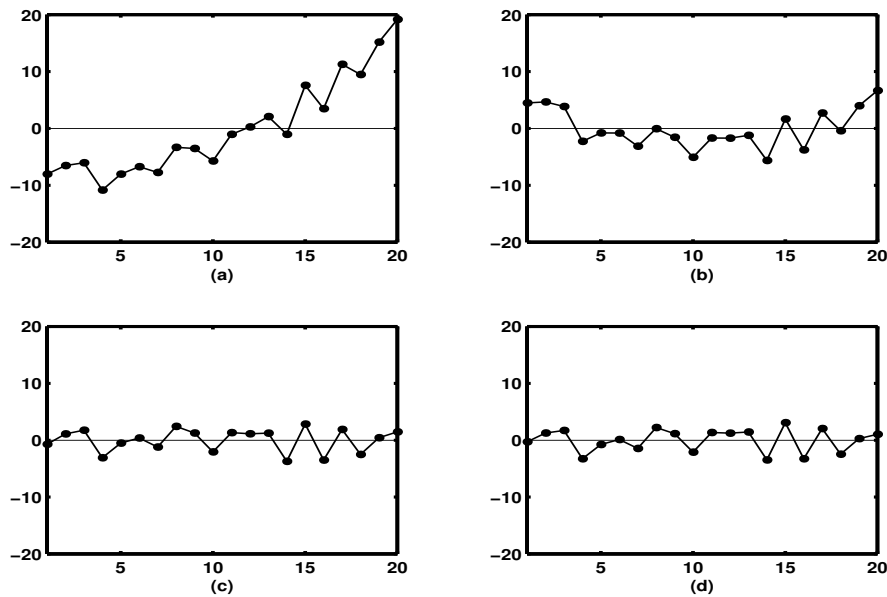


FIG. 7 – Résidus associés ($y_i - f(t_i)$, $1 \leq i \leq n$). (a) $d = 0$, (b) $d = 1$, (c) $d = 2$, (d) $d = 3$.

On peut voir sur les figures (a) et (b) que la série ne fluctue pas de façon irrégulière autour de la courbe (en d'autres termes, les résidus ne fluctuent pas de façon irrégulière autour de 0), alors que c'est clairement le cas dans les figures (c) et (d) : pour que le premier critère soit satisfait, il convient de choisir un degré au moins égal à 2.

D'un autre côté, les ajustements obtenus avec un polynôme de degré 2 ou avec un polynôme de degré 3 sont pratiquement identiques : on "ne gagne rien" à utiliser un polynôme de degré 3 plutôt qu'un polynôme de degré 2. Les résidus obtenus avec des polynômes de degré 2 ou 3 sont d'ailleurs à peu près identiques. Le second critère nous conduit donc à retenir un polynôme de degré 2.

Une méthode numérique peut également nous aider à faire ce choix, en calculant $\sum (y_i - f(t_i))^2$ pour chacune des 4 courbes. On obtient ainsi :

d	0	1	2	3
$\sum (y_i - f(t_i))^2$	1376.82	224.58	78.78	77.84

Le premier critère est d'autant plus satisfait que cette quantité est faible. On voit ici que le choix de $d = 2$ améliore nettement l'ajustement, alors que cette quantité diminue peu avec en passant de $d = 2$ à $d = 3$. Là encore, cela nous conduit à choisir une tendance quadratique.

Remarque : Plutôt qu'étudier la somme des carrés résiduels, on peut calculer la part de variance expliquée pour chacun des modèles :

d	0	1	2	3
Part de variance expliquée	0%	83.69%	94.28%	94.35%

Un polynôme de degré 3 permet d'expliquer 94.35% de la variance de la série. Le gain est donc négligeable devant un polynôme de degré 2, qui explique déjà 94.28% de cette variance.

10.4 La méthode du changement de variable

On peut parfois se ramener à des tendances linéaires ou polynomiales grâce à un simple changement de variables. C'est le cas pour des tendances de la forme $f(t) = 1/(a + bt)$, $f(t) = \exp(a + bt)$, $f(t) = \ln(a + bt + ct^2)$, ...

En effet, ajuster une tendance de la forme $f(t) = 1/(a + bt)$ à une série (y_i) revient à ajuster une tendance de la forme $g(t) = 1/f(t) = a + bt$ à la série $(z_i) = (1/y_i)$. Il suffit donc de calculer cette nouvelle série (z_i) , puis de calculer a et b au moyen d'une des méthodes décrites plus haut (moindres carrés ou 2 points).

Si la tendance f a maintenant la forme $f(t) = \log(a + bt + ct^2)$, on pourra calculer a , b et c en ajustant une tendance polynomiale de la forme $g(t) = \exp(f(t)) = a + bt + ct^2$ à la série $(z_i) = (\exp(y_i))$.

10.5 Courbes de tendance qui ne se réduisent pas à une régression linéaire

Bien sûr, la tendance d'une série chronologique n'est pas nécessairement linéaire ou polynomiale. En particulier, si la tendance semble se stabiliser avec le temps, on aimerait pouvoir la modéliser par une courbe possédant une asymptote horizontale.

La courbe de type exponentiel (ou courbe exponentielle modifiée) est définie par l'équation :

$$f(t) = a e^{bt} + c$$

La figure 8 illustre la forme de cette tendance, suivant le signe de a et de b .

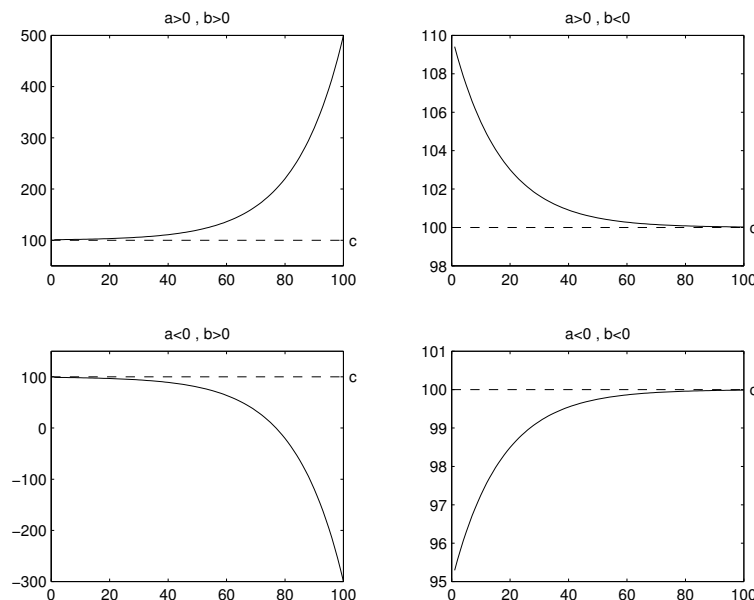


FIG. 8 – exemples de courbes de type exponentiel

Lorsque c est connu, on peut se ramener à une tendance linéaire en posant

$$\begin{aligned} g(t) &= \ln(f(t) - c) \\ &= \ln(a) + bt \end{aligned}$$

On fait alors le changement de variable $z_i = \ln(y_i - c)$ et on ajuste une droite d'équation $g(t) = \ln(a) + bt$ à cette nouvelle série.

Le problème se complique lorsque c est inconnu puisqu'aucun changement de variable ne permet de se ramener à une tendance polynomiale.

La méthode des moindres carrés consiste à déterminer (a, b, c) en minimisant

$$h(a, b, c) = \sum (y_t - ae^{bt} - c)^2.$$

Le minimum de cette fonction n'est pas connu explicitement : en dérivant la fonction h par rapport à a , b et c , puis en utilisant le fait que ces 3 dérivées sont nulles au point recherché, on aboutit à un système de 3 équations à trois inconnues que l'on ne sait pas résoudre.

Néanmoins, il existe des méthodes numériques pour rechercher le minimum de h (algorithme du gradient ou algorithme de Newton). Leur mise en oeuvre n'est pas très simple et sort du cadre de ce cours.

10.6 La fonction d'autocorrélation empirique

Nous venons de voir une façon de procéder afin de déterminer les différentes composantes d'une série chronologique, à savoir *i*) la tendance, *ii*) la composante saisonnière, *iii*) les fluctuations irrégulières.

Cette décomposition ne sera intéressante que si la série des fluctuations irrégulières obtenues ne contient plus ni tendance, ni composante saisonnière : **la série des résidus doit fluctuer irrégulièrement autour de 0.**

Un seul examen graphique de la série des fluctuations irrégulières ne permet pas toujours de vérifier si ce critère est satisfait ou non. La fonction d'autocorrélation empirique va s'avérer être un outil efficace pour une analyse plus fine des fluctuations irrégulières.

Rappelons que le coefficient de corrélation linéaire entre 2 séries $(x_i, 1 \leq i \leq n)$ et $(y_i, 1 \leq i \leq n)$ est défini par :

$$\rho_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2) (\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2)}}$$

Pour une série chronologique $(y_i, 1 \leq i \leq n)$, l'idée consiste à regarder, pour différentes valeurs de k , la corrélation entre la série $(y_i, 1 \leq i \leq n - k)$ et la série "décalée" $(y_{i+k}, k + 1 \leq i \leq n)$. On calcule alors les *coefficients d'autocorrélation* $(r(k), 0 \leq k \leq n - 1)$ au moyen de la formule suivante :

$$\rho_n(k) = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (42)$$

où $\bar{y} = (\sum_{i=1}^n y_i) / n$ est la moyenne de l'ensemble de la série $(y_i, 1 \leq i \leq n)$.

Dans le cas des beaucoup de données, on utilise parfois une formule simplifiée :

$$\rho_n(k) = \frac{\frac{1}{n} \sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (43)$$

Remarque : Par rapport à la définition usuelle (42) du coefficient de corrélation linéaire entre les séries $(y_i, 1 \leq i \leq n - k)$ et $(y_{i+k}, k + 1 \leq i \leq n)$, la formule simplifiée (43) contient trois modifications :

1. Pour simplifier les calculs, les moyennes des séries $(y_i, 1 \leq i \leq n - k)$ et $(y_{i+k}, k + 1 \leq i \leq n)$ ont été remplacées par la moyenne de l'ensemble de la série $(y_i, 1 \leq i \leq n)$.

2. Pour la même raison, les deux variances des séries $(y_i, 1 \leq i \leq n - k)$ et $(y_{i+k}, k + 1 \leq i \leq n)$ ont été remplacées par la variance de l'ensemble de la série $(y_i, 1 \leq i \leq n)$.
3. Le facteur $1/n$ apparaît au numérateur, et non $1/(n - k)$. (Cela entraîne, en particulier, que la série des coefficients d'autocorrélation tend vers 0 lorsque k augmente).

Ces modifications peuvent entraîner des conséquences "illogiques" –voir dernier exercice du TD.

L'importance de la corrélation empirique : Notre critère pour accepter "l'irregularité" d'une série des résidus sera si les testes statistiques confirment que toutes les corrélations $\rho_n(k), k \neq 0$ sont nuls (i.e, assez petites par exemple cf un teste χ^2), et si les résidus ont une distribution Gaussienne.

Voyons maintenant sur différents exemples l'allure de cette fonction d'autocorrélation.

10.7 (*) Autocorrélation d'une chronique possédant une composante périodique

La fonction d'autocorrélation est outil précieux pour détecter des périodicités dans une série. L'exemple présenté Figure 9 illustre bien cette propriété.

La série (y_i) étant purement périodique de période 4, les série décalées $(y_{i+4}), (y_{i+8}), (y_{i+12}) \dots$ coïncident avec la série (y_i) . On peut alors remarquer sur la Figure 9-c, que pour $k = 4$, les points du nuage sont alignés. La série des coefficients d'autocorrélation $(r(k))$ présente donc des maxima pour $k = 4, 8, 12, \dots$

D'autre part, pour toute valeur de $k \geq 0$, les séries (y_{i+k}) et (y_{i+k+4}) coïncident. Donc, $r(k)$ et $r(k + 4)$ sont très proches, c'est-à-dire qu'une périodicité de période 4 apparaît clairement sur la série des coefficients d'autocorrélation.

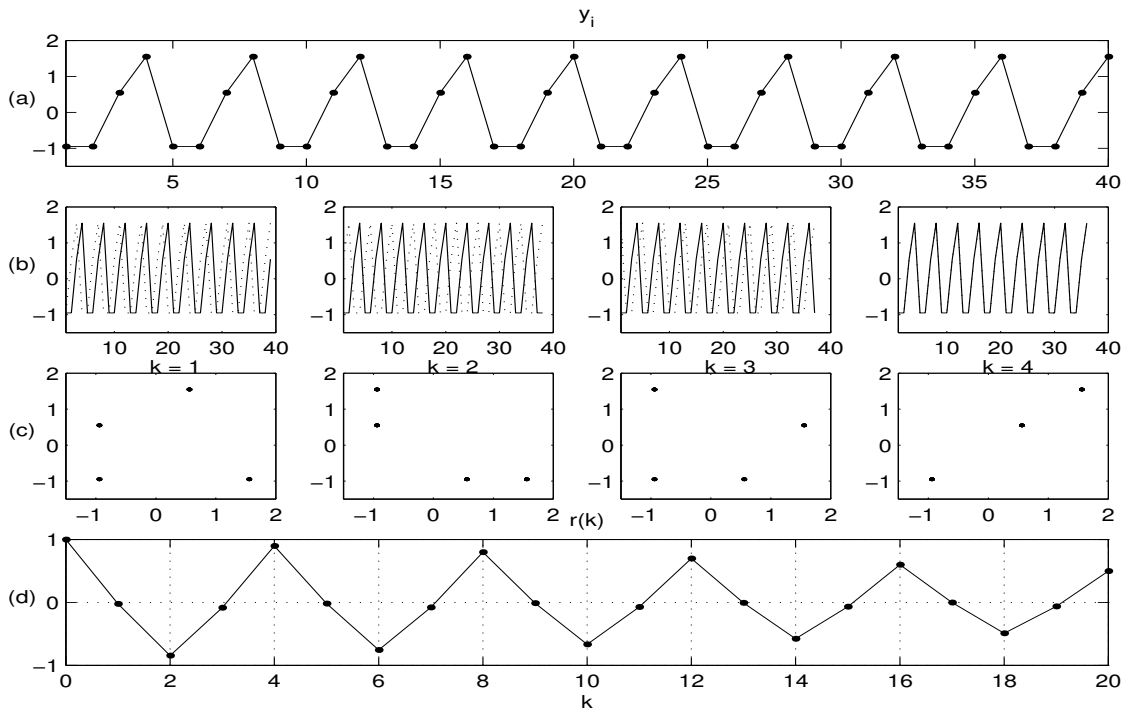


FIG. 9 – (a) une série (y_i) purement périodique de période 4, (b) la série (y_i) et la série décalée (y_{i+k}) pour $k = 1, 2, 3, 4$, (c) le nuage de points obtenu avec (y_i) en abscisse et (y_{i+k}) en ordonnée, (d) la fonction d'autocorrélation $(r(k), 0 \leq k \leq 20)$

La Figure 10 illustre maintenant ce qui se passe pour une série comportant toujours une composante périodique, mais également des fluctuations irrégulières.

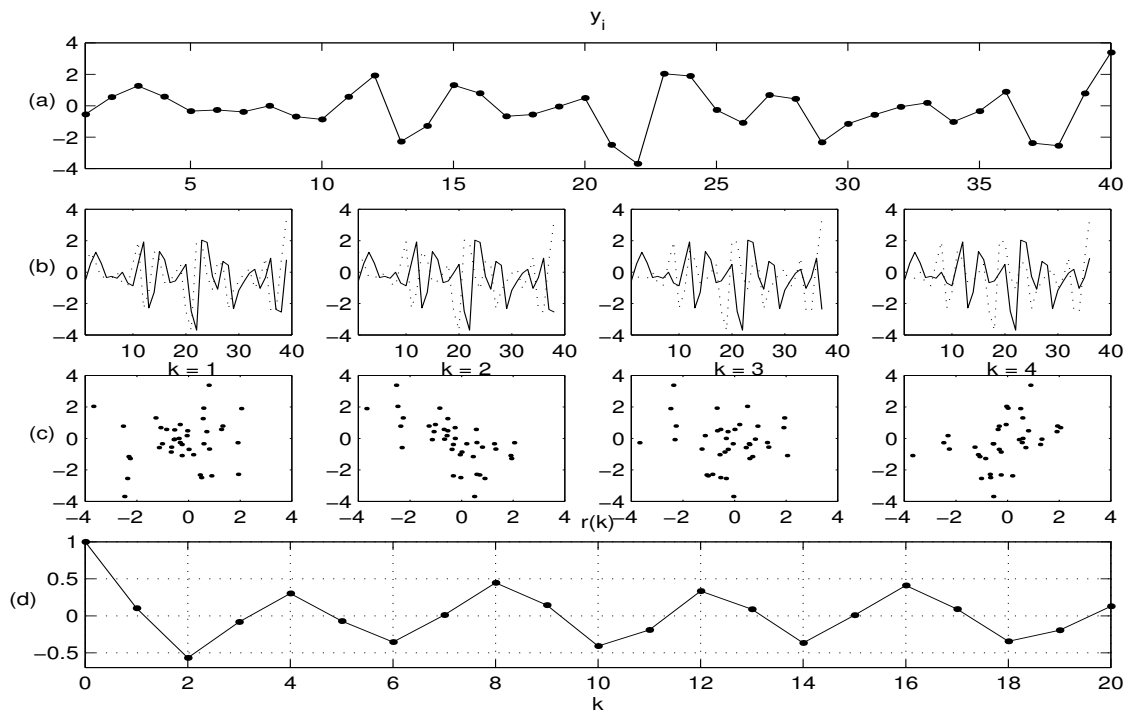


FIG. 10 – (a) une série (y_i) comportant une composante périodique de période 4 et des fluctuations irrégulières, (b) la série (y_i) et la série décalée (y_{i+k}) pour $k = 1, 2, 3, 4$, (c) le nuage de points obtenu avec (y_i) en abscisse et (y_{i+k}) en ordonnée, (d) la fonction d'autocorrélation $(r(k), 0 \leq k \leq 20)$

La composante périodique est difficile à détecter sur la série des observations (y_i) . Par contre, la série des coefficients d'autocorrélation présente très clairement cette périodicité de période 4. La série (y_i) n'étant plus purement périodique de période 4, les séries (y_i) et (y_{i+4}) ne coïncident plus exactement (Figure 10-b, $k = 4$). On peut néanmoins remarquer sur la Figure 10-c, que pour $k = 4$, la forme du nuage met clairement en évidence une corrélation positive importante entre ces deux séries. Au contraire, la corrélation entre (y_i) et (y_{i+2}) semble négative. La série des coefficients d'autocorrélation $(r(k))$ présente donc des maxima pour $k = 4, 8, 12, \dots$, et des minima pour $k = 2, 6, 10, \dots$ et décroît très lentement vers 0.

10.8 (*) Autocorrélation d'une chronique présentant une tendance

La Figure 11 nous montre à présent le comportement de la fonction d'autocorrélation lorsqu'une tendance est présente dans la série.

On peut voir dans les Figures 11-b et 11-c que les séries (y_i) et (y_{i+k}) , pour $k = 1, 2, 3, 4$ sont fortement corrélées positivement. Cela explique la décroissance très lente vers 0 de la fonction d'autocorrélation.

Cette décroissance lente est caractéristique lorsqu'une tendance est présente.

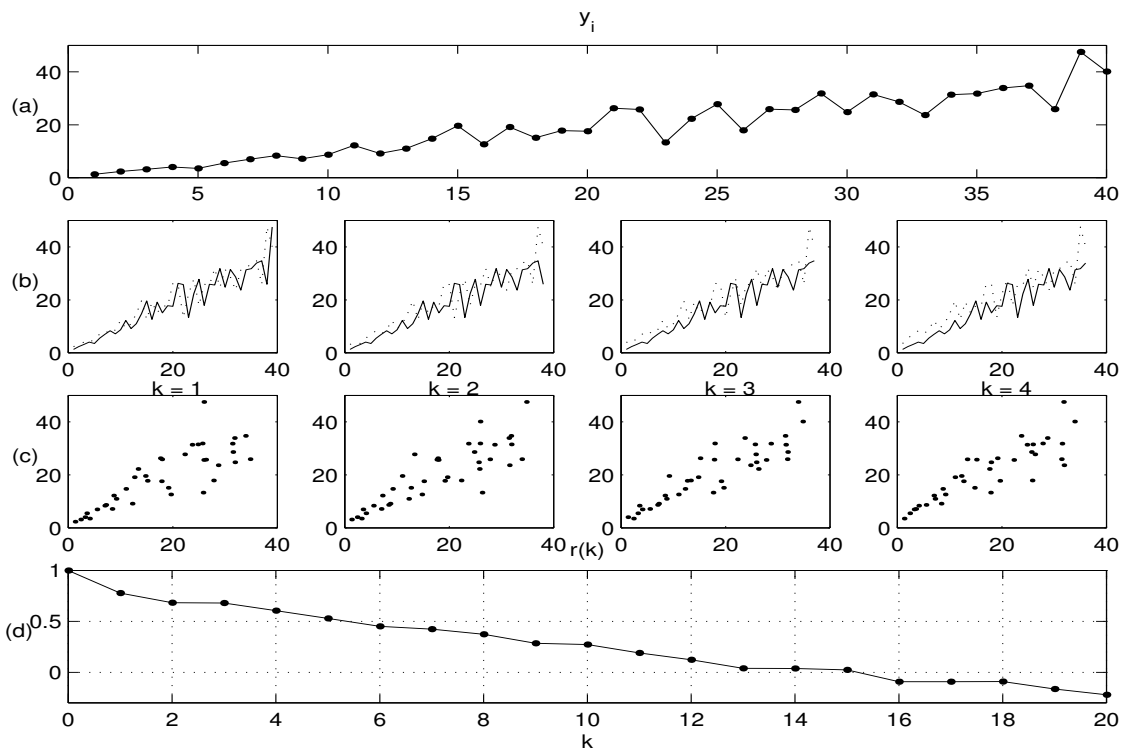


FIG. 11 – (a) une série (y_i) comportant une tendance et des fluctuations irrégulières, (b) la série (y_i) et la série décalée (y_{i+k}) pour $k = 1, 2, 3, 4$, (c) le nuage de points obtenu avec (y_i) en abscisse et (y_{i+k}) en ordonnée, (d) la fonction d'autocorrélation $(r(k), 0 \leq k \leq 20)$

10.9 (*) Autocorrélation d'une série de fluctuations irrégulières

Considérons à présents différents types de fluctuations irrégulières (Figure 12).

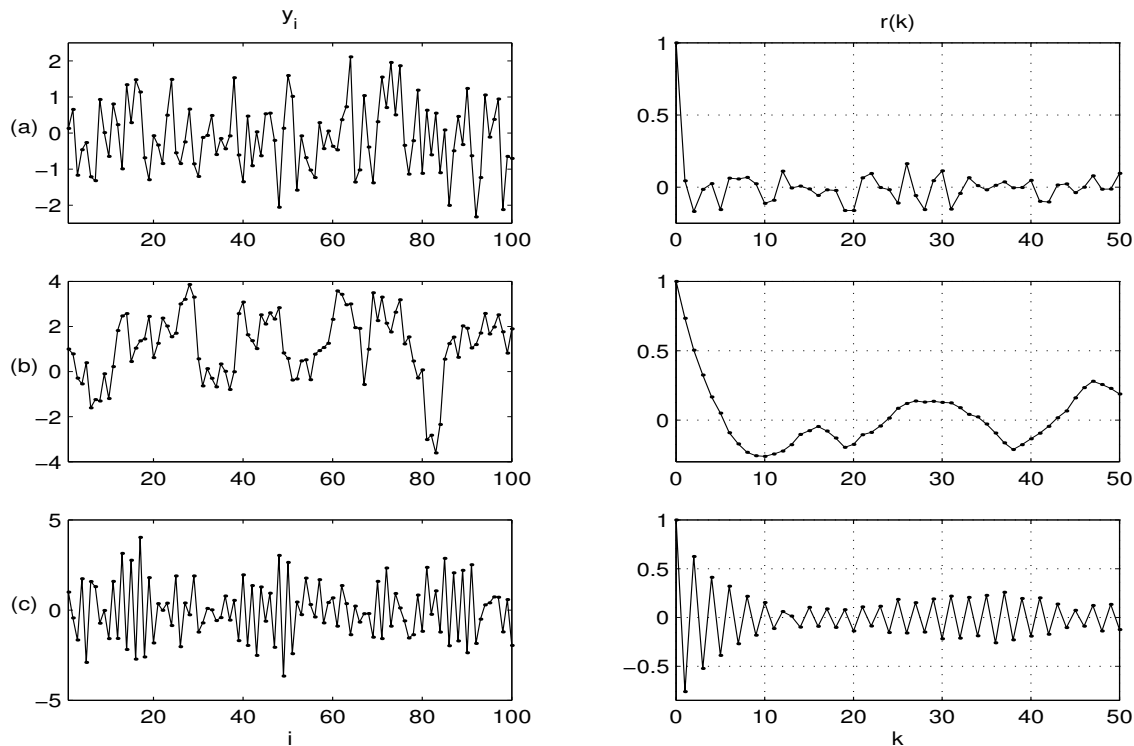


FIG. 12 – à gauche : 3 séries $(y_i, 1 \leq i \leq 100)$ comportant différents types de fluctuations irrégulières, à droite : les 3 fonctions d'autocorrélation $(r(k), 0 \leq k \leq 50)$ associées

Dans le premier exemple (Figure 12-a), la série fluctue très irrégulièrement : une modélisation probabiliste de cette série nous inciterait à la considérer comme une suite de variables indépendantes et identiquement distribuées. La fonction d'autocorrélation ($r(k)$) vaut bien sûr 1 pour $k = 0$, puis elle fluctue ensuite de façon très erratique au voisinage de 0 (c'est-à-dire en ne prenant que des valeurs proches de 0). Cela signifie qu'il ne semble exister aucune structure dans cette série, aucune relation entre observations successives. En particulier, dans un cadre de prévision, la série observée y_1, y_2, \dots, y_n ne semble apporter aucune information utile pour prédire y_{n+1}, y_{n+2}, \dots .

Dans le second exemple (Figure 12-b), la série fluctue toujours irrégulièrement, mais une certaine structure semble présente : dans un cadre probabiliste, cette série ne semblerait pas être une suite de variables i.i.d, mais de variables corrélées positivement puisque des observations voisines ont tendance à être proches. Cela apparaît dans la fonction d'autocorrélation qui décroît toujours vers 0, mais moins vite que pour la série précédente. Les fluctuations de cette fonction d'autocorrélation dans un voisinage de 0 sont également plus régulières que dans l'exemple précédent. Néanmoins, *i*) aucune périodicité n'apparaît : aucune composante périodique n'est présente, *ii*) la fonction d'autocorrélation décroît beaucoup plus vite vers un voisinage de 0 que dans l'exemple proposé dans la Figure 11 : aucune tendance n'est présente dans cette chronique. En résumé, on en conclut donc que la série (y_i) ne comporte que des fluctuations irrégulières, mais qu'une certaine relation semble exister entre les valeurs de cette série. L'analyse de cette relation est importante lorsque l'on cherche à faire de la prévision, puisque maintenant, la série observée y_1, y_2, \dots, y_n comporte très vraisemblablement une information utile pour prédire y_{n+1}, y_{n+2}, \dots . Signalons simplement que les modèles ARMA sont les modèles les plus couramment employés pour modéliser ce type de série.

Les mêmes remarques semblent valables pour la troisième série (Figure 12-c). Alors que dans la série précédente, les séries (y_i) et (y_{i+1}) étaient corrélées positivement, elles sont corrélées négativement dans cet exemple : y_i et y_{i+1} ont tendance à être de signe opposé. Par conséquent, y_i et y_{i+2} ont tendance à être de même signe : les séries (y_i) et (y_{i+2}) sont à nouveau corrélées positivement, etc. Ce comportement se traduit par une fonction d'autocorrélation qui décroît toujours vers 0, mais en oscillant autour de 0. Remarquons qu'il n'existe pas de composante périodique de période 2, car la fonction d'autocorrélation décroîtrait alors beaucoup plus lentement vers 0, comme l'exemple Figure 10-d. Finalement, on en conclut que cette série (y_i) ne comporte que des fluctuations irrégulières, mais que, comme dans l'exemple précédent, une certaine relation semble exister entre les valeurs de cette série.