3M22

Filtrage et Séries Temporelles, première partie

Titre original :
Séries temporelles : régression, modélisation ARIMA(p, d, q),
et modélisation espace-état
29 juin 2006
Enseignant : Florin Avram

Département de Mathématiques
Boulevard Marie et Pierre Curie,
Téléport 2,
BP 30179, F-86962 Chasseneuil–Futuroscope cedex.

Ces notes de cours trouvent leurs origines dans un polycopié de Florin Avram (université de Pau). Elles présentaient l'intérêt d'être assez courtes et de pouvoir être apparemment adaptée à un demi-cours de M2 (Marc Arnaudon et Anthony Phan) destiné à un public mixte : étudiants en Mathématiques et en Informatique. La seconde partie de cet enseignement devait se faire en continuité de la première et donnée par un enseignant plus spécialisé en théorie du signal.

Du côté Mathématiques, on doit regretter que l'enseignement sur les processus à temps discret ne soit mis à profit. De plus il y a une toute petite mathématique qu'il aurait fallu développer de manière plus systématique ou au moins synthétique : l'opérateur de décalage, des éléments de théorie spectrale (des opérateurs), le calcul fonctionnel associé. Ces préalables bien posés, le reste (processus MA jusqu'à SARIMA) n'apparaîtrait ainsi que comme de la zoologie. D'un point de vue statistique, l'estimation des coefficients des modèles est absente. Des résultats ergodiques propres devraient être exposés, quitte à les admettre. Le filtrage est présenté de manière assez anecdotique (fonctions de l'opérateur de décalage). Un cadre plus général serait certainement souhaitable pour inclure notamment le filtre de Kalman–Bucy.

En petits caractères, nous laissons la suite du texte de 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 des 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-temporels 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 opportun).

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. Ensuite, nous examinons l'approche itérative statistique qui consiste à raffiner des modèles ARIMA ou des modèles d'espace-état, choisis conformément aux tests pour les résidus. Éventuellement, la démarche stochastique pourra aussi être mise en œuvre 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 choisis.

Volume horaire. — 12 heures de cours : 1 heure et demi Jeudi 8 à 13h40, S06, pour 8 semaines, et qui se transformeront ensuite en TD pour la neuvième et dixième semaines. 18 heures de TD : 1 heure et demi Vendredi à 8h00, S23, pour 10 semaines (et complétées par trois heures pour les deux dernières semaines, ayant place dans la salle. de cours, Jeudi).

Matériels. — 1. Notes de cours/TD, qui utilisent une parties des notes de M. Lavielle (Université Paris-Sud) et A. Korabinski (Heriot-Watt) sur les séries temporelles (toutes les coquilles sont de ma responsabilité).

2. Notes WEB : A. Charpentier, M. Kratz, J-M. Dufour (en Français) et R.H. Smith, R. Weber (En Anglais), etc.

`http://www.crest.htfr/pageperso/lfa/charpent/charpent.htm#TS`

`http://www.math-info.univ-paris5.fr/~kratz/cours.html`

`http://www.statslab.cam.ac.uk/~rrw1/timeseries/index.html`

3. A. C. Harvey, Time Series Models.
4. J. Durbin and S.J. Koopman, Time series analysis by state space methods.
5. C. Gourieroux et A. Monfort, Cours de séries temporelles.

Nous avons essayé de clarifier les notations et d'annoter le texte. Notamment :

Conventions et notations. — Les ensembles qui indexeront les suites considérées ne seront pas toujours précisés. Il s'agira le plus souvent de $\mathbb{N}$ ou de $\mathbb{Z}$. La suite nulle sera notée $\mathbb{0} = (0)_t$, la suite constante égale à 1 sera notée $\mathbb{1} = (1)_t$, la suite constante égale à k sera notée $\mathbb{k} = (k)_t$.

La transposée d'une matrice ou d'un vecteur u est notée ${}^t u$. L'adjonction $\bar{u}$ est la combinaison de la conjugaison complexe et de la transposée.

Le disque unité du plan complexe est le disque ouvert $\mathbb{D} = \mathbb{D}^2 = \{z \in \mathbb{C} : |z| < 1\}$, le cercle unité est $\mathbb{S} = \mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$, le disque fermé est $\bar{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leqslant 1\}$, et son complémentaire est $\bar{\mathbb{D}}^c = \{z \in \mathbb{C} : |z| > 1\}$.

Les paragraphes annotés par (*) sont réservés à une seconde lecture.

1. **Introduction**

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

Par rapport aux autres types de données statistiques, la particularité des séries chronologiques tient présence d'une relation d'antériorité qui ordonne l'ensemble des informations. Les dates d'observation sont souvent équidistantes les unes des autres : on a des séries mensuelles, trimestrielles, etc. Dans de tels 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, par âge, 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 stochastiques, *i.e.* des familles des variables aléatoires $(U(x))_x$. Pour les séries temporelles/chronologiques, on s'intéresse à $x \in \mathbb{N}$, $\mathbb{Z}$ ou $\mathbb{R}_+$, alors que dans la statistique spatiale (par exemple en géostatistique), on s'intéresse aux cas où $x \in \mathbb{Z}^d$ 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 $i = 1, \ldots, N$. 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 accessibles physiquement, coûteuses, etc.). On veut à la fois :

a) enlever du bruit d'observation éventuel et

b) « extrapoler » du connu à l'inconnu.

Domaines d'application :

— prospection et exploitation pétrolières et minières ;

— traitement du signal ;

— imagerie médicale ;

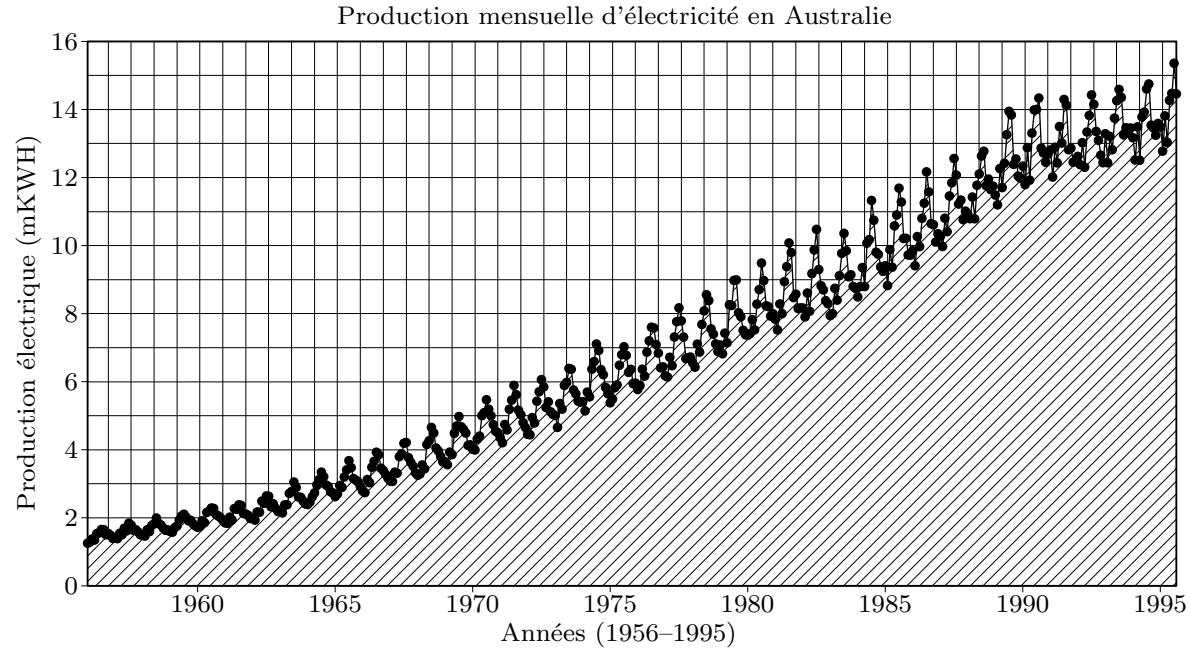
— océanographie, météorologie, hydrogéologie, environnement, ...

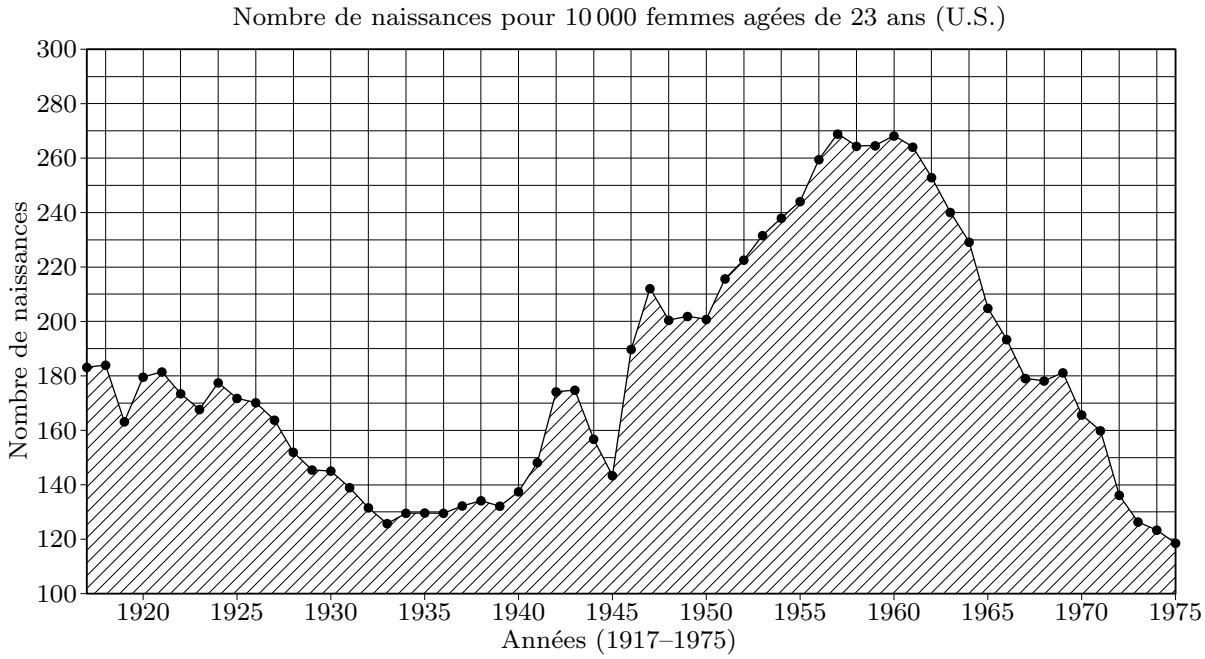
— séries temporelles appliquées en économie, finances, météorologie, médecine, ...

2. **Premier abord des séries temporelles/chronologiques**

Une règle générale en statistique descriptive consiste à commencer par *regarder* ses don-

nées, avant d'effectuer le moindre calcul.





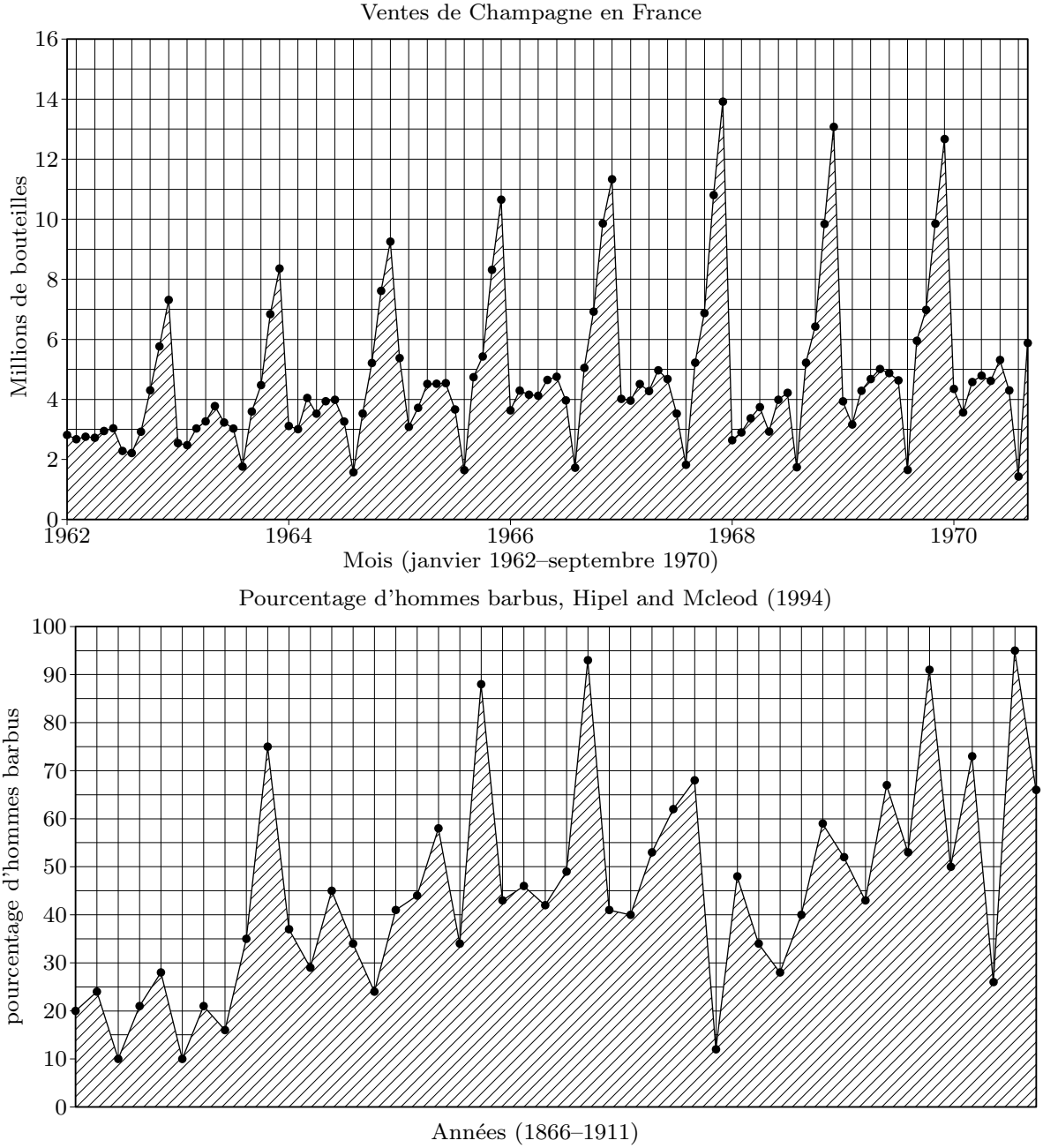


Fig. 1. — Quelques exemples de séries chronologiques

(*) *Notes et commentaires.* — Ce ne sont pas les figures d'origine puisque les données correspondantes n'ont pu être retrouvées... (*)

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 sûre 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 $(y_i, 1 \leqslant i \leqslant n)$ permet de dégager, lorsqu'on envisage une période de temps suffisamment longue, un certain nombre de composantes fondamentales de l'évolution de la grandeur étudiée.

Il faut alors analyser ces composantes, en les *dissociant* les unes des autres, c'est-à-dire en considérant une série comme résultant de la combinaison de différentes composantes, tel que chacune d'elles ait une *évolution simple*.

1. *La tendance* $(f_i, 1 \leqslant i \leqslant n)$ 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* $(s_i, 1 \leqslant i \leqslant n)$ sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz, ...), ou encore par des activités économiques et sociales (fêtes, vacances, soldes, 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 $i \geqslant 1$. Au premier abord, cette composante est entièrement déterminée par ses p premières valeurs $s_1, s_2, \ldots, s_p$. Mais on rencontre souvent aussi des phénomènes pour lesquels la période peut elle-même varier. On parle alors de

3. *Cycles* $(c_i, 1 \leqslant i \leqslant n)$, qui regroupent des variations à période moins precise autour de la tendance, par exemple les phases économiques d'expansion et de récession. 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/résidus/bruit* $(e_i, 1 \leqslant i \leqslant n)$ 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 percevables 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 abérrantes* 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. *Les points de changement* 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 : $(f_i, 1 \leqslant i \leqslant n)$ la tendance (intégrant éventuellement un cycle), $(s_j, 1 \leqslant j \leqslant p)$ les coefficients saisonniers, $(e_i, 1 \leqslant i \leqslant n)$ les fluctuations irrégulières (intégrant éventuellement des accidents).

Exemple. — Trouver l'élément suivant de la série $(y_t)_{t\in\mathbb{N}}$ ci-dessous, puis une équation de récurrence pour

$$(y_t)_{t\in\mathbb{N}} = (1, 3, 7, 13, 21, 31, \ldots)$$

Obtenir une formule analytique pour $(y_t)_{t\in\mathbb{N}}$ en utilisant :

a) la théorie des équations de récurrence à coefficients constants. (*Réponse.* — $n^2 + n + 1$)

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

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

(*Réponse.* — $a(z) = (1+z^2)/(1-z)^3 = 2/(1-z)^3 - 2/(1-z)^2 + 1/(1-z)$.)

2.2. Quelques types de décomposition

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

1° *Le modèle additif.* — Le modèle additif s'écrit

$$y_i = f_i + s_i + e_i, \qquad 1 \leqslant i \leqslant n. \tag{1}$$

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 :

$$y_i = a + b \times i + \sum_{k=1}^{11} s_k \mathbb{1}_{\{i=k \bmod 12\}} - \left(\sum_{k=1}^{11} s_k\right) \mathbb{1}_{\{i=0 \bmod 12\}} + e_i.$$

Les coefficients a, b, $s_1, \ldots, s_{11}$ 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 laissaient prévoir la tendance et l'effet saisonnier pour février 1991).

2° *Le modèle multiplicatif (*).* — Le modèle multiplicatif s'écrit

$$y_i = f_i \times (1 + s_i) \times (1 + e_i), \qquad 1 \leqslant i \leqslant n. \tag{2}$$

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'entrées 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.

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 :

$$y_i = f_i \times (1 + s_i) + e_i, \qquad 1 \leqslant i \leqslant n. \tag{3}$$

(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 ensuite le modèle additif (le modèle multiplicatif revient à un modèle additif pour le logarithme des données).

Une fois un modèle obtenu, il peut être utilisé pour la prédiction des valeurs futures.

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 $(Y_t)_t$ s'appelle une moyenne mobile de $(X_t)_t$ ou filtre si

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

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

Lorsque $\sum \theta_i = 1$, on parle de lissage ; lorsque les θ_i sont tous égaux, on parle de moyenne arithmétique, et le cas d'une moyenne arithmétique avec $k_1 = k_2 = q$ est appelé moyenne arithmétique symétrique ou centrée.

Exemples. — Filtres arithmétiques, causaux, ...

On introduit l'opérateur de retard B (ou encore de « rétro-décalage ») défini sur l'ensemble des suites par

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

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 :

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

(*) *Notes et commentaires.* — Notons $\mathbb{T}$ le temps ($\mathbb{T} = \mathbb{N}$ ou $\mathbb{Z}$), H un espace vectoriel normé ($\mathbb{R}$ dans un cadre déterministe, $L^2(\Omega, \mathcal{A}, \mathbb{P})$ dans le cadre de signaux aléatoires). L'opérateur de retard B est « défini » sur des espaces de suites à valeurs dans H, l'espace de suites le plus intéressant pour nous étant celui des suites bornées $\ell^\infty(\mathbb{T}, H)$, qui lorsque H est un espace de Banach, séparable, est lui aussi un espace de Banach, séparable.

Si $\mathbb{T} = \mathbb{N}$, l'opérateur B est mal défini : si X est une suite à valeurs dans H, que vaut BX_0 ? On pourrait convenir que $BX_0 = 0 \in H$, mais cette convention est arbitraire et serait inconsistante avec une éventuelle extension à $\mathbb{S} \supset \mathbb{T}$, on aurait alors quelque chose du genre $B_{\mathbb{T}} = N_{\mathbb{T}} \circ B_{\mathbb{S}}$ où $N_{\mathbb{T}}$ annule tous les coefficients d'une suite indexée par $\mathbb{S}$ qui sont en dehors de $\mathbb{T}$. De plus, B n'est pas inversible ; son inverse naturel devant être le décalage à droite qui lui est bien défini lorsque $\mathbb{T} = \mathbb{N}$. Les choses ne sont pas simples, il y a des effets de bord, ce sera rappelé ou constaté plus tard.

Si $\mathbb{T} = \mathbb{Z}$, l'opérateur B est bien défini :

$$\begin{aligned} B : \ell^\infty(\mathbb{T}, H) &\longrightarrow \ell^\infty(\mathbb{T}, H) \\ X = (X_t)_{t\in\mathbb{Z}} &\longmapsto BX = (X_{t-1})_{t\in\mathbb{Z}} \end{aligned}$$

ainsi que son inverse $B^{-1} = 1/B$

$$\begin{aligned} B^{-1} : \ell^\infty(\mathbb{T}, H) &\longrightarrow \ell^\infty(\mathbb{T}, H) \\ X = (X_t)_{t\in\mathbb{Z}} &\longmapsto BX = (X_{t+1})_{t\in\mathbb{Z}}. \end{aligned}$$

Ces opérateurs sont linéaires, bornés, de normes égales à 1, ce sont même des isométries. S'il n'y a aucune difficulté à considérer des polynômes d'endomorphismes (les produits d'endomorphismes étant des compositions, l'ensemble des endomorphismes bornés de E étant une algèbre, ... que du classique), on peut vouloir prolonger le calcul fonctionnel à des séries entières d'endomorphismes. Pour ce faire, il faut noter que $E = \ell^\infty(\mathbb{T}, H)$ étant un espace de Banach, l'espace de ses endomorphismes bornés est aussi un espace de Banach. Les calculs auxquels nous sommes habitués, calculs reposant sur la convergence absolue ou la convergence normale, sont licites, en particulier ceux sur les séries entières. Ainsi, si $\theta(z)$ est une série entière de rayon de convergence $R > 0$ et C est un opérateur borné sur E de norme strictement inférieure à R, alors $\theta(C)$ est un opérateur borné sur E. Par la suite, on voudra considérer des expressions du type $\theta_+(B) + \theta_-(1/B)$, c'est-à-dire des séries de Laurent en l'opérateur de retard B. Il sera donc fondamental que les séries entières $\theta_+(z)$ et $\theta_-(z)$ aient des rayons de convergence strictement supérieurs à 1 (ce qui est évidemment vérifié si θ_+ et θ_- sont des polynômes). Pour finir, notons que l'ensemble des valeurs propres de B, ainsi que de B^{-1}, en tant qu'endomorphisme de $\ell^\infty(\mathbb{T}, H)$ est exactement le cercle unité $\{z \in \mathbb{C} : |z| = 1\}$ (considérer les suites $(\lambda^n \times X)_{n\in\mathbb{Z}}$, avec $|\lambda| = 1$ et $X \in H$). $(_*)$

Exemples et applications dans la théorie des suites récurrentes à coefficients constants. — À développer...

L'opérateur $\theta(B)$ 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 $\theta(z) : \mathbb{C} \to \mathbb{C}$.

(*) *Notes et commentaires.* — Rappelons qu'*a priori* k_1 et k_2 peuvent être infinis dans cette définition. L'expression « fonction » de la définition rappelle que $\theta(z)$ n'est peut-être pas définie partout. $(_*)$

Nous travaillerons surtout avec des filtres causaux

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

où k peut être aussi infini. Dans ce cas, les coefficients du filtre seront souvent notés $(\psi_i)_i$, et le symbole $\psi(z)$.

3.1. Fonctions génératrices

Un fait très important est que la fonction génératrice

$$y(z) = \sum_{i=0}^{\infty} Y_i z^i$$

d'une série définie par un filtre causal $Y_t = \psi(B)X_t$ est essentiellement le produit de $x(z) = \sum_{i=0}^{\infty} X_i z^i$, la fonction génératrice de X_t et du symbole $\psi(z)$. Plus précisement, notons $\psi_{\leqslant m}(z) = \sum_{i=0}^{m} \psi_i z^i$ la troncation de n'importe quelle série des puissances aux premiers termes.

Théorème 3.1. — *Pour chaque filtre causal d'ordre fini $\psi(B) = \sum_{i=0}^{k} \psi_i B^i$, on a :*

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

où l'ordre du filtre est inférieur ou égal à $k+1$.

Démonstration. — Nous allons vérifier « formellement »[1] le cas particulier $k = 2$ des filtres quadratiques $\psi_0 + \psi_1 B + \psi_2 B^2$, pour lequel ce théorème s'énonce :

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

(*) *Notes et commentaires.* — Le filtre $\psi(B)$ étant causal d'ordre fini, son symbole $\psi(z)$ est un polynôme et ses coefficients $(\psi_i)_i$ sont nuls en dehors d'une partie finie $\{0, \ldots, k\}$ de $\mathbb{N}$.

Rappelons que si des séries entières $x(z)$ et $\psi(z)$ ont des rayons de convergence strictement positifs, dans le plus petit disque de convergence $\psi(z)x(z)$ est analytique et son expression en série est donnée par le produit de Cauchy :

$$\psi(z)x(z) = \sum_{i=0}^{\infty} \left(\sum_{j=0}^{i} \psi_i X_{i-j} \right) z^i.$$

Comme ici, $\psi(z)$ est un polynôme, cette identité est satisfaite analytiquement sur le disque ouvert de convergence de $x(z)$.

Par ailleurs, sans oublier que la relation de filtrage n'a pas de sens pour les premiers termes

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

de Y, et en notant que $\psi_j = 0$ pour $j > k$,

$$\begin{aligned} y(z) &= \sum_{i=0}^{\infty} Y_i z^i = \sum_{i=k}^{\infty} \Big(\sum_{j=0}^{k} \psi_j X_{i-j} \Big) z^i + \sum_{i=0}^{k-1} Y_i z^i \\ &= \sum_{i=k}^{\infty} \Big(\sum_{j=0}^{i} \psi_j X_{i-j} \Big) z^i - \sum_{i=k}^{\infty} \Big(\sum_{j=k+1}^{i} 0_j X_{i-j} \Big) z^i + \sum_{i=0}^{k-1} Y_i z^i \\ &= \sum_{i=0}^{\infty} \Big(\sum_{j=0}^{i} \psi_j X_{i-j} \Big) z^i - \sum_{i=0}^{k-1} \Big(\sum_{j=0}^{i} \psi_j X_{i-j} \Big) z^i - \sum_{i=k}^{\infty} 0 \times z^i + \sum_{i=0}^{k-1} Y_i z^i \\ &= \psi(z)x(z) - \sum_{i=0}^{k-1} \Big(\sum_{j=0}^{i} \psi_j X_{i-j} \Big) z^i - 0 + y_{\leqslant k-1}(z) \end{aligned}$$

Soit

$$\begin{aligned} y(z) - y_{\leqslant k-1}(z) &= \psi(z)x(z) - \sum_{i=0}^{k-1} \Big(\sum_{j=0}^{i} \psi_j X_{i-j} \Big) z^i = \psi(z)x(z) - \sum_{i=0}^{k-1} \Big(\sum_{j=0}^{i} \psi_j z^j X_{i-j} z^{i-j} \Big) \\ &= \psi(z)x(z) - \sum_{i=0}^{k-1} \Big(\sum_{j=0}^{i} \psi_{i-j} z^{i-j} X_j z^j \Big) = \psi(z)x(z) - \sum_{j=0}^{k-1} \sum_{i=j}^{k-1} \psi_{i-j} z^{i-j} X_j z^j \\ &= \psi(z)x(z) - \sum_{j=0}^{k-1} \Big(\sum_{i=0}^{k-1-j} \psi_i z^i \Big) X_j z^j = \psi(z)x(z) - \sum_{j=0}^{k-1} \psi_{\leqslant k-1-j}(z) X_j z^j \end{aligned}$$

Si le rayon de convergence de $x(z)$ est nul, on invoque le calcul formel sur les séries. $(*)$

Remarques. — *a*) 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 !

b) Pour les séries $(Y_t)_t$, $(X_t)_t$ doublement infinies, *i.e.* avec $t \in \mathbb{Z}$, le résultat est plus simple

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

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

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

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

Mais, le travail avec les séries $(Y_t)_t$, $(X_t)_t$ doublement infinies contient des « pièges » dont on discutera plus tard. D'un autre côté, travailler avec des séries indexées par $t \in \mathbb{N}$ nous force à définir l'égalité d'une série comme égalité des composantes, à partir d'un certain point, *i.e.*

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

Il 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 :*

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

Remarque. — L'approche de décomposition additive demande de rompre une série donnée $Y_t = m_t + \varepsilon_t$ dans un « signal » $(m_t)_t$ et du bruit $(\varepsilon_t)_t$. Il est naturel d'essayer de dépister le signal par un filtre de lissage causal $\ell(B)$, donc $m_t = \ell(B)Y_t$, tel que ce filtre « détruit le bruit mais laisse passer le signal ». Il s'ensuit que le bruit aussi est donné par un filtre causal

$$\varepsilon_t = Y_t - m_t = \big(\mathrm{Id} - \ell(B)\big)Y_t = \pi(B)Y_t$$

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

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

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

3.2. Filtres de lissage

Un filtre de lissage $\theta(B)$ (filtre dont les coefficients vérifient $\sum_i \theta_i = 1$) peut être utilisé pour la prédiction de $(X_t)_t$:

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

Remarquer que le fait que la prédiction est « non-biaisée pour les séries stationnaires », *i.e.*

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

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

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

$$\theta(B)\mathbb{k} = k \times \theta(B)\mathbb{1} = k \times \mathbb{1} = \mathbb{k}.$$

La vérification est très facile pour ça, remarquons que...

Il est possible en fait, en choisissant les coefficients $(\theta_i)_i$ d'un filtre, d'assurer qu'il laisse invariantes toutes les séries polynomiales $(p_t)_t$ d'un degré donné.

Exercice 3.1. — (i) Montrer qu'une moyenne arithmétique symétrique d'ordre $2q + 1 = 3$, donnée par

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

conserve (laisse invariantes) les tendances linéaires $p_t = a + b \times t$.

(ii) Généraliser pour q quelconque.

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

Exercice 3.2. — Montrer 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

$$X_t = p_t + \varepsilon_t$$

où $p_t = a + b \times t$ 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 $(\varepsilon_t)_t$, ramenant à :

$$\hat{X}_t = \theta(B)(p_t + \varepsilon_t) = p_t + \big(\theta(B)\varepsilon_t\big) = p_t + \frac{\varepsilon_{t+q} + \cdots + \varepsilon_t + \cdots + \varepsilon_{t-q}}{2q+1} = p_t + \varepsilon'_t$$

avec un nouveau bruit $\varepsilon'_t = \frac{\varepsilon_{t+q}+\cdots+\varepsilon_t+\cdots+\varepsilon_{t-q}}{2q+1}$ de variance inférieure à celle de $(\varepsilon_t)_t$.

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 cela va réduire (atténuer) le bruit et mettre en évidence la tendance linéaire. L'effet du lissage augmente en augmentant q.

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

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

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 $Z = (Z_t)_t$ satisfaisant $\theta(B)Z = Z$.

Théorème 3.3. — *L'espace invariant d'un filtre contient les polynômes de degré inférieur ou égal à p si et seulement si 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. — Démontrer 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 $Z = (Z_t)_t$ satisfaisant $\theta(B)Z = \mathbb{0}$. Cette question a une portée pratique pour l'enlèvement de composantes saisonnières (et leur détermination).

3.3. Filtres qui enlèvent les composantes saisonnières

Définition 3.3. — (i) Une série $(s_t)_t$ sera appelée périodique de période p si et seulement si

$$s_{t+p} = s_t \iff (1 - B^p)s_t = 0, \qquad \text{pour tout } t. \tag{6}$$

(ii) Une série $(s_t)_t$ sera appelée saisonnière de période p si et seulement si

$$\sum_{i=1}^{p} s_{t+i} = 0 \iff \left(\sum_{i=0}^{p-1} B^i\right)s_t = 0, \qquad \text{pour tout } t. \tag{7}$$

Exercice 3.5. — Montrer qu'un filtre $\theta(z)$ qui est divisible par $1 + z + \cdots + z^{p-1}$, *i.e.* de la forme $\theta(z) = (1 + z + \cdots + z^{p-1})\theta_1(z)$, « enlève » les composantes saisonnières de période p, *i.e.*

$$\theta(B)s(t) = 0, \qquad \text{pour tout } t,$$

pour chaque série $(s_t)_t$ 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 si et seulement si son symbole $\theta(z)$ est divisible par $1 + z + \cdots + z^{p-1}$ (donc si $\theta(z) = 0$, pour toutes les racines 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 déceler ensuite la tendance.

Alternativement, après le choix d'une forme appropriée pour la tendance et une pèriode specifique, selon le graphe, on peut déterminer en 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 (voir TRAVAUX DIRIGÉS N° 1)

4. **Modélisation stochastique des séries temporelles**

4.1. INTRODUCTION

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

$$Y_t = m_t + \varepsilon_t$$

où $(m_t)_t$ représente la « tendance » (intuitivement un « mouvement lisse à long terme »), qui sera la composante la plus importante dans la prévision ; $\varepsilon_t = Y_t - m_t$ sont les « résidus » qui restent après qu'on enlève la partie structurée $(m_t)_t$. 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 la calcul des intervalles de confiance). On s'arrangera toujours pour que les résidus aient une moyenne nulle, mais cela n'est pas suffisant pour qu'ils soient un bruit totalement sans structure, *i.e.* un « bruit blanc » (et s'il y a encore une partie structuré, elle devrait être incluse dans $(m_t)_t$).

Le « bruit blanc » est notre premier exemple de processus stochastique : une formalisation du concept de séries temporelles, ayant des propriétés bien définies (*voir* chapitre suivant). Inspiré par les propriétés de ce processus, on proposera des tests statistiques correspondant a ce modèle qui nous permettrons de décider si $(\varepsilon_t)_t$ ont les propriétés de manque de structure desirées. Pour la tendance, plusieurs modèles se sont averés utiles :

1. régression sur des prédicteurs exogènes (« covariates »), implémenté dans des logiciels comme R par des « formules » :

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

2. modèles de superposition des chocs extérieurs/moyennes mobiles/FIR *inobservables* $(\varepsilon_t)_t$:

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

3. modèles auto-régressifs :

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots) + \varepsilon_t.$$

Dans le manque des prédicteurs exogènes, il est assez naturel d'adopter une modélisation auto-régressive pour la tendance. Sous certaines conditions de regularité, cela ramenera à des prévisions auto-régressives un pas en avant :

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

(*voir* la note en bas de page[2]). Le modèle le plus simple est le processus AR(1) :

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

Ce modèle est recommandable si on envisage une prévision

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

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.

(*) *Notes et commentaires.* — Ce morceau de texte est en bonne partie un copier-coller de l'exercice 4.2. La rédaction est inachevée, ce qui cause des incohérences. $(_*)$

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

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

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

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

et examine solution stationnaire unique si et seulement si $|\phi| < 1$.

Indication. — Vous pouvez le faire en calculant la solution :

(i) par des substitutions répétées ou

(ii) en utilisant des opérateurs, en posant $Y_t = (1 - \phi B) - 1\varepsilon_t$, et en développant la fraction comme une série de puissances de B. Ensuite, calculez les covariances, pour montrer la stationnarité.

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

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 rigourousement les pro- priétés desirées 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.*

2. La modélisation auto-régressive permettra aussi des prédictions k pas en avant :

$$\widehat{Y}_{t+k} = f_k(Y_t, Y_{t-1}, Y_{t-2}, \ldots), \qquad k = 1, 2, \ldots$$

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}, ... Plusieurs choix sont possibles par exemple par extrapolation à partir d'un ajustement/interpolation polynomiale ou par splines.

une certaine uniformité de structure par rapport au temps. Ça nous ramène à considérer les processus stochastiques stationnaires, les distributions jointes desquelles ne varient pas avec le temps.

4.2. Processus stochastiques stationnaires

Définition 4.1. — Soit X un processus aléatoire indexé par $\mathbb{T} = \mathbb{N}$ ou $\mathbb{Z}$. On dit que X est stationnaire (strict) si pour toute famille finie d'instants $t_1, \dots, t_r \in \mathbb{T}$ 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 $\mathbb{T} = \mathbb{N}$ 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 :

$$\mathbb{E}[X_t] = m_t = m, \qquad \forall\, t \in \mathbb{T},$$

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 :

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

Comme la plupart des 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 moyennes théoriques.

Théorème 4.1. — *Pour un processus stationnaire, les covariances empiriques*

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

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

(*) *Notes et commentaires.* — Ça me semble un peu raide ! La simple hypothèse de stationnarité suffirait ? J'en doute (prendre une suite constante $X_n = X$). $(_*)$

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

Remarques. — *a*) 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.).

b) 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 $Y = (Y_n)_n$ de variables aléatoires indépendantes de même moyenne et même variance est toujours stationnaire à l'ordre 2 ; mais si les $(Y_n)_n$ n'ont pas tous la même loi, Y n'est pas stationnaire au sens strict.

c) (*) 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 $(X_k)_k$ considérée comme suite de vecteurs dans l'espace de Hilbert $L^2(\Omega, \mathcal{A}, \mathbb{P})$. 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(∞) :

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \qquad \text{MA}(\infty)$$

avec $(\varepsilon_t)_t$ bruit blanc et leur cas particulier avec un nombre fini q de coefficients $(\psi_i)_i$ non nuls, les « moyennes mobiles » MA(q).

Le troisième exemple étudié seront les « processus auto-régressifs » AR(∞)

$$\varepsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i} \qquad \text{AR}(\infty)$$

avec $(\varepsilon_t)_t$ bruit blanc et leur cas particulier avec un nombre fini p de coefficients $(\pi_i)_i$ non nuls, les processus AR(p).

1° *Le bruit blanc.* — L'exemple le plus simple de modèle stochastique est le bruit blanc discret, la structure « révée » des résidus qui restent après qu'on enlève la tendance/moyenne d'un processus.

Définition 4.3. — Un processus $\varepsilon = (\varepsilon_t)_{t\in\mathbb{T}}$, où $\mathbb{T}$ est un ensemble dénombrable quelconque, est appelé bruit blanc stationnaire si les variables $(\varepsilon_t)_t$ sont i.i.d. (indépendants et identiquement distribuées) à espérance $\mathbb{E}[\varepsilon_t] = 0$. Il sera appelé bruit blanc gaussien si la distribution de chaque variable aléatoire ε_t est gaussienne.

Un bruit blanc ε a pour covariance

$$\gamma(s,t) = \mathbb{E}[\varepsilon_s \varepsilon_t] = 0, \qquad \forall\, s \neq t, \tag{12}$$

et donc pour coefficient de corrélation

$$\varrho(s,t) = \frac{\gamma(s,t)}{\sigma_s \sigma_t} = \delta(s-t) \tag{13}$$

où $\delta(s-t)$ est le symbole de Kronecker.

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

Définition 4.4. — Un processus $\varepsilon = (\varepsilon_t)_{t\in\mathbb{T}}$ ($\mathbb{T} = \mathbb{N}$ ou $\mathbb{Z}$) est appelé bruit blanc de second ordre s'il a pour moyenne 0, une variance constante $\mathbb{E}[\varepsilon_t^2] = \sigma^2$ et une covariance $\gamma(s,t) = \mathbb{E}[\varepsilon_s \varepsilon_t] = 0$, pour tous $s \neq t$ (et donc pour coefficients de corrélation $\varrho(s,t) = \delta(s-t)$).

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 intéressant pour un bruit aléatoire.

2. Le bruit blanc stationnaire est une idéalisation du processus des résidus de la régression linéaire, qu'on aimerait « indépendants ». Mais, comme l'indépendance est un concept probabiliste, et les résidus sont le résultat déterministe d'une régression appliquée à une série observée une seule fois, il est difficile de la vérifier rigoureusemment. Parmi les tests possibles, mentionnons celui du « turning points », qui demande de vérifier que la fréquence de ces points est environ 4/6, et le test qui vérifie si la somme des corrélations empiriques est proche de 0. Si ces deux tests sont positifs, on sait au moins « qu'on ne peut pas repousser l'hypothèse de l'indépendance ». Il y aussi des tests distributionnels des résidus comme ceux de Fisher, de Student, qui testent la gaussiennité (? s'interroge le commentateur...).

3. Quand les tests des données rejettent l'hypothèse du bruit blanc, *i.e.* quand on a du bruit corrélé, la régression classique doit être remplacée par une analyse plus fine, appelée *krigeage* en géostatistique.

2° *Les processus linéaires et les moyennes mobiles* MA(q). — Définition 4.5. — Un processus $Y = (Y_t)_t$ sera appelé linéaire en $(\varepsilon_t)_t$ s'il peut être représenté sous la forme :

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \varepsilon_{t-i} \qquad \text{avec} \qquad \sum_{i=-\infty}^{\infty} |\psi_i|^2 < \infty \tag{14}$$

où $(\varepsilon_t)_t$ est un bruit blanc.

(*) *Notes et commentaires.* — La condition $\sum_{i=-\infty}^{\infty} |\psi_i|^2 < \infty$ est obligatoire pour que Y_t soit dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$ puisque $\|Y_t\|_2^2 = \sigma^2 \sum_{i=-\infty}^{\infty} |\psi_i|^2$, où σ^2 est la variance du bruit blanc. Rappelons que si $1 \leqslant p \leqslant q$ et $\sum_{i=-\infty}^{\infty} |\psi_i|^p < \infty$ alors $\sum_{i=-\infty}^{\infty} |\psi_i|^q < \infty$ (le cas qui pourra nous intéresser est $p = 1$, $q = 2$), ce qui s'écrit

$$\ell^p(\mathbb{Z}) \subset \ell^q(\mathbb{Z}) \quad \text{si} \quad 1 \leqslant p \leqslant q \qquad (\text{on a aussi } \ell^p(\mathbb{N}) \subset \ell^q(\mathbb{N})).$$

En effet si la première série converge, son terme général tend vers 0, auquel cas, le terme général de la seconde série est dominé par celui de la première, ce qui assure la convergence. Ceci est toujours surprenant quand on pense aux intégrales généralisées puisque $L^p(\mathbb{R}, \mathrm{d}x) \not\subset L^q(\mathbb{R}, \mathrm{d}x)$ (une fonction peut être intégrable sans tendre vers 0 aux infinis), et que dans un cadre probabiliste l'inclusion est dans l'autre sens $L^p(\Omega, \mathcal{A}, \mathbb{P}) \supset L^q(\Omega, \mathcal{A}, \mathbb{P})$ si $1 \leqslant p \leqslant q$. $(_*)$

Évidemment, du point de vue pratique (pour la prédiction), on ne s'intéresse qu'au cas — qui sera appelé causal — où la représentation n'utilise pas « le bruit futur » :

Définition 4.6. — Un processus linéaire $Y = (Y_t)_t$ est dit causal s'il peut être représenté sous la forme :

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{15}$$

où $\varepsilon = (\varepsilon_t)_t$ est un bruit blanc et $\sum_{i=1}^{\infty} |\psi_i|^2 < \infty$.

Définition 4.7. — On appelle processus MA(q) un processus linéaire $Z = (Z_t)_{t \in \mathbb{Z}}$ vérifiant une relation du type :

$$Z_t = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}, \qquad \forall\, t \in \mathbb{Z}, \tag{16}$$

où $\varepsilon = (\varepsilon_t)_t$ est un bruit blanc de variance σ^2 et $\theta_0 = 1$.

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

$$Z = \theta(B)\varepsilon.$$

THÉORÈME 4.2. — *Un processus linéaire*

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

où $\sum_{i=-\infty}^{\infty} |\psi_i|^2 < \infty$*, est :*

(i) *bien défini dans* $L^2(\Omega, \mathcal{A}, \mathbb{P})$ *(*i.e. $\mathrm{Var}(Y_t) < \infty$*) ;*

(ii) *à variance constante stationnaire* $\mathrm{Var}(Y_t) = \sigma_\varepsilon^2 \sum_{i=-\infty}^{\infty} |\psi_i|^2$ *;*

(iii) *à auto-covariance donnée par :*

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

(iv) *stationnaire d'ordre deux.*

Démonstration. — (i) En considérant $\mathrm{Var}(Y_t)$, on voit que la condition est nécessaire et suffisante pour la convergence.

(ii, iii, iv) Ensuite, on voit qu'elle suffit pour la stationnarité, car elle assure que la covariance $\mathrm{Cov}(Y_t, Y_{t+k})$ est bien définie par l'inégalité de Cauchy–Schwarz (qui est équivalente à $|\varrho_k| \leqslant 1$) et ne dépend pas de t. □

Exercice 4.1. — (i) Calculer la fonction d'auto-covariance $\gamma(k)$ d'un processus MA(1).

(ii) Calculer la fonction de covariance $\gamma(k)$ d'un processus MA(q). Le fait que les fonctions de covariance et corrélation $\gamma(k)$, respectivement $\varrho(k)$, 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

$$|\varrho_n(k)| \leqslant z_{1-\alpha/2} \sigma_q,$$

où

$$\sigma_q^2 = \frac{1 + 2\big(\hat{\varrho}(1)^2 + \hat{\varrho}(2)^2 + \cdots + \hat{\varrho}(q)^2\big)}{n}$$

(formule de Bartlett) et $z_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la distribution gaussienne centrée réduite, dépend du niveau de confiance $1 - \alpha$ desiré (par exemple, $z_{0.95} = 1{,}6449$ pour $\alpha = 0{,}1$). 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).

3° *Les modèles auto-régressifs* AR(p). — La prédiction d'une série est particulièrement simple quand elle peut être « bien approchée » par un modèle auto-régressif paramétrique :

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots) + \varepsilon_t \tag{18}$$

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

$$\widehat{Y}_t = f(Y_{t-1}, Y_{t-2}, \ldots).$$

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

Définition 4.8. — Un processus stationnaire $Y = (Y_t)_{t\in\mathbb{Z}}$ sera appelé processus auto-régressif linéaire d'ordre p, soit AR(p), s'il existe un bruit blanc $(\varepsilon_t)_t$ et des réels $(\phi_i)_{i=1}^p$ tels qu'une relation de récurrence :

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t, \qquad \forall\, t \in \mathbb{Z}, \tag{19}$$

est vérifiée.

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

$$\phi(B)Y_t = \varepsilon_t \qquad \text{où} \qquad \phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$$

Définition 4.9. — Le polynôme

$$\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$$

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

Remarque. — Les processus auto-régressifs sont définis par une équation, qui *a priori*, peut ne pas avoir des solutions ; pour « solution » de l'équation (19) nous aimerions avoir une représentation du processus $(Y_t)_t$ par rapport au processus $(\varepsilon_t)_t$.

Nous voyons maintenant que le processus AR(1) (qui est markovien) défini par $Y_t = \phi Y_{t-1} + \varepsilon_t$ a une représentation MA(∞) si et seulement si $|\phi| \neq 1$ et cette représentation est causale si et seulement si $|\phi| < 1$. Elle sera obtenue : (i) en résolvant la relation de récurrence ; (ii) par inversion formelle du polynôme $\phi(B) = 1 - \phi B$.

Exercice 4.2 (processus AR(1)*).* — (i) Montrer que l'équation :

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

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

Indication. — On peut le faire en calculant la solution :

a) par des substitutions répétées ou

b) en utilisant des opérateurs, en posant $Y_t = (1-\phi B) - 1\varepsilon_t$, et en développant la fraction comme une série de puissances de B. Ensuite, calculez les covariances, pour montrer la stationnarité.

(ii) Montrer que l'équation (20) a une solution stationnaire unique, qui dépend seulement du bruit futur si et seulement si $|\phi| > 1$.

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 $(\varepsilon_t)_t$ 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 $(\varepsilon_t)_t$ n'est pas un bruit d'innovation.

3. Pour $|\phi| = 1$, l'équation (20) (appelée marche aléatoire) n'a pas de solution stationnaire. Par contre, les accroissements $Y_t - Y_{t-1} = \varepsilon_t$ sont stationnaires ; cette situation plus compliquée sera analysé dans le chapitre sur les processus ARIMA(p, d, q).

En conclusion, on a une représentation causale MA(∞) (en termes du bruit passé) du processus AR(1) si et seulement si le polynôme caractéristique $\phi(z) = 1 - \phi z$ a sa racine $1/\phi$ à l'extérieur du disque unité : $1/\phi \in \bar{\mathbb{D}}^c$, c'est-à-dire $|1/\phi| > 1$ ou encore $|\phi| < 1$.

Exercice 4.3. — Montrer que si un processus AR(2) défini par $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ a une représentation stationnaire causale

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

alors les coefficients $(\psi_i)_i$ satisfont la récurrence de Yule–Walker

$$\psi_t = \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2}, \qquad t \geqslant 2, \qquad \psi(0) = 1, \psi(1) = \phi_1.$$

(*) *Notes et commentaires.* — Dans l'exercice suivant, on est amené à considérer une série entière $\sum_{i=0}^{\infty} \psi_i z^i$ qui converge sur un voisinage du disque unité fermé $\bar{\mathbb{D}}$. Ceci implique sa convergence absolue sur le cercle unité $\mathbb{S}$, et donc $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Dans un commentaire précédent nous avons évoqué le fait qu'alors $\sum_{i=0}^{\infty} |\psi_i|^2 < \infty$ ($p = 1$, $q = 2$). Ceci est important car sinon, le mariage entre variables aléatoires de carré intégrable et analyse complexe ne serait pas aussi simple. $(_*)$

Exercice 4.4. — Montrer que la formule des coefficients $(\psi_j)_j$ de la représentation MA(∞) d'un processus AR(2) défini par $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$, en fonction des racines de « l'équation caractéristique » de la récurrence Yule–Walker

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

est :

$$\psi(j) = \begin{cases} (\lambda_1^{k+1} - \lambda_2^{k+1})/(\lambda_1 - \lambda_2) & \text{si } \lambda_1 \neq \lambda_2 \\ (k+1)\lambda^k & \text{si } \lambda_1 = \lambda_2 \end{cases}$$

Montrer que si l'équation caractéristique a ses racines dans le disque unité, λ_1, $\lambda_2 \in \mathbb{D}$ (et donc le « symbole » $1 - \phi_1 z - \phi_2 z^2$ a ses racines z_1, z_2 en dehors du disque unité), alors la condition

$$\sum_i |\psi_i|^2 < \infty$$

qui assure la convergence des représentations MA(∞), est satisfaite. Donner les coefficients $(\psi_j)_j$ dans les cas particuliers :

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

Indications. — La solution générale de la récurrence d'ordre 2, $\psi(k) = \phi_1 \psi(k-1) + \phi_2 \psi(k-2)$, pour des valeurs initiales arbitraires $\psi(0) = x_0$, $\psi(1) = x_1$ est :

1. lorsque les racines λ_1, λ_2 sont distinctes,
$$\frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} x_1 - \frac{\lambda_1^k \lambda_2 - \lambda_2^k \lambda_1}{\lambda_1 - \lambda_2} x_0 \, ;$$
2. lorsque les racines $\lambda = \lambda_1 = \lambda_2$ sont égales,
$$\psi(k) = x_0 \lambda^k + (x_1 - \lambda x_0) \times k \lambda^{k-1}.$$

Remarquons que les processus AR(1) et AR(2) ont une représentation MA(∞) avec coefficients qui satisfont $\sum_i |\psi_i|^2 < \infty$ si et seulement si l'équation $0 = \phi(z)$ a ses racines en dehors du disque unité z_1, $z_2 \in \bar{\mathbb{D}}^c$ (ou si l'équation caractéristique a ses racines dans le disque unité λ_1, $\lambda_2 \in \mathbb{D}$). Il s'avère que cette situation est typique pour tous les modèles AR(p).

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 :

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}, \qquad \forall\, t \in \mathbb{Z}, \tag{21}$$

où les $(\phi_i)_i$, $(\theta_i)_i$ sont des réels et $(\varepsilon_t)_t$ est un bruit blanc de variance σ^2.

La notation des polynômes de retard ramène (21) à la forme :

$$\phi(B)Y_t = \theta(B)\varepsilon_t \tag{22}$$

Nous verrons ci-dessous que *les processus* ARMA(p,q) *avec des polynômes caractéristiques* $\phi(B)$, $\theta(B)$ *à racines en dehors du disque unité ont deux autres représentations équivalentes :*

1. MA(∞), de $(Y_t)_t$ en termes de $(\varepsilon_t)_t$ (appelée aussi représentation linéaire causale), et
2. AR(∞), de $(\varepsilon_t)_t$ en termes de $(Y_t)_t$ (appelée aussi représentation inverse).

Ces représentations peuvent être obtenues par des inversions formelles de l'équation (22), suivies par un développement de la fraction correspondante en série de puissances :

$$Y_t = \frac{\theta(B)}{\phi(B)}\,\varepsilon_t = \left(\sum_{i=0}^{\infty} \psi_i B^i\right)\varepsilon_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \qquad \varepsilon_t = \frac{\phi(B)}{\theta(B)}\,Y_t = \left(\sum_{i=0}^{\infty} \pi_i B^i\right)Y_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$$

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

$$Y(z) = \psi(z)\varepsilon(z) \tag{23}$$

où $Y(z) = \sum_{t=-\infty}^{\infty} Y_t z^t$, $\varepsilon(z) = \sum_{t=-\infty}^{\infty} \varepsilon_t z^t$ sont les fonctions génératrices doublement infinies, et $\psi(z)$ est la fonction de transfert. En appliquant cette relation aux deux membres de (22), on trouve que :

$$\phi(z)Y(z) = \theta(z)\varepsilon(z)$$

En comparant avec (23), nous voyons que l'hypothèse ARMA(p,q) est équivalente à la rationalité de la fonction de transfert :

$$\psi(z) = \frac{\theta(z)}{\phi(z)}\,.$$

La modélisation ARMA peut être vue aussi comme une approximation Padé.

Hypothèse. — Le besoin de travailler avec des séries stationnaires ARMA ayant représentations causales et inversibles (*voir* ci-dessous) nous force à accepter au début seulement des fonctions de transfert $\psi(z)$ qui n'ont ni des racines ni des pôles dans le disque unité fermé $\bar{\mathbb{D}}$. Plus tard, en renonçant à la stationnarité, nous allons permettre aussi l'existence de pôles sur le cercle unité $\mathbb{S}$ (ces extensions s'appellent modèles ARIMA et SARIMA).

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

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1} + \phi Y_{t-1}.$$

5° *Les modèles* ARIMA(p,d,q). — Définition 4.11. — On appelle processus ARIMA(p,d,q) un processus $X = (X_t)_t$ 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) :

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \iff \phi(B)Y_t = \theta(B)\varepsilon_t, \qquad \forall\, t \in \mathbb{Z}, \tag{24}$$

où les $(\phi_i)_i$, $(\theta_i)_i$ sont des réels et $(\varepsilon_t)_t$ est un bruit blanc de variance σ^2.

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

$$\phi(B)Y_t = \phi(B)(1-B)^d X_t := \tilde{\phi}(B)X_t = \theta(B)\varepsilon_t$$

où $\tilde{\phi}(B) = \phi(B)(1-B)^d$.

Remarquons que ces processus ont à peu près la même définition que les processus ARMA, la seule différence étant qu'on permet au symbole auto-régressif $\tilde{\phi}(B)$ d'avoir la racine 1.

6° *Les modèles* SARIMA(p,d,D,q). — Définition 4.12. — On appelle processus SARIMA (p,d,D,q) un processus $X = (X_t)_t$ pour lequel le processus obtenu en différenciant d fois et en enlevant une saisonalité d'ordre D : $Y_t = (\sum_{i=0}^{D-1} B^i)(1-B)^d X_t$, $t \in \mathbb{Z}$, vérifie une relation de récurrence ARMA(p,q) :

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \iff \phi(B)Y_t = \theta(B)\varepsilon_t, \qquad \forall\, t \in \mathbb{Z}, \tag{24}$$

où les $(\phi_i)_i$, $(\theta_i)_i$ sont des réels et $(\varepsilon_t)_t$ est un bruit blanc de variance σ^2.

La notation des polynômes de retard ramène (25) à la forme :

$$\phi(B)Y_t = \phi(B)\left(\sum_{i=0}^{D-1} B^i\right)(1-B)^d X_t := \tilde{\phi}(B)X_t = \theta(B)\varepsilon_t$$

où $\tilde{\phi}(B) = \phi(B)(\sum_{i=0}^{D-1} B^i)(1-B)^d$.

Remarquons que ces processus ont a peu prés la même définition que les processus ARMA, la seule différence étant qu'on permet au symbole auto-régressif $\tilde{\phi}(B)$ d'avoir la racine 1, et aussi les racines de l'unité d'ordre D. On peut envisager aussi des modèles plus généraux, ayant des racines de l'unité multiples, ce qui revient à permettre des symboles auto-régressifs $\tilde{\phi}(B)$ ayant des racines rationnelles sur le cercle unité $|z| = 1$.

4.4. L'inversion des séries des puissances et des filtres $\phi(B)$ (*)

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

Théorème 4.3. — (i) *Pour un polynôme $\phi(z) = \prod_{i=1}^{p}(1 - z/\lambda_i)$ qui a toutes ses racines λ_i à l'extérieur du disque unité, $1/\phi(z)$ a un développement en série de Taylor*

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

qui est convergent à l'intérieur du disque unité $\bar{\mathbb{D}}$. Dans le cas le plus simple où les racines λ_i sont distinctes, on a

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

où $K_i = -1/\phi'(\lambda_i)$. (Dans le cas où des racines sont confondues, on a des formules similaires qui utilisent des dérivées d'orde supérieur.)

(ii) *Pour un polynôme $\phi(z) = \prod_{i=1}^{p}(1 - z/\lambda_i)$ qui a toutes ses racines λ_i à l'intérieur du disque unité $\mathbb{D}$, $1/\phi(z)$ a un développement en série de Laurent*

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

qui est convergent sur le cercle unité $|z| = 1$. *Dans le cas le plus simple où les racines* λ_i *sont distinctes, on a*

$$\psi_n = -\sum_{i=1}^{p} K_i \times \lambda_i^{n+1}$$

où $K_i = -1/\phi'(\lambda_i)$.

(iii) *Dans le cas mixte avec des racines avec des racine à l'intérieur et aussi à l'extérieur du disque unité, on a un mélange des formules ci-dessus.*

Ce résultat justifie des manipulations formelles analogues à celles qu'on fait avec des fonctions de l'opérateur B.

Notes. — On peut approcher de manière rigoureuse les manipulations formelles comme l'inversion du polynôme $\phi(B)$ par plusieurs démarches.

1. *Les fonctions génératrices.* Cette approche associe à chaque suite $(\psi_n)_n$ 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 « disque de convergence », dans le deuxième cas, la série est convergente dans l'extérieur d'un certain « disque de divergence » et dans le troisième cas, appelé série de Laurent, la série est convergente, mais a des expressions différentes dans l'intérieur des « anneaux de convergence » qui évitent les singularités. Le rôle joué par la convergence dans les calculs n'est pas crucial ; on peut utiliser parfois même des séries divergentes partout, en les définissant commes objets isomorphes à un certain anneau algébrique.

2. *Les matrices Toeplitz.* On s'aperçoit que les opérateurs sur les suites correspondant à des polynômes en B sont représentés par des matrices Toeplitz ; on peut démontrer qu'il y a un isomorphisme entre l'anneau des matrices Toeplitz et celui des fonctions génératrices. Cet isomorphisme explique l'équivalence des deux approches. Formellement, la conclusion est que l'opérateur B doit être traité commme le scalaire $z = 1$ (qui est sa valeur propre), et donc « l'expansion correcte » pour les inversions $1/\phi(z)$ en série de puissances dépendront du positionnement du point $z = 1$ par rapport aux racines.

(*) *Notes et commentaires.* — Il faut revoir les paragraphes précédents qui sont assez brouillons, en se souvenant que le spectre de B et de $1/B$ est le cercle unité. $(_*)$

1° *Causalité des modèles* AR(p). — Rappelons qu'il y a un problème (non-causalité) avec le modèle AR(1) quand la racine $\lambda = \phi^{-1}$ de son polynôme $\phi(z) = 1 - \phi z$ est à l'intérieur du disque unité. Ce problème est lié à l'existence de plusieurs développement possibles pour la fonction $\phi(z)^{-1} = 1/(1 - \phi z)$:

$$\frac{1}{1-\phi z} = \sum_{n=0}^{\infty} \phi^n z^n \qquad \text{si } |\lambda| > 1 \ (|\phi| < 1) \text{ à l'intérieur du disque } \{z \in \mathbb{C} : |z| < |\lambda|\}, \text{ mais}$$

$$\frac{1}{1-\phi z} = -\sum_{n=\infty}^{-1} \phi^n z^n \qquad \text{si } |\lambda| < 1 \ (|\phi| > 1) \text{ à l'extérieur du disque } \{z \in \mathbb{C} : |z| > |\lambda|\}.$$

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

Théorème 4.4. — *Un processus* AR(p) *est causal,* i.e. *il peut être représenté sous la forme* $Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$, *où* $\sum_{i=0}^{\infty} |\psi_i| < \infty$, *si et seulement si toutes les racines de son polynôme caractéristique* $\phi(z)$ *sont à l'extérieur du disque unité. Les coefficients* $(\psi_i)_i$ *sont dans ce cas les coefficients de la série Taylor de* $\pi(z) = 1/\phi(z)$.

(*) *Notes et commentaires.* — Évidemment, si $\sum_{i=0}^{\infty} |\psi_i| < \infty$, alors $\sum_{i=0}^{\infty} |\psi_i|^2 < \infty$. $(_*)$

2° *Inversibilité des processus* MA(q). — Définition 4.13. — Une représentation causale

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

d'un processus stationaire $(Y_t)_t$ est dite inversible si on peut aussi représenter le bruit par une représentation causale :

$$\varepsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i} \tag{28}$$

où $\sum_{i=0}^{\infty} |\pi_i|^2 < \infty$.

Exemple 4.2. — Le processus MA(1) défini par $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ est inversible si et seulement si $|\theta| < 1$. En effet, comme dans la résolution de la récurrence AR(1), on voit que :

$$\varepsilon_t = Y_t - \theta Y_{t-1} + \cdots + (-\theta)^{t-1} Y_1 + (-\theta)^t \varepsilon_0.$$

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

(*) *Notes et commentaires.* — La rédaction initiale est un peu confuse. L'absence de valeurs absolues par-ci, par-là est notable. À vérifier. (∗)

Théorème 4.5. — *Un processus* MA(q) *pour lequel les racines du polynôme caractéristique* $\theta(z)$ *sont à l'extérieur du disque unité est inversible,* i.e. *le bruit peut être représenté sous la forme :* $\varepsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$, *où* $\sum_{i=1}^{\infty} |\pi_i| < \infty$. *Les coefficients* $(\pi_i)_i$ *sont dans ce cas les coefficients de la série Taylor de* $\pi(z) = 1/\theta(z)$.

(*) *Notes et commentaires.* — Évidemment, si $\sum_{i=0}^{\infty} |\pi_i| < \infty$, alors $\sum_{i=0}^{\infty} |\pi_i|^2 < \infty$. (∗)

Note. — La 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 en z » des suites. Cet isomorhisme explique quelques manipulations formelles avec les filtres (mais pas celles liés a l'inversion).

Remarque 4.1. — Donc, $(\varepsilon_t)_t$ appartient à l'espace linéaire engendré par le passé du signal observé

$$\varepsilon_t \in \text{vect}\{Y_{t-i} : i = 0, 1, \ldots\}$$

et les espaces engendrés par $\{Y_{t-i} : i = 0, 1, \ldots\}$ et $\{\varepsilon_{t-i} : i = 0, 1, \ldots\}$ coïncident.

3° *Causalité et inversibilité des modèles* ARMA(p, q). — Les problèmes de non-causabilité et non-inversibilité des modèles ARMA(p, q) disparaissent quand toutes les racines de $\phi(z)$ et $\theta(z)$ sont à l'extérieur du disque unité :

Théorème 4.6. — (i) *Un processus* ARMA(p, q) *avec toutes les racines du polynôme caractéristique* $\phi(z)$ *sont à l'extérieur du disque unité est causal,* i.e. *il peut être représenté sous la forme :* $Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$, *où* $\sum_{i=0}^{\infty} |\psi_i| < \infty$, *donc* Y_t *appartient à l'espace vectoriel engendré par le passé du bruit*

$$Y_t \in \text{vect}\{\varepsilon_{t-i} : i = 0, 1, \ldots\}.$$

Les coefficients $(\psi_i)_i$ *sont dans ce cas les coefficients de la série Taylor de* $\psi(z) = \theta(z)/\phi(z)$.

(ii) *Un processus* ARMA(p,q) *avec toutes les racines du polynôme caractéristique* $\theta(z)$ *sont à l'extérieur du disque unité est inversible,* i.e. *le bruit peut être représenté sous la forme :* $\varepsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$*, où* $\sum_{i=0}^{\infty} |\pi_i| < \infty$*, donc* ε_t *appartient à l'espace vectoriel engendré par le passé du signal observé*

$$\varepsilon_t \in \text{vect}\{Y_{t-i} : i = 0, 1, \ldots\}.$$

Les coefficients $(\pi_i)_i$ *sont dans ce cas les coefficients de la série Taylor de* $\pi(z) = \phi(z)/\theta(z)$.

(*) *Notes et commentaires.* — Évidemment, si $\sum_{i=0}^{\infty} |\psi_i| < \infty$, alors $\sum_{i=0}^{\infty} |\psi_i|^2 < \infty$. De même, si $\sum_{i=0}^{\infty} |\pi_i| < \infty$, alors $\sum_{i=0}^{\infty} |\pi_i|^2 < \infty$. Je crois qu'il n'est plus la peine de faire ce type de remarques pour la suite. $(_*)$

Corollaire 4.1. — *Pour un processus* ARMA(p,q) *avec toutes les racines des polynômes caractéristiques* $\phi(z)$ *et* $\theta(z)$ *à l'extérieur du disque unité, les espaces vectoriels engendrés par le bruit et le passé du signal coïncident :*

$$\text{vect}\{Y_{t-i} : i = 0, 1, \ldots\} = \text{vect}\{\varepsilon_{t-i} : i = 0, 1, \ldots\}$$

et

$$\mathbb{E}[Y_t \varepsilon_{t+k}] = 0, \qquad \forall\, k \geqslant 1.$$

Remarque. — Ce corollaire permettra un développement immédiat d'une approche de prévision (*i.e.* régression) par projection dans l'espace de Hilbert engendré par le passé.

En conclusion, comme du point de vue pratique les développements de Laurent sont inacceptables pour la prévision (parce qu'ils impliquent les valeurs futures, imprévisibles du bruit), nous allons considérer désormais surtout les modèles ARMA(p,q) avec toutes les racines de $\phi(z)$ et $\theta(z)$ à l'extérieur du disque unité, qui sont causaux et inversibles, et on s'appuyera sur le corollaire ci-dessus.

Exercice 4.5. — Soit $(Y_t)_t$ un processus ARMA$(1,1)$ vérifiant l'équation

$$Y_t - 0.5 Y_{t-1} = \varepsilon_t + 0.4 \varepsilon_{t-1}$$

avec $(\varepsilon_t)_t$ un bruit blanc.

(i) Préciser si le processus est stationnaire, causal et inversible, et calculer sa fonction d'autocovariance.

(ii) Trouver les coefficients $(\psi_j)_j$ de sa représentation comme processus MA(1) et les coefficients $(\pi_j)_j$ de sa représentation comme processus AR(1) et préciser si ces représentations sont convergentes.

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

$$Y_t - 0.7 Y_{t-1} + 0.1 Y_{t-2} = \varepsilon_t + 2\varepsilon_{t-1}$$

Remarque. — Dans le cas le plus simple où les racines $(\lambda_i)_{i=1}^p$ de l'équation $\phi(z) = 0$ sont distinctes, on obtient facilement des formules générales pour les coefficients en commençant par un développement en fractions simples

$$\pi(z) = \phi(z)/\theta(z) = \sum_{i=1}^{p} \frac{K_i}{1 - z/\lambda_i}$$

où les λ_i sont les racines du polynôme $\theta(z)$ et donc $K_i = -\theta(\lambda_i)/\phi'(\lambda_i)$. On arrive à

$$\psi_n = \sum_{i=1}^{p} \frac{K_i}{\lambda_i^{n+1}}.$$

Des formules similaires existent pour $(\pi_n)_n$, et dans le cas des racines non distinctes.

Dans le cas des racines non distinctes et complexes, il est préferable d'aborder le développement de Taylor $\theta(z)/\phi(z) = \psi(z)$ directement, en obtenant des équations de récurrence pour ψ_k, à partir des coefficients du développement $\phi(z)\psi(z) = \theta(z)$. Cette méthode, est en effet toujours applicable, comme on le voit avec l'exercice qui suit.

Exercice 4.6 ARMA$(2,1)$. — (i) Trouver par la méthode directe la représentation MA(∞) d'un processus ARMA$(1,1)$ causal $Y_t = \phi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$.

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

Cet exercice peut être generalisé :

Théorème 4.7 (*). — (i) *Pour un processus* ARMA(p,q) *vérifiant* $\phi(B)Y_t = \theta(B)\varepsilon_t$ *tel que toutes les racines du polynôme caractéristique* $\phi(z)$ *sont à l'extérieur du disque unité, les coefficients* $\psi_i = \sigma^{-2}\,\mathbb{E}[Y_t\varepsilon_{t-i}]$ *de la représentation causale* $Y_t = \sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i}$ *satisfont la relation de récurrence*

$$\psi_0 = 1, \qquad \psi_k = \theta_k + \sum_{i=1}^{\min(k,p)} \phi_i\psi_{k-i}, \quad 1 \leqslant k \leqslant q, \qquad \psi_k = \sum_{i=1}^{\min(k,p)} \phi_i\psi_{k-i}, \quad k > q.$$

Rappel. — La dernière équation est appelée récurrence de Yule–Walker.

Notes. — *a*) Remarquer 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 $\psi(B)\phi(B) = \theta(B)$ pour $k \leqslant q$ changent par l'ajout de θ_k du coté droit.

b) Pour un processus ARMA(p,q), $\phi(B)Y_t = \theta(B)\varepsilon_t$ tel que toutes les racines du polynôme caractéristique $\theta(z)$ sont à l'extérieur du disque unité, les coefficients $(\pi_i)_i$ de la représentation inverse $\varepsilon_t = \sum_{i=1}^{\infty}\pi_i Y_{t-i}$ satisfont la relation de récurrence

$$\pi_0 = 1, \qquad \pi_k = -\phi_k + \sum_{i=1}^{\min(k,q)} \theta_i\pi_{k-i}, \quad 1 \leqslant k \leqslant p, \qquad \pi_k = \sum_{i=1}^{\min(k,q)} \theta_i\pi_{k-i}, \quad k > p.$$

Exemple 4.3. — Ainsi, $\psi_1 = \phi_1 + \theta_1$, $\psi_2 = \phi_2 + \phi_1^2 + \theta_1\phi_1 + \theta_2$, $\psi_3 = \phi_3 + 2\phi_1\phi_2 + \phi_1^3 + (\phi_2 + \phi_1^2)\theta_1 + \phi_1\theta_2 + \theta_3$, ...

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

4.5. Exercices (voir travaux dirigés n° 2)

4.6. La positivité : caractérisation des suites de covariance (∗)

Les suites de nombres qui peuvent être des suites de covariances sont uniquement caracterisées par leurs transformées Fourier.

Théorème 4.8 (Bochner). — *Une suite paire* $(\gamma_k)_{k\in\mathbb{Z}}$ *dans* $\ell^2(\mathbb{Z})$ *peut représenter la suite des covariances d'une série stationnaire si et seulement si la transformé Fourier (appelée aussi densité spectrale)*

$$f(\omega) = \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k\cos(\omega k)$$

est non négative pour chaque $\omega \in \mathbb{R}$.

Il y a une caractérisation équivalente en termes de matrices de covariance. Soit X un processus p-dimensionnel, stationnaire à l'ordre 2, supposé centré. On note γ la fonction d'auto-covariance du processus X. Les propriétés suivantes sont satisfaites :

— $\gamma(k) = \mathbb{E}[X_n {}^t\bar{X}_t]$ est une matrice carrée d'ordre p ;

— $\gamma(k) = {}^t\bar{\gamma}(-k)$, en particulier, la matrice de variance-covariance du processus X est une matrice hermitienne (symétrique dans le cas réel) puisque $\gamma(0) = {}^t\bar{\gamma}(0)$;

— dans le cas $p = 1$, $|\gamma(k)| \leqslant \gamma(0)$;

— $(\gamma(k))_{k\in\mathbb{Z}}$ est une famille de type positif, c'est-à-dire que pour tous $A_1 \ldots, A_k$ dans $\mathbb{C}^p$ et tous $n_1, \ldots, n_k$,

$$\sum_{i=1}^{k}\sum_{j=1}^{k} {}^tA_i\gamma(n_i - n_j)\bar{A}_j \geqslant 0.$$

Démonstration. — Soit $W = \sum_{i=1}^{k} O^tA_iX_{n_i}$. On a alors

$$\begin{aligned}
\mathrm{Var}(W) &= \mathbb{E}\left[\sum_{i=1}^{k} {}^tA_iX_{n_i}\overline{\sum_{j=1}^{k} {}^tA_jX_{n_j}}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} {}^tA_iX_{n_i}\overline{{}^tX_{n_j}A_j}\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} {}^tA_i\gamma(n_i - n_j)X_{n_i}\bar{A}_j \geqslant 0. \qquad \square
\end{aligned}$$

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

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

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

$${}^taC_n\bar{a} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\gamma(i-j)\bar{a}_j \geqslant 0.$$

Au lieu de considérer la fonction d'auto-covariance, on peut choisir d'utiliser la fonction d'auto-corrélation

$$\varrho(k) = (\varrho_{i,j}(k))_{i,j} = \left(\frac{\gamma_{i,j}(k)}{\sqrt{\gamma_{i,i}(0)\gamma_{j,j}(0)}}\right)_{i,j}.$$

Cela revient à considérer non pas le processus X mais un processus $Y = (Y_n)_n$ 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 $\widehat{X}_t(k)$ 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. $\widehat{X}_t(k) = \widehat{X}(t + k \,|\, t) = \mathbb{E}[X_{t+k} \,|\, \mathcal{F}_t]$ (toutes ces notations sont rencontrées dans la littérature) sont les prévisions k pas en avant ;
5. et $e_t(k) = X_{t+k} - \widehat{X}_t(k)$ sont les erreurs de prévision.

Comme les processus ARIMA(p, d, q) satisfont des contraintes linéaires, il est naturel de chercher une prévision linéaire $\widehat{X}_t(k)$ par une combinaison linéaire de valeurs passées ou du bruit blanc, à variance minimale, c'est-à-dire,

$$\widehat{X}_t(k) = \sum_{i=0}^{\infty} \pi_k(i) X_{t-i} \qquad \text{ou} \qquad X_t(k) = \sum_{i=0}^{\infty} a_k(i) \varepsilon_{t-i}, \quad k = 1, 2, \ldots$$

Toutes les trois représentations AR(∞), MA(∞) et ARMA(p, q) nous aideront dans la prévision, notamment la première. Par exemple, la représentation auto-régressive AR(∞)

$$X_{t+1} = \sum_{i=0}^{\infty} \pi_i X_{t-i} + \varepsilon_{t+1}$$

nous fournit directement une formule explicite pour la prévision un pas en avant :

$$\widehat{X}_t(1) = \widehat{X}(t + 1 \,|\, t) = \sum_{i=0}^{\infty} \pi_i X_{t-i} = \big(1 - \pi(B)\big) X_t$$

(*voir* note en bas de page[3]). La prédiction un pas en avant est donc très aisée : on applique simplement la relation d'auto-régression, « en effaçant le bruit » ! La représentation MA(∞) est aussi utile, à cause du théorème suivant :

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

$$\widehat{X}_t(k) = \mathbb{E}[X_{t+k} \,|\, \mathcal{F}_t]$$

où $\mathcal{F}_t = \sigma\{X_t, X_{t-1}, \ldots, X_0, X_{-1}, \ldots\}$.

Autrement dit, la prévision linéaire à variance minimale satisfait les équations de projection :

$$\mathbb{E}\big[X_s \big(X_{t+k} - \widehat{X}_t(k)\big)\big] = 0, \qquad \forall\, s \leqslant t.$$

Nous allons pouvoir profiter du fait que l'opérateur d'espérance conditionnelle $\widehat{X}_t(k)$ est linéaire.

Note. — La différence $Y_{t+1} - \widehat{Y}_t(1)$ a covariance 0 avec les variables déjà observées Y_t, Y_{t-1}, Y_{t-2}, … Un concept peu plus général est celui de de bruit d'innovation.

---

3. ainsi, le problème de prévision 1 pas en avant se réduit formellement a travers la décomposition $X_{t+1} = \widehat{X}(t + 1 \,|\, t) + \varepsilon_{t+1} = \pi(B) X_t + (1 - \pi(B)) X_t$ au calcul des deux filtres $1 - \pi(B)$ et $\pi(B))$; des formules similaires existent pour des écarts plus grands que 1.

Définition 5.1. — Si un processus $(Y_t)_t$ défini par la relation d'auto-régression (18), $Y_t = f(Y_{t-1}, Y_{t-2}, \ldots) + \varepsilon_t$, est tel que le bruit blanc $\varepsilon_t = Y_{t+1} - f(Y_t, Y_{t-1}, Y_{t-2}, \ldots) = Y_t - \widehat{Y}_t$ est indépendant du passé $\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, \ldots\}$, alors le bruit $(\varepsilon_t)_t$ sera appelée bruit d'innovation (par rapport au passé). Dans ce cas,

$$\mathbb{E}[\varepsilon_t \,|\, \mathcal{F}_{t-1}] = 0 \tag{29}$$

et est non correlé avec $Y_{t-1}, Y_{t-2}, \ldots$

Remarque. — Dans le cas des équations auto-régressives causales, le bruit blanc a la propriété très convenable de coïncider avec l'erreur de prédiction un pas en avant $Y_{t+1} - \widehat{Y}_t(1)$.

5.1. La prévision des processus ARIMA$(p, d, 0)$

Nous considérons maintenant plus en détail la prévision des processus auto-régressifs ARIMA$(p, d, 0)$, en utilisant la notation ϕ_i au lieu de π_i, même quand $p = \infty$, et en permettant aussi aux racines du symbole d'être sur le cercle unité (mais pas dans le disque ouvert).

Les démonstrations sont plus aisées dans le cas stationnaire AR(p) causal. La méthode principale est « d'appliquer l'opérateur chapeau » dans l'équation linéaire définissant le modèle, en tenant compte de la représentation causale MA(∞).

Théorème 5.2. — *Pour un modèle* ARIMA$(p, d, 0)$

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

tel que le symbole $\phi(z)$ ne s'annule pas dans l'intérieur du disque unité, le bruit $(\varepsilon_t)_t$ satisfait $\mathbb{E}[\varepsilon_{t+k} \,|\, \mathcal{F}_t] = 0$ si $k > 0$ (étant donc un « bruit d'innovation »). Les prévisions satisfont la relation de récurrence de Yule–Walker :

$$\widehat{X}_t(k) = \mathbb{E}[X_{t+k} \,|\, X_t, X_{t-1}, \ldots] = \sum_{i=1}^{p} \phi_i \widehat{X}_t(k-i)$$

En particulier,

$$\widehat{X}_t = \widehat{X}_t(1) = \mathbb{E}[X_{t+1} \,|\, X_t, X_{t-1}, \ldots] = \sum_{i=1}^{p} \phi_i \widehat{X}_t(1-i) = \sum_{i=1}^{p} \phi_i X_{t-i}$$

Démonstration. — (i) Dans le cas AR(p), on utilise le théorème 5.1. Avec pour écart 1, par exemple, ca revient à vérifier que $\mathbb{E}[X_s \varepsilon_{t+1}] = 0$, pour tous $s \leqslant t$, ce qui est immédiat par la causalité et donc l'indépendance de ε_{t+1} de $\mathcal{F}_t = \sigma\{X_t, X_{t-1}, \ldots\}$.

(ii) Le cas des processus auto-régressifs ARIMA$(p, d, 0)$ d'ordre d fini est obtenu en l'approchant par des processus AR$(p + d)$. □

Exemple 5.1. — La prévision linéaire $\widehat{X}_t(k)$ pour un processus AR(1) a moyenne 0 satisfait la relation de récurrence de Yule–Walker

$$\widehat{X}_t(k) = \phi \widehat{X}_t(k-1)$$

et donc est simplement

$$\widehat{X}_t(k) = X_t \phi^k$$

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

$$\widehat{X}_t(k) - \mu = (X_t - \mu)\phi^k$$

Exemple 5.2. — La prévision linéaire $\widehat{X}_t(k)$ pour le processus ARIMA$(0,1,0)$ à moyenne μ satisfait la relation de récurrence de Yule–Walker

$$\widehat{X}_t(k) = \widehat{X}_t(k-1)$$

et est donc constante

$$\widehat{X}_t(k) = X_t$$

(c'est un cas limite de la formule dans l'exercice antérieur sur la prévision AR(1)).

Théorème 5.3. — *Pour le processus* AR(2)

$$\phi(B)X_t = (1-\lambda_1 B)(1-\lambda_2 B)X_t = \varepsilon_t$$

(λ_1, λ_2 étant les inverses des racines de $\phi(z)$), les prévisions de type Box–Jenkins $X_t(k)$ au temps t satisfont la relation de récurrence :

$$\phi(B)X_t(k) = X_t(k) - \phi_1 X_t(k-1) - \phi_2 X_t(k-2) = X_t(k) - (\lambda_1+\lambda_2)X_t(k-1) + \lambda_1\lambda_2 X_t(k-2) = 0$$

Les prévisions sont de la forme :

$$\widehat{X}_t(k) = A_1(t)\lambda_k 1 + A_2(t)\lambda_k \tag{30}$$

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

$$\widehat{X}_t(k) = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} X_t - \frac{\lambda_1^{k+1}\lambda_2 - \lambda_2^{k+1}\lambda_1}{\lambda_1 - \lambda_2} X_{t-1} \qquad \textit{si } \lambda_1 \neq \lambda_2$$
$$\widehat{X}_t(k) = \lambda^{k+1} X_{t-1} + (k+1)\lambda^k (X_t - \lambda X_{t-1}) \qquad \textit{si } \lambda_1 = \lambda_2 = \lambda.$$

Démonstration. — Les prévisions $\widehat{X}_t(k)$ satisfont l'équation de Yule–Walker

$$\widehat{X}_t(k) = \phi_1 \widehat{X}_t(k-1) + \phi_2 \widehat{X}_t(k-2)$$

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

La solution générale de la récurrence d'ordre deux $\psi(k) = \phi_1\psi(k-1) + \phi_2\psi(k-2)$, avec des valeurs initiales arbitraires $\psi(0) = x_0$, $\psi(-1) = x_{-1}$ est :

1. avec des racines distinctes λ_1, λ_2 :

$$\psi(k) = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} x_0 - \frac{\lambda_1^{k+1}\lambda_2 - \lambda_2^{k+1}\lambda_1}{\lambda_1 - \lambda_2} x_{-1}$$

2. avec des racines confondues $\lambda_1 = \lambda_2 = \lambda$:

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

ce qui donne le résultat. □

Notes. — *a)* Dans le cas causal, $\lambda_i < 1$, et donc la prévision *éventuelle*

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

est toujours 0.

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

Exemple 5.3. — Déduire la formule de prévision de type Box–Jenkins pour un processus ARIMA$(1, 1, 0)$. Calculer la limite $\lim_{k\to\infty} \widehat{X}_t(k)$ pour un processus ARIMA$(1, 1, 0)$.

Exemple 5.4. — On considère un processus $(X_t)_t$ pour lequel la série différenciée deux fois est un bruit blanc, c'est-à-dire que $(X_t)_t$ est un processus ARIMA$(0, 2, 0)$. Montrer que la fonction de prévision de type Box–Jenkins est donnée par

$$\widehat{X}_t(k) = X_t + k(X_t - X_{t-1}), \qquad k - 0,$$

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

Définition 5.2. — Les dernières $p + d$ valeurs $\widehat{X}_t(q)$, $\widehat{X}_t(q-1)$, …, $\widehat{X}_t(q-d-p+1)$ qui précédent $\widehat{X}_t(q)$ (donc avant le point où la relation de récurrence de Yule–Waker devient valable) s'appellent les valeurs pivots.

En conclusion, nous voyons que le « type » de la fonction de prévision $\widehat{X}_t(k)$ dans le cas des processus auto-régressifs (sans partie MA) est déterminé complètement par la fonction $\phi(z)$ et les valeurs pivots. Dans le prochain chapitre nous verrons que ça reste vrai dans un certain sens pour les processus ARIMA(p, d, q).

5.2. Prévision linéaire des modèles ARIMA(p, d, q)

Pour la prévision linéaire $\widehat{X}_t(k)$ des processus ARIMA(p, d, q), nous aurons besoin aussi d'une estimation de ε_{t-1}, ε_{t-2}, …, ε_{t-q}, *i.e.* du « bruit inobservable passé » du modèle, ou au moins de ε_{-1}, ε_{-2}, …, ε_{-q}, qu'on approche parfois par 0. On peut aussi recourir à la représentation auto-régressive ARIMA$(1, d, 0)$, auquel cas on aura besoin de X_{-1}, X_{-2}, …, X_{-q}, qui sont aussi inobservables. Dans tous les cas, le résultat final demandera une approximation des valeurs précédant le début d'observations 0 ; l'approximation la plus simple en l'absence des moyennes est $\varepsilon_k = X_k = 0$ pour $k < 0$.

Exemple 5.5. — Calculer les prévisions $\widehat{X}_t(k)$ pour un processus MA(1) défini par $X_{t+1} = \varepsilon_{t+1} + \theta\varepsilon_t$.

Solution. — Pour le processus MA(1), on vérifie facilement que $\widehat{X}_t(k) = 0$ pour $k = 2, 3, \ldots$, (pour une généralisation, voir le théorème ??). Pour $k = 1$, la définition :

$$X_t + 1 = \varepsilon_t + 1 + \theta\varepsilon_t$$

donne :

$$\widehat{X}_t(1) = \theta\varepsilon_t$$

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

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

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

$$\widehat{X}t(1) = \sum_{i=0}^{\infty} (-1)^i \theta^{i+1} X_{t-i}, \qquad \widehat{X}_t(k) = 0, \quad k = 2, 3, \cdots$$

Si on suppose que « l'information » est finie, *i.e.* $\mathcal{F}_t = \sigma\{X_t; X_{t-1}, \dots, X_1\}$, la formule de prévision, obtenue par récurrence, est :

$$\widehat{X}_t(1) = \mathbb{E}[X_{t+1} \,|\, \mathcal{F}_t] = \sum_{i=0}^{t-1} (-1)^i \theta^{i+1} X_{t-i} - (-\theta)^{t+1} \varepsilon_0.$$

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

Donc, cet exemple montre déjà qu'une estimation du « bruit inobservable » $(\varepsilon_t)_t$ ou au moins de ε_0 est incontournable pour les modèles ARMA avec $q \geqslant 1$.

Théorème 5.4. — *Dans le cas d'un modèle* ARIMA(p, d, q)*, la meilleure « prévision linéaire de Box–Jenkins » au temps* t*,* $\widehat{X}_t(k)$*, satisfait :*

$$\widehat{X}_t(k) = \mathbb{E}[X_{t+k} \,|\, \mathcal{F}_t] = \sum_{i=1}^{p} \tilde{\phi}_i \widehat{X}_t(k-i) + \sum_{i=k}^{q} \theta_i \hat{\varepsilon}_{t+k-i} \tag{33}$$

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

Les inconnues $(\hat{\varepsilon}_{t-i})_{i \geqslant 0}$ *peuvent être enlevées en utilisant la représentation inverse « π » du bruit en fonction de la série, ou en utilisant* $\hat{\varepsilon}_t = Y_t - \widehat{Y}_t - 1(1)$ *(les dernières se calculent récursivement). Une estimation arbitraire de* ε_0*,* ε_{-1}*, …, ou de* X_0*,* $X-1$*, … sera nécessaire.*

Remarque. — Pour $k > q$, la formule (33) est exactement la récurrence homogène de Yule–Walker $\phi(B)\widehat{X}_t(k) = 0$, et donc la prévision sera donnée par la solution de cette équation qui passe par les $p + d$ points pivots $X_{t+q}, \dots, X_{t+q-p-d+1}$, ajustés en tenant compte du bruit $(\varepsilon_t)_t$, comme indiqué en (33).

Par exemple, pour les processus ARIMA$(0, d, q)$ la fonction de prévision $\widehat{X}_t(k)$ est un polynôme d'ordre $d - 1$ en k.

Corollaire 5.1. — *Dans le cas des processus* ARIMA$(p, d, 0)$ *la fonction de prévision* $\widehat{X}_h(k)$ *est la solution de la récurrence* $\phi(B)X_t(k) = 0$ *qui passe par les valeurs pivots* $X_t, \dots, X_{t-p-d+1}$*.*

Exemple 5.6. — Déduire la formule de prévision de type Box–Jenkins pour un processus ARIMA$(1, 2, 0)$. Calculer la limite $\lim_{k\to\infty} \widehat{X}_t(k)$ pour un processus ARIMA$(1, 2, 0)$.

Corollaire 5.2. — *La prévision linéaire* $\widehat{X}_t(k)$ *pour le processus* ARIMA$(0, d, 0)$ *est donnée par le polynôme d'ordre* $d - 1$ *qui passe par les* d *derniers points.*

Exercice 5.1. — On considère le processus ARMA$(1, 1)$ à moyenne 0 $(1 - \phi B)Y_t = (1 + \theta B)\varepsilon_t$ où $-1 < \phi < 1$ et $-1 < \theta < 1$.

(i) Montrer que la fonction de prévision de type Box–Jenkins est donnée par

$$\widehat{Y}_t(k) = \widehat{Y}_t(1)\phi^{k-1}, \qquad k \geqslant 1,$$

et que

$$\widehat{Y}_t(1) = \phi Y_t + \theta\varepsilon_t = (\phi + \theta)Y_t - \theta Y_{t-1}(1) = (\phi + \theta)\big(Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} + \dots\big)$$

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

(ii) On utilise ce modèle pour ajuster une série et on obtient comme estimations des paramètres $\phi = 0.8$, $\theta = 0.3$ 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

Donner les prévisions des trois valeurs suivantes de la série. Quelle parmi les trois formules pour $\widehat{Y}_t(1)$ 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 :

$$(1 - B)Y_t = (1 + \theta B)\varepsilon_t.$$

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

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

(à vérifier).

(i) Montrer qu'on peut le représenter sous la forme

$$Y_t = \varepsilon_t + (1 + \theta) \sum_{k=1}^{t-1} \varepsilon_{t-k} + \theta\varepsilon_0.$$

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

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 ponderée $\bar{Y}_t = \alpha Y_t + (1 - \alpha)\bar{Y}_{t-1}$. Le nombre α s'appelle constante de lissage.

Remarques. — *a*) 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.

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

Exemple 5.7 (*). — La prévision des processus stationnaires ARMA$(p, 1)$ et des processus ARIMA$(0, d, 1)$... Nous développons maintenant plus en détail la prévision des processus ARMA$(p, 1)$ et ARIMA$(0, d, 1)$. On conclut que ...

5.3. Exercices (voir travaux dirigés n° 3)

5.4. Contrôle continu en séries temporelles (voir travaux dirigés)

6. **Résumé des définitions et résultats sur les séries temporelles**

6.1. Filtres

Définition 6.1. — (i) Une série $(s_t)_t$ est dite saisonnière de période p si et seulement si

$$\sum_{i=1}^{p} s_{t+i} = 0 \iff \left(\sum_{i=0}^{p-1} B^i\right) s_t = 0, \qquad \forall\, t \tag{34}$$

(ii) Une série $(s_t)_t$ est dite périodique de période p si et seulement si

$$s_{t+p} = s_t \iff (1 - B^p)s_t = 0, \qquad \forall\, t \tag{35}$$

THÉORÈME 6.1. — *Un filtre $\psi(B)$ annule (ou « enlève ») les composantes saisonnières d'ordre p si et seulement si $\psi(z)$ est divisible par $1+z+\cdots+z^{p-1}$ (donc si $\psi(z)=0$, pour toutes les racine d'ordre p de l'unité, sauf $z=1$).*

THÉORÈME 6.2. — *L'espace invariant d'un filtre contient les polynômes de degré inférieurs ou égaux à p si et seulement si 1 est une racine d'ordre au moins $p+1$ de l'équation $\psi(z)=1$,* i.e. *$\psi(1)=1$, $\psi'(1)=0$, $\psi''(1)=0$, $\psi^{(p)}(1)=0$.*

6.2. Causalité et inversibilité des modèles ARMA(p,q)

THÉORÈME 6.3. — *Un processus* ARMA(p) *est causal (resp. inversible),* i.e. *il peut être représenté sous la forme : $Y_t=\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i}$ où $\sum_i|\psi_i|^2<\infty$ (resp. $\varepsilon_t=\sum_{i=0}^{\infty}\pi_iY_{t-i}$ où $\sum_i|\pi_i|^2<\infty$) si et seulement si toutes les racines de son polynôme caractéristique $\phi(z)$ (resp. de $\theta(z)$) sont à l'extérieur du disque unité. Les coefficients $(\psi_i)_i$ (resp. $(\pi_i)_i$) sont dans ce cas les coefficients de la série Taylor de $\psi(z)=\theta(z)/\phi(z)$ (resp. de $\pi(z)=\phi(z)/\theta(z)$).*

THÉORÈME 6.4 (*). — (i) *Pour un processus* ARMA(p,q), *$\phi(B)Y_t=\theta(B)\varepsilon_t$, avec toutes les racines du polynôme caractéristique $\phi(z)$ à l'extérieur du disque unité, les coefficients $\psi_i=\mathbb{E}[Y_t\varepsilon_{t-i}]$ de la représentation causale $Y_t=\sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i}$ satisfont la récurrence*

$$\psi_0=1,\qquad \psi_k=\theta_k+\sum_{i=1}^{\min(k,p)}\phi_i\psi_{k-i},\quad 1\leqslant k\leqslant q,\qquad \psi_k=\sum_{i=1}^{\min(k,p)}\phi_i\psi_{k-i},\quad k>q.$$

(Cette dernière équation est appelée récurrence de Yule–Walker).

(ii) *Pour un processus* ARMA(p,q), *$\phi(B)Y_t=\theta(B)\varepsilon_t$, avec toutes les racines du polynôme caractéristique $\theta(z)$ à l'extérieur du disque unité, les coefficients $(\pi_i)_i$ de la représentation inverse $\varepsilon_t=\sum_{i=0}^{\infty}\pi_iY_{t-i}$ satisfont la récurrence*

$$\pi_0=1,\qquad \pi_k=-\phi_k+\sum_{i=1}^{\min(k,q)}\theta_i\pi_{k-i},\quad 1\leqslant k\leqslant p,\qquad \pi_k=\sum_{i=1}^{\min(k,q)}\theta_i\pi_{k-i},\quad k>p.$$

Exemple 6.1. — Ainsi, $\psi_1=\phi_1+\theta_1$, $\psi_2=\phi_2+\phi_1^2+\theta_1\phi_1+\theta_2$, $\psi_3=\phi_3+2\phi_1\phi_2+\phi_1^3+(\phi_2+\phi_1^2)\theta_1+\phi_1\theta_2+\theta_3$, ...

6.3. Prévision linéaire

$$\widehat{Y}_t(k)=\mathbb{E}[Y_{t+k}\,|\,\mathcal{F}_t]=\sum_{i=1}^{p}\hat{\phi}_i\widehat{Y}_t(k-i)+\sum_{i=k}^{q}\theta_i\hat{\varepsilon}_{t+k-i}.$$

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

Par exemple, pour le processus AR(2)

$$\phi(B)X_t=(1-\lambda_1B)(1-\lambda_2B)X_t=\varepsilon_t$$

(avec λ_1, λ_2 étant les inverses des racines de $\phi(z)$), les prévisions de type Box–Jenkins $\widehat{X}_t(k)$ au temps t satisfont la relation de récurrence :

$$\begin{aligned}\phi(B)\widehat{X}_t(k)&=\widehat{X}_t(k)-\phi_1\widehat{X}_t(k-1)-\phi_2\widehat{X}_t(k-2)\\&=\widehat{X}_t(k)-(\lambda_1+\lambda_2)\widehat{X}_t(k-1)+\lambda_1\lambda_2\widehat{X}_t(k-2)=0.\end{aligned}$$

La solution générale de la récurrence d'ordre deux $\psi(k) = \phi_1(k-1) + \phi_2(k-2)$, pour valeurs initiales arbitraires $\psi_0 = x_0$, $\psi_{-1} = x_{-1}$, est :

1. avec des racines distinctes, λ_1, λ_2,

$$\psi(k) = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2}\, x_0 - \frac{\lambda_1^{k+1}\lambda_2 - \lambda_2^{k+1}\lambda_1}{\lambda_1 - \lambda_2}\, x_{-1}\,;$$

2. avec des racines confondues, $\lambda_1 = \lambda_2 = \lambda$,

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

Des lors, les prévisions sont de la forme :

$$\widehat{X}t(k) = A_1(t)\lambda_1^k 1 + A_2(t)\lambda_2^k$$

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

$$\widehat{X}_t(k) = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} X_t - \frac{\lambda_1^{k+1}\lambda_2 - \lambda_2^{k+1}\lambda_1}{\lambda_1 - \lambda_2} X_{t-1}.$$

6.4. Le système Yule–Walker pour les corrélations

1. Les premières p corrélations $\varrho = (\varrho(1), \ldots, \varrho(p))$ peuvent être obtenues du système réduit à p équations et p inconnues :

$$R\phi = \varrho \tag{36}$$

ou R est la matrice Toeplitz symétrique :

$$R = \begin{pmatrix} 1 & \varrho(1) & \varrho(2) & \cdots & \varrho(p-1) \\ \varrho(1) & 1 & \varrho(1) & \cdots & \varrho(p-2) \\ \varrho(2) & \varrho(1) & 1 & \cdots & \varrho(p-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \varrho(p-1) & \varrho(p-2) & \varrho(p-3) & \cdots & 1 \end{pmatrix}.$$

2. Ensuite, pour $k > p$ on utilise la reccurence : $\varrho_k = \sum_{i=1}^{p} \phi_i \varrho_{k-i}$.

3. La variance est $\gamma_0 = \sigma^2/(1 - \sum_i \phi_i \varrho_i)$ (et ensuite, on obtient les covariances par $\gamma_k = \varrho_k \gamma_0$, $k > 1$).

*Exemple (*AR(2)*).* — Pour le processus AR(2), $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$, avec racines λ_1, λ_2 de $0 = \lambda^2 - \phi_1\lambda - \phi_2 = \lambda^2\phi(\lambda^{-1})$ à l'intérieur du disque unité (pour assurer la causalité), on obtient, en résolvant (36) :

$$\varrho_1 = \frac{\phi_1}{1-\phi_2} = \frac{\lambda_1 + \lambda_2}{1 + \lambda_1\lambda_2}, \qquad \varrho_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_2} = \cdots$$

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

$$\varrho_k = \varrho_1 \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} - \lambda_1\lambda_2 \frac{\lambda_1^{k-1} - \lambda_2^{k-1}}{\lambda_1 - \lambda_2}\,.$$

Finalement, en utilisant $\phi_1 = \lambda_1 + \lambda_2$, $\phi_2 = -\lambda_1\lambda_2$, on arrive à

$$\varrho_k = \frac{(1-\lambda_2^2)\lambda_1^{k+1} - (1-\lambda_1^2)\lambda_2^{k+1}}{(\lambda_1 - \lambda_2)(1 + \lambda_1\lambda_2)} = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{(\lambda_1 - \lambda_2)(1 + \lambda_1\lambda_2)} - \lambda_1^2\lambda_2^2 \frac{\lambda_1^{k-1} - \lambda_2^{k-1}}{(\lambda_1 - \lambda_2)(1 + \lambda_1\lambda_2)},$$

$k \geqslant 0$, en termes de racines seulement.

b) Dans le cas de racines confondues $\lambda_1 = \lambda_2 = \lambda$, on obtient :

$$\varrho_k = \left(1 + \frac{1-\lambda^2}{1+\lambda^2}\, k\right)\lambda^k, \qquad k \geqslant 0.$$

Table des matières

1. Introduction . . . 1
2. Premier abord des séries temporelles/chronologiques . . . 1
   2.1. *Les composantes d'une chronique* . . . 4
   2.2. *Quelques types de décomposition* . . . 5
      1° Le modèle additif . . . 5
      2° Le modèle multiplicatif (*) . . . 5
      3° Les modèles mixtes (*) . . . 6
3. Filtres/moyennes mobiles . . . 6
   3.1. *Fonctions génératrices* . . . 8
   3.2. *Filtres de lissage* . . . 10
   3.3. *Filtres qui enlèvent les composantes saisonnières* . . . 11
   3.4. *Exercices (*voir *travaux dirigés n° 1)* . . . 12
4. Modélisation stochastique des séries temporelles . . . 12
   4.1. *Introduction* . . . 12
   4.2. *Processus stochastiques stationnaires* . . . 14
   4.3. *Exemples des processus stationnaires* . . . 15
      1° Le bruit blanc . . . 15
      2° Les processus linéaires et les moyennes mobiles MA(q) . . . 16
      3° Les modèles auto-régressifs AR(p) . . . 17
      4° Les modèles ARMA(p, q) . . . 20
      5° Les modèles ARIMA(p, d, q) . . . 20
      6° Les modèles SARIMA(p, d, D, q) . . . 21
   4.4. *L'inversion des séries des puissances et des filtres $\phi(B)$ (*)* . . . 21
      1° Causalité des modèles AR(p) . . . 22
      2° Inversibilité des processus MA(q) . . . 23
      3° Causalité et inversibilité des modèles ARMA(p, q) . . . 23
   4.5. *Exercices (*voir *travaux dirigés n° 2)* . . . 25
   4.6. *La positivité : caractérisation des suites de covariance* $(*)$ . . . 25
5. La prévision linéaire . . . 27
   5.1. *La prévision des processus* ARIMA($p, d, 0$) . . . 28
   5.2. *Prévision linéaire des modèles* ARIMA(p, d, q) . . . 30
   5.3. *Exercices (*voir *travaux dirigés n° 3)* . . . 32
   5.4. *Contrôle continu en séries temporelles (*voir *travaux dirigés)* . . . 32
6. Résumé des définitions et résultats sur les séries temporelles . . . 32
   6.1. *Filtres* . . . 32
   6.2. *Causalité et inversibilité des modèles* ARMA(p, q) . . . 33
   6.3. *Prévision linéaire* . . . 33
   6.4. *Le système Yule–Walker pour les corrélations* . . . 34