

Series temporelles, Régression, Interpolation, et Géostatistique

Florin Avram

Objectif : La régression et l'interpolation déterministe sont parmi les méthodes les plus importantes en statistique et dans les mathématiques appliquées.

Leur but est d'estimer la valeur d'un "signal" $g(x)$ en un point x quelconque, en connaissant des observations Y_i bruitées du signal, observées dans un ensemble fini des points x_i :

$$Y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, I.$$

On veut à la fois : a) diminuer le bruit et b) "extrapoler" du connu à l'inconnu.

Domaines d'application :

- Estimation des densités (il suffit d'appliquer le modèle ci dessus aux observations d'une fonction de survie empirique, en ajoutant la contrainte que $g(x)$ est non-croissante).
- Séries temporelles, appliquées en économétrie, économie, finances, météo, médecine
- Traitement du signal
- Statistiques spatiales appliquées à la prospection et exploitation pétrolière et minière
- Imagerie médicale
- Sciences environnementales : océanographie, météorologie, géologie, ...

Références :

1. Notes de cours/TD distribuées en classe
2. Nonparametric regression by P.J. Green and B.W. Silverman
3. Spline Models for Observational Data, by Wahba, G., (1990), SIAM.
4. Chiles and Delfiner, 1999. Geostatistics, Modeling spatial uncertainty. Wiley series on probability and statistics.
5. Notes WEB de Diggle et Ribeiro sur geoR, Nychka sur fields.R et notes sur R : <http://pbil.univ-lyon1.fr/R/tdr.html>
6. Notes WEB de X. Guyon http://www.math.u-psud.fr/stafav/IMG/pdf/Statistique_Spatiale.pdf
7. Notes de M. Gumpertz : <http://www.stat.ncsu.edu/st733/info/oldindex.html> (avec des TP SAS), et notes de R.H. Smith, Gneiting, Besag et Wackernagel.
8. Reproducing Kernel Hilbert Spaces in Probability and Statistics, by Alain Berlinet, Christine Thomas-Agnan.
9. Hierarchical modeling by Bradley Carlin, S. Banerjee et A.E. Gelfand.
10. Philipp Pluch : Some Theory for the Analysis of Random Fields - With Applications to Geostatistics <http://front.math.ucdavis.edu/0701.5323>

Table des matières

1	Introduction	5
2	Modélisation statistique des séries temporelles	6
2.1	Les composantes d'une série temporelle	6
2.2	Quelques types de décomposition	8
2.2.1	Le modèle additif	8
2.2.2	Le modèle multiplicatif	8
2.2.3	Modèles mixtes	8
3	Filtres/moyennes mobiles	10
3.1	Filtres qui enlèvent/laissent invariantes des composantes périodiques, saisonnières et polynômiales	11
3.2	La reduction du bruit des observations	13
3.3	A quoi servent les filtres ?	13
3.4	Récurrances et équations différentielles linéaires	14
3.4.1	L'équation de récurrence linéaire à coefficients constants	14
3.4.2	La méthode des fonctions génératrices(*)	17
3.5	Exercices : TD 1	19
3.6	TP : L'approche de decomposition sur R	20
4	Modélisation stochastique des séries temporelles	29
4.1	Introduction	29
4.2	Processus stochastiques stationnaires	30
4.3	Exemples des processus stationnaires	31
4.3.1	Le bruit blanc	31
4.3.2	Les moyennes mobiles MA(q) : l'exemple le plus simple de processus a composantes dépendantes	32
4.3.3	Les moyennes mobiles MA(∞)	33
4.3.4	Les modèles autorégressifs AR(p) : des processus faciles à prédire	34
4.4	Fonctions de transfert et applications	35
4.4.1	Récurrances infinies et la causalité des modèles AR(p)	36
4.4.2	La récurrence de Yule-Walker pour les coefficients de la fonction de transfert	37
4.5	Équations de Yule-Walker pour les covariances et corrélations d'un processus AR(p) causal	38
4.5.1	Inversibilité des processus MA(q)	41

4.6	L'estimation des modèles AR(p) par le système de Yule-Walker	42
5	Les modèles ARMA(p,q)	45
5.1	Causalité et inversibilité des modèles ARMA(p,q)	45
5.1.1	Équations de Yule-Walker pour les covariances/corrélations des processus ARMA(p,q)	48
5.2	Les modèles ARIMA(p,d,q)	49
5.3	Contrôle continu d'entraînement	50
5.4	Contrôle continu	52
6	Prévision linéaire des modèles autorégressifs ARIMA(p,d,q)	54
6.1	Prévision des processus ARIMA(p,d,0)	54
6.2	Prévision des processus MA(1)	55
6.3	Prévision des processus AR(p) par la récurrence de Yule-Walker	56
6.4	Prévision des processus ARMA(p,1)	58
6.5	Prévision des processus ARIMA(p,d,q)	59
6.6	Exercices	59
6.7	Intervales de confiance pour la prédiction	62
6.8	La prévision linéaire des processus ARMA(p,q)	63
6.9	Exercices	64
7	Sommaire des définitions et résultats dans les séries temporelles	66
7.1	Filtres	66
7.2	Causalité et inversibilité des modèles ARMA(p,q)	66
7.3	Équations Yule Walker	67
7.4	Prévision linéaire	68
8	Arima et Regression en R	69
8.1	Projets	70
9	Examens	72
9.1	Examen 1	72
9.2	Examen 2	74
9.3	Examen 3	78
10	Approximation par splines	80
10.1	Splines cubiques naturelles	83
10.2	Le calcul des splines naturelles d'interpolation exacte (*)	85
10.3	Splines d'ajustement	89
10.4	Calcul des splines naturelles par interpolation de noyaux	90
10.5	Exercices	91
10.6	Interpolation par splines de degré $2m - 1, m \geq 1$	92
10.7	Splines cubiques en Scilab	93
11	Régression et interpolation par noyaux	96

12 L'analyse des données spatiales	101
12.1 Premier aperçu : la visualisation des données	101
12.2 Rudiments de statistique bivariée	101
12.3 La distribution gaussienne/normale multivariée	102
12.4 Les distributions Gaussiennes conditionnelles, la régression linéaire et l'interpolation	103
12.5 Processus/champs gaussiens	105
12.6 Processus Gaussiens stationnaires et isotropiques	107
12.7 Le corrélogramme et le variogramme spatiaux empiriques	107
12.8 Le corrélogramme et variogramme théoriques	109
12.9 Les noyaux positifs (*)	111
12.10 Exercices	112
13 Exemples de krigeage	113
13.1 Le krigeage simple	113
13.2 Krigeage de la moyenne	114
13.3 Le krigeage ordinaire	115
13.4 Resultat auxiliaire : l'inverse d'une matrice partitionnée en blocs	116
13.5 Exercices	118
14 Interpolation optimale/régression à tendance paramétrique/krigeage avec dérive	121
14.1 L'interpolation par noyaux de base à reproduction polynomiale	121
14.2 Le système d'interpolation/krigeage dual	122
14.3 Interpolation Hermite-Birkhoff	126
14.4 Exercices : Krigeage et interpolation	127
15 Examens d'entraînement	129
15.1 Examen 1	129
16 Rudiments de R	132
16.1 Le Variogramme	134
16.2 variog	137
16.3 trend.spatial	139
16.4 likfit	139
16.5 cov.spatial	139
16.6 krige.conv	139
16.7 image.kriging	140
16.8 Données de type "geodata" disponibles	140
16.9 Projet : Krigeage d'un champs deterministe, avec symmetrie radiale par rapport a un certain centre	140

Chapitre 1

Introduction

Nous allons étudier la prédiction des champs des variables aléatoires $Y(x)$, où $x \in \mathbb{R}^d$ dans la géostatistique, et $x \in \mathbb{N}, \mathbb{Z}$ ou \mathbb{R}_+ pour les séries temporelles/chronologiques. Le cas le plus intéressant est quand $x \in \mathbb{R}^d \times \mathbb{R}_+$, où une des dimensions de x représente le temps (et les autres spatiales), dans quel cas on parle d'interpolation/régression **spatio-temporelle**.

En utilisant des valeurs observées

$$Y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, I \quad (1.1)$$

et leurs coordonnées x_i , $i = 1, \dots, I$, on se propose d'estimer la fonction $g(x)$ et le champ $Y(x)$ dans un point x arbitraire. La fonction "signal" $g(x)$ devrait appartenir à un espace vectoriel de dimension finie, et la première question est comment choisir cette espace. Pour un des choix favoris pour modéliser le signal $g(x)$, les splines, voir Chapitre 10. Nous allons aborder notre sujet dans le cas le plus simple des observations unidimensionnelles, où on parle des séries temporelles.

Définition 1.0.1 *Une série temporelle (ou chronologique) est une succession des valeurs $(Y_1, Y_2, \dots, Y_n, \dots)$ observées aux moments $(T_1 < T_2 < \dots < T_n, \dots)$. On a donc une suite des paires $(T_1, Y_1), (T_2, Y_2), \dots, (T_n, Y_n), \dots$*

Par rapport aux autres types de données statistiques, la particularité des séries chronologiques tient à la présence d'une relation d'antériorité entre les temps, qui ordonne l'ensemble des informations. Les dates d'observations sont souvent équidistantes les unes des autres : on a des séries mensuelles, trimestrielles, etc, dans quel cas on peut supposer $T_i \in \mathbb{N}$. **Exemples** : a) Nombre des moutons par année en Angleterre, entre 1867 et 2003. b) Nombre de voyageurs par mois (SNCF) entre 1990 et 2003. c) Nombre de voitures vendues par un garage, par trimestre entre 1995 et 1999. d) Taux de mortalité, per âge, entre 55 et 104 (c'est le premier exemple d'utilisation de splines, par Whittaker (1923)).

Chapitre 2

Modélisation statistique des séries temporelles

La modélisation statistique des séries temporelles commence en observant leur graphique et en proposant un modèle pour le signal perçu. Jetons donc un coup d'oeil sur quelques données disponibles en R [voir toutes les données avec `data()`].

```
bj<-BJsales
is.ts(bj)
plot.ts(bj)
```

```
sun<-sunspots
is.ts(sun)
plot.ts(sun)
```

Ces deux séries semblent contenir des composantes périodiques, et la première aussi une tendance.

2.1 Les composantes d'une série temporelle

1. *La tendance* ($f_i, 1 \leq i \leq n$) représente l'évolution à long terme de la grandeur étudiée, et traduit l'aspect général de la série. C'est une fonction monotone, souvent polynomiale.
2. *Les variations saisonnières* ($s_i, 1 \leq i \leq n$) sont liées au rythme imposé par les saisons météorologiques (production agricole, consommation de gaz, ...), ou encore par des activités économiques et sociales (fêtes, vacances, solde, etc). Mathématiquement, il s'agit des fonctions périodiques, c'est-à-dire qu'il existe un entier p , appelé *période*, tel que

$$s_i = s_{i+p}, \forall i \geq 1 \quad (2.1)$$

Cette composante est entièrement déterminée par ses p premières valeurs s_1, s_2, \dots, s_p . 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 :

$$\sum_{j=1}^p s_j = 0. \quad (2.2)$$

Opérateurs de retard et d'avance. Pour manipuler les récurrences comme (2.1) et (2.2), il est convenable d'introduire un **opérateur de retard** B (ou encore de "rétro-décalage") défini sur l'ensemble des suites par

$$BX_t := X_{t-1} \quad \text{donc } B^i X_t = X_{t-i},$$

Alternativement, on peut travailler avec $F = B^{-1}$. En utilisant l'opérateur de retard, les équations (2.1), (2.2) deviennent

$$(I - B^p)\mathbf{s} = 0, (I + B + B^2 + \dots + B^{p-1})\mathbf{s} = 0$$

Définition 2.1.1 a) Une série s_t sera appelée *périodique de période p* ssi

$$s_{t+p} = s_t \iff (1 - B^p)s_t = 0, \quad \forall t \quad (2.3)$$

b) Une série s_t sera appelée *saisonnnière de période p* ssi

$$\sum_{i=1}^p s_{t+i} = 0 \iff \left(\sum_{i=0}^{p-1} B^i\right)s_t = 0, \quad \forall t \quad (2.4)$$

Outrement dit, les opérateurs $1 - B^p, \sum_{i=0}^{p-1} B^i$ élimine les séries périodiques/saisonnnières de degré p .

Remarque 2.1.1 Dans la notation d'opérateurs de retard, on voit immédiatement que la deuxième équation (2.4) implique la première (7.2) (en multipliant par $(I - B)$, et donc les séries saisonnières (périodiques de somme nulle) sont périodiques.

3. *Les fluctuations irrégulières/résidues/bruit* ($e_i, 1 \leq i \leq n$) sont des variations de faible intensité et de courte durée, et de nature aléatoire (ce qui signifie ici, dans un cadre purement descriptif, qu'elles ne sont pas complètement expliquables). En effet, elles ne sont pas clairement apercevables dans les graphiques, à cause de leur faible intensité par rapport aux autres composantes. Elles apparaissent clairement seulement après "l'enlèvement du signal" ; la question qui se posera alors sera : est-ce qu'ils contiennent encore du signal, ou est-ce que c'est vraiment du "bruit" ?
4. *Les variations accidentelles/observations aberrantes* sont des valeurs isolées anormalement élevées ou faibles de courte durée. Ces variations brusques de la série sont généralement explicables (Mai 68, réunification de l'Allemagne, tempête, ...). La plupart du temps, ces accidents sont intégrés dans la série des bruits (les fluctuations irrégulières).
5. *Points de changement* Ce sont des points où la série change complètement d'allure, par exemple de tendance. Ils sont normalement explicables, et imposent une analyse séparée de la série, par morceaux.

Les premières deux composantes constituent le signal, et les deux suivantes constituent le bruit. Dans les analyses à suivre, nous nous concentrerons sur les premières 3 composantes : ($f_i, 1 \leq i \leq n$) la tendance, ($s_j, 1 \leq j \leq p$) les coefficients saisonniers, ($e_i, 1 \leq i \leq n$) les fluctuations irrégulières (intégrant éventuellement des accidents).

2.2 Quelques types de décomposition

Après avoir détecté graphiquement quelles sont les composantes présentes, on propose un modèle de décomposition :

2.2.1 Le modèle additif

$$y_i = f_i + s_i + e_i, \quad 1 \leq i \leq n. \quad (2.5)$$

2.2.2 Le modèle multiplicatif

revient à un modèle additif pour le log des données.

$$\text{Log}(y_i) = f_i + s_i + e_i \iff y_i = e^{f_i} e^{s_i} e^{e_i}, \quad 1 \leq i \leq n. \quad (2.6)$$

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

Remarque 2.2.1 *On choisit ce modèle si la courbe qui unit les maximums des observations s'éloigne ou s'approche de celle qui unit les minimums des observations, i.e. si les amplitudes des fluctuations dépendent du niveau.*

Remarque 2.2.2 *Le modèle multiplicatif est souvent utilisé pour des données de type économique.*

2.2.3 Modèles mixtes

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

$$y_i = f_i(1 + s_i) + e_i = f_i + f_i s_i + e_i, \quad 1 \leq i \leq n. \quad (2.7)$$

Remarque 2.2.3 *Une fois un modèle est obtenu, il peut être utilisé pour la prédiction des valeurs futurs.*

La décomposition peut être accomplie directement avec la fonction `decompose` de R.

```
## example taken from Kendall/Stuart
x <- c(-50, 175, 149, 214, 247, 237, 225, 329, 729, 809, 530, 489, 540, 457, 195, 176)
x <- ts(x, start = c(1951, 1), end = c(1958, 4), frequency = 4)
m <- decompose(x)
plot(m)
```

Toutes les décompositions peuvent être implémentées en effectuant une régression avec `lm` sur des fonctions de base appropriées.

Remarque 2.2.4 Nous venons de voir que les composantes périodiques (2.1) et saisonnières (2.2) satisfont des récurrences à coefficients constants, et par conséquent ont des formules explicites comme sommes des puissances de racines de l'unité. Il est quand même plus intéressant de fournir des bases de décomposition réelles.

- Exercice 2.2.1**
1. Trouvez une base réelle de l'espace vectoriel des séries périodiques réelles d'ordre p .
 2. Donnez une formule explicite décrivant une série arbitraire appartenant à l'espace d'états vectoriel des séries périodiques. Quelle est la dimension de cet espace ?
 3. Même questions pour une série arbitraire appartenant à l'espace d'états vectoriel des séries saisonnières.
 4. Montrez que chaque série périodique p_t est la somme d'une série saisonnière et d'une série constante.
 5. Trouvez une base réelle de l'espace vectoriel des séries saisonnières d'ordre p , et ensuite une base de l'espace des séries périodiques, qui la contient.
 6. Formulez en forme matricielle un problème de régression (moindre carrés) qui permet de décomposer une série comme une somme d'une tendance linéaire, une partie saisonnière de période 4, et des résidus.

Remarque 2.2.5 Après le choix d'une forme appropriée pour la tendance et une période spécifique (selon le graphe), on peut déterminer au même temps les coefficients de la tendance et de la partie périodique par une régression linéaire.

Exercice 2.2.2 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. Comme les données sont mensuelles, la période est de un an, et donc $p = 12$. Le modèle additif est :

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

Les coefficients a, b, s_1, \dots, s_{11} et les résidus peuvent être déterminés en minimisant la somme des carrés des résidus $\sum_i e_i^2$, c.-à-d. par régression. Remarquons que :

- la série (f_i) représente la tendance générale (réchauffement ? cycle ?).
- Des valeurs comme $s_1 = -10$ 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 comme $e_{14} = -2$ signifiera 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).

Chapitre 3

Filtres/moyennes mobiles

Souvent il convient de remplacer l'information locale fournie par un "pixel" d'un champs aleatoire/série par l'information obtenue en ajoutant ses voisins, ce qui suggère de construire des "moyennes mobiles".

Définition 3.0.1 La série Y_t s'appelle une **moyenne mobile** de X_t ou **filtre**[§] si

$$Y_t = \sum_{i=-k_1}^{k_2} \theta_i X_{t-i} \quad (3.1)$$

où $k_1, k_2 \geq 0$. Le **degré** du filtre est $k_1 + k_2$, et son **ordre**/nombre des coefficients est $k_1 + k_2 + 1$.

Le cas des θ_i positifs à somme égale à 1 s'appelle **lissage**, celui avec des θ_i égaux s'appelle **moyenne arithmétique**, et le cas d'une moyenne arithmétique avec $k_1 = k_2 = q$ sera appelé **moyenne arithmétique symétrique ou centré**. Un filtre de lissage est supposé réduire le bruit, mais laisser invariantes des tendances constantes ou polynomiales.

Remarque 3.0.6 La notation des polynômes de retard ramène (3.1) à la forme :

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

où $\theta(B)$ est un **opérateur de récurrence** (sur l'espace des séries). a) Le polynôme de Laurent associé $\theta(z) = \sum_{i=-k_1}^{k_2} \theta_i z^i, : \mathbb{C}^- \rightarrow \mathbb{C}$ sera appelée le symbole de l'opérateur $\theta(B)$. b) En écrivant le filtre comme

$$\theta(z) = z^{-k_1} p(z)$$

ou $p(z)$ est un polynôme on obtient le **polynôme caractéristique**. c) En renversant les coefficients du polynôme caractéristique on obtient le **polynôme dual**.

§. En examinant des graphes des séries temporelles comme ci-dessus, les pères fondateurs du domaine (Wiener, Kolmogorov, ...) ont eu l'intuition que le souci de dégager les composantes périodiques et la tendance était similaire au problème d'enlèvement du bruit dans les appareils de radio, qui avait été résolu par l'utilisation des filtres ("low-pass" et "high-pass"). En effet, les prochains chapitres montrerons les multiples utilisations des filtres/moyennes mobiles pour la prédiction des séries temporelles.

Un filtre de lissage

$$Y_t = \sum_{i=1}^k \theta_i X_{t-i}, \quad \sum_i \theta_i = 1$$

peut être utilisé pour calculer la tendance et pour la prédiction de X_t . Remarquez que le fait que la prédiction est "non-biaisée pour une série stationnaires", c.-à-d. :

$$\mathbb{E}Y_t = \mathbb{E} \sum_{i=1}^k \theta_i X_{t-i} = \left(\sum_{i=1}^k \theta_i \right) \mathbb{E}X_t$$

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

Remarque 3.0.7 Cette condition est aussi équivalente au fait que la série égale à $\mathbf{1}$ (et en fait chaque série constante $X_t = c$) reste invariante après l'application du filtre, c.-à-d. $\theta(B)\mathbf{1} = \mathbf{1}$.

3.1 Filtres qui enlèvent/laissent invariante des composantes périodiques, saisonnières et polynômiales

Lemme 3.1.1 Un filtre $\theta(B)$ laisse invariante une suite x_t ssi le filtre $\hat{\theta}(B) := I - \theta(B)$ annule la suite x_t .

Des lors, les deux problèmes étudiés ici coïncident.

Exercice 3.1.1 a) Montrez qu'une moyenne arithmétique symétrique d'ordre $2q + 1 = 3$, donné par

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

conserve (laisse invariante) les tendances linéaires $p_t = a + bt$, et élimine les séries saisonnières de période 3. b) Généraliser pour q quelconque. c) Étudier la première propriété pour une moyenne symétrique d'ordre 3 arbitraire.

Exercice 3.1.2 Moyennes mobiles modifié au bord. Trouver un filtre polynômiale symétrique d'ordre $2k + 1$ $\theta(z) = \theta_0(1 + z^{2k}) + \theta_1(\sum_{i=1}^{k-1} z^i + z^{2k-i})$ qui laisse invariants les polynômes de premier degré et élimine les séries saisonnières de période $2k$.

Exercice 3.1.3 Montrez qu'un filtre $\theta(z)$ qui est divisible par $1 + z + \dots + z^{p-1}$, c.-à-d. de la forme $\theta(z) = (1 + z + \dots + z^{p-1})\theta_1(z)$, élimine les composantes saisonnières de période p , c.-à-d. :

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

pour chaque série s_t satisfaisant (2.4).

En effet, la réciproque est aussi vraie :

Théorème 3.1.1 *Un filtre $\theta(B)$ annule (ou enlève) les composantes saisonnières d'ordre p ssi son symbole $\theta(z)$ est divisible par $1 + z + \dots + z^{p-1}$ (donc si $\theta(z) = 0$, pour toutes les racine d'ordre p de l'unité, sauf $z = 1$).*

Dem : Soit $\theta_1(z)$ le reste de la division de $\theta(z)$ par $1 + z + \dots + z^{p-1}$. Alors, $\theta_1(z)$ annulent l'espace vectoriel de dimension $p - 1$ des séries saisonnières. Comme $\theta_1(z)$ a au plus $p - 1$ coefficients, il doit être identiquement 0.

Remarque 3.1.1 *Pour enlever les composantes saisonnières d'ordre 4, on peut utiliser donc la moyenne mobile arithmétique d'ordre 4, ou encore mieux, une moyenne mobile modifié au bord d'ordre 5, pour une périodicité mensuelle on peut utiliser la moyenne mobile arithmétique d'ordre 12, etc... Ainsi, en éliminant la partie saisonnière, on obtient la tendance.*

Un autre opérateur intéressant est l'opérateur différence qui fait la différence entre le processus et sa version décalée d'une unité de temps :

$$\Delta X_t := X_t - X_{t-1} = (1 - B)X_t$$

Exercice 3.1.4 *Montrer que l'opérateur différence élimine les constantes, et que ses puissances $\Delta^d = (1 - B)^d$ élimine les tendances polynomiales de degré $d - 1$. Réciproquement, un filtre polynomiale $\theta(B)$ élimine les tendances polynomiales de degré $d - 1$ ssi son symbole $\theta(z)$ est divisible par $(1 - z)^d$, c.-à-d. admet 1 comme racine multiple de degré d .*

Remarque 3.1.2 *Si un opérateur de récurrence annule une suite caractérise la suite x_t , alors x_t est une combinaison des puissances des racines du polynôme caractéristique, ce qui rend évidente l'importance d'identifier des tels opérateurs.*

Le théorème suivant nous donne un critère pour identifier les filtres $\theta(B)$ qui laissent invariants tous les polynômes de degré donné.

Théorème 3.1.2 *Soit $\theta(B) = \sum_{i=-k_1}^{k_2} \theta_i B^i = B^{-k_1} \sum_{i=-k_1}^{k_2} \theta_i B^{i+k_1} = B^{-k_1} p(B)$ un filtre (typiquement symétrique), déterminé par k_1 et par le polynôme $p(B)$. Alors, l'espace invariant des séries Z_t satisfaisant $\theta(B)Z_t = Z_t$ contient les polynômes de degré $\leq n$ ssi a) 1 est une racine d'ordre au moins $n + 1$ de l'équation $\theta(z) = 1$, c.-à-d. $\theta(1) = 1, \theta'(1) = 0, \theta''(1) = 0, \theta^{(n)}(1) = 0$, où $\theta(z)$ est le symbole. b) 1 est une racine d'ordre au moins $n + 1$ de l'équation polynomiale $p(z) - z^{k_1} = 0$, c.-à-d. $p(1) = 1, p'(1) = k_1, p''(1) = k_1(k_1 - 1), p^{(n)}(1) = (k_1)_n$, où $p(z)$ est la partie polynomiale du symbole.*

Exercice 3.1.5 *Démontrez le théorème pour $p = 0, 1, \dots$*

Exercice 3.1.6 *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.*

Exercice 3.1.7 La meilleure moyenne mobile modifié au bord. *Trouver un filtre polynomiale symétrique d'ordre $2k + 1$ $\theta(z) = \theta_0(1 + z^{2k}) + \theta_1(\sum_{i=1}^{k-1} z^i + z^{2k-i})$ qui laisse invariants les polynômes de deuxième degré et élimine les séries saisonnières de période $2k$.*

R : $\theta_1 = \frac{3k}{4k^2 - 1}, \theta_0 = -\frac{1+k}{2(2k-1)}.$

3.2 La réduction du bruit des observations

Un résultat désirable de l'application des filtres de lissage : la réduction de la variance des observations.

Exercice 3.2.1 a) Montrez qu'une moyenne arithmétique symétrique d'ordre $2q+1$ diminue la variance σ^2 d'un bruit blanc (=série i.i.d. de moyenne 0) par $2q+1$, et que ce résultat est optimal. b) **La meilleure moyenne mobile arithmétique modifiée au bord ?** Trouver un filtre polynomiale symétrique d'ordre $2k+1$

$$\theta(z) = \theta_0(1 + z^{2k}) + \theta_1 \left(\sum_{i=1}^{k-1} z^i + z^{2k-i} \right)$$

qui laisse invariants les polynômes de premier degré et minimise la variance σ^2 d'un bruit blanc.

Remarque 3.2.1 Si la série observée est de la forme

$$X_t = p_t + \epsilon_t$$

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

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

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

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

3.3 A quoi servent les filtres ?

1. Les **filtres causaux**

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

sont importants dans la prediction

2. Les **filtres symétriques**

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

sont importants dans le lissage (l'enlèvement du bruit). Les **filtres arithmétiques**

$$\theta(B) = (2k+1)^{-1} \sum_{i=-k}^k B^i$$

minimisent en plus la variance (i.e. sont optimales pour enlever le bruit).

Aussi, les filtres fournissent un langage adéquat pour décrire les séries temporelles. Rappelons que l'approche de décomposition additive suggère de rompre une série donnée

$$Y_t = s_t + m_t + \epsilon_t$$

dans un "signal" $s_t + m_t$ et un bruit stationnaire ϵ_t . La partie la plus importante d'une série est bien sur le signal, et son dépistage peut se faire par l'approche classique de régression. Une fois dépisté, la prédiction du signal est évidente (par sa formule!). Par contre, la prédiction du bruit (corrélé) demande des approches plus sophistiquées, comme la modélisation ARMA, où l'utilisation des filtres devient essentielle.

Il est naturel d'essayer de dépister le signal par un filtre de lissage causal $\theta(B)$, donc $m_t = \theta(B)Y_t$, tq ce filtre "détruit le bruit mais laisse passer le signal". Il en suit que le bruit aussi est donné par un filtre causal

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

où $\pi(B)$ "détruit le signal mais laisse passer le bruit". On peut voir donc l'analyse des séries temporelles comme la recherche pour un filtre qui transformera notre série en bruit. Pour cela, il va être important d'appliquer des tests pour décider quand une série est un bruit. Nous allons examiner plus tard des statistiques comme les corrélations, corrélations partielles, le periodogramme, etc., issues de la modélisation probabiliste des séries temporelles.

3.4 Récurrences et équations différentielles linéaires

L'étude des marches aléatoires et des processus Markoviens ramène souvent à des équations différentielles ou des récurrences linéaires. Le cas des coefficients constants est assez simple, car toutes les solutions peuvent être construites à partir des solutions basiques exponentielles e^{rx} . Comme le cas des équations différentielles à coefficients constants est très bien connu, on rappelle ici seulement le cas de récurrences linéaires.

3.4.1 L'équation de récurrence linéaire à coefficients constants

Les deux équations de récurrence linéaire de deuxième ordre ci-dessous

$$au_{n+2} + bu_{n+1} + cu_n = 0, \tag{3.2}$$

$$av_{n+2} + bv_{n+1} + cv_n = d_n, \tag{3.3}$$

sont appelées **homogène** et **nonhomogène** respectivement.

L'équation

Si les coefficients a , b et c sont constants, on sait qu'ils existent des solutions de la forme $u_n = x^n$ pour tout n (fonctions exponentielles). Pour trouver x on remplace x^n en (3.2) et on trouve que x doit satisfaire **l'équation auxiliaire** :

$$ax^2 + bx + c = 0. \tag{3.4}$$

Soient x_1 et x_2 les deux racines de l'équation de deuxième degré (3.4). On en déduit que la solution *générale* de (3.2) est toujours de la forme

1. Si $x_1 \neq x_2$

$$u_n = Ax_1^n + Bx_2^n,$$

2. Si $x_1 = x_2$,

$$u_n = Ax_1^n + Bnx_1^n,$$

avec des constantes A et B .

Dans les deux cas A et B doivent être déterminées à partir des conditions supplémentaires sur la frontière.

L'équation nonhomogène

La résolution du problème nonhomogène (3.3) comporte quatre pas :

1. Trouver une base pour l'espace vectoriel des solutions de l'équation auxiliaire homogène (3.2), et donc la solution générale u_n pour cette équation.
2. Déterminer une solution particulière de (3.3), par exemple en utilisant une expression «essai» \tilde{v}_n qui a la même forme générale que le membre droit d_n , mais des coefficients non déterminés. Par exemple, si d_n est un polynôme d'ordre k , on essaie un polynôme général d'ordre k .
3. Néanmoins, si votre expression d'essai a des termes qui sont inclus dans l'espace vectoriel des solutions de l'équation homogène obtenue au pas 1 (et donc qui vont être annihilés par l'opérateur des différences), il faut multiplier l'expression d'essai par n, n^2, \dots jusqu'à ce qu'il n'y a plus des termes inclus dans cet'espace .
4. Après la décision de la forme d'essai, on trouve les valeurs des coefficients de \tilde{v}_n à partir de (3.3), par la méthode des coefficients non déterminés.
5. La solution générale de (3.3) est de la forme $v_n = \tilde{v}_n + u_n$. On trouve finalement les coefficients encore non déterminés en u_n , en utilisant les conditions sur la frontière pour v_n .

Exemple 3.4.1 On considère l'ensemble E des suites $(u_n)_{n \in \mathbb{N}}$ qui vérifient la relation suivante :

$$(R) \quad \forall n \in \mathbb{N}, \quad u_{n+2} + \frac{3}{2}u_{n+1} - u_n = 0.$$

1. Rechercher les suites géométriques qui vérifient cette relation (R).
2. On note r_1 et r_2 leurs raisons et on admet que E est un espace vectoriel de dimension 2, c.-à-d. toute suite de E s'écrit sous la forme

$$u_n = \alpha r_1^n + \beta r_2^n, \quad \forall n \in \mathbb{N}.$$

Soit (a_n) la suite de E qui vérifie $a_0 = 1$ et $a_1 = 0$. Calculer a_n .

Exemple 3.4.2 On considère l'ensemble E' des suites (v_n) qui vérifient la relation :

$$(R') \quad \forall n \in \mathbb{N}, \quad v_{n+2} + \frac{3}{2}v_{n+1} - v_n = 4n + 1.$$

1. On pose $\forall n \in \mathbb{N}, \quad u_n = an + b$. Déterminer a et b pour que (u_n) soit une solution particulière de (R').
2. Soit (v_n) une suite de E' .

- (a) Pour toute suite (t_n) de E' on définit la suite (u_n) par $\forall n \in \mathbb{N}, u_n = v_n - t_n$.
Vérifier que $(u_n) \in E$.
- (b) En déduire que $\forall n \in \mathbb{N}, v_n = \alpha r_1^n + \beta r_2^n + an + b$.
- (c) Déterminer v_n pour $v_0 = -\frac{5}{9}$ et $v_1 = -\frac{26}{9}$.

Exemple 3.4.3 Obtenez les formules analytiques des suites décrites par les relations de récurrence ci-dessous, et vérifiez-les en utilisant les premiers termes de la suite t_2, t_3 .

1. $t_i = 2t_{i-1} + i - 1, t_0 = 0$
2. $t_i = 2t_{i-1} + 5 \cdot 2^i, t_0 = 0$
3. $t_i = 3t_{i-1} - 2t_{i-2} + 2, t_0 = 0, t_1 = 2$
4. $t_i = 2t_{i-1} - t_{i-2} + 2, t_0 = 0, t_1 = 2$

Solution :

1. C'est une équation nonhomogène, alors nous aurons :

$$t_i = \tilde{t}_i + A2^i, \quad \tilde{t}_i = c_1 i + c_2 \quad \text{avec} \quad c_1 i + c_2 = 2(c_1 i | c_1 + c_2) + i - 1$$

$$\begin{aligned} \text{et alors} \quad c_1 &= 2c_1 + 1 \quad \text{et} \quad c_1 = -1 \\ c_2 &= |2c_1 + 2c_2 - 1 \quad \text{et} \quad c_2 = 2c_1 + 1 = -1 \\ \tilde{t}_i &= |i - 1 \quad \text{Finalement,} \\ t_0 &= = 0 = -1 + A \quad \text{et} \quad A = 1 \\ t_i &= |i - 1 + 2^i \end{aligned}$$

2. C'est une équation nonhomogène, alors :

$$t_i = \tilde{t}_i + A2^i, \quad \tilde{t}_i = ci2^i \quad \text{avec} \quad ci2^i = 2(c|i - 1|2^i/2) + 52^i$$

$$\begin{aligned} \text{et alors} \quad c &= 5, \quad t_i = 5i2^i + A2^i \quad \text{et finalement,} \\ t_0 &= = 0 = A \quad \text{et} \quad A = 0 \\ t_i &= 5i2^i \end{aligned}$$

3. C'est une équation de différences nonhomogène et l'équation quadratique attachée a les racines 1, 2, alors nous aurons :

$$t_i = \tilde{t}_i + A_1 2^i + A_2, \quad \tilde{t}_i = ci \quad \text{avec} \quad ci = 3(ci|c)|2(ci|2c) + 2$$

$$\begin{aligned} \text{et alors} \quad c &= -2 \quad \text{et} \quad c_2 = 2c_1 + 1 = -1 \\ \tilde{t}_i &= |i - 1 \quad \text{Finalement,} \\ t_0 &= = 0 = -1 + A \quad \text{et} \quad A = 1 \\ t_i &= |i - 1 + 2^i \end{aligned}$$

4. C'est une équation de différences nonhomogène dont les racines de l'équation quadratique attachée sont confondues égales à 1 donc nous aurons :

$$t_i = \tilde{t}_i + A_1 + A_2 i, \quad \tilde{t}_i = c_1 i + c_2 \quad \text{avec} \quad c_1 i + c_2 = 2(c_1 i | c_1 + c_2) + i - 1$$

$$\begin{aligned} \text{et alors} \quad c_1 &= 2c_1 + 1 \quad \text{et} \quad c_1 = -1 \\ c_2 &= |2c_1 + 2c_2 - 1 \quad \text{et} \quad c_2 = 2c_1 + 1 = -1 \\ \tilde{t}_i &= |i - 1 \quad \text{Finalement,} \\ t_0 &= = 0 = -1 + A \quad \text{et} \quad A = 1 \\ t_i &= |i - 1 + 2^i \end{aligned}$$

3.4.2 La méthode des fonctions génératrices(*)

Exercice 3.4.1 Calculez, par la méthode des fonctions génératrices : a) $T(z) = \sum_{n \geq 0} T_n z^n$, où

$$T_0 = 0, \quad T_n = 2T_{n-1} + 1, n \geq 1$$

Trouvez T_n . b) Calculez $T(z) = \sum_{n \geq 0} T_n z^n$, où

$$T_0 = 1, \quad T_n = 2T_{n-1} + n - 1, n \geq 1$$

Trouvez T_n .

Sol : b) Par la méthode de décomposition des équations linéaires : $T_n = 2^{n+1} - (n + 1)$. La fonction génératrice de T_n est $T(z) = 2/(1-2z) - 1/(1-z)^2 = (2z^2 - 2z + 1)/(1-2z)(1-z)^2$. En appliquant directement la méthode des fonctions génératrices à $T_n = 2T_{n-1} + n - 1$ on trouve l'équation : $T(z) = 1 + 2zT(z) + z/(1-z)^2 - (1/(1-z) - 1) = 2zT(z) + (2z^2 - 2z + 1)/(1-z)^2$, et on retrouve la même réponse. L'avantage de cette méthode plus compliquée est qu'elle réussit parfois quand la première méthode échoue.

Exercice 3.4.2 a) Obtenez une formule explicite pour la suite décrite par la relation de récurrence ci-dessous (et vérifiez-la en utilisant les premiers termes de la suite)

$$T_n = 4T_{n-1} - 3T_{n-2} + n, n \geq 1, \quad T_0 = a, T_{-1} = 0 \quad (3.5)$$

b) Calculez la fonction génératrice $T(z) = \sum_{n \geq 0} T_n z^n$ c) (*) Obtenez la fonction génératrice $T(z)$ directement, en appliquant la méthode des fonctions génératrices à l'équation (3.5).

Sol : a)

$$\begin{aligned} &(a + 3/4) 3^n - \frac{1}{4} (3 + 2n) \\ &\frac{a + 3/4}{1 - 3z} - \frac{1/4}{1 - z} - \frac{1/2}{(z - 1)^2} = \frac{a(z - 1)^2 + z}{(z - 1)^2(3z - 1)} \end{aligned}$$

Exercice 3.4.3

a) Obtenez une formule explicite pour la suite décrite par la relation de récurrence ci-dessous (et vérifiez-la en utilisant les premiers termes de la suite)

$$T_n = 2T_{n-1} - T_{n-2} + 2, n \geq 2, \quad T_0 = T_1 = 0 \quad (3.6)$$

b) Calculez la fonction génératrice $T(z) = \sum_{n \geq 0} T_n z^n$ c) Obtenez la fonction génératrice $T(z)$ directement, en appliquant la méthode des fonctions génératrices à l'équation (3.6)

Exercice 3.4.4 a) Obtenez la fonction génératrice $T(z) = \sum_{n \geq 0} T_n z^n$ pour la récurrence

$$T_n = \lambda T_{n-1} + \lambda_2 T_{n-2}, n \geq 2, \quad T_0 = 1, T_1 = \lambda T_0 \quad (3.7)$$

b) Calculez (directement, ou par développement limité) les premiers termes, et vérifiez qu'ils sont des combinaisons linéaires de puissances des racines caractéristiques.

Exercice 3.4.5 a) Obtenez une formule explicite pour la suite décrite par la relation de récurrence ci-dessous (et vérifiez-la en utilisant les premiers termes de la suite)

$$T_n = 4T_{n-1} - 3T_{n-2} + 2, n \geq 2, \quad T_0 = T_1 = 0 \quad (3.8)$$

b) Calculez la fonction génératrice $T(z) = \sum_{n \geq 0} T_n z^n$ c) (*) Obtenez la fonction génératrice $T(z)$ directement, en appliquant la méthode des fonctions génératrices à l'équation (3.9).

Solution :

1. a)

$$T_n = \frac{1}{2} (3^n - 1) - n$$

b)

$$T(z) = \frac{1}{2(1-3z)} - \frac{1}{2(1-z)} - \frac{z}{(1-z)^2} = \frac{1}{2(1-3z)} + \frac{1}{2(1-z)} - \frac{1}{(1-z)^2} = \frac{2z^2}{(1-z)^2(1-3z)}$$

La dernière expression peut être obtenue directement, à partir de la récurrence.

Exercice 3.4.6 a) Obtenez une formule explicite pour la suite décrite par la relation de récurrence ci-dessous (et vérifiez-la en utilisant les premiers termes de la suite)

$$T_n = 4T_{n-1} - 3T_{n-2} + n, n \geq 1, \quad T_0 = a, T_{-1} = 0 \quad (3.9)$$

b) Calculez la fonction génératrice $T(z) = \sum_{n \geq 0} T_n z^n$ c) (*) Obtenez la fonction génératrice $T(z)$ directement, en appliquant la méthode des fonctions génératrices à l'équation (3.9).

Solution : a)

$$(a + 3/4) 3^n - \frac{1}{4} (3 + 2n)$$

b)

$$\frac{a + 3/4}{1 - 3z} - \frac{1/4}{1 - z} - \frac{1/2}{(z - 1)^2} = \frac{a(z - 1)^2 + z}{(z - 1)^2(3z - 1)}$$

3.5 Exercices : TD 1

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

$$2, 6, 12, 20, 30, 42, \dots \quad (3.10)$$

$$4, 10, 20, 36, 62, 104, \dots \quad (3.11)$$

$$3, 2, 1, 6, 3, 2, 1, \dots \quad (3.12)$$

$$0, -1, -2, 3, 0, -1, -2, \dots \quad (3.13)$$

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

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

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

Obtenez une formule analytique pour y_t , en utilisant : a) la théorie des équations de récurrence à coefficients constants. R : $n^2 + n + 1$ b) la méthode des fonctions génératrices, décomposition en fractions partielles et l'expansion en série des puissances :

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

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

3. Une série vérifie la récurrence $y_t - 2y_{t-1} + y_{t-2} = (-1)^{t-1}, t \geq 2, y_0 = 1, y_1 = 3$
 Obtenez une formule analytique pour y_t , a) en éliminant le terme inhomogène **Sol** :
 On arrive à $(B+1)(B-1)^2 y_t = 0$ La solution particulière doit être de la forme $y_t = A(-1)^{t-1}$. Comme $(B-1)^2(-1)^{t-1} = -4$, il suit que $A = \frac{-1}{4}$.
 b) (*) en utilisant la méthode des fonctions génératrices, décomposition en fractions partielles, et l'expansion en série des puissances :

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

Sol : $a(z) = \frac{1+2z}{(1+z)(1-z)^2} = \frac{3}{2(1-z)^2} - \frac{1/4}{(1-z)} - \frac{1/4}{(1+z)}, a(n) = (5 - (-1)^n + 6n)/4$ c) en utilisant le calcul (formel) opérationnel de différences de Arbogast et Heaviside (formel) http://en.wikipedia.org/wiki/Operational_calculus **Sol** : $y_t = (1-B)^{-2}(-1)^{t-1} = \sum_{n=0}^{\infty} (n+1)B^n(-1)^{t-1}$

4. (a) Montrez que le filtre $P(B) = \frac{1}{3}(2 + B + B^2 - B^3)$ "enlève" les composantes saisonnières de période 3, c.-à-d. qu'il transforme chaque fonction de période 3 dans une fonction constante.
 (b) Trouvez l'ordre de la tendance polynomiale maximale conservée (laissée invariante) par ce filtre.

Sol : Ordre 1.

5. Trouver un filtre $1 + \alpha B + \beta B^2 + \gamma B^3$ qui laisse passer une tendance affine sans distorsion et élimine les périodicités d'ordre 2. Indication : Trouver un système des 2 équations avec deux inconnues, et résoudre.
6. Trouvez un filtre $f(B)$ qui conserve les polynômes de degré ≤ 1 , et qui enlève les composantes saisonnières d'ordre 4, et déduisez que pour une série ayant une composante périodique d'ordre 4 et une tendance linéaire m_t , la tendance est donnée par $m_t = f(B)Y_t$. **Sol** : $\frac{1+B+B^2+B^3}{4} \frac{5-3B}{2}$
7. a) Trouvez les conditions pour qu'un filtre causal $\theta(B) = \sum_{i=0}^q \theta_i B^i$ d'ordre q laisse passer une tendance linéaire sans distorsion. b) Trouvez le filtre de cette forme qui minimise la variance de $X_t = \sum_{i=0}^n \theta_i \epsilon_{t-i}$. Formuler ce problème comme un problème d'optimisation avec des contraintes, qu'on résoudra par la méthode des multiplicateurs de Lagrange. **Ind** : On rappelle que $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$. **R** : a) $\sum_i \theta_i = 1, \sum_i i\theta_i = 0$. b) $\text{Var } X_t = \sum_i \theta_i^2$. Elle doit être minimisée sous ces contraintes. On trouve $\theta_i = \frac{\lambda + i\mu}{2}$ et les multiplicateurs de Lagrange satisfont

$$\begin{aligned} \lambda + \frac{n\mu}{2} &= \frac{2}{n+1} \\ \lambda + \frac{(2n+1)\mu}{3} &= 0 \end{aligned}$$

8. On considère un modèle simple où la tendance est une constante ($f(t) = a$).
- a) On considère tout d'abord le modèle sans composante saisonnière. Comment choisir a si le modèle est additif? que peut-on alors dire sur les fluctuations irrégulières? que se passe-t-il si le modèle est multiplicatif?
- b) On considère maintenant qu'une composante saisonnière ($s_j, 1 \leq j \leq p$) est présente. On suppose que le nombre d'observations n est un nombre entier L de périodes : $n = Lp$. Comment choisir a et (s_j) si le modèle est additif? que peut-on alors dire sur les fluctuations irrégulières? que se passe-t-il si le modèle est multiplicatif?
- c)* Reprendre la question b) lorsque le nombre d'observations n'est pas un nombre entier de périodes : $n = Lp + m$.
9. On considère une série ($y_i, 1 \leq i \leq n$) périodique, de période p . On suppose que le nombre d'observations n est un multiple de p : $n = Lp$. Montrer alors que les corrélations suivantes sont :

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

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

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

3.6 TP : L'approche de decomposition sur R

getwd()#Directoire courant

```
?stl # Aide
data() # jeux de données disponibles
library()
data(package = .packages(all.available = TRUE))
data(package = "MASS")
```

Quelques commandes : decompose, summary, diff, filter, stl, acf, tapply, window, layout, aggregate, boxplot, pacf, ar,

Rappels. L'approche générale pour la modélisation des séries chronologiques est la suivante :

1. Tracer la série et examiner les caractéristiques du graphique (par la commande stl ou/et en regardant les fonctions d'autocovariances et d'autocovariances partielles estimées. Vérifier en particulier s'il existe 1. une tendance polynômiale/exponentielle 2. une composante saisonnière, 3. des changements nets, 4. des données exceptionnelles.

```
is.ts(BOD)
as.ts(BOD)
plot.ts(BOD)
```

```
is.ts(CO2)
as.ts(CO2)
plot.ts(CO2)
```

2. Enlever la tendance et la composante saisonnière afin d'obtenir une série stationnaire, par différenciation (commande diff ou filter)

Nous allons illustrer la décomposition en tendance, saisonnier et résidus sur une jeux des données analysé par A. Lagnoux :

```
X<-scan("http://www.math.univ-toulouse.fr/~lagnoux/trafic.dat")
is.ts(X) #Vérifier
X=as.ts(X) #Transformer X en un objet du type ts
X=ts(X,start=1963,frequency=12) # assigner un nouveau début pour enlever des N.A.
X #Vérifier
plot.ts(X) #Représenter la série
```

Commençons avec decompose, et comparons les résultats obtenues avec les options type="add" ou "mult", qui permettent de spécifier si on souhaite utiliser un modèle additif (par défaut) ou multiplicatif :

```
X.add= decompose(X,type="add")
plot(X.add)
```

Sur la représentation graphique, quatre courbes sont représentées : de haut en bas, figurent la série initiale, la tendance, la composante saisonnière et la partie résiduelle. Ces quatre parties correspondent aux différentes composantes de l'objet ainsi créé. La décomposition repose sur l'application de moyennes mobiles dont on peut préciser le filtre éventuellement. Par défaut, une moyenne mobile symétrique est employée.

?decompose

```
res=X.add$random
summary(res)
hist(res) #Histogramme des résidus
sq<-sqrt(var(res,na.rm=T))
res=res/sq #Définition des résidus réduits
acf(res,na.action=na.pass) #Corrélogramme des résidus

qqnorm(res) # Comparaison des quantiles des résidus
abline(0,1) #et des quantiles d'une loi normale

X.mul= decompose(X,type="mult")
plot(X.mul)
```

```
res=X.add$random
summary(res)
hist(res) #Histogramme des résidus
```

```
sq<-sqrt(var(res,na.rm=T))
res=res/sq #Définition des résidus réduits
acf(res,na.action=na.pass)
```

On peut aussi décomposer avec "lm", en spécifiant les fonctions de base du modèle linéaire désiré, ou, plus pédagogiquement, un pas après l'autre, comme ci-dessous :

```
#Estimation de la tendance
filt=rep(1/12,11) #Définition des coefficients
filt=c(1/24,filt,1/24) #Définition de la moyenne mobile
Z=filter(X,filter=filt,sides=2) #Calcul de la série filtrée par moyenne arithmétique
d'ordre 13 modifiée aux extrémités
Z=ts(Z,start=1963,frequency=12)
ts.plot(X,Z) #Superposition de la série brute et de la série filtrée
```

Estimation du saisonnier, en éliminant la tendance

```
S=X-Z
s=tapply(S,cycle(S),mean,na.rm=T) #Estimation des coefficients du saisonnier
sd=s-mean(s)
```

Série corrigée des variations saisonnières CVS

```

CVS=matrix(1,18,12)
for (i in 1 :18) {for (j in 1 :12) {CVS[i,j]=t(matrix(X,12,18))[i,j]-s[j]}}
#Calcul de CVS = X -s; t donne la transposé
CVS=as.vector(t(CVS))
CVS=as.ts(CVS)
CVS=ts(CVS,start=1963,frequency=12)
ts.plot(X,CVS,col = c("blue", "red"), lwd = c(1, 2))

```

Regression polynômiale de la tendance

```

y=time(CVS)
z=time(CVS)^2
CVS.lm=lm(CVS ~ y+z) #Estimation de la tendance : f(t) = a + bt + ct2
v<-CVS.lm$coefficients#Valeurs des coefficients a,b,c
ts.plot(time(CVS),CVS.lm$fitted.values)

```

Analyse des résidus

```

res=CVS-CVS.lm$fitted.values #Définition des résidus
res=res/sqrt(var(res)) #Définition des résidus réduits
acf(res) #Corrélogramme des résidus

```

Prévision

```

X1=rep(1,12)
for (i in 1 :12)
{X1[i]=v[1]+v[2]*(1981+(i-1)/12)+v[3]*(1981+(i-1)/12)^2+s[i]} #Prévoir le trafic en 1981
X2=c(as.vector(X),X1)#Juxtaposition de la série
#chronologique (jusqu'en 1980)et de la prévision pour #1981
X2=as.ts(X2) #Transformation en SC
X2=ts(X2,start=1963,frequency=12) #Renouvellement des temps
ts.plot(X2)

```

Exercice 3.6.1 On considère la série suivante

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

- Représenter graphiquement cette série.
- Quel modèle proposeriez-vous pour cette série (justifier) ?
- Calculer les facteurs saisonniers ($s_j, 1 \leq j \leq p$) ainsi que leur somme $\sum_{j=1}^p s_j$, en supposant une tendance constante $m_t = a$.
- En notant ($e_i, 1 \leq i \leq n$) la série des fluctuations irrégulières, calculer e_1, e_2, \dots, e_n .
- Proposer une méthode pour l'estimation des paramètres, en supposant cette fois une tendance linéaire $m_t = at + b$ (en plus de la composante saisonnière). Implementez le calcul en utilisant un logiciel. Proposez un teste pour choisir entre les deux modèles.

R :

```

y=c(7.5,4.4,3.3,7.6,3.9,2.4,6.9,4.5,2.7,8.2,4.1,3.0,7.5,3.5,2.8)
t=1:15
plot(t,y,type="b")
#Calculs des facteurs saisonnières:
s1=0.2*(y[1]+y[4]+y[7]+y[10]+y[13])-mean(y)
s2=0.2*(y[2]+y[5]+y[8]+y[11]+y[14])-mean(y)
s3=0.2*(y[3]+y[6]+y[9]+y[12]+y[15])-mean(y)
s1+s2+s3
#Calcul bruit
e=y-mean(y)-rep(c(s1,s2,s3),5)
e
mean(e)
#Régression
z=ts(y, frequency = 3)
cz<-cycle(z)
fz<-factor(cz)
class(fz) #?factor
zlm<-lm(z~ time(z)+ fz)
summary(zlm)

```

Comme l'intervalle de confiance pour le coefficient de "time" inclu le 0 et sa t-value n'est pas plus grande que 3 en valeur absolue, ce coefficient n'est pas nécessaire[§].

Exercice 3.6.2 On considère la série suivante :

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

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

Exercice 3.6.3 Analyse visuelle de séries chronologiques. Décrire les trois séries chronologiques uspop, AirPassengers et sunspots, dans cet ordre. On rappelle que "décrire une série chronologique" veut dire : tracer la série, et puis commenter les aspects suivants : - tendance (linéaire ? exponentielle ? etc.) - phénomène saisonnier (amplitude variable ?) - bruit (amplitude variable ? Corrélation ?) Enfin, cela signifie transformer la série de la manière la plus simple possible afin d'obtenir des résidus stationnaires .

Exemple 3.6.1 Passons au jeu des données fourni en R AirPassengers, qui vient du livre Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) Time Series Analysis, et illustre la méthode de prédiction ARIMA.

§. il semble y avoir un problème avec factor, car summary(zlm) donne seulement deux niveaux saisonniers au lieu de 3

```

AP<-AirPassengers
AP
is.ts(AP) #Vérifier si ap objet R de type time ser
frequency(AP)
class(AP)
plot(AP)
plot.ts(AP)

plot(log(AP))

layout(1:2)
plot(aggregate(AP)) #calcul tendance par le filtre "somme per période"
boxplot(AP~cycle(AP)) #variabilité de chaque group
?boxplot

layout(1:2)
AP.add <- decompose(AP)
AP.mult <- decompose(AP, "multiplicative")
plot(ts(AP.add$random[7:138]))
plot(ts(AP.mult$random[7:138]))

layout(1:2)
acf(AP.add$random[7:138])
pacf(AP.add$random[7:138])

layout(1:2)
acf(AP.mult$random[7:138])
pacf(AP.mult$random[7:138])

lg<-log10(AirPassengers)
diff(lg,differences=1)
(fit <- arima(lg, c(0, 1, 1),
              seasonal = list(order = c(0, 1, 1), period = 12)))
pred <- predict(fit, n.ahead = 24)
tl <- pred$pred - 1.96 * pred$se
tu <- pred$pred + 1.96 * pred$se
ts.plot(AirPassengers, 10^tl, 10^tu, log = "y", lty = c(1, 2, 2))

install.packages(forecast)
library(forecast)
auto.arima(lg)

```

On peut aussi simuler des séries

```

eps=rnorm(100) #Simuler un bruit blanc gaussien de variance 1 et de longueur 100
y=sort(eps) # Ordonner le vecteur eps
hist(eps,freq=FALSE) #Tracer l'histogramme

```

```

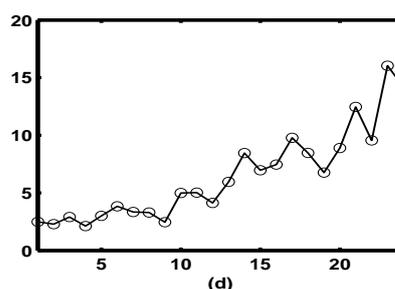
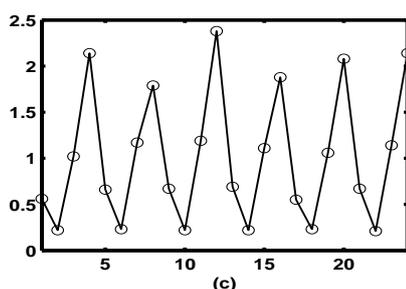
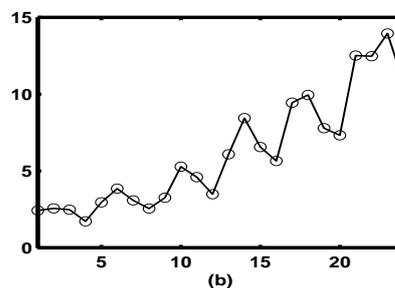
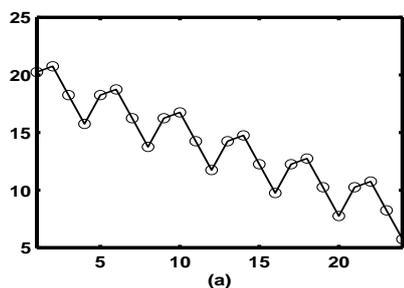
lines(y,dnorm(y)) #Comparer avec la densité d'une N (0, 1)
qqnorm(eps) #Tracer eps en fonction des quantiles d'une loi N (0, 1)
qqline(eps) #Tracer la droite de régression sur les données
abline(0,1) #Tracer la première bissectrice et commenter.
help(qqnorm) #Pour avoir un descriptif de la fonction ? qqline #Même chose

```

Exercice 3.6.4 a) Comment générer un vecteur gaussien de moyenne m et de variance σ^2 à partir de eps ? Et avec la commande rnorm ? b) Recommencer cette simulation mais avec un bruit blanc de longueur 1000 puis 10000. Vérifier que l'histogramme devient de plus en plus proche de la densité de la gaussienne (courbe en cloche). De même, vérifier que la droite de régression sur les données est de plus en plus proche de la première bissectrice. Pourquoi ?

Exercice 3.6.5 Simuler une série chronologique $Y_t, t = 1, \dots, 100$ suivant le modèle $Y_t = .01t + 1 + 2\sin(2\pi t/5) + \epsilon_t$, où ϵ_t est un bruit blanc gaussien de variance $1/100$. 1. Déterminer la tendance, la saisonnalité (période) de cette série chronologique. Les tracer. 2. Utiliser la méthode des moyennes mobiles ci-dessus pour éliminer la saisonnalité puis estimer les coefficients du saisonnier. 3. Utiliser une régression linéaire par moindres carrés pour estimer les coefficients de la tendance. 4. Comparer les estimateurs avec les vrais coefficients. 5. Proposer une prévision à l'horizon 3. 6. Analyser les résidus. Les représenter. 7. Appliquer la fonction decompose et comparer avec les vraies valeurs.

Exercice 3.6.6 Pour chacune des quatre séries suivantes,



- écrire un modèle qui vous semble convenir, en précisant le type du modèle, la tendance et la période
- Exprimez le modèle choisi sous la forme d'une équation vectorielle **linéaire** dans les paramètres inconnues, et donnez la formule de la régression qui permet à déterminer ces paramètres.

Exercice 3.6.7 La figure 3.1 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 Etats-Unis ne semble pas évoluer globalement, mais présente également cette alternance de baisses et de hausses.*
- *Les ventes de champagnes, tout comme la production industrielle semblent exhiber un caractère périodique (ventes importantes de champagne en fin d'année, baisse de la production industrielle en été, ...).*
- *D'autre part, les variations de ces 2 séries (indice de production industrielle et ventes de champagne) ont une amplitude qui semble augmenter au cours du temps.*

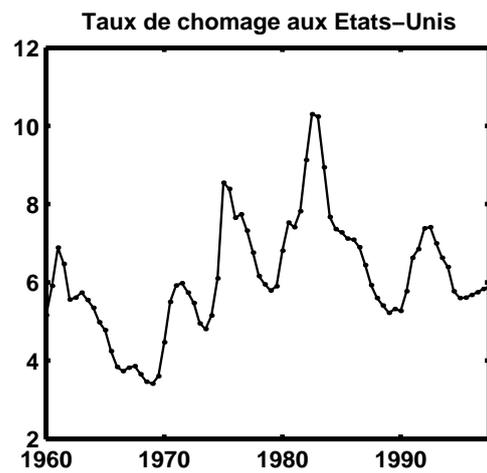
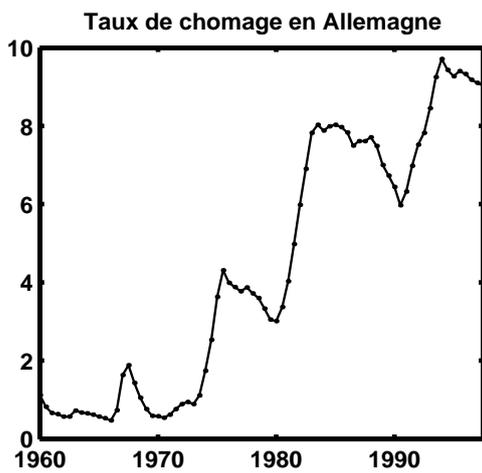
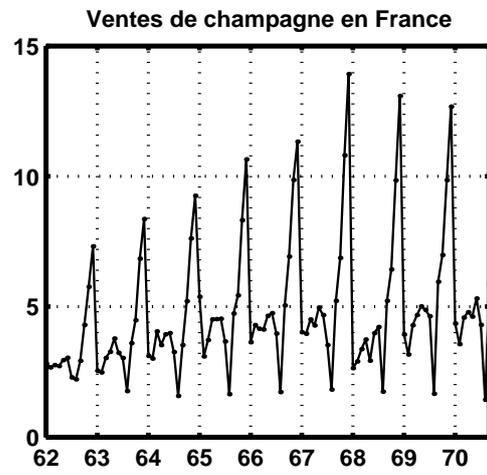
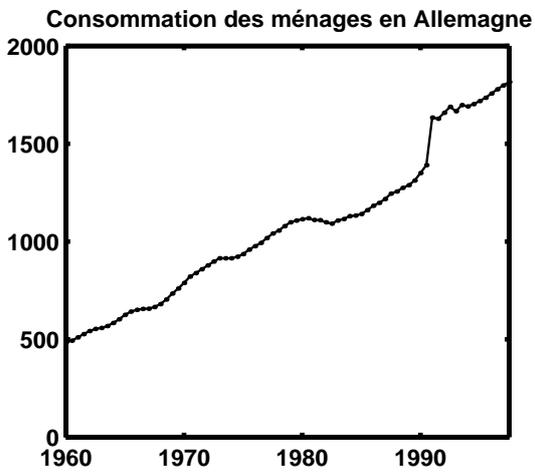
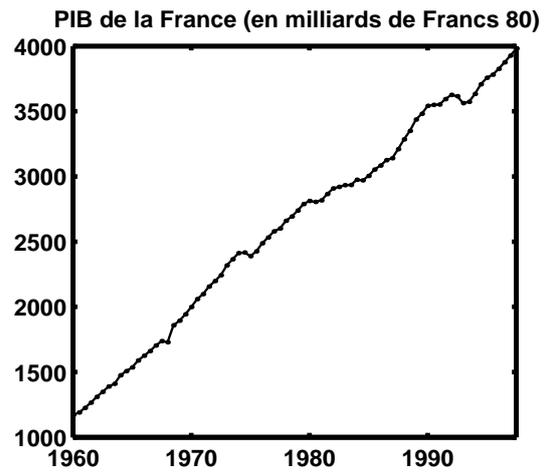
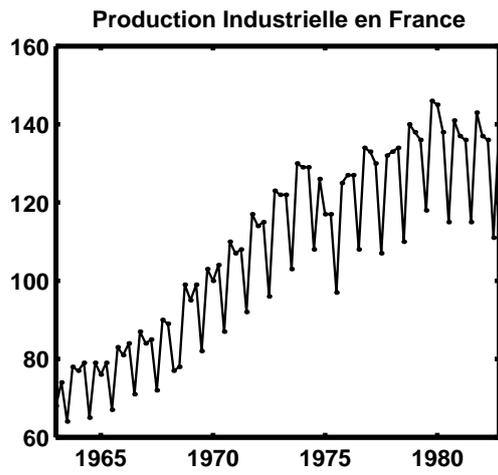


FIGURE 3.1 – Quelques exemples de séries chronologiques

Chapitre 4

Modélisation stochastique des séries temporelles

4.1 Introduction

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

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

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

On s'arrangera toujours tel que les résidus ont la moyenne 0, mais ça n'est pas suffisant pour qu'ils soient un bruit totalement sans structure="bruit blanc" (et s'il y a encore une partie structurée, elle devrait être incluse en m_t). Le "bruit blanc" est notre premier exemple d'un processus stochastique : une formalisation du concept de séries temporelles, ayant des propriétés bien définies (voir prochain chapitre). Inspirés par les propriétés de ce processus, on proposera des tests statistiques correspondant à ce modèle, qui nous permettront de décider si ϵ_t ont les propriétés de manque de structure désirées.

Pour tendance, plusieurs modélisations se sont avérées utiles :

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

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

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

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

3. modèles autoregressifs :

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

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

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

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

4.2 Processus stochastiques stationnaires

Définition 4.2.1 Soit X un processus aléatoire indexé par $T = \mathbb{N}$ ou \mathbb{Z} . On dit que X est **stationnaire** (strict) si pour toute famille finie d'instantants $t_1 \dots t_r \in T$ et tout entier s , les lois jointes de $(X_{t_1} \dots X_{t_r})$ et de $(X_{t_1+s} \dots X_{t_r+s})$ sont les mêmes.

Définition 4.2.2 Soit X un processus aléatoire indexé par $T = \mathbb{N}$ ou \mathbb{Z} . On dit que X est **stationnaire à l'ordre 2** si la moyenne $m(t)$ et la covariance $\Gamma(s, t)$ sont invariantes par translation dans le temps, c.-à-d. si la moyenne est constante :

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

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

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

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

Théorème 4.2.1 Pour un processus stationnaire, les covariances empiriques

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

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

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

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

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

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

Remarques :

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

4.3 Exemples des processus stationnaires

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

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

4.3.1 Le bruit blanc

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

Définition 4.3.1 *Un processus $\epsilon_t, t \in \mathcal{T}$, où \mathcal{T} est un ensemble dénombrable quelconque, est appelé **bruit blanc stationnaire** si les variables ϵ_t sont **i.i.d.** (indépendants et identiquement distribués) à espérance $\mathbb{E}\epsilon_t = 0$. Il sera appelé **bruit blanc Gaussien** si la distribution de chaque variable aléatoire ϵ_t est Gaussienne.*

Un bruit blanc a la covariance

$$\gamma(s, t) = \mathbb{E}[\epsilon_s \epsilon_t] = 0, \forall s \neq t \quad \text{et donc le coefficient de corrélation} \quad (4.1)$$

$$\rho(s, t) = \frac{\gamma(s, t)}{\sigma_s \sigma_t} = \delta(s - t) \quad (4.2)$$

où $\delta(s - t)$ est le symbole du Kronecker). Comme les tests d'indépendance et Gaussianité demandent beaucoup de données, qui ne sont pas toujours disponibles, il faut faire parfois avec un "idéale probabiliste moins structuré" : le "bruit blanc de second ordre" défini par les deux dernières formules équivalentes (4.1), (4.2).

Définition 4.3.2 Un processus $\epsilon_t, t \in \mathbb{N}$ ou $t \in \mathbb{Z}$ est appelé **bruit blanc de second ordre** s'il a la moyenne 0, la variance constante $\mathbb{E}\epsilon_t^2 = \sigma^2$ et une covariance $\gamma(s, t) = \mathbb{E}[\epsilon_s \epsilon_t] = 0, \forall s \neq t$ (et donc les coefficients de corrélation $\rho(s, t) = \delta(s - t)$).

Notes :

1. Le bruit blanc Gaussien est une structure probabiliste très naturelle, car la distribution Gaussienne possède plusieurs propriétés importantes, comme celle d'être invariante par rapport aux rotations, ce qui est évidemment une réquise pour un bruit 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 déterminist d'une régression appliqué a une série observée une seule fois, il est difficile de la vérifier rigoureusement. Parmi les tests possibles, mentionnent celui de "turning points", qui demande de vérifier que la fréquence de ces points est environ 4/6, et le teste qui vérifie si la somme des corrélations empiriques est proche de 0. Si ces deux testes sont positives, on sait au moins "qu'on ne peut pas repousser l'hypothèse de l'indépendance". Il y aussi des tests distributionnels des résidus comme Fisher, Student, qui testent la Gaussianité.
3. Quand les tests des données rejettent l'hypothèse du bruit blanc, c.-à-d. quand on a du **bruit corrélé**, la régression classique doit être remplacée par une analyse plus fine, appelée krigeage en géostatistique.

4.3.2 Les moyennes mobiles MA(q) : l'exemple le plus simple de processus a composantes dépendantes

Définition 4.3.3 On appelle processus MA(q) un processus linéaire $Z_t, t \in \mathbb{Z}$ vérifiant une relation :

$$Z_t = \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (4.3)$$

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

La notation des polynômes de retard ramène (4.3) à la forme :

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

Exercice 4.3.1 1. Calculer la fonction d'autocovariance $\gamma(k)$ d'un processus MA(1).

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

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

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

où

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

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

4.3.3 Les moyennes mobiles MA(∞)

Définition 4.3.4 Un processus Y_t sera appelé MA(∞) ou linéaire en ϵ_t s'il peut être représenté dans la forme :

$$Y_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i} \quad \text{avec} \quad \sum \psi_i^2 < \infty \quad (4.4)$$

où ϵ_t est un bruit blanc.

Théorème 4.3.1 Un processus linéaire

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

où $\sum \psi_i^2 < \infty$ est : a) bien défini dans L_2 (c.-à-d. $\text{Var} Y_t < \infty$), b) à variance constante stationnaire $\text{Var} Y_t = \sigma_\epsilon^2 \sum_{i=-\infty}^{\infty} \psi_i^2$ c) à autocovariance donnée par :

$$\gamma(t, t+k) = \sigma_\epsilon^2 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+k} < \infty \quad (4.5)$$

d) stationnaire à ordre deux.

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

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

Définition 4.3.5 Un processus linéaire Y_t s'appelle **causal** s'il peut être représenté dans la forme :

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (4.6)$$

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

4.3.4 Les modèles autorégressifs AR(p) : des processus faciles à prédire

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

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + \epsilon_t \quad (4.7)$$

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

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

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

Définition 4.3.6 *Un processus stationnaire $Y_t, t \in \mathbb{Z}$ sera appelé **processus autorégressif linéaire d'ordre p : AR(p)** s'il existe un bruit blanc ϵ_t et des réels $\varphi_i, i = 1, \dots, p$ tels qu'une relation de récurrence :*

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \epsilon_t, \forall t \in \mathbb{Z} \quad (4.8)$$

est vérifiée.

La notation des polynômes de retard ramène (4.8) à la forme :

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

Définition 4.3.7 *Le polynôme*

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

*sera appelé **symbole** du modèle (4.8). Le polynôme avec toutes les puissances renversées sera appelé **polynôme caractéristique**.*

Rq : Les processus autorégressifs sont définis par une équation, qui à priori, peut ne pas avoir des solutions ; comme "solution" de l'équation (4.8) nous aimerions avoir une représentation du processus Y_t par rapport au processus ϵ_t .

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

$$Y_t = \phi Y_{t-1} + b + \epsilon_t, \forall t \in \mathbb{Z}$$

Ce modèle est recommandable si on envisage une prévision

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

où $b = a(1 - \phi)$. On vérifie que si la moyenne de Y_t est 0 on a $a = b = 0$; pour simplifier, on supposera souvent qu'on a déjà enlevé la moyenne de Y_t . Pour utiliser ce modèle, on peut estimer le paramètre ϕ par une régression linéaire des points

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

Le fait d'avoir enlevé la moyenne ramène à une droite passant par l'origine $y = \phi x$. En suite, on utilise la valeur trouvée pour résoudre l'équation. Avec valeur initiale Y_0 , on trouve

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

Exercice 4.3.2 Processus AR(1) stationnaire. a) Montrez que l'équation :

$$Y_t = \phi Y_{t-1} + \epsilon_t, \forall t \in \mathbb{Z} \quad (4.9)$$

a une solution stationnaire unique **causale** (qui dépend seulement du **bruit présent et passé**) si : $|\phi| < 1$.

Indication : Calculer la solution de (4.9) par des substitutions répétées, ou en utilisant le calcul opérationnel de Heaviside (Arbogast), en posant $Y_t = (1 - \phi B)^{-1} \epsilon_t$, et en développant la fraction comme une série de puissances en B . En suite, montrez que la série est convergente dans le sens L_2 , et calculez les covariances, pour vérifier la stationnarité. b) Montrez que l'équation : (4.9) a une solution stationnaire unique, qui dépend seulement du **bruit futur** si : $|\phi| > 1$.

Remarque 4.3.1 Dans le cas $|\phi| = 1$, les solutions de l'équation : (4.9) (appelée marches aléatoires) ne peuvent pas être stationnaires. Par contre, les **incrément** $Y_t - Y_{t-1} = \epsilon_t$ sont stationnaires; cette situation plus compliquée sera analysée dans le chapitre sur les processus ARIMA(p, d, q).

En conclusion

Théorème 4.3.2 Le processus (de Markov) AR(1) $Y_t = \phi Y_{t-1} + \epsilon_t$ a une représentation causale ssi $|\phi| \neq 1$ et cette représentation est causale ssi $|\phi| < 1$. Aussi

1. pour $|\phi| < 1$, on vérifie que ϵ_t est un bruit d'innovation, c.-à-d. $E[\epsilon_t Y_{t-k}] = 0, \forall k \geq 1$.
2. pour $|\phi| > 1$, l'équation : (4.9) a une solution stationnaire unique, qui dépend seulement du bruit futur, et ϵ_t n'est pas un bruit d'innovation.

4.4 Fonctions de transfert et applications

Définition 4.4.1 La fonction $\psi(z) = \frac{1}{\varphi(z)}$ est appelée fonction de transfert du modèle AR(p).

4.4.1 Récurrences infinies et la causalité des modèles AR(p)

Nous venons de voir qu'il y a des problèmes (manque de solutions stationnaires et non-causalité) avec le modèle AR(1) quand φ n'est pas à l'intérieur du cercle unitaire (est la racine $\lambda = \varphi^{-1}$ du symbole $\varphi(z) = 1 - \varphi z$ n'est pas à l'extérieur du cercle unitaire. Ces problèmes sont liés à l'existence des plusieurs développements possibles pour la fonction de transfert $\varphi(z)^{-1} = \frac{1}{1-z\varphi}$.

Exercice 4.4.1 Donner tous les développements en série des puissances Laurent de la fonction de transfert du modèle AR(1).

$$\frac{1}{1-z\varphi} = \begin{cases} \sum_{n=0}^{\infty} \varphi^n z^n & \text{à l'intérieur du cercle } |z| < \varphi^{-1}, \text{ mais} \\ -\sum_{n=-\infty}^{-1} \varphi^n z^n & \text{à l'extérieur du cercle, } |z| > \varphi^{-1} \\ \text{pas de développement} & |z| = \varphi^{-1} \end{cases}$$

Nous verrons maintenant qu'on a une situation pareille pour tous les modèles AR(p).

Remarque 4.4.1 La fonction de transfert intervient quand on essaie de résoudre formellement le modèle AR(p), "en multipliant" l'équation

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

par $\varphi(B)^{-1}$. On arrive à la représentation

$$Y_t = \psi(B)\epsilon_t.$$

Mais quelle sens donner à la multiplication par $\varphi(B)^{-1}$? En fait, il s'agit d'un problème délicat, la solution du quelle n'est pas toujours possible.

Considerons un développement en série des puissances Laurent

$$\psi(z) = \varphi(z)^{-1} = \sum_{k=-\infty}^{\infty} \psi_k z^k$$

(il existent plusieurs), qui nous suggère l'interprétation

$$Y_t = \psi(B)\epsilon_t = \sum_{k=-\infty}^{\infty} \psi_k B^k \epsilon_t = \sum_{k=-\infty}^{\infty} \psi_k \epsilon_{t-k}. \quad (4.10)$$

Une telle représentation du AR(p) comme modèle MA(∞) sera valable au cas où la série ci-dessus est convergente. Il nous reste le dylème du choix du développement en série des puissances Laurent.

1. Començons par le cas le plus simple, avec toutes les racines de $\varphi(z) = 0$ dehors le cercle unitaire. Alors, on peut vérifier que les coefficients ψ_k satisfont une récurrence avec polynôme caractéristique $z^p \varphi(1/z)$, qui a toutes les racines à l'intérieur du cercle unitaire. Dans ce cas, le développement de Laurent devient une série de Taylor $\psi(z) =$

$\varphi(z)^{-1} = \sum_{k=0}^{\infty} \psi_k z^k, \forall |z| \leq 1$ convergente. Finalement, cela assure la convergence en L_2 de (4.10), on a

$$Y_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k},$$

et notre processus AR(p) a une représentation MA(∞) causale.

2. Le deuxième cas facile à résoudre est quand il existe des racines satisfaisant $|z_i| = 1$. Dans ce cas, la stationnarité est impossible.
3. Dans le cas restant $|z_i| \neq 1$, mais avec certaines racines du polynôme caractéristique à l'extérieur du cercle unitaire, on peut encore choisir un développement de Laurent tel que l'expansion MA(∞) fournisse une solution stationnaire, mais non causale de l'équation $\varphi(B)Y_t = \epsilon_t$. Ce cas n'a pas d'application en statistiques. En conclusion, on accepte seulement les AR(p) avec toutes les racines du polynôme caractéristique à l'intérieur du cercle unitaire, dans quel cas la fonction de transfert a un développement de Taylor valable dans le cercle unitaire, y compris autour de $z = 1$. Dans ce cas, les coefficients ψ_k de l'expansion de Taylor de la fonction de transfert peuvent être calculés par identification des coefficients de z^k dans $\psi(z)\varphi(z) = 1$.

Théorème 4.4.1 a) *Un processus AR(p) est causal, c.-à-d. il peut être représenté sous la forme :*

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (4.11)$$

où $\sum \psi_i^2 < \infty$ ssi toutes les racines du polynôme caractéristique $z^p \varphi(1/z)$ sont à l'intérieur du cercle unitaire.

Exercice 4.4.2 *Donnez une solution stationnaire de l'équation $(1 - 2B)(1 - B/2)Y_t = \epsilon_t$.*

On se contentera désormais à étudier le cas causal. En appliquant la représentation des modèles AR(p) comme des modèles MA(∞), nous découvrirons que tout est basé sur la "réurrence fondamentale" satisfaite par les coefficients $\psi(k)$ de la fonction de transfert.

4.4.2 La récurrence de Yule-Walker pour les coefficients de la fonction de transfert

Exercice 4.4.3 *Soit $\sum_{i=0}^{\infty} \psi_i z^i$ le développement limité autour de 0 de $\frac{1}{1 - \varphi_1 z - \varphi_2 z^2}$.*

1. *Montrez que les coefficients ψ_i satisfont les équations Yule-Walker*

$$\begin{aligned} \psi(0) &= 1, & \psi(1) &= \phi_1 \quad (\text{ou } \psi(-1) = 0) \\ \psi_i &= \phi_1 \psi_{i-1} + \phi_2 \psi_{i-2} \iff \varphi(B)\psi_i = 0, & i &\geq 2 \end{aligned}$$

2. *Donnez les formules explicites de $\psi_i, i = 0, 1, \dots, 3$. Sol : $\{1, \phi(1), \phi(1)^2 + \phi(2), \phi(1)^3 + 2\phi(2)\phi(1)\}$.*

3. *Généraliser au développement limité autour de 0 de $\frac{1}{1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p}$*

Sol : Les solutions explicites se compliquent, mais la récurrence

$$\boxed{\varphi(B)\psi_i = 0, \quad i \geq p, \psi_0 = 1, \psi_{-1} = 0, \dots, \psi_{-p+1} = 0} \quad (4.12)$$

reste vraie.

Remarque 4.4.2 Il suit que si un processus $AR(p)$ $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \epsilon_t$ a une représentation stationnaire causale

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

(ce qui est le cas si le polynôme caractéristique $\varphi(z) = (1 - \lambda_1 z)(1 - \lambda_2 z) \dots (1 - \lambda_p z)$ a ses racines $\lambda_1^{-1}, \dots, \lambda_p^{-1}$ dehors le cercle unitaire, ou si le polynôme dual $\lambda^p \varphi(\lambda^{-1})$ a ses racines λ_i dedans le cercle unitaire), alors les coefficients ψ_i satisfont la **récurrence Yule-Walker** (4.12).

Exercice 4.4.4 Montrez que les coefficients $\psi_k = A_1 \lambda_1^k + A_2 \lambda_2^k$ d'un processus $AR(2)$ causal (avec des racines du polynôme caractéristique dans le cercle unitaire) on a

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

Vérifier que $\sum_i \psi_i^2 \leq \infty$.

Corollaire 4.4.1 Pour un processus $AR(p)$ causal Y_t on a $\epsilon_{t+k} \perp Y_t$ si $k \geq 0$. Si le processus est centré, il suit que $E[\epsilon_{t+k} Y_t] = 0$.

Exercice 4.4.5 (*) Trouvez les solutions stationnaires des processus $AR(2)$ avec au moins une des racines caractéristiques λ_1, λ_2 strictement dehors le cercle unitaire $|\lambda| < 1$.

4.5 Équations de Yule-Walker pour les covariances et corrélations d'un processus $AR(p)$ causal

Remarque 4.5.1 Les covariances et les corrélations d'un processus sont des fonctions paires, i.e. $\gamma_k = \gamma_{-k}, \rho_k = \rho_{-k}$.

Le corrélogramme des processus autorégressifs $AR(p)$ n'est pas aussi facile à calculer que celle des processus $MA(q)$.

Théorème 4.5.1 Les covariances et les corrélations d'un processus $AR(p)$ causal sont liées aux coefficients $\varphi = (\varphi_1, \dots, \varphi_p)$ par les équations de Yule-Walker.

1. Pour les covariances, en multipliant la formule autoregressive de Y_{t+k} par Y_t , en prenant l'espérance et en utilisant le Corollaire 4.4.1, on obtient :

$$\gamma_k = \sum_{i=1}^p \varphi_i \gamma_{k-i} \quad \text{pour } k \geq 1 \quad (4.13)$$

$$\gamma_0 = \sum_{i=1}^p \varphi_i \gamma_i + \mathbb{E}Y_t \epsilon_t = \sum_{i=1}^p \varphi_i \gamma_i + \sigma^2 \quad \text{pour } k = 0$$

2. Pour les corrélations, on remarque d'abord, en divisant par γ_0 , qu'elles satisfont aussi la récurrence (4.13) :

$$\rho_k = \sum_{i=1}^p \varphi_i \rho_{k-i} \quad \text{pour } k \geq 1 \quad (4.14)$$

Par conséquent, il suffit de déterminer les premières p corrélations. En appliquant la récurrence (4.14) pour $k = 1, \dots, p$, et en tenant compte aussi de la symétrie de ρ_k , on arrive au **système Yule-Walker** :

$$\boxed{\mathbf{R}_p \boldsymbol{\varphi} = \boldsymbol{\rho}} \quad (4.15)$$

où \mathbf{R}_p est la matrice Toeplitz symétrique :

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

Ces équations permettent de calculer les coefficients $\boldsymbol{\varphi}$ à partir des corrélations et viceversa (on calcule d'abord les premières p corrélations ; en suite, on trouve aussi les autres corrélations en utilisant la récurrence).

3. En prenant γ_0 comme facteur commun dans la deuxième equation en (4.13), on trouve γ_0 en fonction des corrélations ρ_i :

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

ça permet en suite d'obtenir les autres covariances, en partant des corrélations.

Remarque 4.5.2 Les équations (4.13) sont un cas particulier des "équations normales" de la régression.

Exemple 4.5.1 Dans le cas $p = 2$, les premières deux équations pour $k = 1, 2$ sont

$$\rho_1 = \varphi_1 + \varphi_2 \rho_{-1} = \varphi_1 + \varphi_2 \rho_1$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2$$

et la première donne $\rho_1 = \frac{\varphi_1}{1-\varphi_2}$. Pour trouver les φ_i à partir de ρ_i , on écrit le système comme :

$$\begin{cases} \varphi_1 + \varphi_2\rho_1 = \rho_1 \\ \varphi_1\rho_1 + \varphi_2 = \rho_2 \end{cases} \iff \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

Dans le cas $p = 3$, les premières trois équations pour $k = 1, 2, 3$ sont

$$\begin{cases} \varphi_1 + \varphi_2\rho_1 + \varphi_3\rho_2 = \rho_1 \\ \varphi_1\rho_1 + \varphi_2 + \varphi_3\rho_1 = \rho_2 \\ \varphi_1\rho_2 + \varphi_2\rho_1 + \varphi_3 = \rho_3 \end{cases} \iff \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix}$$

et on voit que la structure Toeplitz précisée dans le théorème est correcte.

Remarque 4.5.3 Les matrices de corrélations \mathbf{R}_p sont positive définies et jouent un rôle fondamental dans l'estimation des modèles $AR(p)$ et $ARMA(p, q)$.

Finalement, il est intéressant de comparer trois solutions possibles pour obtenir les corrélations :

1. en résolvant le système et récurrence de Yule-Walker
2. par le développement limité de la fonction génératrice des covariances $\gamma(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k = \frac{1}{\varphi(z)\varphi(z^{-1})}$.
3. en représentant d'abord le processus $AR(p)$ comme un processus $MA(\infty)$, par le développement limité $\psi(z) = 1/\varphi(z)$, et en utilisant en suite la formule (4.5) du corrélogramme des processus $MA(\infty)$.

Exercice 4.5.1 Calculer le corrélogramme de $AR(1)$ avec $\sigma_\epsilon^2 = 1$ par :

1. les équations Yule-Walker
2. par le développement limité de la fonction génératrice $\gamma(z)$
3. la représentation $MA(\infty)$.

Tracez le corrélogramme $\{\rho_k\}$ pour (a) $\varphi = 0.5$, (b) $\varphi = 0$, and (c) $\varphi = -0.5$.

R : 1. Les équations Yule-Walker sont :

$$\rho_n = \varphi\rho_{n-1}, n \geq 1 \implies \rho_n = \varphi^n$$

et $\gamma_0 = \frac{1}{1-\varphi^2}$. 2. Décomposant en fractions simples :

$$\begin{aligned} f(z) &= \frac{z}{(1-\varphi z)(z-\varphi)} = \frac{1}{1-\varphi^2} \left(\frac{1}{1-z\varphi} + \frac{\varphi}{z-\varphi} \right) = \frac{1}{1-\varphi^2} \left(\frac{1}{1-z\varphi} + \frac{\varphi}{z} \frac{1}{1-\varphi/z} \right) \\ &= \frac{1}{1-\varphi^2} \left(\sum_{k=0}^{\infty} z^k \varphi^k + \frac{\varphi}{z} \left(\sum_{k=0}^{\infty} z^{-k} \varphi^k \right) \right) = \frac{1}{1-\varphi^2} \left(\sum_{k=0}^{\infty} z^k \varphi^k + \sum_{k=1}^{\infty} z^{-k} \varphi^k \right) \end{aligned}$$

Théorème 4.5.2 Formule générale des corrélations pour AR(2) : Pour le processus AR(2)

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

le système de Yule-Walker (4.15) donne :

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

(et les inégalités $|\rho_i| \leq 1, \rho_1^2 \leq \frac{1+\rho_2}{2}$ implique les restrictions $|\varphi_1| \leq 2, -1 \leq \varphi_2 < 1 - |\varphi_1|$).

1. Si les racines λ_1, λ_2 de $0 = \lambda^2 - \phi_1 \lambda - \phi_2 = \lambda^2 \varphi(\lambda^{-1})$ (qui sont dedans le cercle unitaire, par la causalité) sont distinctes, on obtient

$$\rho_k = \frac{\rho_1 - \lambda_2}{\lambda_1 - \lambda_2} \lambda_1^k + \frac{\lambda_1 - \rho_1}{\lambda_1 - \lambda_2} \lambda_2^k = \rho_1 \psi_{k-1} + \varphi_2 \psi_{k-2}, \quad k \geq 1$$

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

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

termes de racines seulement.

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

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

Exercice 4.5.2 AR(2) : Vérifiez si les procesus AR(2) : a) $Y_t = -0.5Y_{t-1} + 0.14Y_{t-2} + \epsilon_t$ et b) $Y_t = -0.6Y_{t-2} + \epsilon_t$ sont stationnaires causals. Montrez, en partant directement du système de Yule-Walker que leur corrélogrammes sont : a) $\rho_k = \frac{17}{129}(0.2)^k + \frac{112}{129}(-0.7)^k, \quad k = 0, 1, 2, \dots$ et b) $\rho_k = \frac{1}{2}k(0.6)^{k/2} \{1 + (-1)^k\} = (0.6)^{k/2} \cos(k\pi/2), \quad k = 0, 1, 2, \dots$

Exercice 4.5.3 Tracez les corrélogrammes pour les processus AR(2) avec : (i) $\varphi_1 = 0.2, \quad \varphi_2 = 0.35$ et (ii) $\varphi_1 = -0.8, \quad \varphi_2 = -0.16$

4.5.1 Inversibilité des processus MA(q)

Définition 4.5.1 Une représentation causale

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

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

$$\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i} \tag{4.16}$$

où $\sum \pi_i^2 < \infty$

Exemple 4.5.2 Le processus MA(1) $Y_t = \epsilon_t + \theta\epsilon_{t-1}$ est inversible ssi $|\theta| < 1$. En effet, comme dans la résolution de la récurrence AR(1), on voit que :

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

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

Exercice 4.5.4 On considère le processus aléatoire MA(2) suivant :

$$X_t = \epsilon_t - \frac{1}{3}\epsilon_{t-1} + \frac{1}{4}\epsilon_{t-2}$$

ou ϵ_t est BB(0, $\sigma^2 = 1$).

1. Calculer la variance de X_t , $\sigma^2 = \mathbb{E}X_t$, et les covariances.
2. Proposer une représentation autoregressive de ce processus. Est-ce qu'elle est causale est convergente ?

Théorème 4.5.3 Un processus MA(q) avec les racines du symbole $\theta(z)$ à l'extérieur du cercle unitaire est inversible, c.-à-d. le bruit peut être représenté sous la forme : $\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ où $\sum |\pi_i| < \infty$ (alternativement, les racines du polynôme caractéristique sont à l'intérieur du cercle unitaire). Dans ce cas :

1. Les coefficients π_i sont les coefficients de la série Taylor de $\pi(z) = \frac{1}{\theta(z)}$.
2. ϵ_t appartient à l'espace linéaire engendré par le passé du signal observé

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

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

4.6 L'estimation des modèles AR(p) par le système de Yule-Walker

Le système de Yule-Walker (4.15) permet aussi l'estimation, à partir des corrélations observées $\rho_i, i = 1, \dots, p$, des coefficients $\varphi_i, i = 1, \dots, p$. L'idée la plus naturelle est d'adopter une approche recursive, en commençant avec $p = 1, 2, \dots$. Pour chaque p , on denotera par

$$\varphi'_i := \varphi_{p,i}, i = 1, \dots, p$$

le coefficient de Y_{t-i} dans l'équation $Y_t = \sum_{i=1}^p \varphi_{p,i} Y_{t-i} + \epsilon_t$, c.-à-d. la solution du problème de moindres carrés de trouver les valeurs $\varphi_1, \varphi_2, \dots, \varphi_p$ qui minimisent l'objectif $E[\epsilon_t^2] = E(Y_t - \sum_{i=1}^p \varphi_i Y_{t-i})^2$. On calcule la solution $\varphi_{p,i}, i = 1, \dots, p$ du système de Yule-Walker, en s'arrêtant quand le dernier coefficient introduit $\varphi_{p,p}$ est "petit", l'inconnue essentielle étant en effet p .

Définition 4.6.1 Les coefficients

$$\tilde{\rho}_p := \varphi_{p,p}$$

sont appelés autocorrélations partielles et aussi coefficients de Verblunsky.

Exercice 4.6.1 Soit Y_t, Y_{t-1}, \dots, Y_0 une suite des observations à allure stationnaire, qu'on aimerait modéliser comme $AR(p)$, avec p à déterminer.

1. On considère d'abord une modélisation :

$$Y_t = \varphi Y_{t-1} + \epsilon_t, t = 1, 2, \dots, n \quad (4.17)$$

où on propose de prendre pour φ la valeur qui minimise la somme des erreurs carrées $SC(\varphi) = \sum_{i=1}^n \epsilon_i^2$. Quelle est la valeur de φ qui minimise cet objectif ?

2. Supposons maintenant que Y_t sont en effet des observations d'un processus stationnaire $AR(1)$ satisfaisant (9.3), où ϵ_t est un bruit blanc de variance $\sigma^2 = 1$. Quelle est la valeur de φ qui minimise l'objectif $E[\epsilon_t^2]$? Comment estimer cette valeur à partir des données observées ? Comparer avec la réponse précédente.

3. On suppose en suite que Y_t sont des observations d'un processus stationnaire $AR(2)$,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \epsilon_t, t = 2, 3, \dots, n \quad (4.18)$$

où ϵ_t est un bruit blanc de variance $\sigma^2 = 1$. Trouver et résoudre un système d'équations satisfaites par les valeurs φ_1, φ_2 qui minimisent l'objectif $E[\epsilon_t^2]$.

4. Reprenons maintenant la première question, i.e. essayons d'approximer notre suite par le modèle (4.18) par le principe des moindres carrés, sans des hypothèses statistiques sur le bruit.

Trouver un système d'équations satisfaites par les valeurs φ_1, φ_2 qui minimisent l'objectif $SC = \sum_{i=2}^n \epsilon_i^2$. Résoudre le système et comparer avec la réponse précédente.

5. Calculer les 2 premières autocorrélations partielles $\tilde{\rho}_i = \varphi_{i,i}, i = 1, 2$ du processus Y_t , en fonction de coefficients φ_1, φ_2 .

R : 1) Pour $p = 1$, on trouve $\tilde{\rho}_1 = \varphi_{1,1} = \rho_1$.

3) Après avoir déterminé que $\tilde{\rho}_1 = \rho_1$, passons à $p = 2$. La solution du système YW :

$$\begin{aligned} \rho_1 &= \varphi'_1 + \varphi'_2 \rho_1 \\ \rho_2 &= \varphi'_1 \rho_1 + \varphi'_2 \end{aligned}$$

est

$$\varphi'_1 = \varphi_{2,1} = \frac{\begin{vmatrix} \rho_1 & \rho_1 \\ \rho_2 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \quad \varphi'_2 = \varphi_{2,2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \tilde{r}_2$$

Exercice 4.6.2 On accepte finalement un modèle $AR(3)$ avec bruit blanc $\epsilon_t \sim N(0, \sigma^2 = 1)$ (avec coefficients $\varphi_1, \varphi_2, \varphi_3$ déterminés par les données. Donner les équations de Yule-Walker liant les autocorrélations ρ_1, ρ_2, ρ_3 d'un processus $AR(3)$ aux coefficients $\varphi_1, \varphi_2, \varphi_3$. Calculer les 2 premières autocorrélations $\rho_i, i = 1, 2$ du processus Y_t , ainsi que la variance, en fonction de coefficients $\varphi_1, \varphi_2, \varphi_3$.

Calculer la troisième autocorrélation partielle $\tilde{\rho}_3$ du processus Y_t , en fonction de coefficients $\varphi_1, \varphi_2, \varphi_3$.

R : Pour $p = 3$,

$$\varphi_{3,3} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}.$$

Cette expression permet déjà de démontrer que :

Exercice 4.6.3 Pour un processus $AR(2)$, le coefficient $\varphi_{3,3} = 0$, et aussi $\varphi_{4,4} = \varphi_{5,5} = \dots = 0$. Plus généralement, pour un processus $AR(p)$, on a $\varphi_{p+k,p+k} = 0, \forall k \geq 1$.

Remarque 4.6.1 Les équations de Yule-Walker sont juste un cas particulier des équations de la régression multivariée, et φ'_i sont exactement les coefficients de l'autorégression de Y_t sur Y_{t-1}, Y_{t-2}, \dots

Un autre exercice intéressant est de montrer la formule recursive de Szego-Levinson-Durbin :

$$\varphi_{k,k} = \frac{\rho_k - \sum_{i=1}^{k-1} \rho_{k-i} \varphi_{k-1,i}}{1 - \sum_{i=1}^{k-1} \rho_i \varphi_{k-1,i}}, \forall k \geq 1 \quad (4.19)$$

Exercice 4.6.4 (*) Démontrer (4.19), en commençant avec le cas $k = 3$.

Ind : Cherchez une équation pour $\varphi'_3 = \varphi_{3,3}$, en éliminant $\varphi'_1 = \varphi_{3,1}, \varphi'_2 = \varphi_{3,2}$ du système YW :

$$\begin{aligned} \rho_1 &= \varphi'_1 + \varphi'_2 \rho_1 + \varphi'_3 \rho_2 \\ \rho_2 &= \varphi'_1 \rho_1 + \varphi'_2 + \varphi'_3 \rho_1 \\ \rho_3 &= \varphi'_1 \rho_2 + \varphi'_2 \rho_1 + \varphi'_3 \end{aligned}$$

Les deux premières équations donnent

$$\begin{pmatrix} \varphi'_1 \\ \varphi'_2 \end{pmatrix} = R_2^{-1} \begin{pmatrix} \rho_1 - \varphi'_3 \rho_2 \\ \rho_2 - \varphi'_3 \rho_1 \end{pmatrix} = R_2^{-1} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} - \varphi'_3 R_2^{-1} \begin{pmatrix} \rho_2 \\ \rho_1 \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} - \varphi'_3 R_2^{-1} \begin{pmatrix} \rho_2 \\ \rho_1 \end{pmatrix}$$

Finalement,

$$\varphi'_3 (1 - (\rho_2, \rho_1) R_2^{-1} \begin{pmatrix} \rho_2 \\ \rho_1 \end{pmatrix}) = \rho_3 - (\rho_2, \rho_1) \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \iff \varphi'_3 (1 - (\rho_1, \rho_2) \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}) = \rho_3 - (\rho_2, \rho_1) \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$$

et

$$\varphi'_3 = \frac{\rho_3 - (\rho_2 \varphi_1 + \rho_1 \varphi_2)}{1 - (\rho_1 \varphi_1 + \rho_2 \varphi_2)}$$

Les coefficients $\varphi_{k,k}, \forall k \geq 1$ étaient bien connues aux mathématiciens sous le nom des coefficients de Verblunsky. En statistique, ils sont appelés coefficients d'autocorrélation partielle, due à la propriété suivante :

Exercice 4.6.5 Montrer que pour un processus stationnaire

$$\varphi_{k,k} = E[(Y_k - E[Y_k/Y_{k-1}, Y_{k-2}, \dots, Y_1])(Y_0 - E[Y_0/Y_{k-1}, Y_{k-2}, \dots, Y_1])] / \gamma_0$$

Vérifiez que cette formule donne la même valeur pour $\varphi_{2,2}$ comme celle de l'exercice 4.6.1.

Chapitre 5

Les modèles ARMA(p,q)

Définition 5.0.2 On appelle processus ARMA(p,q) un processus stationnaire $Y_t, t \in \mathbb{Z}$ vérifiant une relation de récurrence :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (5.1)$$

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

La notation des polynômes de retard ramène (5.1) à la forme :

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

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

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

Ces représentations peuvent être obtenues par des **inversions formelles** de l'équation (5.1), suivies par un développement de la fraction correspondante dans une série des puissances :

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

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

5.1 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 $\varphi(z)$ et $\theta(z)$ sont à l'extérieur du cercle unitaire :

Théorème 5.1.1 a) Un processus ARMA(p,q) avec toutes les racines du polynôme caractéristique $\varphi(z)$ à l'extérieur du cercle unitaire est causal, c.-à-d. il peut être représenté sous la forme : $Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ où $\sum |\psi_i| < \infty$ et donc Y_t appartient au espace linéaire engendré par le passé du bruit

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

Les coefficients ψ_i sont dans ce cas les coefficients de la série Taylor de $\psi(z) = \frac{\theta(z)}{\varphi(z)}$ b) Un processus ARMA(p,q) avec les racines du polynôme caractéristique $\theta(z)$ à l'extérieur du cercle unitaire est inversible, c.-à-d. le bruit peut être représenté sous la forme : $\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ où $\sum |\pi_i| < \infty$ et donc ϵ_t appartient au espace linéaire engendré par le passé du signal observé

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

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

Corollaire 5.1.1 Pour un processus ARMA(p,q) avec toutes les racines des polynômes caractéristiques $\varphi(z), \theta(z)$ à l'extérieur du cercle unitaire, les espaces linéaires engendrés par le bruit et le passé du signal coïncident :

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

et

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

Rémarque : Ce corollaire permettra un développement immédiat d'une approche de prévision (=régression) par projection dans l'espace de Hilbert engendré par le passé. En conclusion, comme du point de vue pratique les développements Laurent sont inacceptables pour la prévision (parce-qu'elles impliquent les valeurs futures, imprévisibles du bruit), nous allons considérer désormais surtout les modèles ARMA(p,q) avec toutes les racines de $\varphi(z)$ et $\theta(z)$ à l'extérieur du cercle unitaire, qui sont causales et inversibles, et on s'appuyera sur le corollaire ci-dessus.

Exercice 5.1.1 ARMA(1,1)

1. Trouver la représentation MA(∞) d'un processus ARMA(1,1) causal

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

2. Trouver la représentation AR(∞) d'un processus invertible ARMA(1,1).
3. Soit Y_t un processus ARMA(1,1) vérifiant l'équation $Y_t - 0.5Y_{t-1} = \epsilon_t + 0.4\epsilon_{t-1}$ avec ϵ_t un bruit blanc. Précisez si le processus est stationnaire, causal et inversible, et calculez sa fonction d'autocovariance.
4. Trouvez les coefficients ψ_j de sa représentation comme processus MA(∞) et les coefficients π_j de sa représentation comme processus AR(∞) et précisez si ces représentations sont convergentes.

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

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

Sol : $\{1, -\theta(1) - \phi(1), \theta(1)^2 + \phi(1)\theta(1), -\theta(1)^3 - \phi(1)\theta(1)^2\}$

Rémarque : Dans le cas le plus simple avec les racines λ_i de l'équation $\varphi(z) = 0$ distinctes, on obtient facilement des formules générales pour les coefficients ψ_n en commençant par un développement en fractions simples $\pi(z) = \frac{\varphi(z)}{\theta(z)} = \sum_i K_i \frac{1}{1-z/\lambda_i}$ où λ_i sont les racines du $\theta(z)$ et donc $K_i = -\frac{\theta(\lambda_i)}{\varphi'(\lambda_i)}$. On arrive à : $\psi_n = \sum_{i=1}^p \frac{K_i}{\lambda_i^{n+1}}$. Des formules pareilles existent pour π_n , et dans le cas des racines non-distinctes. Dans le cas des racines non-distinctes et complexes, il est préférable d'aborder le développement Taylor $\frac{\theta(z)}{\varphi(z)} = \psi(z)$ directement, en obtenant des équations de récurrence pour ψ_k , à partir des coefficients du développement $\varphi(z) \psi(z) = \theta(z)$. Cette méthode, est en effet applicable toujours :

L'exercice ci-dessus peut être généralisé :

Théorème 5.1.2 (*) a) Pour un processus ARMA(p,q) $\varphi(B)Y_T = \theta(B)\epsilon_t$ avec toutes les racines du polynôme caractéristique $\varphi(z)$ à l'extérieur du cercle unitaire, les coefficients $\psi_i = \sigma^{-2}\mathbb{E}Y_t\epsilon_{t-i}$ de la représentation causale $Y_t = \sum \psi_i\epsilon_{t-i}$ satisfont la récurrence

$$\varphi(B)\psi_k = \theta_k \iff$$

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

$$\iff \varphi(B)\psi(k) = \theta(k)$$

Note : Cette dernière équation est une **récurrence de Yule-Walker généralisé** Ainsi,

$$\begin{cases} \psi_1 = \theta_1 + \varphi_1 \\ \psi_2 = \theta_2 + \varphi_1(\theta_1 + \varphi_1) + \varphi_2 \\ \psi_3 = \theta_3 + \varphi_3 + 2\varphi_1\varphi_2 + \varphi_1^3 + (\varphi_2 + \varphi_1^2)\theta_1 + \varphi_1\theta_2 \\ \dots \end{cases} \quad \text{b) Pour un processus ARMA(p,q) } \varphi(B)Y_T =$$

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

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

$$\iff \theta(B)\pi(k) = -\varphi(k)$$

$$\text{Ainsi, } \begin{cases} \pi_1 = -\phi(1) - \theta(1) \\ \pi_2 = -\phi(2) + \theta(1)^2 + \phi(1)\theta(1) - \theta(2) \\ \pi_3 = -\phi(3) - \theta(1)^3 - \phi(1)\theta(1)^2 + 2\theta(2)\theta(1) + \phi(2)\theta(1) - \theta(3) + \theta(2)\phi(1) \end{cases}$$

Dem : Formellement, il suffit de projeter les relations $\varphi(B)\psi(B) = \theta(B)$, $\theta(B)\pi(B) = \varphi(B)$ sur la composante k .

Note : Les représentations inverse/causale permettent d'appliquer aux processus ARMA(p,q) les méthodes adaptés aux modèles AR(∞)/MA(∞). En particulier, on a la formule de prédiction

$$\begin{aligned}\hat{X}_t &:= E[X_t/F_{t-1}] = - \sum_{i=1} \pi_i X_{t-i} = \\ &(\phi(1) + \theta(1))X_{t-1} + (\phi(2) - \theta(1)^2 - \phi(1)\theta(1) + \theta(2))X_{t-2} + \\ &(\phi(3) + \theta(1)^3 + \phi(1)\theta(1)^2 - 2\theta(2)\theta(1) - \phi(2)\theta(1) - \theta(2)\phi(1) + \theta(3))X_{t-3} + \dots\end{aligned}$$

La formule est plus compliquée à lire, mais, finalement, pour un logiciel symbolique, pas plus compliquée à dériver que son cas particulier AR(p)

$$\hat{X}_t = \phi(1)X_{t-1} + \phi(2)X_{t-2} + \phi(3)X_{t-3} + \dots$$

Théorème 5.1.3 (*) La fonction génératrice des covariances $\gamma(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$ d'un processus ARMA est

$$\gamma(z) = \frac{\theta(z)\theta(z^{-1})}{\varphi(z)\varphi(z^{-1})}$$

Pour $z = e^{itx}$, $x \in \mathbb{R}$ on obtient la transformée Fourier des corrélations aussi appelée densité spectrale

$$f(\theta) = \frac{1}{2\pi} \gamma(e^{itx})$$

La densité spectrale est nonnégative, ce qui est équivalent au fait que les matrices Toeplitz \mathbf{R} des corrélations sont positives définies.

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

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

$$\gamma_k = \sum_{i=1}^p \varphi_i \gamma_{k-i} + \sum_{0 \leq j \leq q} \theta_j \gamma_{k-j}^{(Y,\epsilon)} = \sum_{i=1}^p \varphi_i \gamma_{k-i} + \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{k-j} \quad (5.2)$$

Pour appliquer la recursion, il faut obtenir les coefficients ψ_j , $j = 1, \dots, p$ et aussi les p valeurs initiales $\gamma_q, \dots, \gamma_{q-p+1}$, qu'on trouvera en utilisant les équations (5.2) et la symétrie de γ_k . Rappel : Pour un processus ARMA(p,q) causal, les coefficients $\psi_i = \sigma^{-2} \mathbb{E}Y_t \epsilon_{t-i}$ de la représentation causale $Y_t = \sum \psi_i \epsilon_{t-i}$ satisfont la récurrence

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

Exercice 5.1.2 ARMA(1,1)

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

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

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

En général, nous trouvons :

Théorème 5.1.4 (*) Les premières $p + 1$ covariances s'obtiennent du système à $p + 1$ équations et $p + 1$ inconnues :

$$\mathbf{\Gamma} \begin{pmatrix} 1 \\ -\varphi_1 \\ -\varphi_2 \\ \dots \\ -\varphi_p \end{pmatrix} = \sigma^2 \mathbf{\Psi} \begin{pmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_q \end{pmatrix} \quad (5.3)$$

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

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

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

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

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

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

5.2 Les modèles ARIMA(p,d,q)

Définition 5.2.1 On appelle processus ARIMA(p,d,q) un processus non stationnaire X_t pour lequel le processus différencié d'ordre d, $Y_t = (1 - B)^d X_t, t \in \mathbb{Z}$ est stationnaire, et vérifie une relation de récurrence ARMA(p,q) :

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=0}^q \theta_i \epsilon_{t-i}, \forall t \in \mathbb{Z} \quad (5.4)$$

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

La notation des polynômes de retard ramène (5.4) à la forme :

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

où $\phi(B), \theta(B)$ sont des polynômes relativement primes dans l'opérateur de retard B à ordres p, q avec coefficient libre 1, et avec racines dehors le cercle unitaire.

Formellement, il s'agit des processus ARMA ayant aussi la racine 1, et nous verrons qu'en effet, la prévision des processus ARIMA(p,d,q) est donnée par les mêmes formules que celle des processus stationnaires ARMA(p,q).

5.3 Contrôle continu d'entraînement

1. Trouvez les coefficients ψ_j de la représentation comme processus $MA(\infty)$ du processus $AR(2)$ défini par : $Y_t - \frac{1}{2}Y_{t-1} - \frac{3}{16}Y_{t-2} = \epsilon_t$ Est-ce que le processus est causal ?
2. Déterminer un filtre (une moyenne mobile) causal $\theta(B) = \sum_{i=0}^q \theta_i B^i$ d'ordre q minimal, qui laisse passer une tendance quadratique sans distortion et qui enlève les composantes saisonnières d'ordre 3.
3. Soit $X_t = \sum_{i=0}^k \theta_i \epsilon_{t-i}$ un processus moyenne mobile, utilisé pour lisser un bruit blanc $\epsilon_t = BB(0, \sigma^2 = 1)$. a) Quelle est la variance de X_t ? b) Trouvez le filtre (c.-à-d. les coefficients $(\theta_0, \dots, \theta_k)$) qui ramènent à un processus X_t à variance minimale, parmi toutes les moyennes mobiles d'ordre k qui laissent passer une tendance constante sans distortion.
4. On considère le processus aléatoire suivant :

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

ou ϵ_t est $BB(0, \sigma^2 = 1)$.

- (a) Calculer l'espérance de X_t , $m = \mathbb{E}X_t$, en supposant que le processus est stationnaire.
 - (b) Quelle est l'équation de récurrence satisfaite par le processus $Y_t = X_t - m$?
 - (c) Est-ce que notre processus Y_t est stationnaire causal ? Si oui, donner les coefficients ψ_k de sa représentation causale $Y_t = \sum_k \psi_k \epsilon_{t-k}$.
 - (d) Donner les prévisions $\hat{X}_t(k)$, en fonctions des racines λ_i du polynôme dual de Y_t (définies par $\varphi(B) = (1 - \lambda_1 B)(1 - \lambda_2 B)$).
 - (e) Donner les équations de Yule-Walker pour les autocorrélations du processus Y_t , calculer les 3 premières autocorrélations, ainsi que la variance.
 - (f) Calculer les 3 premières autocorrélations partielles.
5. La série binomiale négative
- (a) Calculer les développements limités autour de $x = 0$ de $\frac{1}{(1-x)^2}$ et de $\frac{1}{(1-x)^3}$, en différenciant $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$.

(b) Quelle est le developpement limité de la "fonction binomiale negative" $\frac{1}{(1-x)^3}$?

6. Est-ce que le processus Y_t défini par :

$$\left(1 - \frac{3}{2}B + \frac{3B^2}{4} - \frac{B^3}{8}\right)Y_t = \epsilon_t$$

est stationnaire causal ? Si oui, donner une formule générale pour les coefficients ψ_k de sa représentation comme processus $MA(\infty)$.

Solutions :

1. $1 - \frac{1}{2}z - \frac{3}{16}z^2 = (1 - \frac{3}{4}z)(1 + \frac{1}{4}z)$. $\psi_k = (\frac{3}{4})^{k+1} - (-\frac{1}{4})^{k+1}$ Oui, le processus est causal, car les racines duales $\lambda_1 = \frac{3}{4}, \lambda_2 = -\frac{1}{4}$ sont dedans le cercle unitaire $|\lambda| \leq 1$.
2. En demandant que $\theta(B) = \frac{1+B+B^2}{3}(a_0 + a_1B + a_2B^2)$ satisfait $\theta(1) = 1, \theta'(1) = 0, \theta''(1) = 0$, on trouve : $\theta(B) = 1/9(8 - 7B + 2B^2)(1 + B + B^2) = 8/9 + B/9 + B^2/3 - 5B^2/9 + 2B^4/9$.
3. a) $\text{Var } X_t = \sum_i \theta_i^2$. b) Elle est minimisé sous la contrainte $\sum_i \theta_i = 1$ par la moyenne arithmétique à coefficients égaux.
4. (a) Ce processus AR(2) pas centré peut s'écrire $(1 - .2B - .35B^2)X_t = (1 - .7B)(1 + .5B)X_t = 45 + \epsilon_t$ En prenant espérance on trouve $\mathbb{E}(X_t) = \frac{45}{1-.2-.35} = \frac{40}{.45} = 100$.
 (b) Soit $Y_t = X_t - \mathbb{E}X_t$. Alors Y_t est un processus AR(2) centré satisfaisant $(1 - .2B - .35B^2)Y_t = (1 - .7B)(1 + .5B)Y_t = \epsilon_t$
 (c) Pour un processus stationnaire causal (avec les racines du polynôme $\varphi(B)$ á l'extérieur du disque unité), les coefficients $\psi(k)$ satisfont les 'équations de Yule Walker

$$\psi_k = \varphi_1\psi_{k-1} + \varphi_2\psi_{k-2}, \psi_{-1} = 0, \psi_0 = 1.$$

Avec des racines duales distinctes λ_1, λ_2 , la solution est :

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

- (d) En resolvant la récurrence (6.1), où par (6.3), on trouve ...
- (e) La fonction d'autocovariance de Y_t (qui est la même que celle de X_t) est obtenue de l'équation : $\mathbb{E}(Y_t Y_{t-h}) = 0.2\mathbb{E}(Y_{t-1} Y_{t-h}) + 0.35\mathbb{E}(Y_{t-2} Y_{t-h}) + \mathbb{E}(\epsilon_t Y_{t-h})$, ce qui donne l'équation Yule Walker

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

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

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

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

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

5.4 Contrôle continu

1. (a) Trouvez un filtre causal $\theta(B) = \sum_{i=0}^2 \theta_i B^i$ qui conserve les polynômes de degré ≤ 1 et qui enlève les composantes saisonnières d'ordre 2.
 - (b) Trouvez un filtre causal $\theta(B) = \sum_{i=0}^3 \theta_i B^i$ qui conserve les polynômes de degré ≤ 1 , qui enlève les composantes saisonnières d'ordre 2, et qui minimise la variance des moyennes mobiles $Y_t = \theta(B)\epsilon_t$.
2. Est-ce que le processus Y_t défini par :

$$(1 - B + \frac{B^2}{4})Y_t = \epsilon_t$$

est stationnaire causal? Si oui, donner les coefficients ψ_k de sa représentation comme processus $MA(\infty)$.

3. On considère le processus aléatoire suivant :

$$X_t = 20 - \frac{2}{5} X_{t-1} + \frac{3}{25} X_{t-2} + \epsilon_t$$

ou ϵ_t est $BB(0, \sigma^2 = 1)$.

- (a) Calculer l'espérance $m = \mathbb{E}X_t$ de X_t , en supposant que le processus est stationnaire. Quelle est l'équation de récurrence satisfaite par le processus $Y_t = X_t - m$?
 - (b) Donner les équations de Yule-Walker pour les autocorrélations d'un processus AR(2). Calculer les 3 premières autocorrélations $\rho_i, i = 1, 2, 3$ du processus X_t , ainsi que la variance.
4. Soit Y_n, Y_{n-1}, \dots, Y_0 une suite qu'on suppose être bien approximée comme AR(1)

$$Y_t = \varphi Y_{t-1} + \epsilon_t, t = 1, 2, \dots, n$$

- (a) On propose de prendre pour φ la valeur qui minimise la somme des erreurs $E = \sum_{t=1}^n \epsilon_t^2$. Quelle est la valeur de φ qui minimise cet objectif?
- (b) On suppose maintenant que notre suite peut être bien approximée comme AR(2)

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \epsilon_t, t = 2, 3, \dots, n$$

Trouver un système d'équations satisfaites par les valeurs φ_1, φ_2 qui minimisent l'objectif $E = \sum_{i=2}^n \epsilon_t^2$?

Solutions :

1. (a) $\varphi(z) = 1 - a(1-z)^2$ et $\varphi(-1) = 0$ donnent $\varphi(z) = 1 - \frac{(1-z)^2}{4} = \frac{1}{2} \left(\frac{3}{2} - \frac{z}{2} \right) (z+1) = -\frac{z^2}{4} + \frac{z}{2} + \frac{3}{4}$

- (b) Il faut résoudre
$$\begin{cases} \min \sum_{i=0}^3 \theta_i^2 \\ \sum_{i=0}^3 \theta_i = 1 \\ \sum_{i=1}^3 i\theta_i = 0 \\ \sum_{i=0}^3 \theta_i (-1)^i = 0 \end{cases}$$
 Il est plus simple d'utiliser une des deux paramétrisations utilisant les factorisations connues

$$\theta(z) = (1+z)(a_0 + a_1z + a_2z^2) \iff \theta(-1) = 0, \quad \theta(z) = 1 + (z-1)^2(b + b_1z),$$

surtout la deuxième. Alors, $\theta(-1) = 0 \implies b_1 = b + 1/4 \implies \theta(z) = 1 + (z-1)^2(b + (b+1/4)z) = (b + \frac{1}{4})z^3 - (b + \frac{1}{2})z^2 - (b - \frac{1}{4})z + b + 1$. Il faut minimiser $(b + \frac{1}{4})^2 + (b + \frac{1}{2})^2 + (b - \frac{1}{4})^2 + (b+1)^2$, le minimum est quand $2b + b + \frac{1}{2} + b + 1 = 0$, $b = -\frac{3}{8}$, et le filtre est $\frac{1}{2}(z+1)\left(\frac{5}{4} - \frac{z^2}{4}\right) = -\frac{z^3}{8} - \frac{z^2}{8} + \frac{5z}{8} + \frac{5}{8}$. On peut vérifier que la variance $\frac{13}{16}$ a été réduite par rapport à $\frac{7}{8}$ dans la question précédente.

2. Un processus est stationnaire causal (avec les racines du polynôme $\varphi(B)$ à l'extérieur du disque unité) ssi les coefficients $\psi(k)$ définies par l'équation Yule Walker

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

avec valeurs initiales $\psi_{-1} = 0, \psi_0 = 1$ satisfont $\sum_k \psi^2(k) < \infty$. La solution dans le cas des racines distinctes est

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

et avec racines confondues

$$\psi(k) = (k+1)\lambda^k$$

(en résolvant la récurrence YW, ou en différenciant $\frac{1}{(1-x)^1} = \sum_{k=0}^{\infty} x^k$). Ici, $\psi_k = \frac{k+1}{2^k}$

3. (a) Ce processus AR(2) pas centré peut s'écrire $(1 + \frac{2}{5}B - \frac{3}{25}B^2)X_t = (1 - \frac{1}{5}B)(1 + \frac{3}{5}B)X_t = 20 + \epsilon_t$. En prenant espérance on trouve $\mathbb{E}(X_t) = \frac{20}{1 + \frac{2}{5} - \frac{3}{25}} = \frac{125}{8}$. Soit $Y_t = X_t - \mathbb{E}X_t$. Alors Y_t est un processus AR(2) centré satisfaisant $(1 + \frac{2}{5}B - \frac{3}{25}B^2)Y_t = (1 - \frac{1}{5}B)(1 + \frac{3}{5}B)Y_t = \epsilon_t$
- (b) La fonction d'autocorrélation de Y_t (qui est la même que celle de X_t) satisfait l'équation Yule Walker

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

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

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

La première équation donne $\rho_1 = -\frac{5}{11} = -.45$, et la rec. YW donne $\rho_2 = \frac{83}{275} = .302$, $\rho_3 = -\frac{241}{1375} = -.175$. Finalement, $\gamma_0 = \frac{1}{1 - \sum \varphi_i \rho_i} = 1.2788$.

Chapitre 6

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

6.1 Prévision des processus ARIMA(p,d,0)

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

Exemple 6.1.1 La prévision linéaire $X_t(k)$ pour le processus ARIMA(0,1,0) à moyenne c satisfait la recursion Yule Walker

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

et est donc constante

$$X_t(k) = X_t$$

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

Exemple 6.1.2 Déduisez la formule de prévision Box-Jenkins pour un processus ARIMA(1,1,0) Calculez la limite $\lim_{k \rightarrow \infty} X_t(k)$ pour un processus ARIMA(1,1,0)

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

Proposition 6.1.1 La fonction de prévision "eventuelle" de Box-Jenkins pour les processus ARIMA(p,d,q) est un élément de l'espace linéaire des solutions de la recursion $\varphi(B)X_t(k)$, pour $k > q$.

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

Exemple 6.1.3 On considère un processus $\{X_t\}$ pour lequel la série différenciée deux fois est un bruit blanc, c'est à dire $\{X_t\}$ est un processus ARIMA(0,2,0). Montrez que la fonction de prévision Box-Jenkins est donnée par

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

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

Définition 6.1.1 Les dernières $p + d$ valeurs $(X_t(q), X_t(q - 1), \dots, X_t(q - d - p + 1))$ qui précèdent $X_t(q)$ (donc avant le point où la recursion de Yule Waker devient valable) s'appellent les **valeurs pivots**.

Il suit clairement que :

Proposition 6.1.2 La prévision "eventuelle" de Box-Jenkins pour les processus ARIMA(p, d, q) est la fonction dans l'espace linéaire des solutions de la recursion $\varphi(B)X_t(k)$ qui passe par les valeurs pivots.

Corollaire 6.1.1 La prévision linéaire $X_t(k)$ pour le processus ARIMA($0, d, 0$) est donnée par le polynôme d'ordre $d - 1$ qui passe par les d dernières points.

6.2 Prévision des processus MA(1)

Pour la prévision linéaire $\hat{X}_t(k)$ des processus MA(q), l'idée est de convertir à la représentation autoregressive AR(∞) $\epsilon_t = X_t + \sum_{i=1}^{\infty} \pi_i X_{t-i}$, qui nous fourni une formule explicite :

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

Exemple 6.2.1 Pour le processus MA(1), on verifie facilement que $X_t(k) = 0$ pour $k = 2, 3, \dots$, (pour une généralisation, voir le théorème 6.7.1). Pour $k = 1$, la définition :

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

donne :

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

Pour se debarasser de ϵ_t , on peut utiliser

$$\epsilon_t = X_t - \theta\epsilon_{t-1},$$

qui ramène a la représentation :

$$\epsilon_t = \sum_{i=0}^{t-1} (-1)^i \theta^{i+1} X_{t-i} - (-\theta)^{t+1} \epsilon_0 = \dots = \sum_{i=0}^{\infty} (-1)^i \theta^i X_{t-i} = X_t + \sum_{i=1}^{\infty} (-1)^i \theta^i X_{t-i}$$

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

$$\begin{aligned} X_t(1) &= \hat{X}_t = \sum_{i=0}^{\infty} (-1)^i \theta^{i+1} X_{t-i} \\ X_t(k) &= 0, k = 2, 3, \dots \end{aligned}$$

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

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

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

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

6.3 Prédiction des processus AR(p) par la récurrence de Yule-Walker

Nous considérons maintenant la prédiction des processus stationnaires dans la forme AR(p). La méthode sera "d'appliquer l'opérateur chapeau" dans l'équation linéaire définissant le modèle.

Théorème 6.3.1 Pour un modèle AR(p)

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

tel que le symbole $\varphi(z)$ ne s'annule pas dans le cercle unitaire, le bruit ϵ_t est un bruit d'innovation, c.-à-d. $\mathbb{E}[\epsilon_{t+k}|\mathcal{F}_t] = 0$ si $k > 0$, et les prévisions satisfont la récurrence Yule-Walker :

$$\hat{X}_t(k) := \mathbb{E}[X_{t+k}/\{X_t, X_{t-1}, \dots\}] = \sum_{i=1}^p \varphi_i \hat{X}_t(k-i), \quad \hat{X}_t(k) = X_{t+k}, \quad \text{si } k \in \{0, -1, -2, \dots\} \quad (6.1)$$

En particulier,

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

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

Exemple 6.3.1 La prédiction linéaire $X_t(k)$ pour un processus AR(1) à moyenne 0 satisfait la recursion Yule Walker

$$X_t(k) = \varphi X_t(k-1), \quad X_t(0) = X_t,$$

étant donc simplement

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

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

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

Théorème 6.3.2 *Soit*

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

un processus $AR(2)$, avec $\lambda_1 \neq \lambda_2$ étant les inverses des racines de $\varphi(z) = 0$. Montrez que les prévisions Box-Jenkins $X_t(k)$ au temps t sont de la forme :

$$X_t(k) = A_1(t)\lambda_1^k + A_2(t)\lambda_2^k = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2}X_t - \lambda_1\lambda_2 \frac{(\lambda_1^k - \lambda_2^k)}{\lambda_1 - \lambda_2}X_{t-1} \quad (6.2)$$

où encore

$$X_t(k) = \psi_k X_t - \lambda_1 \lambda_2 \psi_{k-1} X_{t-1}, k = 0, 1, \dots \quad (6.3)$$

Démonstration : Les prévisions $\hat{X}_t(k)$ satisfont la récurrence Yule Walker :

$$\hat{X}_t(k) - \varphi_1 \hat{X}_t(k-1) - \varphi_2 \hat{X}_t(k-2) = \hat{X}_t(k) - (\lambda_1 + \lambda_2) \hat{X}_t(k-1) + \lambda_1 \lambda_2 \hat{X}_t(k-2) = 0 \quad (6.4)$$

et donc

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

avec valeurs initiales $\hat{X}_t(0) = X_t, \hat{X}_t(-1) = X_{t-1}$. La solution générale d'une récurrence d'ordre 2 $\psi(k) = \varphi_1 \psi(k-1) + \varphi_2 \psi(k-2)$, pour valeurs initiales arbitraires $\psi(0) = x_0, \psi(1) = x_1$ est :

1. avec des racines distinctes λ_1, λ_2 :

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

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

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

Aussi, en termes des valeurs initiales $\psi(0) = x_0, \psi(-1) = x_{-1}$, la solution quand $\lambda_1 \neq \lambda_2$ est : $\psi_k = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} x_0 - \frac{(\lambda_1^{k+1} \lambda_2 - \lambda_2^{k+1} \lambda_1)}{\lambda_1 - \lambda_2} x_{-1}$, d'où le résultat.

Remarque 6.3.1 *Dans le cas causal, $\lambda_i < 1$, et donc la prévision eventuelle*

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

est toujours 0.

Il est facile d'étendre cette approche pour tous les processus autorégressifs $AR(p)$ à ordre p fini, et d'obtenir des formules explicites de prévision en termes de racines de l'équation $\varphi(z) = 1$, ou des dernières valeurs observées $X_t, X_{t-1}, \dots, X_{t-p+1}$.

Théorème 6.3.3 *La fonction génératrice $X(z) := \sum_0^\infty z^k \hat{X}_t(k)$ des prévisions d'un processus $AR(p)$ est*

$$X(z) = \frac{X_t + z(\varphi_2 X_{t-1} + \varphi_3 X_{t-2} + \dots) + z^2(\varphi_3 X_{t-1} + \varphi_4 X_{t-2} + \dots) + \dots z^{p-1} \varphi_p X_{t-1}}{1 - \varphi(z)}$$

Exemple 6.3.2 Obtenir les prédictions $\hat{X}_t(k)$ pour le AR(2) $(1 - .2B - .35B^2)X_t = (1 - .7B)(1 + .5B)X_t = \epsilon_t$

Sol : on a $\hat{X}_t(1) = \varphi_1 X_t + \varphi_2 X_{t-1}$, $\hat{X}_t(2) = (\varphi_1^2 + \varphi_2)X_t + \varphi_1 \varphi_2 X_{t-1}, \dots$ et en général is s'agit des coefficients du developpement limité du théorème 6.3.3

$$\frac{X_t + z\varphi_2 X_{t-1}}{1 - \varphi(z)} = X_t + z(\varphi_1 X_t + \varphi_2 X_{t-1}) + z^2((\varphi_1^2 + \varphi_2)X_t + \varphi_1 \varphi_2 X_{t-1}) + \dots$$

On peut aussi résoudre les équations YW, ou utiliser la formule (6.3).

6.4 Prédiction des processus ARMA(p,1)

Exercice 6.4.1 On considère le processus ARMA(1,1) à moyenne 0 $(1 - \varphi B)Y_t = (1 + \theta B)\epsilon_t$ où $-1 < \varphi < 1$ et $-1 < \theta < 1$.

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

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

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

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

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

Exercice 6.4.2 Est-ce que les résultats ci-dessus restent vrais si $\varphi = 1$, donc pour le processus ARIMA(0,1,1) (appelé aussi IMA(1,1)), défini par :

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

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

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

(à vérifier).

- Montrez qu'on peut le représenter :

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

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

Note : La dernière formule est appelée lissage exponentiel, au moins quand $\theta \in (-1, 0)$ et donc $\alpha = 1 + \theta \in (0, 1)$. La formule donne une moyenne pondérée : $\bar{Y}_t = \alpha Y_t + (1 - \alpha)\bar{Y}_{t-1}$ α s'appelle constante de lissage. **Rémarques :** 1) Plus α est petit, plus la nouvelle série est lisse et les valeurs passées ont un plus grand poids dans la prévision. Quand α est proche de 1 les valeurs les plus récentes ont le poids le plus important. 2) On peut voir la prévision Box-Jenkins comme une généralisation du lissage exponentiel, en utilisant des paramètres estimés à partir des données (au-lieu de ad-hoc).

6.5 Prévision des processus ARIMA(p,d,q)

En conclusion, pour la prévision linéaire $\hat{X}_t(k)$ des processus ARIMA(p,d,q), on aura toujours besoin d'une estimation de $\epsilon_{t-1}, \epsilon_{t-2}, \dots$, ou au moins de $\epsilon_{-1}, \epsilon_{-2}, \dots$, c.-à-d. du "bruit inobservable passé" du modèle. On peut aussi recourir à la représentation AR(∞), dans quel cas on aura besoin de X_{-1}, X_{-2}, \dots , qui sont aussi inobservables. En plus, le resultat final demandera une approximation des valeurs precedant le debut d'observations 0 ; l'approximation la plus simple dans l'absence des moyennes est $\epsilon_k = Y_k = 0$ pour $k < 0$.

Théorème 6.5.1 Dans le cas d'un modèle ARIMA(p,d,q), la meilleure prévision linéaire au temps t est :

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

où les $\tilde{\varphi}_i$ sont les coefficients du polynôme $\varphi(B)(1-B)^d$ (dans le cas d'un modèle ARMA(p,q) $\tilde{\varphi}_i = \varphi_i$). Pour $k > q$, cette formule est exactement la récurrence homogène Yule-Walker $\varphi(B)\hat{X}_t(k) =$, et donc la prévision sera donnée par la solution de cette équation qui passe par les $p + d$ points pivots. Les inconnues $\hat{\epsilon}_{t-i}, i \geq 0$ peuvent être enlevés en utilisant la représentation inverse "π" du bruit en fonction de la série, ou en utilisant $\hat{\epsilon}_t = Y_t - \hat{Y}_{t-1}(1)$ (les dernières se calculent récursivement). Une estimation arbitraire de $\epsilon_0, \epsilon_{-1}, \dots$ sera nécessaire.

6.6 Exercices

1. Basé sur http://www.statok.wiso.uni-goettingen.de/veranstaltungen/zeitreihen/sommer03/ts_r_in

```
data(LakeHuron)
LH<-arima(LakeHuron,order=c(1,0,1))
tsdiag(LH)
Box.test(LH$residuals,lag=1)
LH.pred<-predict(fit,n.ahead=8)
plot(LakeHuron,xlim=c(1875,1980),ylim=c(575,584))
LH.pred<-predict(fit,n.ahead=8)
lines(LH.pred$pred,col="red")
```

```
lines(LH.pred$pred+2*LH.pred$se,col="red",lty=3)
lines(LH.pred$pred-2*LH.pred$se,col="red",lty=3)
```

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

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

où ε_n est un bruit blanc standard.

3. *Restrictions sur les valeurs des coefficients d'autocorrélation pour les processus MA.* Trouvez, pour le processus $MA(1)$, les valeurs maximales et minimales de la corrélation ρ_1 et les valeurs de θ pour les quelles ces valeurs sont atteintes.

4. a) Déterminez la corrélogramme des processus suivants :

- (i) le processus $MA(2)$ $Y_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$
(ii) le processus $MA(3)$ $Y_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3}$

- b) Calculez et tracez la correlogramme pour les cas :

- (i) $MA(2)$: $\theta_1 = -5/6, \theta_2 = 1/6$
(ii) $MA(2)$: $\theta_1 = 0.8, \theta_2 = 0.5$
(iii) $MA(3)$: $\theta_1 = 0.8, \theta_2 = -0.4, \theta_3 = -0.3$

5. Investiguez si les deux processus $MA(2)$ sont inversibles, en examinant la récurrence obtenue par la méthode des coefficients indéterminés.

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

- (a) Précisez si le processus est stationnaire, causal et inversible.
(b) Trouvez les coefficients ψ_j de sa représentation comme processus $MA(\infty)$ et les coefficients π_j de sa représentation comme processus $AR(\infty)$ et précisez si ces représentations sont convergentes.

7. Mêmes questions pour les processus $ARMA(2,1)$ et $ARMA(2,2)$ définies par : a) $Y_t - \frac{1}{2}Y_{t-1} - \frac{3}{16}Y_{t-2} = \varepsilon_t + 1.25\varepsilon_{t-1}$ b) $(1 - B + \frac{B^2}{4})Y_t = (1 + B + B^2)\varepsilon_t$ *Indication :* Rémarquez que pour le processus $ARMA(p,q)$, la récurrence pour $k > q$ est exactement comme pour le processus $AR(p)$; mais, les premières équations obtenues en comparant les coefficients des puissances en $\psi(B)$ $\varphi(B) = \theta(B)$ pour $k \leq q$ changent par l'ajout de θ_k de côté droite.

8. Soit le processus :

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

- (a) Est-ce que ce processus est stationnaire causal ? Si oui, obtenez la "représentation ψ " du Y_t par rapport au bruit ε_t .
(b) Sous quelles conditions est ce processus inversible ? Obtenez la "représentation π " du bruit ε_t en termes de la série. De quel problème on s'aperçoit si le processus n'est pas inversible ?

9. a) Trouvez les inégalités (il y en a trois) qui définissent la région (triangulaire) du plan (θ_1, θ_2) pour laquelle un processus $MA(2)$ est inversible. Tracez la région sur un graphe. Indiquez le domaine des racines réelles et celui des racines complexes. *Indication* : Les conditions pour avoir racines de module plus grand que 1 sont différentes pour le cas des racines complexes et celui des racines réelles, et pour un polynôme $\theta(z) = 1 + \theta_1 z + \theta_2 z^2$, la condition pour racines réelles de module plus grand que 1 sont plus compliquées que les conditions (équivalentes) que le polynôme "reciproque" $\tilde{\theta}(z) = z^2 \theta(1/z) = z^2 + \theta_1 z + \theta_2$ ait des racines réelles z_i de module plus petit que 1. Pour ce dernier polynôme, les conditions sont :

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

(b) racines réelles : $\tilde{\theta}(1) = 1 + \theta_1 + \theta_2 > 0$, $\tilde{\theta}(-1) = 1 - \theta_1 + \theta_2 > 0$

- b) (*) Pour le processus $MA(2)$, trouvez un domaine S contenant toutes les valeurs possibles des coefficients d'autocorrélation ρ_1, ρ_2 tel que le processus soit inversible, et les valeurs de θ_1, θ_2 pour les quelles les valeurs sur la frontière de S sont atteintes.

Solution : La région d'inversibilité dans le domaine (θ_1, θ_2) :

$$\theta_2 > -\theta_1 - 1$$

$$\theta_2 > \theta_1 - 1$$

$$\theta_2 < 1$$

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

Les racines sont réelles/complexes dessous/dessus la parabole $\theta_2 = \frac{\theta_1^2}{4}$. b) Pour passer de (θ_1, θ_2) à (ρ_1, ρ_2) on utilise

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

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

$$\text{dessous } \rho_1 = 2\sqrt{\rho_2(1 - 2\rho_2)}$$

$$\rho_2 + 1/2 \geq \rho_1$$

$$\rho_2 + 1/2 \geq -\rho_1$$

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

10. (*) Trouver le domaine de causalité dans le plan (φ_1, φ_2) d'un processus $AR(2)$.
11. Obtenez, en partant directement du système de Yule-Walker, les premières cinq corrélations pour un processus $AR(2)$ avec : a) $\phi_1 = 0.6, \phi_2 = -0.2$ b) $\phi_1 = -0.6, \phi_2 = 0.2$ Calculez aussi la variance $\gamma(0)$. Tracez les corrélations. **Sol** : Le domaine de causalité d'un processus $AR(2)$

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

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

12. a) Vérifiez si le processus $AR(2)$ $Y_t = -0.3Y_{t-1} + 0.10Y_{t-2} + \epsilon_t$ est stationnaire causal. Calculez son corrélogramme, en partant directement du système de Yule-Walker, et tracez le. b) Même questions pour le processus $AR(2)$ $Y_t = -Y_{t-1} - 0.34Y_{t-2} + \epsilon_t$.
13. Calculez la fonction d'autocovariance et la fonction d'autocorrélation des processus dans les exercices antérieurs.
14. *Une question d'unicité - est-ce que deux processus distincts peuvent avoir la même FAC (fonction d'autocovariance) ?* Soient $\{u_t, t \in \mathbb{Z}\}$ et $\{v_t, t \in \mathbb{Z}\}$ deux bruit blancs de variances respectives σ^2 et $\theta^2\sigma^2$, $0 < |\theta| < 1$. On considère alors les processus aléatoires $\{X_t, t \in \mathbb{Z}\}$ et $\{Y_t, t \in \mathbb{Z}\}$ tels que :

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

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

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

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

et ϵ_t est bruit blanc.

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

6.7 Intervalles de confiance pour la prédiction

Théorème 6.7.1 (*) Soit $e_t(k) = X_{t+k} - X_t(k)$ l'erreur de prévision de la prédiction linéaire optimale $X_t(k) = \sum_{i=0}^{\infty} a_{t,k}(i) \epsilon_{t-i}$ d'un processus stationnaire causal X_t , et soit $V(k) = \mathbb{E}[X_{t+k} - X_t(k)]^2$ sa variance. Alors, les coefficients de prévision ont donnés par

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

et donc

$$\begin{aligned} X_t(k) &= \sum_{i=0}^{\infty} \psi_{k+i} \epsilon_{t-i}, \text{ l'erreur est} \\ e_t(k) &= \sum_{i=0}^{k-1} \psi_i \epsilon_{t+k-i} \end{aligned}$$

et la variance de ce prédicteur est :

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

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

1. Ce résultat fournit des **intervalles de confiance pour la prévision**.
2. Pour $k=1$, $e_t(1) = \epsilon_t(1)$ et $V(1) = \sigma^2$
3. Ce résultat suggère que pour un processus non stationnaire, la variance de l'erreur de prévision converge vers infini :

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

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

6.8 La prévision linéaire des processus $ARMA(p, q)$

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

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

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

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

Théorème 6.8.1 *La prévision linéaire à variance minimale des processus $ARMA(p, q)$ avec du bruit blanc Gaussien coïncide avec l'espérance conditionnelle*

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

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

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

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

Définition 6.8.1 *Si un processus Y_t défini par une autorégression (4.7) $Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + \epsilon_t$ est tel que la différence $\epsilon_t = Y_t - f(Y_{t-1}, Y_{t-2}, \dots) = Y_t - \hat{Y}_t$ est indépendante du passé $\mathcal{F}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots)$, il sera appelée **bruit d'innovation** (par rapport au passé). Dans ce cas,*

$$\mathbb{E}[\epsilon_t | Y_{t-1}, Y_{t-2}, \dots] = 0 \tag{6.5}$$

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

Rq : On montre facilement que pour un bruit d'innovation, les innovations ϵ_t satisfont aussi $\mathbb{E}\epsilon_t = 0$ et $\mathbb{E}\epsilon_t\epsilon_{t-k} = 0$, pour $k \neq 0$. Donc, les innovations constituent un bruit blanc de deuxième ordre. Le bruit blanc d'une équation autorégressive causale a la propriété très convenable de coïncider avec l'erreur de prédiction par rapport au passé $Y_t - \hat{Y}_t = Y_t - \mathbb{E}[Y_t|Y_{t-1}, Y_{t-2}, \dots]$:

Un résultat fondamental assure que le bruit d'un modèle (4.8) est d'innovation ssi le modèle est causal, et ssi le polynôme caractéristique a seulement des racines plus grandes en valeur absolue que 1. Dans ce cas, on a :

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

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

6.9 Exercices

1. Soit le processus :

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

- Est-ce que ce processus est stationnaire causal ? Si oui, obtenez la "représentation ψ " du Y_t par rapport au bruit Z_t et trouvez la fonction d'autocorrélation de Y_t .
 - Sous quelles conditions est ce processus inversible ? Obtenez la "représentation π " du bruit Z_t en termes de la série. De quel problème on s'aperçoit si le processus n'est pas inversible ?
 - Donnez la prévision k pas en avant $\hat{Y}_t(k)$, en utilisant les valeurs Y_t, Y_{t-1}, Z_t .
2. Trouvez les coefficients d'autocorrélation et tracez la corrélogramme pour le processus ARMA(1,2) $Y_t = 0.6Y_{t-1} + \epsilon_t - 0.3\epsilon_{t-1} - 0.1\epsilon_{t-2}$
3. *Prévision du modèle ARIMA(2,1,0)*

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

- Verifiez que :
 - $X_t(k) = (1 + \phi_1 + \phi_2)X_t(k-1) - (\phi_1 + \phi_2 + \phi_1\phi_2)X_t(k-2) + \phi_1\phi_2 X_t(k-3)$, $k \geq 1$
 - $X_t(k) = A_t + B_t\phi_1^k + (X_t - A_t - B_t)\phi_2^k$, $k = 0, 1, 2, \dots$ pour certaines A_t et B_t qui dépendent seulement du t .
 - Trouvez les poids des valeurs pivots X_t, X_{t-1}, X_{t-2} dans la prévision Box-Jenkins de X_t
 - Trouvez la limite $\lim_{k \rightarrow \infty} X_t(k)$
4. *Prévisions sous le modèle ARIMA(1,1,1)* Considerons le processus ARIMA(1,1,1) $(1 - \varphi B)(1 - B)Y_t = (1 + \theta B)\epsilon_t$, avec $-1 < \varphi < 1$ et $-1 < \theta < 1$.
- Montrez que $Y_t(1) = (1 + \varphi)Y_t - \varphi Y_{t-1} + \theta\epsilon_t$ et $Y_t(k) = (1 + \varphi)Y_t(k-1) - \varphi Y_t(k-2)$, $k \geq 2$.
 - Montrez que $Y_t(k) = A_t + B_t\varphi^k$ pour $k \geq 0$, et trouvez des expressions pour A_t et B_t en termes de $Y_t, Y_{t-1}, \epsilon_t, \varphi$ et θ , en utilisant $Y_t(0)[= Y_t]$ et $Y_t(1)$ du (a) ci-dessus. Montrez que :

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

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

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

$t :$	1	2	3	4	5	6	7	8	9	10
$y_t :$	14.8	12.4	9.4	7.7	7.3	9.0	10.5	11.2	10.4	11.1

trois termes suivants de la série :

$t :$	11	12	13	14	15	16	17	18	19	20
$y_t :$	12.1	11.6	9.9	8.1	6.6	5.4	4.2	5.3	6.8	8.1

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

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

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

- (a) Montrez que $Y_t(1) = (1 + \alpha)Y_t - \alpha Y_{t-1} + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}$ et trouvez les expressions correspondantes pour $Y_t(2)$ et $Y_t(k)$ pour $k \geq 3$
- (b) Montrez que la fonction de prévision peut s'exprimer sous la forme $Y_t(k) = a_t + b_t \alpha^k$, $k \geq 1$, et donnez la formule de a_t, b_t comme fonctions de $Y_t, Y_{t-1}, \epsilon_t, \epsilon_{t-1}$.
- (c) Montrez que $Y_t(k)$ peut s'exprimer en fonction seulement des valeurs passées de la série.
- (d) Un statisticien a utilisé le modèle ARIMA (1,1,2) décrit ci-dessus pour une série (dénommée prix) qui exprime le prix d'une action à la bourse pour 100 jours consécutifs. En sachant que $Y_{98}(1) = 686,996$ et $Y_{99}(1) = 659,416$ et $\sigma_\epsilon = 2$, calculer les prévisions $Y_{101|100}, Y_{102|100}$ de Y_{101} et Y_{102} et donner les 95% intervalles de confiance associés avec ces prévisions.
- (e) Soit $X_m = \sin(2\pi s_1 m) + \sin(2\pi s_2 m) + \epsilon_m, m = 0, 1, \dots, N$. Donner une représentation autoregressive de ce processus. Calculer la densité spectrale, la variance $\sigma^2 = \mathbb{E}(X_t)^2$, et les covariances.
6. **Projet** : Il serait intéressant de déterminer analytiquement "la tendance asymptotique", c.-à-d. le polynôme d'ordre $d - 1$ vers le quel les prévisions convergent asymptotiquement pour les processus ARIMA(p,d,q). Considérons par exemple ARIMA(p,2,0) ; ce modèle inclue une tendance linéaire, pour la quelle le premier candidat est la droite par les deux derniers points pivots (comme dans le cas $p = 0$, quand les pivots coïncident avec les valeurs X_t, X_{t-1}). En général, les prévisions doivent encore converger asymptotiquement vers une droite. Pour $p = 0$, on commence déjà exactement sur la "droite de tendance" (due à l'absence d'autres racines dans la partie autorégressive) ; mais, pour $p \geq 1$, nous serons obligé de tenir compte d'autres valeurs pivots et donc de X_{t-2}, X_{t-3}, \dots . A priori donc, les p points qui précèdent les 2 derniers points auront aussi une influence sur la "droite de tendance".

Chapitre 7

Sommaire des définitions et résultats dans les séries temporelles

7.1 Filtres

Définition 7.1.1 a) Une série s_t sera appelée saisonnière de période p ssi

$$\sum_{i=1}^p s_{t+i} = 0 \iff \left(\sum_{i=0}^{p-1} B^i \right) s_t = 0, \quad \forall t \quad (7.1)$$

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

$$s_{t+p} = s_t \iff (1 - B^p) s_t = 0, \quad \forall t \quad (7.2)$$

Théorème 7.1.1 Un filtre $\psi(B)$ annule (ou "enlève") les composantes saisonnières d'ordre p ssi $\psi(z)$ est divisible par $1 + z + \dots + z^{p-1}$ (donc si $\psi(z) = 0$, pour toutes les racine d'ordre p de l'unité, sauf $z = 1$).

Théorème 7.1.2 L'espace invariant d'un filtre contient les polynômes de degré $\leq p$ ssi 1 est une racine d'ordre au moins $p + 1$ de l'équation $\psi(z) = 1$, i.e. $\psi(1) = 1, \psi'(1) = 0, \psi''(1) = 0, \dots, \psi^{(p)}(1) = 0$.

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

Théorème 7.2.1 a) Un processus ARMA(p) $\varphi(B)Y_t = \theta(B)\epsilon_t$ est causal(inversible), i.e. il peut être représenté sous la forme : $Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ où $\sum \psi_i^2 < \infty$ ($\epsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i}$ où $\sum \pi_i^2 < \infty$) ssi toutes les racines de son polynôme caractéristique $\varphi(z)$ ($\theta(z)$) sont à l'extérieur du cercle unitaire. Les coefficients ψ_i (π_i) sont dans ce cas les coefficients de la série Taylor de $\psi(z) = \frac{\theta(z)}{\varphi(z)}$ ($\pi(z) = \frac{\varphi(z)}{\theta(z)}$)

Théorème 7.2.2 (*) a) Pour un procesus ARMA(p,q) $\varphi(B)Y_T = \theta(B)\epsilon_t$ avec toutes les racines du polynôme caractéristique $\varphi(z)$ à l'extérieur du cercle unitaire, les coefficients

$\psi_i = \sigma^{-2} \mathbb{E}Y_t \epsilon_{t-i}$ de la représentation causale $Y_t = \sum \psi_i \epsilon_{t-i}$ satisfait la récurrence

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

Note : a) Cette dernière équation est appelée **recursion de Yule-Walker**. b) Pour un processus $ARMA(p,q)$ $\varphi(B)Y_T = \theta(B)\epsilon_t$ avec toutes les racines du polynôme caractéristique $\theta(z)$ à l'extérieur du cercle unitaire, les coefficients π_i de la représentation inverse $\epsilon_t = \sum \pi_i Y_{t-i}$ satisfait la récurrence

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

Exemple 7.2.1 Ainsi, $\psi_1 = \varphi_1 + \theta_1, \psi_2 = \varphi_2 + \varphi_1^2 + \theta_1 \varphi_1 + \theta_2, \psi_3 = \varphi_3 + 2\varphi_1 \varphi_2 + \varphi_1^3 + (\varphi_2 + \varphi_1^2) \theta_1 + \varphi_1 \theta_2 + \theta_3, \dots$

7.3 Équations Yule Walker

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

$$\mathbf{R} \boldsymbol{\varphi} = \boldsymbol{\rho} \tag{7.3}$$

ou R est la matrice Toeplitz symétrique :

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

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

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

avec racines λ_1, λ_2 de $0 = \lambda^2 - \varphi_1 \lambda - \varphi_2 = \lambda^2 \varphi(\lambda^{-1})$ dedans le cercle unitaire (pour assurer la causalité), on obtient, en résolvant (7.3) :

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

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

$$\rho_k = \rho_1 \frac{\lambda_1^k - \lambda_2^k}{\lambda_1 - \lambda_2} - \lambda_1 \lambda_2 \frac{\lambda_1^{k-1} - \lambda_2^{k-1}}{\lambda_1 - \lambda_2} = \boxed{\rho_1 \psi_{k-1} + \varphi_2 \psi_{k-2}}, k \geq 1, \rho_1 = \frac{\varphi_1}{1 - \varphi_2}$$

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

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

en termes de racines seulement.

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

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

7.4 Prévision linéaire

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

Dans le cas d'un modèle ARIMA(p,d,q), les $\tilde{\varphi}_i$ sont les coefficients du polynôme $\varphi(B)(1-B)^d$, et dans le cas d'un modèle ARMA(p,q) $\tilde{\varphi}_i = \varphi_i$. Les $\hat{\epsilon}_t$ peuvent être enlevés en utilisant la représentation "π" du bruit en fonction de la série. La recurrence satisfaite par les π_i est $\varphi(B)\hat{Y}_t(k) = \theta(B)\epsilon_{t+k}$ - voir Thm 5.1.2.

Chapitre 8

Arima et Regression en R

Commandes : `lm`, `cycle`, `factor`, `resid`, `factor`, `confint`, `arima`

Ce chapitre est basé sur le livre *Introductory Time Series with R*, by Paul S.P. Cowpertwait and Andrew V. Metcalfe. Par contre, l'adresse `http`

`http://ebookey.org/go/?u=http://www.massey.ac.nz/~pscowper/ts`

n'est plus maintenue, et par conséquent les jeux de données doivent être remplacés.

```
Exemple 8.0.1 www<-scan("http://www.massey.ac.nz/~pscowper/ts/Maine.dat")
Maine.month<-read.table(www,header=TRUE)
#attach(Maine.month)
class(Maine.month)
Mts<-ts(unemploy,start=c(1996,1),freq=12)
Mtsa<-aggregate(Mts)/12 ##filtre "moy. arithm"
layout(1:2)
plot(Mts)
plot(Mtsa)
M.feb<-window(Mts,start=c(1996,2),freq=TRUE)
Febratio<-mean(M.feb)/mean(Mts)
Febratio
```

```
www <-"http://www.massey.ac.nz/~pscowper/ts/global.dat"
Glob <- scan(www)
Glob.ts <- ts(Global, st = c(1856, 1), end = c(2005, 12), fr = 12)
plot(Glob.ts)
Gl.an<- aggregate(Glob.ts,FUN= mean)##"moy.arithm"
plot(Gl.an)
temp <- window(Glob.ts, start = 1970)
td<-decompose(temp)
tdr<-ts(td$random,start=c(1970,7),end=c(2004,12),fr=12)
acf(tdr)
pacf(tdr)
tds.ar <- ar(as.ts(tdr), method = "mle")
tdrr<-tds.ar$resid
tempc<-temp-tdrr
```

```

m<-decompose(tempc)
plot(m)
temp.lm <- lm(temp ~ time(temp))
coef(temp.lm)
summary(temp.lm) ## SE=.12, R^2=.67
acf(resid(temp.lm))
pacf(resid(temp.lm))
Seas <- cycle(temp)
fs<-factor(Seas)
Time <- time(temp)
temp.lm <- lm(temp ~ 0 + Time + factor(Seas))
coef(temp.lm)
confint(temp.lm)
new.t <- seq(2006, len = 2 * 12, by = 1/12)
new.dat<-data.frame(Time= new.t,Seas =rep(1:12,2))
pr<-predict(temp.lm, new.dat)[1:24]
plot(pr)

require(graphics)
m <- decompose(co2)
names(m)
plot(m)
mr<-window(m$random,start=c(1960,1),end=c(1996,4))
summary(mr)
layout(1:2)
acf(mr)
pacf(mr)
mrs<-as.ts(mr)
mr.ar <- ar(mrs, method = "mle")
mr.ar

```

8.1 Projets

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

à bruit petit, c) ARMA(2,2) (en passant la sortie du premier program au deuxième), et d) ARIMA(2,1,2) (en appliquant cumsum à la sortie du program anterieur). Enoncez les théorèmes satisfaits par l'acf et le pacf des premiers deux cas, et vérifiez ensuite que vos programmes produisent des résultats adéquats. Pour le troisième et quatrième cas, **estimez** le modèle par la commande arima, avec les ordres simulés, et aussi avec des ordres plus grands. Est-ce que la commande retrouve les coefficients que vous aviez choisi quand le bruit est très petit (en supposant que l'analyse est bonne, et donc que l'acf et pacf des résidus indiquent un bruit blanc) ?

4. Interpretation de l'acf

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

Chapitre 9

Examens

9.1 Examen 1

1. Trouvez un filtre causal $\theta(B) = \sum_{i=0}^3 \theta_i B^i$ qui conserve les polynômes de degré ≤ 2 et qui enlève les composantes saisonnières d'ordre 2.
2. Soit Y_t, Y_{t-1}, \dots, Y_0 une suite d'observations à allure stationnaire, qu'on aimerait modéliser comme AR(p), avec p à déterminer.

(a) On considère d'abord une modélisation :

$$Y_t = \varphi Y_{t-1} + \epsilon_t, t = 1, 2, \dots, n \quad (9.1)$$

où on propose de prendre pour φ la valeur $\hat{\varphi}(Y_0, \dots, Y_n)$ qui minimise la somme des erreurs au carré $SC(\varphi) = \sum_{i=1}^n \epsilon_i^2$. Quelle est la valeur $\hat{\varphi}(Y_0, \dots, Y_n)$?

- (b) Supposons maintenant que Y_t sont en effet des observations d'un processus stationnaire AR(1) satisfaisant (9.3), où ϵ_t est un bruit blanc (centré, non-corrélé) de variance $\sigma^2 = 1$. Quelle est la valeur de φ qui minimise l'objectif $E[\epsilon_t^2]$? Comment estimer cette valeur à partir des données observées ? Comparer avec la réponse précédente.
- (c) On suppose en suite que Y_t sont des observations d'un processus stationnaire AR(2),

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \epsilon_t, t = 2, 3, \dots, n$$

où ϵ_t est un bruit blanc de variance $\sigma^2 = 1$. Trouver et résoudre le système d'équations satisfaites par les valeurs φ_1, φ_2 qui minimisent l'objectif $E[\epsilon_t^2]$.

- (d) On accepte finalement un modèle AR(3) avec bruit blanc ϵ_t ($\sigma^2 = 1$) (avec coefficients $\varphi_1, \varphi_2, \varphi_3$ déterminés par les données). Donner les équations liant les autocorrélations ρ_1, ρ_2, ρ_3 d'un processus AR(3) aux coefficients $\varphi_1, \varphi_2, \varphi_3$ (on ne demande pas de justifier). Calculer les 2 premières autocorrélations $\rho_i, i = 1, 2$ du processus Y_t , en fonction des coefficients $\varphi_1, \varphi_2, \varphi_3$.
- (e) Calculer les autocorrélations partielles $\tilde{\rho}_i = \varphi_{(i,i)}, i = 3, 4$ du processus Y_t , en fonction des coefficients $\varphi_1, \varphi_2, \varphi_3$.

3. On considère des processus X_t, Y_t, Z_t satisfaisant respectivement :

$$(1 - B/2)X_t = \epsilon_t, \quad (1 - B)Y_t = \epsilon_t, \quad (1 - 2B)Z_t = \epsilon_t, \quad t = T, T - 1, \dots, 1, \dots \quad (9.2)$$

où ϵ_t est un bruit blanc de loi $N(0, \sigma^2 = 1)$.

- Donner l'expression de X_T , de Y_T , et de Z_T en fonction de $\epsilon_t, t = 1, 2, \dots, T$, et des valeurs au temps $t = 0$ (denotées par X_0, Y_0, Z_0).
- Est-ce que l'un de ces trois processus, en commençant de 0 (c'est-à-dire avec $X_0 = 0, Y_0 = 0, Z_0 = 0$) est stationnaire ?
- Est-ce que l'une des équations (9.4) admet une solution stationnaire ? Le cas échéant, donner cette solution.

Solutions :

1. $\varphi(z) = 1 - a(1 - z)^3$ et $\varphi(-1) = 0$ donnent $a = \frac{1}{8}$,

$$\varphi(z) = 1 - \frac{(1 - z)^3}{8} = \frac{z^3}{8} - \frac{3z^2}{8} + \frac{3z}{8} + \frac{7}{8} = \frac{1}{8}(z + 1)(z^2 - 4z + 7)$$

2. (a) $\hat{\varphi}(Y_0, \dots, Y_n) = \frac{\sum_{t=1}^n Y_t Y_{t-1}}{\sum_{t=0}^{n-1} Y_t^2}$

(b) $\hat{\varphi} = \frac{\gamma_1}{\gamma_0}$

(c) Pour $p = 2$, on arrive au système YW :

$$\rho_1 = \varphi_1 + \varphi_2 \rho_1$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2$$

avec solution

$$\varphi_{2,1} = \frac{\begin{vmatrix} \rho_1 & \rho_1 \\ \rho_2 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} = \frac{\gamma_0 \gamma_1 - \gamma_1 \gamma_2}{\gamma_0^2 - \gamma_1^2} \quad \varphi_{2,2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\gamma_0 \gamma_2 - \gamma_1^2}{\gamma_0^2 - \gamma_1^2}$$

(remarquer les déterminants de Hankel dans les expressions par rapport au γ_i). On s'aperçoit en résolvant cet exercice que les $\varphi_i, i = 1$, sont exactement les coefficients de la régression de Y_t sur Y_{t-1}, Y_{t-2} (d'où le nom autoregression), et que pour $p = 1$, y a coïncidence entre les valeurs du coefficient de régression φ et de la corrélation ρ_1 , mais pour $p \geq 2$, la situation se complique.

(d) Pour $p = 3$, le système YW est :

$$\rho_1 = \varphi_1 + \varphi_2 \rho_1 + \varphi_3 \rho_2$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2 + \varphi_3 \rho_1$$

$$\rho_3 = \varphi_1 \rho_2 + \varphi_2 \rho_1 + \varphi_3$$

Les deux premières équations

$$\begin{pmatrix} (1 - \varphi_2) & -\varphi_3 \\ -(\varphi_1 + \varphi_3) & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$$

donnent

$$\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \frac{1}{1 - \varphi_2 - \varphi_3(\varphi_1 + \varphi_3)} \begin{pmatrix} 1 & \varphi_3 \\ \varphi_1 + \varphi_3 & 1 - \varphi_2 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \\ = \left\{ \frac{\varphi(1) + \varphi(2)\varphi(3)}{1 - (\varphi(2) + \varphi(3)(\varphi(1) + \varphi(3)))}, \frac{\varphi(1)(\varphi(3) + \varphi(1)) - \varphi(2)^2 + \varphi(2)}{1 - (\varphi(2) + \varphi(3)(\varphi(1) + \varphi(3)))} \right\}$$

(e) On trouve que le troisième coefficient d'autocorrélation partielle est

$$\varphi_{3,3} = \frac{\rho(1)^3 - \rho(3)\rho(1)^2 + \rho(2)^2\rho(1) - 2\rho(2)\rho(1) + \rho(3)}{(1 + \rho(2) - 2\rho(1)^2)(1 - \rho(2))},$$

et une formule encore plus compliquée pour $\varphi_{4,4}$. Mais, en supposant un modèle AR(3), ça simplifie à

$$\varphi_{3,3} = \varphi_3, \varphi_{4,4} = 0$$

3. (a) Résolvant les récurrences donne
- (b) Non
- (c) $X_t = \sum_{k=0}^{\infty} 2^{-k}\epsilon_{t-k}$, $Z_t = \sum_{k=0}^{\infty} 2^{-k}\epsilon_{t+k}$, $t = 0, 1, 2, \dots$ sont des processus stationnaires.

9.2 Examen 2

1. Est-ce que le processus AR(2) :

$$Y_t = -\frac{1}{2}Y_{t-1} + \frac{7}{50}Y_{t-2} + \epsilon_t, \quad \epsilon_t \text{ bruit blanc } (\sigma^2 = 1)$$

est stationnaire causal? Si oui, calculer ses corrélations.

2. a) Calculer le développement limité autour de $x = 0$ de $\frac{1}{(1-x)^3}$, en différenciant $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$. b) Est-ce que le processus Y_t défini par :

$$\left(1 - \frac{3}{2}B + \frac{3B^2}{4} - \frac{B^3}{8}\right)Y_t = \epsilon_t$$

est stationnaire causal? Si oui, donner une formule générale pour les coefficients ψ_k de sa représentation comme processus MA(∞).

3. On cherche un filtre causal $\theta(B) = \sum_{i=0}^n \theta_i B^i$ d'ordre n qui laisse passer une tendance linéaire sans distortion et qui minimise la variance de moyennes mobiles

$$X_t = \sum_{i=0}^n \theta_i \epsilon_{t-i}, \quad \epsilon_t \text{ bruit blanc } (\sigma^2 = 1).$$

Formuler ce problème comme un problème d'optimization avec des contraintes, qu'on résoudra par la méthode des multiplicateurs de Lagrange. **Ind** : On rappelle que $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.

4. Soit Y_t, Y_{t-1}, \dots, Y_0 une suite d'observations à allure stationnaire, qu'on aimerait modéliser comme AR(p), avec p à déterminer. a) On considère d'abord une modélisation :

$$Y_t = \varphi Y_{t-1} + \epsilon_t, t = 1, 2, \dots, n \quad (9.3)$$

où on propose de prendre pour φ la valeur $\hat{\varphi}(Y_0, \dots, Y_n)$ qui minimise la somme des erreurs au carré $SC(\varphi) = \sum_{i=1}^n \epsilon_i^2$. Quelle est la valeur $\hat{\varphi}(Y_0, \dots, Y_n)$? b) Supposons maintenant que Y_t sont en effet des observations d'un processus stationnaire AR(1) satisfaisant (9.3), où ϵ_t est un bruit blanc (centré, non-corrélé) de variance $\sigma^2 = 1$. Quelle est la valeur de φ qui minimise l'objectif $E[\epsilon_t^2]$? Comment estimer cette valeur à partir des données observées? Comparer avec la réponse précédente. c) On suppose en suite que Y_t sont des observations d'un processus stationnaire AR(2),

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \epsilon_t, t = 2, 3, \dots, n$$

où ϵ_t est un bruit blanc de variance $\sigma^2 = 1$. Trouver et résoudre le système d'équations satisfaites par les valeurs φ_1, φ_2 qui minimisent l'objectif $E[\epsilon_t^2]$. d) On accepte finale-

ment un modèle AR(3) avec bruit blanc ϵ_t ($\sigma^2 = 1$) (avec coefficients $\varphi_1, \varphi_2, \varphi_3$ déterminés par les données). Donner les équations liant les autocorrélations ρ_1, ρ_2, ρ_3 d'un processus AR(3) aux coefficients $\varphi_1, \varphi_2, \varphi_3$ (on ne demande pas de justifier). Calculer les 2 premières autocorrélations $\rho_i, i = 1, 2$ du processus Y_t , en fonction des coefficients $\varphi_1, \varphi_2, \varphi_3$.

5. Soit $(\varepsilon_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{N}(0, \sigma^2)$ où $\sigma^2 > 0$. Soit $a \in]0, 1[$ et $(X_n)_{n \geq 0}$ une suite de variables aléatoires telle que $X_0 = x_0$ presque sûrement et $X_n = aX_{n-1} + \varepsilon_n$ pour $n \geq 1$.

- (a) Montrer que pour $n \geq 1$

$$X_n = a^n x_0 + \sum_{k=0}^{n-1} a^k \varepsilon_{n-k}.$$

Solution : Pour $n = 1$ on a $X_1 = ax_0 + \varepsilon_1 = aX_0 + \varepsilon_1$. Supposons la relation vraie au rang $n \geq 1$ alors $X_{n+1} = aX_n + \varepsilon_{n+1} = a(a^n x_0 + \sum_{k=0}^{n-1} a^k \varepsilon_{n-k}) + \varepsilon_{n+1} = a^{n+1} x_0 + \sum_{k=0}^{n-1} a^{k+1} \varepsilon_{n+1-(k+1)} + \varepsilon_{n+1} = a^{n+1} x_0 + \sum_{k'=1}^n a^{k'} \varepsilon_{n+1-k'} + \varepsilon_{n+1} = a^{n+1} x_0 + \sum_{k'=0}^n a^{k'} \varepsilon_{n+1-k'}$. Donc relation établie par récurrence.

- (b) Déterminer la loi de X_n et montrer que $X_n \xrightarrow{\mathcal{L}} X$ où $X \sim \mathcal{N}(0, \sigma^2/(1-a^2))$.

Solution : X_n étant une combinaison linéaire de variables aléatoires normales indépendantes, elle est normale de moyenne et variance : $\mathbb{E}(X_n) = a^n x_0$ et $\text{var}(X_n) = \sigma^2(1 + a^2 + \dots + a^{2(n-1)}) = \sigma^2(1 - a^{2n})/(1 - a^2)$. Donc si Φ est la fdr d'une loi $\mathcal{N}(0, 1)$ on a pour tout $x \in \mathbb{R}$

$$P(X_n \leq x) = \Phi \left(\frac{x - a^n x_0}{\sigma \sqrt{(1 - a^{2n})/(1 - a^2)}} \right) \rightarrow \Phi \left(\frac{x \sqrt{1 - a^2}}{\sigma} \right)$$

qui est la fdr de la loi $\mathcal{N}(0, \sigma^2/(1 - a^2))$.

- (c) On suppose maintenant que $X_n = aX_{n-1} + b + \varepsilon_n$ pour $n \geq 1$ où $b \in \mathbb{R}$. Montrer que $(X_n)_{n \geq 0}$ converge en loi et déterminer sa loi asymptotique.

Solution : On établit par récurrence que

$$X_n = a^n x_0 + b \left(\frac{1 - a^n}{1 - a} \right) + \sum_{k=0}^{n-1} a^k \varepsilon_{n-k}$$

X_n étant une combinaison linéaire de variables aléatoires normales indépendantes, elle est normale de moyenne et variance : $\mathbb{E}(X_n) = a^n x_0 + b(1 - a^n)/(1 - a)$ et $\text{var}(X_n) = \sigma^2(1 + a^2 + \dots + a^{2(n-1)}) = \sigma^2(1 - a^{2n})/(1 - a^2)$. Donc si Φ est la fdr d'une loi $\mathcal{N}(0, 1)$ on a pour tout $x \in \mathbb{R}$

$$P(X_n \leq x) = \Phi \left(\frac{x - a^n x_0 - b(1 - a^n)/(1 - a)}{\sigma \sqrt{(1 - a^{2n})/(1 - a^2)}} \right) \rightarrow \Phi \left(\frac{(x - b/(1 - a))\sqrt{1 - a^2}}{\sigma} \right)$$

qui est la fdr de la loi $\mathcal{N}(b/(1 - a), \sigma^2/(1 - a^2))$.

Solutions :

- $\rho_k = \rho_1 \psi_{k-1} + \varphi_2 \psi_{k-2}, k \geq 1$
- $\text{Var } X_t = \sum_i \theta_i^2$. Elle doit être minimisée sous les contraintes $\sum_i \theta_i = 1, \sum_i i\theta_i = 0$. On trouve $\theta_i = \frac{\lambda + i\mu}{2}$ et les multiplicateurs de Lagrange satisfont

$$\begin{aligned} \lambda + \frac{n\mu}{2} &= \frac{2}{n+1} \\ \lambda + \frac{(2n+1)\mu}{3} &= 0 \end{aligned}$$

- $\hat{\varphi}(Y_0, \dots, Y_n) = \frac{\sum_{t=1}^n Y_t Y_{t-1}}{\sum_{t=0}^{n-1} Y_t^2}$
 - $\hat{\varphi} = \frac{\gamma_1}{\gamma_0}$
 - Pour $p = 2$, on arrive au système YW :

$$\begin{aligned} \rho_1 &= \varphi_1 + \varphi_2 \rho_1 \\ \rho_2 &= \varphi_1 \rho_1 + \varphi_2 \end{aligned}$$

avec solution

$$\varphi_{2,1} = \frac{\begin{vmatrix} \rho_1 & \rho_1 \\ \rho_2 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \quad \varphi_{2,2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

On s'aperçoit en résolvant cet exercice que les $\varphi_i, i = 1$, sont exactement les coefficients de la régression de Y_t sur Y_{t-1}, Y_{t-2} (d'où le nom autoregression), et que pour $p = 1$, y a coïncidence entre les valeurs du coefficient de régression φ et de la corrélation ρ_1 , mais pour $p \geq 2$, la situation se complique.

(d) Pour $p = 3$, le système YW est :

$$\rho_1 = \varphi_1 + \varphi_2\rho_1 + \varphi_3\rho_2$$

$$\rho_2 = \varphi_1\rho_1 + \varphi_2 + \varphi_3\rho_1$$

$$\rho_3 = \varphi_1\rho_2 + \varphi_2\rho_1 + \varphi_3$$

Les deux premières équations

$$\begin{pmatrix} (1 - \varphi_2) & -\varphi_3 \\ -(\varphi_1 + \varphi_3) & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$$

donnent

$$\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \frac{1}{1 - \varphi_2 - \varphi_3(\varphi_1 + \varphi_3)} \begin{pmatrix} 1 & \varphi_3 \\ \varphi_1 + \varphi_3 & 1 - \varphi_2 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \\ = \left\{ \frac{\varphi(1) + \varphi(2)\varphi(3)}{1 - (\varphi(2) + \varphi(3)(\varphi(1) + \varphi(3)))}, \frac{\varphi(1)(\varphi(3) + \varphi(1)) - \varphi(2)^2 + \varphi(2)}{1 - (\varphi(2) + \varphi(3)(\varphi(1) + \varphi(3)))} \right\}$$

(e) Les coefficients d'autocorrélation partielle $\varphi_{k,k}, \forall k \geq 1$ sont

$$\varphi_{1,1} = \rho_1 = \frac{\varphi(1) + \varphi(2)\varphi(3)}{1 - (\varphi(2) + \varphi(3)(\varphi(1) + \varphi(3)))}, \\ \varphi_{2,2} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\varphi(2) + \varphi(1)\varphi(3)}{1 - \varphi(3)^2}, \varphi_{3,3} = \varphi_3, \varphi_{4,4} = 0$$

9.3 Examen 3

1. Trouvez un filtre causal $\theta(B) = \sum_{i=0}^3 \theta_i B^i$ qui conserve les polynômes de degré ≤ 2 et qui enlève les composantes saisonnières d'ordre 2.

2. On considère des processus X_t, Y_t, Z_t satisfaisant respectivement :

$$(1 - B/2)X_t = \epsilon_t, \quad (1 - B)Y_t = \epsilon_t, \quad (1 - 2B)Z_t = \epsilon_t, \quad t = T, T - 1, \dots, 1, \dots \quad (9.4)$$

où ϵ_t est un bruit blanc de loi $N(0, \sigma^2 = 1)$.

- Donner l'expression de X_T , de Y_T , et de Z_T en fonction de $\epsilon_t, t = 1, 2, \dots, T$, et des valeurs au temps $t = 0$ (denotées par X_0, Y_0, Z_0).
- Est-ce que l'une des équations (9.4) admet une solution stationnaire? Le cas échéant, donner cette solution.
- Est-ce que l'un de ces trois processus, en commençant de 0 (c'est-à-dire avec $X_0 = 0, Y_0 = 0, Z_0 = 0$) est stationnaire?

3. a) Est-ce que le processus Y_t défini par :

$$\left(1 - \frac{2B}{3} + \frac{B^2}{9}\right)Y_t = \epsilon_t$$

est stationnaire causal?

b) Si oui, donner les coefficients ψ_k de sa représentation comme processus $MA(\infty)$ et calculer les 3 premières autocorrélations $\rho_i, i = 1, 2, 3$ du processus, ainsi que la variance.

Solutions :

1. $\varphi(z) = 1 - a(1 - z)^3$ et $\varphi(-1) = 0$ donnent $a = \frac{1}{8}$,

$$\varphi(z) = 1 - \frac{(1 - z)^3}{8} = \frac{z^3}{8} - \frac{3z^2}{8} + \frac{3z}{8} + \frac{7}{8} = \frac{1}{8}(z + 1)(z^2 - 4z + 7)$$

2. (a) Résolvant les récurrences donne
(b) $X_t = \sum_{k=0}^{\infty} 2^{-k} \epsilon_{t-k}$, $Z_t = \sum_{k=0}^{\infty} 2^{-k} \epsilon_{t+k}$, $t = 0, 1, 2, \dots$ sont des processus stationnaires.
(c) Non

Chapitre 10

Approximation par splines

Nous aimerions approximer un "nuage" des points $(x_i, y_i) \in \mathbb{R}^2, i = 1, \dots, n$ par une courbe unidimensionnelle $\xi(x), x \in \mathbb{R}$. Comment la choisir? Une des méthodes les plus simples est l'interpolation exacte par des polynômes (Wallis/Lagrange/Newton), qui fait passer un polynôme de degré n $p(x) = \sum_{i=0}^n a_i x^i$ par $n + 1$ points. Il s'agit d'un problème linéaire dans les inconnues a_i . L'interpolation polynomiale à haut degré n'est pas très utile, car elle donne naissance à beaucoup d'oscillations (par exemple quand on utilise un polynôme de degré n avec n racines réelles, et n grand). **Idée** : Passer par les points, c'est peut-être inutile, si les points sont déjà incertains; dans telles conditions, l'interpolation exacte n'est pas raisonnable. Il devient plus sensé d'ajuster approximativement les données, avec le moindre nombre des paramètres possibles; passer près des points, peut être mieux, parce qu'on enlève ainsi du "bruit". Remarquons que cette approche est aussi très simple à implémenter, juste comme cas particulier de la régression de moindres carrés du modèle linéaire (obtenu en R par la commande `lm`).

```
library(MASS);
mcycle
tm<-mcycle$times
am<-mcycle$accel
plot(tm,am)
tm2=tm^2
tm3=tm^3
tm4=tm^4
tm5=tm^5
reg <- lm( am ~tm + tm2 + tm3 + tm4 + tm5)
print( summary( reg ) )
#
##### Look at the fitted model:
#
reg.mean <- predict( reg )
#
##### Plot it!
#
lines( tm, reg.mean, col="blue", lwd=2 )
```

Une méthode alternative à la régression polynomiale est la régression par **splines**, c.-à-d. des polynômes de petit degré, par morceaux. On propose de couper $[a, b] = [x_1, x_n]$ en petits morceaux, et utiliser un polynôme de petit degré dans chaque morceau, en imposant aussi **lissité** dans les points de contact. Le choix du degré des polynômes et de la lissité ($\xi(x) \in \mathcal{C}^1$, $\xi(x) \in \mathcal{C}^2$, ...) n'est pas évident, mais un des choix les plus populaires est degré 3, et lissité $\xi(x) \in \mathcal{C}^2$.

```
sp<-smooth.spline(tm,am, all.knots=T)
lines( sp, col="yellow" )

sp5<-smooth.spline(tm,am, all.knots=F,nknots=5)
lines( sp5, col="green" )

sp10<-smooth.spline(tm,am, all.knots=F,nknots=10)
lines( sp10, col="violet" )

sp8<-smooth.spline(tm,am, all.knots=F,nknots=8)
lines( sp8, col="red" )
```

Voilà plusieurs exemples de régression polynomiale et par splines en R, offerts par : <http://faculty.washington.edu/heagerty/Courses/b571/homework/spline-tutorial.q>

```
#### Simulation des donnees
x <- c(1:24)
mu <- 10 + 5 * sin( x * pi / 24 ) - 2 * cos( (x-6)*4/24 )
#
##### erreurs
#
set.seed(2010)
eee <- rnorm( length(mu) )
#
##### donnees observees
#
y <- mu + eee
#
plot( x, y )
lines( x, mu, col="red" )
title("Data and mean curve")

#### Approximation polynômiale cubique
x.squared <- x^2
x.cubed <- x^3
#
fit <- lm( y ~ x + x.squared + x.cubed )
#
print( summary( fit ) )
#
```

```
##### Look at the fitted model:
#
fitted.mean <- predict( fit )
#
##### Plot it!
#
plot( x, y )
lines( x, mu, col="red" )
lines( x, fitted.mean, col="blue", lwd=2 )
title("Data, true mean curve (red), and fitted (blue) using cubic")
```

Le cas le plus simple d'interpolation par splines est l'interpolation linéaire par morceaux.

Voilà un programme de régression des données par une spline linéaire avec trois noeuds donnés 6, 12, 18 :

```
x6 <- ( x - 6 )
x6[ x6<0 ] <- 0
#
x12 <- ( x - 12 )
x12[ x12<0 ] <- 0
#
x18 <- ( x - 18 )
x18[ x18<0 ] <- 0
#
##### Let's look at these new predictors:
#
print( cbind( x, x6, x12, x18 ) )
#
#####
##### Now we can fit the linear spline
#####
#
fit <- lm( y ~ x + x6 + x12 + x18 )
#
##### Look at the fitted coefficients:
#
print( summary( fit ) )
#
##### Look at the fitted model:
#
fitted.mean <- predict( fit )
#
##### Plot it!
#
plot( x, y )
lines( x, mu, col="red" )
```

```
lines( x, fitted.mean, col="blue", lwd=2 )
title("Data, true mean curve (red), and fitted (blue) using linear spline")
```

Passons maintenant à une régression des nôt données par une spline cubique avec noeuds 6,12,18 (la somme d'un polynôme cubique et des puissances cubiques tronquées dans les noeuds) en utilisant encore une fois le modèle linéaire :

```
go.pdf <- T
#### Approximation par une spline cubique
x6.cubed <- x6^3
x12.cubed <- x12^3
x18.cubed <- x18^3
#
##### Fit the model:
#
fit <- lm( y ~ x + x.squared + x.cubed + x6.cubed + x12.cubed + x18.cubed )
#
print( summary( fit ) )
#
##### Look at the fitted model:
#
fitted.mean <- predict( fit )
#
##### Plot it!
#
if( go.pdf) pdf( file="TheCubicSpline.pdf" )
plot( x, y )
lines( x, mu, col="red" )
lines( x, fitted.mean, col="blue", lwd=2 )
title("Data, true mean curve (red), and fitted (blue) using cubic spline")
if( go.pdf ) graphics.off()
```

10.1 Splines cubiques naturelles

Une des solutions les plus utilisées est fournie par les "splines cubiques naturelles", qui sont des polynômes de degré 3 par morceaux, dont la dérivée seconde est continue, et nulle aux bords.

Définition 10.1.1 *On appellera spline cubique naturelle d'interpolation exacte des points $(x_i, y_i), i = 1, \dots, n, n \geq 2$ une fonction qui vérifie :*

- $\xi(x_i) = y_i, i = 1, \dots, n$
- $\xi, \xi',$ et ξ'' sont définies et continues sur \mathbb{R}
- Sur chaque $]x_i, x_{i+1}[, i = 1, \dots, n - 1, \xi$ est un polynôme de degré 3
- Sur $(-\infty, a]$ et sur $]b, +\infty[, \xi$ est un polynôme de degré 1; en particulier, $\xi''(a) = \xi''(b) = 0$

La spline cubique naturelle par $n = 2$ points est la droite. Par $n = 1$ points il y a une famille infinie $\xi(x) = \xi(x_1) + (x - x_1)\xi'(x_1)$, où $\xi'(x_1)$ est arbitraire. Pour $n \geq 3$, on peut montrer que la spline cubique naturelle qui interpole n points donnés est unique.

Remarque 10.1.1 Une spline sur $[a, b] = [x_1, x_n]$ comporte $n-1$ intervalles et donc $4(n-1)$ coefficients inconnus. Les 3 conditions de lissité \mathcal{C}^2 dans les $n-2$ noeuds intérieurs, nous fournissent un nb. d'équations égal au $3(n-2)$. Encore 2 équations sont fournies par les conditions de bord $\xi''(a) = \xi''(b) = 0$. Il restent donc précisément n inconnues, qui seront fournies par les valeurs à interpoler $\xi(x_j) = y_j, j = 1, \dots, n$. Il est facile de vérifier que le système qui détermine $\xi'(x_j), \xi''(x_j), j = 1, \dots, n$ est linéaire. Par conséquent, si la matrice de ce système linéaire n'est pas singulière (comme c'est le cas pour $n \geq 2$), il y aura une spline cubique naturelle unique passant par $n \geq 2$ points.

La régression par une spline cubique naturelle est plus difficile à implémenter comme cas particulier du modèle linéaire, mais une base pour cela peut être obtenue avec la commande ns :

```
library( splines )
#
fit <- lm( y ~ ns( x, knots=c(6,12,18) ) )
#
print( summary( fit ) )
#
##### Look at the fitted model:
#
fitted.mean <- predict( fit )
#
##### Plot it!
#
plot( x, y )
lines( x, mu, col="red" )
lines( x, fitted.mean, col="blue", lwd=2 )
title("Data, true mean curve (red), and fitted (blue) using natural spline")

fit1 <- lm( y ~ ns( x, df = 5 ) )#utilise df -1 noeuds(knots)
fit1.mean <- predict( fit1 )
lines( x, fit1.mean, col="green")
attr(terms(fit1), "predvars")
```

Remarque 10.1.2 Il y a plusieurs types des splines, qui diffèrent surtout par les conditions frontière au bout du domaine, et par le respect ou non-respect de la forme des données (monotonie, convexité, etc.). Par exemple, pour les splines périodiques, en supposant

$$\xi(a) = \xi(b), \text{ on ajoute : } \begin{cases} \xi'(a) = \xi'(b) \\ \xi''(a) = \xi''(b) \end{cases} \text{ et pour les splines "not_a_knot" on ajoute : } \\ \begin{cases} \xi'''(x_2) = \xi'''(x_{2,+}) \\ \xi'''(x_{n-1}) = \xi'''(x_{n-1,+}) \end{cases}$$

10.2 Le calcul des splines naturelles d'interpolation exacte (*)

Idée : Une spline a une l'écriture "locale", dans chaque intervalle $[x_j, x_{j+1}]$ où ξ est

$$\text{un polynôme : } \boxed{\begin{aligned} \forall j \in [1 : n-1], \forall x \in [x_j, x_{j+1}], \\ \xi(x) &= \xi(x_j) + (x - x_j)\xi'(x_j) + \frac{(x-x_j)^2}{2} \xi''(x_j) + \frac{(x-x_j)^3}{6} \xi'''(x_j) \\ &= \xi_j + (x - x_j)\xi'_j + \frac{(x-x_j)^2}{2} \xi''_j + \frac{(x-x_j)^3}{6} \xi'''_j, \end{aligned}} \quad \text{Les pa-}$$

ramètres $\xi'(x_j), \xi''(x_j), \xi'''(x_j)$ représentent les dérivées associées à l'extrémité gauche de l'intervalle $[x_j, x_{j+1}]$. Cette paramétrisation permet de calculer les dérivées associées à l'extrémité droite de l'intervalle, dont les premières deux devront correspondre à ceux spécifier par le prochain intervalle $[x_{j+1}, x_{j+2}]$.

Soit $h_i = x_{i+1} - x_i, i \in \{1, \dots, n-1\}$. Les troisièmes dérivées sont déterminées très simplement, par le fait que la deuxième dérivée est linéaire en chaque intervalle, et donc :

$$\boxed{\xi'''_i = \frac{(\xi''_{i+1} - \xi''_i)}{h_i}, i = 1, \dots, n-1} \quad (10.1)$$

Exercice 10.2.1 1) Déterminer la forme de la spline naturelle cubique $\xi(x)$ qui passe par 3 points équidistants à distance 1, $(x_i, y_i), i = 0, 1, 2$, en sachant que $\xi(x) \in C^2, \xi''(x_0) = \xi''(x_2) = 0$ et que la fonction $\xi(x)$ est polynomiale d'ordre au plus 3 sur chaque intervalle déterminé par les "noeuds".

2) Verifier que pour la fonction obtenue "l'énergie" $\int_0^2 (\xi''(x))^2 dx$ est plus petite que celle obtenue pour a) l'interpolation linéaire par morceaux de ces trois points, et b) l'interpolation par une parabole.

Sol : Avec trois points, on a deux morceaux, et il est convenable de choisir comme paramètres les valeurs communes des dérivés dans le point du milieu $x_1 : \xi_1, \xi'_1, \xi''_1$, qui coïncident pour les deux morceaux, et qui déterminent, ensemble avec ξ_0, ξ_2 , la spline (comme les troisièmes dérivées sont déterminées par (10.1)). Les deux morceaux sont :

$$\begin{aligned} \xi(x) &= \xi_1 + (x - x_1)\xi'_1 + \frac{(x - x_1)^2}{2} \xi''_1 + \frac{(x - x_1)^3}{6} \xi'''_1 \\ \tilde{\xi}(x) &= \xi_1 + (x - x_1)\xi'_1 + \frac{(x - x_1)^2}{2} \xi''_1 + \frac{(x - x_1)^3}{6} (-\xi'''_1) \end{aligned}$$

Les valeurs spécifiés aux extrémités déterminent les deux dérivées :

$$\begin{aligned} \xi_0 &= \xi_1 - \xi'_1 + \frac{1}{3} \xi''_1 \\ \xi_2 &= \xi_1 + \xi'_1 + \frac{1}{3} \xi''_1 \end{aligned}$$

$$\xi''_1 = \frac{3}{2} (\xi_0 - 2\xi_1 + \xi_2), \quad \xi'_1 = \xi_1 - \xi_0 + \frac{1}{2} (\xi_0 - 2\xi_1 + \xi_2) = \frac{1}{2} (\xi_2 - \xi_0)$$

Remarque 10.2.1 Pour la première dérivée on obtient une réponse assez prévisible, mais cela n'est pas aussi le cas avec le facteur $3/2$ dans la deuxième dérivée.

Exemple : **Spline d'interpolation des 3 points (0,0), (1,1), (2,0)** Calcul des ξ_i'' : $\xi_0'' = 0$; $\xi_2'' = 0$ $4\xi_1'' = 6(\xi_2 - 2\xi_1 + \xi_0) = -12$ alors $\xi_1'' = -3$ Calcul des ξ_i''' : $\xi_0''' = -3$; $\xi_1''' = 3$

Calcul des ξ_i' : $\xi_0' = 1 + \frac{3}{6} = 1,5$; $\xi_1' = -1 + 1 = 0$; $\xi_2' = \xi_1' + \xi_1'' + \frac{\xi_1'''}{2} = -1,5$

On obtient donc :
$$\xi(x) = \begin{cases} \forall x \in]-\infty, 0], & \xi(x) = 1,5x \\ \forall x \in [0, 1], & \xi(x) = 1,5x - \frac{x^3}{2} \\ \forall x \in [1, 2], & \xi(x) = 1 - \frac{3(x-1)^2}{2} + \frac{(x-1)^3}{2} \\ \forall x \in [2, \infty[, & \xi(x) = -\frac{3}{2}(x-2) \end{cases}$$
 ξ s'écrit

aussi : $\forall x \in \mathbb{R}, \xi(x) = 1,5x - \frac{x^3}{2} + (x-1)_+^3 - (x-2)_+^3/2$ Energie de flexion : $= 1,5x - |x|^3/4 + |x-1|^3/2 - |x-2|^3/4$

$\int_0^2 (\xi''(x))^2 dx = 6 \rightarrow$ c'est plus faible que tous les autres ; par exemple, $g(x) = \sin(\pi x/2)$

rend $(\frac{\pi^2}{4})^2 = \frac{\pi^4}{16}$. Comparaison des valeurs et dérivées : $\xi'(0) = \frac{3}{2} g'(0) =$ $\xi'(1) = 0 \quad g'(1) = 0 \quad \xi''(1) =$

Exercice 10.2.2

Déterminer la spline cubique naturelle ξ d'interpolation des 3 points $(0,0), (1,1), (2,-1)$, en utilisant les formules :

$$\xi_{1,+}''' = \frac{(\xi_2'' - \xi_1'')}{h}, \quad \xi_{1,-}''' = \frac{(\xi_1'' - \xi_0'')}{h}, \quad \xi_1'' = \frac{3(\xi_2 - 2\xi_1 + \xi_0)}{2h^2}, \quad \xi_1' = \frac{\xi_2 - \xi_0}{2h}$$

b) La continuité de ξ en x_{i+1} détermine la première dérivée en x_i , par : $\xi_{i+1} = \xi_i + h_i \xi_i' + \frac{h_i^2}{2} \xi_i'' + \frac{h_i^3}{6} \xi_i''' \Rightarrow \xi_i' = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{2} \xi_i'' - \frac{h_i^2}{6} (\xi_{i+1}'' - \xi_i'') \frac{1}{h_i^2} \Rightarrow \xi_i' = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{2} \xi_i'' - \frac{h_i^2}{6} (\xi_{i+1}'' - \xi_i'')$

Calculer $\int_0^2 (\xi''(x))^2 dx$, et comparer cette valeur avec celles d'autres interpolants de ces trois points. Tracer sur un même graphe ξ, ξ', ξ'', ξ''' .

Exercice 10.2.3 On propose la méthode suivante pour calculer la spline cubique naturelle qui interpole n points équidistants.

1. On calcule en chaque point intérieur des dérivées numériques, en utilisant par exemple :

$$\xi_i' = \frac{(y_{i+1} - y_{i-1}))}{h}, \quad \xi_i'' = \frac{(y_{i+1} - 2y_i + y_{i-1}))}{h^2}, \quad i = 2, \dots, n-1 \text{ où } h = x_{i+1} - x_i, i \in \{1, \dots, n-1\}.$$

2. On calcule sur chaque intervalle intérieur le polynôme de degré minimal, avec les valeurs de y_i, ξ_i', ξ_i'' données/trouvées dans les deux extrémités.

Est-ce que cette méthode produit une spline cubique ?

Exercice 10.2.4 Développer un programme qui calcule la spline d'interpolation exacte $\xi(x)$ qui passe par n points, en la testant sur le cas des 3 points $(-1,0), (0,2), (1,0)$ quand la

réponse est :
$$\xi(x) = \begin{cases} \forall x \in]-\infty, -1], & \xi(x) = 3(x+1) \\ \forall x \in [-1, 0], & \xi(x) = 2 - 3x^2 - x^3 = 3(x+1) - (x+1)^3 \\ \forall x \in [0, 1], & \xi(x) = 2 - 3x^2 + x^3 = (x-1)^3 - 3(x-1) \\ \forall x \in [1, \infty[, & \xi(x) = -3(x-1) \end{cases}$$

Sol : Méthode générale : paramétrisation par $y_i = \xi(x_i)$ et $\xi_i'' = \xi''(x_i)$. Soit

1. $h_i = x_{i+1} - x_i, i \in \{1, \dots, n-1\}$
2. $d_i = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \approx \xi_i'', i \in \{2, \dots, n-1\}$
3. R la matrice tridiagonale **carré, indexée par $i = 1, \dots, n-2$** , définie par $R_{i,i} = (h_{i+1} + h_i)/3, i = 1, \dots, n-2, R_{i,i+1} = R_{i+1,i} = h_{i+1}/6, i = 1, \dots, n-1$:

$$R = \begin{pmatrix} (h_1 + h_2)/3 & h_2/6 & 0 & \dots & \dots & \dots \\ h_2/6 & (h_2 + h_3)/3 & h_3/6 & 0 & \dots & \dots \\ 0 & h_3/6 & (h_3 + h_4)/3 & h_4/6 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & h_{n-2}/6 \\ \vdots & \ddots & \ddots & \ddots & h_{n-2}/6 & (h_{n-2} + h_{n-1})/3 \end{pmatrix} \quad (10.2)$$

Théorème 10.2.1 Les vecteurs $\xi_i, i = 1, \dots, n$ et $\xi_i'', i = 2, \dots, n-1$ spécifient les valeurs $\xi_{i+1}''', i = 1, \dots, n-1$ et $\xi_i', i = 1, \dots, n$ par les formules : a) **troisième dérivée :**

$$\xi_i''' = \frac{(\xi_{i+1}'' - \xi_i'')}{h_i}, \quad i = 1, \dots, n-1$$

b) **première dérivée, "à gauche" d'un intervalle intérieur** (en utilisant Taylor à droite) :

$$\xi_i' = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{6}(\xi_{i+1}'' + 2\xi_i''), \quad i = 1, \dots, n-1$$

c) **première dérivée, "à droite" d'un intervalle intérieur** (en utilisant Taylor à gauche) :

$$\xi_i' = \frac{\xi_i - \xi_{i-1}}{h_{i-1}} + \frac{h_{i-1}}{6}(\xi_{i-1}'' + 2\xi_i''), \quad i = 2, \dots, n$$

d) Les vecteurs $\xi_i, i = 1, \dots, n$ et $\xi_i'', i = 1, \dots, n$, avec $\xi_1'' = \xi_n'' = 0$, définissent une spline naturelle ssi $R\xi'' = \vec{d} \iff$

$$\frac{h_{i-1}}{6} \xi_{i-1}'' + \frac{h_{i-1} + h_i}{3} \xi_i'' + \frac{h_i}{6} \xi_{i+1}'' = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{\xi_i - \xi_{i-1}}{h_{i-1}}, i = 2, \dots, n-1 \quad (10.3)$$

e) L'énergie de flexion est :

$$\int_{\mathbb{R}} (\xi''(x))^2 dx = (\xi'')^t R \xi'' = \xi^t K \xi \quad (10.4)$$

où $K = S^t R^{-1} S$ et où S est la matrice tridiagonale **rectangulaire** $(n-2) \times (n)$, **indexée par $i = 1, \dots, n-2, j = 1, \dots, n$**

$$S = \begin{pmatrix} \frac{1}{h_1} & -\frac{1}{h_1} - \frac{1}{h_2} & \frac{1}{h_2} & 0 & \dots & \dots \\ 0 & \frac{1}{h_2} & -\frac{1}{h_1} - \frac{1}{h_2} & \frac{1}{h_2} & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots \\ \vdots & \dots & \frac{1}{h_{i-1}} & -\frac{1}{h_{i-1}} - \frac{1}{h_i} & \frac{1}{h_i} & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \frac{1}{h_{n-2}} \\ \vdots & \dots & \dots & \dots & \frac{1}{h_{n-2}} & -\frac{1}{h_{n-1}} - \frac{1}{h_{n-2}} \end{pmatrix} \quad (10.5)$$

10.3 Splines d'ajustement

Ce but peut être atteint par la minimisation des objectives ¹ :

$$E = \min_{\xi(x) \in C^{2m-2}} \mu \sum_i (y_i - \xi(x_i))^2 + \lambda \int_{\mathbb{R}} (\xi^{(m)}(x))^2 dx \quad (10.7)$$

Définition 10.3.1 ξ est dite *spline d'ajustement (ou de lissage) d'ordre $2m - 1$ des points $(x_i, y_i)_{i \in [1:n]}$, et de poids λ .*

L'objectif (10.7) essaie de trouver un équilibre entre deux buts opposés : 1) celui d'un bon ajustement de données, obtenu en minimisant la somme des écarts au carré et 2) la "lissité", obtenu en minimisant la dérivé d'ordre m au carré (on voit par exemple que avec $m = 2$ ce terme est minimisé par la droite, qui est "très lisse"). Pour mieux comprendre l'objectif de minimisation (10.7) quand $m = 2$, nous examinons maintenant les cas limite $\lambda \rightarrow \infty$ et $\mu \rightarrow \infty$. Dans ces deux cas, le problème devient une minimisation "du objectif moins important, avec l'objectif plus important imposé d'être 0", c.-à-d.

1. Dans la limite $\lambda \rightarrow \infty$, la deuxième contrainte $\int_{\mathbb{R}} (\xi^{(m)}(x))^2 dx = 0$ est imposé. Donc, le problème devient :

$$\begin{aligned} \min_{\xi(x) \in C^{2m-2}} \sum_i (y_i - \xi(x_i))^2 \\ \int_{\mathbb{R}} (\xi^{(m)}(x))^2 dx = 0 \end{aligned}$$

La solution est le polynôme $\xi(x)$ d'ordre $m - 1$ ajustant n points $(x_i, y_i), i = 1, \dots, n, n \geq m$. Pour $m = 2$ par exemple, le problème devient

$$\begin{aligned} \min_{\xi(x) \in C^2} \sum_i (y(x_i) - \xi(x_i))^2 \\ \int_{\mathbb{R}} (\xi''(x))^2 dx = 0 \end{aligned}$$

et la solution est la **droite des moindres carrés** ajustant $n > 2$ points.

2. Dans la limite $\mu \rightarrow \infty$, la première contrainte est imposée. Donc, le problème devient :

$$\begin{aligned} \min_{\xi(x) \in C^{2m-2}} \int_{\mathbb{R}} (\xi^{(m)}(x))^2 dx \\ \sum_i (y_i - \xi(x_i))^2 = 0 \end{aligned}$$

et la solution est la **spline naturelle d'interpolation exacte** d'ordre $2m - 1$.

1. En ajoutant des poids ρ_i dépendant du point i , on peut accorder plus d'importance au fait de passer près des points pour lesquels ρ_i est plus grand (par exemple, dans une évolution temporelle, les dernières points sont les plus importants aux niveaux de l'extrapolation vers le futur).

Théorème 10.3.1 a) Pour toute famille $(x_i, y_i)_{i \in [0, n]}$, $n \geq 2$, et $\lambda > 0$, il existe une et une seule fonction $\xi \in \mathcal{C}^{2m-2}(\mathbb{R})$ minimisant la somme des carrés ajustés (10.7). b) Pour $m = 2$, cette fonction est une spline cubique naturelle qui interpole exactement les noeuds $(x_i, \tilde{y}_i)_{i \in [1: n]}$, où $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n) = (I + \lambda K)^{-1} \mathbf{y}$, avec $K = S^t R^{-1} S$, où R, S sont définies en (10.2), (10.5). Elle est souvent déterminée par l'algorithme de Reinsch, en commençant par les dérivées secondes données par :

$$(\mathbf{R} + \lambda \mathbf{S} \mathbf{S}^t) \boldsymbol{\xi}'' = \left(\frac{\xi_{i+1} - \xi_i}{h_i} - \frac{\xi_i - \xi_{i-1}}{h_{i-1}}, i = 1, \dots, n-1 \right)$$

Démonstration : Conséquence de (10.4).

10.4 Calcul des splines naturelles par interpolation de noyaux

Exercice 10.4.1 L'exercice suivant sur le calcul des splines naturelles par une interpolation de noyaux ouvre le chemin pour une généralisation multidimensionnelle des splines. Soit la fonction u_+ définie par $u_+ = u$ si $u \geq 0$ et $u_+ = 0$ si $u \leq 0$. Montrez que chaque spline naturelle cubique peut-être écrit comme : $\forall x \in \mathbb{R}, \xi(x) = ax + b + (1/2) \sum_{i \in [1: n]} \alpha_i (x - x_i)_+^3$

Écrivez, dans les deux cas, les systèmes linéaires de $n + 2$ équations qui déterminent les $n + 2$ paramètres inconnus $(\alpha_1, \dots, \alpha_n, a, b)$, si $x_i = x_0 + i$.

Sol :

$$\left\{ \begin{array}{l} \xi(x) = ax + b + \sum_{i \in [1: n]} \alpha_i (x - x_i)_+^3 \\ \sum_{i \in [1: n]} \alpha_i = \sum_{i \in [1: n]} \alpha_i x_i = 0 \\ \forall j \in [1 : n], \xi(x_j) = y_j \end{array} \right\} \Rightarrow$$

$$\left\{ \begin{array}{l} \forall j \in [1 : n], y_j = ax_j + b + \sum_{i \in [1: n]} \alpha_i (x_j - x_i)_+^3 \\ \sum_{i \in [1: n]} \alpha_i = \sum_{i \in [1: n]} \alpha_i x_i = 0 \end{array} \right.$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \dots & \dots & \dots & 0 & 0 & 1 \\ 1^3 & 0 & 0 & 0 & \dots & \dots & \dots & 0 & 1 & 1 \\ 2^3 & 1^3 & 0 & 0 & \dots & \dots & \dots & 0 & 2 & 1 \\ 3^3 & 2^3 & 1^3 & 0 & \dots & \dots & \dots & 0 & 3 & 1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & & & \vdots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots & \vdots & \vdots \\ (n-1)^3 & (n-2)^3 & \dots & \dots & 2^3 & 1^3 & 0 & 0 & n-1 & 1 \\ n^3 & (n-1)^3 & \dots & \dots & \dots & 2^3 & 1^3 & 0 & n & 1 \\ 0 & 1 & 2 & \dots & \dots & \dots & n-1 & n & 0 & 0 \\ 1 & 1 & 1 & \dots & \dots & \dots & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ \alpha_n \\ a \\ b \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ \vdots \\ y_n \\ 0 \\ 0 \end{pmatrix}$$

10.5 Exercices

- Par analogie avec les splines cubiques, formulez et résolvez les cas limites $\lambda \rightarrow \infty, \mu \rightarrow \infty$ de l'interpolation ajustés d'ordre $m = 1$.
- Déterminer la spline cubique s d'interpolation des 4 points $(0,0)$, $(1,1)$, $(2,0)$, $(3,0)$. Tracer sur un même graphe s, s', s'', s''' . Comparer ξ et s (graphiquement, et par leurs dérivées aux nœuds,...). Dans quelle zone s est-elle la plus proche de ξ ?
- Déterminer la spline cubique s d'interpolation des 4 points $(0,0)$, $(1,1)$, $(2,0)$, $(3,-3/2)$.
- Démontrer le théorème suivant : **Théorème** : Soient ξ et s les splines d'interpolation des points $(x_i, \xi_i)_{i \in [1:n]}$ et $(x_i, s_i)_{i \in [1:n]}$, alors $\xi - s$ est la spline d'interpolation des points $(x_i, \xi_i - s_i)_{i \in [1:n]}$. En déduire que, ξ et s étant les splines de 2 et 3 ci-dessus, $\xi - s$ est la spline d'interpolation des points $(0, 0)$; $(1, 0)$; $(2, 0)$; $(3, \frac{3}{2})$
- Comparaison de divers interpolants.** On considère plusieurs interpolants des 4 points $(0,0)$, $(1,0)$, $(2,0)$, $(3,3/2)$:
 - La spline cubique naturelle w de ces points. Comme vu aux exercices 1, 2, 3 ci-dessus, on a $w = \xi - s$ où s et ξ sont les splines définies en 1 et 2 ci-dessus.
 - Le polynôme p d'interpolation de degré le plus bas possible, interpolant ces quatre points. De quel degré est ce polynôme ?
 - La fonction f_1 deux fois continument dérivable, identiquement nulle sur $] - \infty .. 2]$, polynomiale de degré le plus faible possible sur $\forall 2\infty$, telle que $f(3) = 3/2$.
 - La fonction f_2 deux fois continument dérivable, identiquement nulle sur $[0 .. 2]$, polynomiale de degré le plus faible possible sur $\forall 2\infty$, telle que $f(3) = 3/2$ et $f''(3) = 0$.
 - La fonction f_3 une fois continument dérivable, identiquement nulle sur $[0 .. 2]$, polynomiale de degré le plus faible possible sur $\forall 2\infty$, telle que $f(3) = 3/2$ et $f''(3) = 0$.

Donner une forme explicite des fonctions ci-dessus. Expliquez pourquoi le polynôme p et les fonctions f_1, f_2 , et f_3 , bien que polynomiales ou polynomiales par morceaux, ne sont pas des fonctions spline cubiques naturelles. Calculez $\int_0^3 (f''(x))^2 dx$ pour les fonctions déterminées en 2, 3, 4, 5. Constatez que ces valeurs ne sont pas les mêmes. Que pouvez-vous dire de $\int_0^3 (w''(x))^2 dx$ à partir de ces résultats ? Éventuellement, tracez ces courbes sur ordinateur. Éventuellement, calculez $\int_0^3 (w''(x))^2 dx$.

- Montrez, que, pour toute famille $(x_i, y_i)_{i \in [1,n]}$, $n \geq 3$, et $\lambda > 0$, la fonction $\xi \in \mathcal{C}^2(\mathbb{R})$ minimisant la somme des carrés pondérés et ajustés :

$$\min_{\xi(x) \in \mathcal{C}^2} \sum_i \rho_i (y(x_i) - \xi(x_i))^2 + \lambda \int_{\mathbb{R}} (\xi''(x))^2 dx$$

est une spline cubique naturelle qui interpole exactement les noeuds $(x_i, \tilde{y}_i)_{i \in [1:n]}$, où $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n) = (I + \lambda K)^{-1} \mathbf{y}$, et déterminez K .

7. Obtenez la formule du filtre correspondant à une spline d'ajustement correspondant à un opérateur différentiel arbitraire à coefficients constants.

Solutions :

2 **Spline d'interpolation des 4 points (0,0) , (1,1) , (2,0) , (3,0)**

$$\left. \begin{array}{l} 4s_1'' + s_2'' = -12 \\ s_1'' + 4s_2'' = 6 \end{array} \right\} \implies \left\{ \begin{array}{l} s_1'' = -3,6 \\ s_2'' = 2,4 \end{array} \right.$$

On en déduit : $\left\{ \begin{array}{l} s_0''' = -3,6 \\ s_1''' = 6 \\ s_2''' = -2,4 \end{array} \right. , \left\{ \begin{array}{l} s_0' = 1,6 \\ s_1' = -0,2 \\ s_2' = -0,8 \end{array} \right. , s_3' = 0,4$ D'où : $\left\{ \begin{array}{l} \forall x \in]-\infty .. 0] , \\ \forall x \in [0 .. 1] , \\ \forall x \in [1 .. 2] , \\ \forall x \in \mathbb{R} \setminus \mathbb{Z} , \\ \forall x \in \mathbb{Z} + \infty , \end{array} \right. \xi(x)$

10.6 Interpolation par splines de degré $2m - 1, m \geq 1$

Définition 10.6.1 Les splines naturelles de degré $2m - 1$: sont des fonctions \mathcal{P}_{2m-1} par morceaux, \mathcal{C}^{2m-2} , et avec les deux morceaux extérieures \mathcal{P}_{m-1} , c.-à-d. satisfaisant :

$$\xi^{(m)}(x_0) = \dots = \xi^{(2m-1)}(x_0) = 0$$

$$\xi^{(m)}(x_n) = \dots = \xi^{(2m-1)}(x_n) = 0$$

Théorème 10.6.1 Les splines naturelles de degré $2m - 1$ minimisent $\int_{\mathbb{R}} (f^{(m)}(x))^2 dx$ sous la contrainte d'interpoler les (x_i, y_i) .

(D'autres familles de splines, minimisant d'autres "normes", sont intéressantes aussi).

Théorème 10.6.2 ()** Dans le cas équidistant, la solution de la récurrence (10.6) est donné, en posant $\lambda = \sqrt{3} - 2 \simeq -0,27$, par :

$$\xi_j'' = 6h^{-2}(1 - \lambda^{2n})^{-1} \sum_{i \in [0:n]} A_{ji} \xi_i , \text{ avec, pour tout } j \in [1 : n-1],$$

$$\begin{aligned} A_{j0} &= -\lambda^j(1 - \lambda^{2n-2j}) \\ \forall i \in [1 : j-1] , \quad A_{ji} &= -\sqrt{3}\lambda^{j-i}(1 - \lambda^{2i})(1 - \lambda^{2n-2j}) \\ A_{jj} &= 1 - \lambda^{2n} - \sqrt{3}(1 - \lambda^{2j})(1 - \lambda^{2n-2j}) \\ \forall i \in [j+1 : n-1] , \quad A_{ji} &= -\sqrt{3}\lambda^{i-j}(1 - \lambda^{2j})(1 - \lambda^{2n-2i}) \\ A_{jn} &= -\lambda^{n-j}(1 - \lambda^{2j}) \end{aligned}$$

Note : Cette "forme explicite" est plutôt d'intérêt théorique, car elle est moins aisée à utiliser que la résolution numérique du système tridiagonal. Retenons de ce chapitre la désirabilité des prédictions lisses, pour découvrir des "signals" qu'on suppose cachés dans les données, et passons à un autre aspect, l'évaluation du "bruit". Remarquez qu'en ajustant au gré le paramètre de pondération λ des splines d'ajustement, on arrive à un point où le bruit disparaît totalement. Mais, pt-etre que le bruit peut aider dans la prédiction ? Pour arriver pt-etre à un choix rationnel du λ ?

10.7 Splines cubiques en Scilab

Voilà aussi des exemples de la documentation Scilab, où les splines cubiques sont implémentées par la commande SMOOTH : http://help.scilab.org/docs/5.3.0/fr_FR/smooth.html

```
x=[1 10 20 30 40];
y=[1 30 -10 20 40];
plot2d(x',y',[3],"011"," ",[-10,-40,50,50]);

yi=smooth([x;y],0.1);
plot2d(yi(1,:)',yi(2,:)',[5],"000");
```

Le problème du manque de convergence de l'interpolation par polynômes de degré n quand $n \rightarrow \infty$, même pour des fonctions analytiques, a été illustré dans un exemple fameux de Runge, qui utilise des points d'interpolations equidistants.[§] Par contre, l'interpolation par splines cubiques ne rencontre pas ce problème. Voilà un programme de la documentation Scilab, qui utilise les commande SPLIN (qui calcule la dérivée numérique nécessaire pour caractériser les splines cubiques) et INTERP.

```
deff("y=runge(x)","y=1./(1+x.^2)")
a = -5; b = 5; n = 11;
x = linspace(a, b, n)';
y = runge(x);
clf()
plot2d(x, y, -9)

d = splin(x, y)
m = 400;
xx = linspace(a, b, m)';
yye = runge(xx);
yyi = interp(xx, x, y, d);
plot2d(xx, [yyi yye], style=[2 5], leg="interpolation spline@exact function")
xtitle("interpolation of the Runge function")
```

Exercice 10.7.1 *Ajouter des commentaires qui explique le programme, en utilisant le Help de Scilab.*

Voilà un programme de la documentation Scilab, qui illustre plusieurs types de splines, appliquées aux données uniformes :

```
// behavior of différent splines on random data
a = 0; b = 1;           // intervalle of interpolation
n = 10;                 // nb of interpolation points
                        // discretisation for evaluation
x = linspace(a,b,n)'; // abscissae of interpolation points
y = rand(x);           // ordinates of interpol points
clf()
plot2d(x,y,-9, strf="000") // show interpol points
```

§. Par contre, l'interpolation polynomiale avec des points choisis, comme les points de Chebyshev, est remarquablement efficace, comme démontré par le logiciel recent Chebfun.

```

m = 800;
xx = linspace(a,b,m)';
yk = interp(xx, x, y, splin(x,y,"not_a_knot"));
yf = interp(xx, x, y, splin(x,y,"fast"));
ym = interp(xx, x, y, splin(x,y,"monotone"));
plot2d(xx, [yf ym yk], style=[5 2 3], strf="121", ...
        leg="fast@monotone@not a knot spline")
xlabel("Various spline and sub-splines on random data")
show_window()

// an example showing C2 and C1 continuity of spline and subspline
a = -8; b = 8;
x = linspace(a,b,20)';
y = sinc(x);
dk = splin(x,y); // not_a_knot
df = splin(x,y, "fast");
xx = linspace(a,b,800)';
[yyk, yy1k, yy2k] = interp(xx, x, y, dk);
[yyf, yy1f, yy2f] = interp(xx, x, y, df);
clf()
subplot(3,1,1)
plot2d(xx, [yyk yyf])
plot2d(x, y, style=-9)
legends(["not_a_knot spline","fast sub-spline","interpolation points"],...
        [1 2 -9], "ur",%f)
xlabel("spline interpolation")
subplot(3,1,2)
plot2d(xx, [yy1k yy1f])
legends(["not_a_knot spline","fast sub-spline"], [1 2], "ur",%f)
xlabel("spline interpolation (derivatives)")
subplot(3,1,3)
plot2d(xx, [yy2k yy2f])
legends(["not_a_knot spline","fast sub-spline"], [1 2], "lr",%f)
xlabel("spline interpolation (second derivatives)")

// here is an example showing the different extrapolation possibilities
x = linspace(0,1,11)';
y = cosh(x-0.5);
d = splin(x,y);
xx = linspace(-0.5,1.5,401)';
yy0 = interp(xx,x,y,d,"C0");
yy1 = interp(xx,x,y,d,"linear");
yy2 = interp(xx,x,y,d,"natural");
yy3 = interp(xx,x,y,d,"periodic");
clf()
plot2d(xx,[yy0 yy1 yy2 yy3],style=2:5,frameflag=2,leg="C0@linear@natural@periodic")

```

```

xtitle(" différent way to evaluate a spline outside its domain")

// this is an artificial example where the data xd and yd
// are build from a perturbed sin function
a = 0; b = 2*%pi;
sigma = 0.1; // standard deviation of the gaussian noise
m = 200; // number of expérimental points
xd = linspace(a,b,m)';
yd = sin(xd) + grand(xd,"nor",0,sigma);
n = 6; // number of breakpoints
x = linspace(a,b,n)';
// compute the spline
[y, d] = lsq_splin(xd, yd, x); // use equal weights
// plotting
ye = sin(xd);
ys = interp(xd, x, y, d);
clf()
plot2d(xd,[ye yd ys],style=[2 -2 3], ...
        leg="exact function@expérimental mesures (gaussian perturbation)@fitted spline
xtitle("a least square spline")
show_window()

```

Chapitre 11

Régression et interpolation par noyaux

La démarche habituelle dans la régression est d'estimer la valeur dans un point inconnu x par une moyenne pondérée des valeurs observées donc :

$$\hat{Y}(x) = \sum_i w_i(x) Y_i \quad (11.1)$$

où les poids $w_i(x)$ dépendront de x et de x_i . Il semble une bonne idée d'utiliser pour chaque point x seulement ses voisins, et dans une mesure inversement proportionnelle à leur proximité. Ce but pourrait être achevé en choisissant des poids de la forme

$$w_i = k(x)^{-1} \rho(\|x - x_i\|)$$

où $k(x) = \sum_i \rho(\|x - x_i\|)$, avec $\rho(r), r \geq 0$ une fonction décroissante dépendant de la distance $r = \|x - x_i\|$ (et peut être à support compact, donc $\rho(r) = 0$, pour r plus grand qu'un certain "rayon d'influence" R^1). La formule de régression (11.1) peut aussi être écrite dans la forme alternative :

$$\hat{Y}(x) = \sum_{i=1}^N \frac{\rho(\|x - x_i\|)}{k(x)} Y_i = \sum \alpha_i \rho(\|x - x_i\|) \quad (11.2)$$

où $\alpha_i = k^{-1} Y_i$.

Définition 11.0.1 On appellera **interpolation exacte par un noyau** $\rho(r), r > 0$, une interpolation de la forme :

$$\hat{Y}(y) = \sum_{i=1}^I \alpha_i \rho(|y - x_i|) \quad (11.3)$$

1. L'hypothèse $w_i(x) = \rho(|x_i - x|)$ que "l'influence" de x_i sur x est exercée seulement à travers une "fonction radiale" $\rho(r)$ de la distance $r = |x_i - x|$ est une simplification. Mais, on rencontre souvent de telles formules en physique, où la fonction $\rho(r)$ est un potentiel électrique ou gravitationnel, sous conditions d'homogénéité du champ, et aussi en statistique, si $g(x)$ est un champ Gaussien isotropique (voir chapitre 12.5) avec covariance $\text{Cov}(Y_i, Y) = \varphi(|x_i - y|)$ et **moyenne** 0.

où les coefficients **constants** α_i sont choisis tels que l'interpolation soit exacte en $x_i, i = 1, \dots, I^2$.

Une fois la "base" $\rho(r)$ choisie, l'interpolation (11.3) est facile à implémenter, car la demande qu'elle soit exacte dans les points observés nous fournit un système d'équations linéaires :

$$\mathbf{C} \boldsymbol{\alpha} = \mathbf{X}, \quad (11.4)$$

où $C_{ij} = \rho(|x_i - x_j|)$.

Remarquons, cf. (11.2), qu'on a deux formules (11.1) et (11.3) pour effectuer la même prévision $\hat{Y}(x)$. Pour les distinguer, nous les appellerons respectivement formules de régression et interpolation. Bien sûr, c'est la même formule (et la relation de "dualité" entre les deux formules sera encore discutée plus tard).

Exercice 11.0.2 Supposons un champ déterministe $g(x) = 4x^2$, observé en $x_1 = 0, x_2 = 1/2, x_3 = 1$. On a :

$$\lambda_1[y] := y(0) = 0, \quad \lambda_2[y] := y(1/2) = 1, \quad \lambda_3[y] := y(1) = 4$$

a) Calculez l'interpolation $\hat{Y}(x)$ basée sur le noyau Gaussien $\varphi(r) = e^{-r^2}$, c.-à-d. l'approximation de la forme :

$$\hat{Y}(x) = \sum_i^3 \alpha_i e^{-(x-x_i)^2} = \alpha_1 e^{-x^2} + \alpha_2 e^{-(x-1/2)^2} + \alpha_3 e^{-(x-1)^2}$$

b) Donnez aussi la forme générale du système d'interpolation exacte obtenu pour les α , pour une fonction $\varphi(r)$ arbitraire.

c) Calculez l'interpolation $\hat{Y}(x)$ basé sur le noyau $\varphi(r) = r^2$, c.-à-d. l'approximation de la forme :

$$\hat{Y}(x) = \sum_i^3 \alpha_i (x - x_i)^2 = \alpha_1 x^2 + \alpha_2 (x - 1/2)^2 + \alpha_3 (x - 1)^2$$

en identifiant sur la route le "système d'interpolation".

Notes :

1. Pour une fonction $\rho(r)$ décroissante à support compact de rayon R , les coefficients de régression sont

$$w_i(x) = \frac{\rho(\|x - x_i\|)}{\sum_{j:\|x-x_j\|\leq R} \rho(\|x - x_j\|)}$$

2. En choisissant $R \geq \max_{i,j} \|x_i - x_j\|$, tous les points inclus dans l'enveloppe convexe subiront l'influence de toutes les observations.

2. La formule d'interpolation (11.3) apparaît souvent dans le contexte de l'interpolation déterministe. Par exemple, supposons que Y_i sont les intensités des quelques sources d'un champ déterministe électromagnétique, situées dans les positions $x_i, i = 1, \dots, I$ et qu'on veut prédire l'effet de ce champ dans la position y . Le champ produit par la source de x_i en y est proportionnel au $\rho(|y - x_i|)$, ou $\rho(r) = r^{-2}$ et donc le champ en y sera de la forme $Y = \sum_i \alpha_i \varphi(y - x_i)$.

3. Par contre, en choisissant $R \leq \frac{\min_{i,j} \|x_i - x_j\|}{2}$, chaque point reçoit la valeur de l'unique voisin le plus proche, s'il se trouve à une distance plus petite que R (et 0 autrement). L'interpolation (11.2) est dans ce cas exacte, c.-à-d. on trouve que pour $x = x_i$ on a $w_j(x) = \delta_i(j)$ et donc $\hat{Y}(x_i) = Y_i$. Par contre, les prévisions sont nonlisses, données par la fonction

$$\hat{Y}(x) = \sum_{i=1}^I Y_i 1_{x \in S_R(x_i)}$$

où $S_R(x_i)$ est la sphère de rayon R autour de x_i .

4. Dans le cas $R \geq \frac{\min_{i,j} \|x_i - x_j\|}{2}$, la formule (11.2) devient lisse et plus intéressante, mais elle n'est plus exacte dans les points observés. Mais, on peut encore obtenir une interpolation exacte dans ce cas, en choisissant les coefficients α_i tels que le système $\hat{Y}(x_i) = Y_i, i = 1, \dots, I$ soit satisfait.

Une information importante pour la prévision de $\hat{Y}(x)$ est fournie par les coordonnées $x_i, i = 1, \dots, d$. Plus généralement, on pourrait ajouter une famille des fonctions des coordonnées $p_j(y), j = 1, \dots, J$ pour servir comme des prédicteurs exogènes. On arrive ainsi à prédire le champs par une combinaison linéaire :

$$\hat{Y}(y) = \sum_{i=1}^I \alpha_i \rho(|y - x_i|) + \sum_{j=1}^J \beta_j p_j(y) \quad (11.5)$$

où $p_j(y), j = 1, \dots, J$ sont des prédicteurs exogènes (fonctions des coordonnées, par exemple polynômes), et $\rho(|y - x_i|), i = 1, \dots, I$ sont des poids de pondération à accorder aux observations Y_i (qui seront incluses en α_i). Cette fois ci, on calcule α, β par le **système de krigeage dual** :

$$\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix}, \quad (11.6)$$

où \mathbf{P} dénote la matrice définie par $\mathbf{P} = (p_j(x_i) : i = 1, \dots, I, j = 1, \dots, M)$ (i.e les fonctions de base évaluées dans les points d'observation). Cette formule d'interpolation peut-etre justifiée dans un cadre déterministe : par exemple, les premières équations ne sont que les conditions d'interpolation exacte. Mais, la démarche adoptée par les géostatisticiens et les météorologues est de justifier cette formule dans un cadre stochastique, en supposant que le champs $g(x)$ est une réalisation d'un processus stochastique (c.-à-d. une famille des variables aléatoires), et de lier le noyau $\rho(y - x)$ à la covariance du champs. Remarquons d'emblée que l'avantage de la modélisation stochastique sur l'approche déterministe n'est pas évident, surtout avec peu des données.

Cette approche décompose le processus en :

$$g(x) = m(x) + Z(x) \quad \text{où :} \quad (11.7)$$

1. $m(x)$ représente la moyenne $\mathbb{E}g(x)$, ou une première prévision basée sur la "tendance". La fonction $m(x)$ est **inconnue** et nonisotropique, mais supposée d'appartenir à un espace vectoriel (des fonctions) de dimension finie, c.-à-d. d'admettre une décomposition

paramétrique

$$\boxed{m(x) = \sum_{j=1}^M \beta_j p_j(x) = \boldsymbol{\beta}^t \mathbf{p}(x)} \quad \text{où } p_1(x) = 1 \quad (11.8)$$

Ici, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_M)$, $\mathbf{p}(x) = (p_1(x), \dots, p_M(x))$ dénotent des vecteurs colonne, représentant respectivement des coefficients à déterminer et une base de l'espace. D'habitude, on prend pour $p_j(x)$ une base des polynômes à degré inférieur à un nombre donné m ; une autre possibilité est d'utiliser la base du noyau d'un opérateur différentiel (quand on a des raisons à croire que l'opérateur concerné intervient dans la physique du phénomène observé). Donc, on peut écrire le vecteur des moyennes dans les points d'observation comme

$$\mathbf{m} = \mathbf{P}\boldsymbol{\beta}$$

- $Z(x) = g(x) - m(x)$ est supposé un processus stochastique (qui essaie de capter des corrélations inconnues locales) **stationnaire** et **isotrope** à moyenne 0 et covariance \mathbf{C} normalement non-diagonale, au moins dans un sous-ensemble "homogène". Dans le cas le plus simple de covariance diagonale, on parle d'un bruit blanc (=non-corrélé), la variance du quel fourni des intervalles des confiance pour les predictions, un premier pas vers la modélisation stochastique. Dans la présence des corrélations non-nulles, elles devront être prises en compte dans la prediction. **Projet** : Sur votre jeu des données préféré, tester l'hypothèse du bruit blanc, et de la Gaussianité du bruit. Quels sont les tests les plus populaires? Pour inspiration sur des projets possibles, voir <http://robjhyndman.com/TSDL/> <http://www.tsi.enst.fr/moulines/enseignement/Projet/FinancialTimeSeries/index.html>, <http://www.physionet.org/>
- Dans le cas de champs spatiales, on fait souvent l'hypothèse d'isotropie, qui rend l'estimation beaucoup plus simple. Pour les champs Gaussiens, ça revient à supposer que les covariances entre les variables observées sont de la forme :

$$\tilde{\mathbf{C}}_{i,j} := \mathbb{E}[(Y(x_i) - m(x_i))(Y(x_j) - m(x_j))] = \sigma^2 \rho(\|(x_i - x_j)/R\|)$$

où $\sigma^2 = \text{Var}(Z(x))$, R est appelé la **portée** ou "échelle de corrélation" et $\rho(r)$ est une fonction uni-dimensionnelle "radiale de base" / "noyau radial positif", qui mesure la décroissance de la **corrélation** en fonction de distance.

Le champs "sous-jacent" $g(x)$ est observé en plusieurs points x_1, \dots, x_I avec des erreurs d'observation ϵ_i de **variance** τ^2 , appelée aussi **effet de pépite**. Les observations sont donc :

$$\boxed{Y_i = Y(x_i) + \epsilon_i = m(x_i) + Z(x_i) + \epsilon_i}, i = 1, \dots, I \quad (11.9)$$

En tenant compte des erreurs d'observation ϵ_i , les covariances entre les variables observées deviennent :

$$\mathbf{C}_{i,j} = \tau^2 \delta_{i,j} + \sigma^2 \varphi(\|(x_i - x_j)/R\|)$$

On arrive ainsi à un modèle linéaire :

$$\boxed{\mathbf{X} = \mathbf{P}\boldsymbol{\beta} + \mathbf{Z} + \boldsymbol{\epsilon}} \quad (11.10)$$

avec des "résidus" Z corrélés et mal observés. Pour sa prediction, il faudra estimer au moins $\beta, \sigma^2, \tau^2, R$, et idéalement aussi la corrélation $\rho(r)$.

Le problème de prédiction de telles variables spatiales $g(x)$ est donc un problème de régression compliqué, qui sort de la trame de la régression classique, à cause de :

1. "bruits colorés" $Z(x)$ ayant une matrice de covariance \tilde{C} qui n'est pas diagonale et
2. la présence des moyennes $m(x)$ qui dépendent de x .

La formule de régression correspondante, appelée aussi "krigeage", sera développée dans les chapitres 3-4. Notons ici que les difficultés d'application ne résident pas bien sur dans l'application de la formule de krigeage, mais dans :

1. le choix du noyau $\rho(r)$
2. l'estimation des paramètres de la moyenne, ainsi que de la portée, pépite et variance.

Le premier problème ne sera pas discuté dans ces notes. Pour le deuxième, on verra dans le chapitre 2 l'approche de Mathéron par l'estimation du "variogramme" (une transformation linéaire de la corrélation).

Chapitre 12

L'analyse des données spatiales

12.1 Premier aperçu : la visualisation des données

Comme chaque analyse statistique, l'analyse des données spatiales commence par une visualisation des données. Les commandes en geoR et fields sont `points()`, `plot()`. Une première question est si le jeu de données est stationnaire, c.-à-d. si il semble homogène dans toute la région; si non, il faudra analyser séparément plusieurs régions¹. En supposant ces manipulations préalables déjà faites, on arrive au dilemme de l'estimation de la covariance/sémivariance – voir prochain chapitre.

12.2 Rudiments de statistique bivariée

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

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

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

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

Exercice 12.2.1 Montrez que le coefficient de corrélation des deux variables X, Y est nécessairement compris entre -1 et $+1$. a) Pour des variables discrètes qui prennent des valeurs $(x_i, y_i), i = 1, \dots, n$ avec des probas égales. b) Pour des variables \tilde{X}, \tilde{Y} , ayant la moyenne 0 et la variance 1. c) Le cas général.

Le coefficient de corrélation indique le degré de dépendance linéaire entre deux variables : - s'il est proche de ± 1 , il indique l'existence d'une relation linéaire entre variables; - s'il est proche de 0, il indique que les deux variables ne sont pas linéairement dépendantes (l'indépendance complète entre les variables correspond à une hypothèse plus forte, soit en

1. Sans homogénéité, il est impossible d'estimer les paramètres globaux qui sont à la base de tous les modèles probabilistes.

terme de probabilités : $f_{XY}(x, y) = f_X(x)f_Y(y)$, ce qui implique que la covariance et le coefficient de corrélation correspondants sont nuls - la réciproque n'étant pas nécessairement vraie). Les géostatisticiens utilisent également la **sémivariance**

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

où (X, Y) est une pair des v.a.

Exercice 12.2.2 Montrez que pour des variables X, Y avec la même variance $\text{Var}(X, X) = \text{Var}(Y, Y) = \sigma^2$, la covariance et sémivariance satisfont :

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

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

12.3 La distribution gaussienne/normale multivariée

Définition 12.3.1 Un vecteur colonne $\mathbf{X} = (X_1, \dots, X_I)$ a une distribution gaussienne/normale multivariée $N(\mathbf{m}, \mathbf{C})$, où \mathbf{m}, \mathbf{C} denotent un vecteur ligne et une **matrice positive définie**, ssi sa densité est de la forme

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^I \det(\mathbf{C})}} e^{-\frac{(\mathbf{X}-\mathbf{m})^{(t)}\mathbf{C}^{-1}(\mathbf{X}-\mathbf{m})}{2}}$$

La matrice positive définie $\mathbf{Q} = \mathbf{C}^{-1}$ dans l'exposant est appelée **matrice de précision**.

On voit que la distribution gaussienne multivariée est uniquement caractérisé par son espérance \mathbf{m} et matrice de covariance \mathbf{C} (ou, alternativement, par la matrice de précision \mathbf{Q}), et que le log de sa vraisemblance ("LOGLIK") est une fonction quadratique en \mathbf{X} , à paramètres \mathbf{Q}, \mathbf{m} (ou $\mathbf{C}^{-1}, \mathbf{m}$).

Les combinaisons linéaires des vecteurs Gaussiens sont aussi Gaussiens.

Exercice 12.3.1 Soit (X_1, \dots, X_4) un vecteur Gaussien centré, avec matrice de covariance

$$\mathbf{C} = \begin{pmatrix} 1 & \rho & 0 & \rho \\ \rho & 1 & \rho & 0 \\ 0 & \rho & 1 & \rho \\ \rho & 0 & \rho & 1 \end{pmatrix}$$

- a) Trouver la lois du vecteur $X_1 - X_3$ b) Trouver la loi du vecteur $Y = (X_1 + X_2, X_1 - X_3)$
c) Déterminer des conditions à imposer sur ρ pour que le pb. soit bien défini.

R : c) Il y a deux méthodes pour vérifier la positivité définie de la matrice \mathbf{C} :

1. en vérifiant la positivité des valeurs propres
2. en vérifiant la positivité des déterminants principaux

La première est faisable ici, car il s'agit d'une matrice circulante (ou ciclyque) de la forme :

$$C = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ a_3 & a_0 & a_1 & a_2 \\ a_2 & a_3 & a_0 & a_1 \\ a_1 & a_2 & a_3 & a_0 \end{pmatrix}$$

Les valeurs propres d'une telle matrice sont données par $a(w), a(w^2), a(w^3), \dots, a(w^n)$ où w, \dots, w^n sont les racines de l'unité, et $a(x) = \sum_{i=0}^{n-1} a_i x^i$ est la fonction génératrice de la première ligne. Cette formule rend ici $a(x) = 1 + \rho x(1 + x^2)$, et les valeurs propres sont $1 + 2\rho, 1 - 2\rho, 1, 1$. La positivité définie demande donc $|\rho| \leq 1/2$.

Proposition 12.3.1 Soit \mathbf{X} un vecteur aléatoire à valeurs dans \mathbb{R}^I . Pour que la loi de \mathbf{X} soit gaussienne, il faut et il suffit que la fonction caractéristique de \mathbf{X} (ou transformée de Fourier de la loi de \mathbf{X}) s'écrive

$$\phi^{\mathbf{X}}(\mathbf{z}) = E(e^{i\langle \mathbf{z}, \mathbf{X} \rangle}) = e^{i\langle \mathbf{z}, \mathbf{m} \rangle - \frac{1}{2}\langle \mathbf{z}, \mathbf{C}\mathbf{z} \rangle}, \quad \mathbf{z} \in \mathbb{R}^I$$

avec $\mathbf{m} \in \mathbb{R}^I$ et \mathbf{C} une matrice symétrique d'ordre I , positive.

Rq : La fonction génératrice des moments est

$$E(e^{\langle \mathbf{z}, \mathbf{X} \rangle}) = e^{\langle \mathbf{z}, \mathbf{m} \rangle + \frac{1}{2}\langle \mathbf{z}, \mathbf{C}\mathbf{z} \rangle}, \quad \mathbf{z} \in \mathbb{R}^I$$

Exercice 12.3.2 L'espérance et la matrice de covariance de la distribution gaussienne multivariée sont

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

c.-à-d. on a $\text{Cov}(Y_i, X_j) = \mathbf{C}_{i,j}$.

L'indépendance de deux variable aléatoire X et Y entraîne toujours $\mathbf{C}(X, Y) = 0$. La réciproque, fautive en général, est vraie si la loi jointe de X et Y est gaussienne :

Proposition 12.3.2 Soient $X_1 \dots X_r$ des vecteurs à valeurs dans \mathbb{R}^p . Supposons la loi jointe de $X_1 \dots X_r$ gaussienne. Alors les $X_1 \dots X_r$ sont indépendants si et seulement si les matrices $\mathbf{C}(X_i, X_j)$, $i \neq j$, $1 \leq i, j \leq r$ sont nulles.

(La preuve est immédiate en utilisant la fonction caractéristique de $X_1 \dots X_r$).

12.4 Les distributions Gaussiennes conditionnelles, la régression linéaire et l'interpolation

Il s'avère que pour la distribution gaussienne multivariée, l'espérance conditionnelles

$$\hat{Y} = E(Y/X_1 \dots X_I)$$

est une combinaison linéaire de (X_1, \dots, X_I) , par conséquent aussi Gaussienne, et facilement calculable par ses propriétés de variance minimale, et d'orthogonalité par rapport à l'erreur de prédiction, ce qui explique l'intérêt d'utiliser cette distribution dans la modélisation, quand permis par les testes de normalité. Révisons le résultat fondamental de la prédiction Gaussienne, en vue de l'appliquer à un champ Gaussien $g(x), x \in \mathbb{R}^d$.

Proposition 12.4.1 Soient $\mathbf{X} = (X_1, X_2)$ un vecteur Gaussien de moyenne (m_1, m_2) est covariance $C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$. Alors, la loi de $X_1|X_2$ est Gaussienne, de moyenne et variance :

$$m_1 + C_{12}C_{22}^{-1}(X_2 - m_2), \quad C_{11} - C_{12}C_{22}^{-1}C_{21}$$

Considerons maintenant le cas centré avec une seule variable à prévoir, pour mettre l'accent sur les deux formules alternatives pour exprimer l'espérance conditionnelle (12.3).

Proposition 12.4.2 Soient $\mathbf{X} = (X_1 \dots X_I)$ (un vecteur colonne), et Y des variable aléatoire réelles de moyennes nulles, définies sur (Ω, P) . Supposons la loi jointe de $X_1 \dots X_I, Y$ gaussienne. Posons $\mathbf{C}_{X,Y} = (\text{Cov}(Y_i, Y), i = 1, \dots, I)$ (un vecteur ligne), et soit $\mathbf{C} = \mathbf{C}_{X,X} = (\text{Cov}(Y_i, X_j), i, j = 1, \dots, I)$ la matrice des covariances des "regresseurs". Alors

$$\hat{Y} = E(Y|X_1 \dots X_I)$$

est la projection orthogonale dans $L^2(\Omega, P)$ de Y sur l'espace engendré par $X_1 \dots X_I$ et a une distribution gaussienne, à espérance :

$$\hat{Y} = (\mathbf{C}_{YX} (\mathbf{C}_{XX})^{-1}) \mathbf{X} = \sum_{i=1}^I w_i Y_i \quad \text{Regression} \quad (12.1)$$

où le vecteur ligne $\mathbf{w} = (w_i, i = 1, \dots, I) = \mathbf{C}_{YX} (\mathbf{C}_{XX})^{-1}$ vérifie le système d'équations normales de régression :

$$\mathbf{w} \mathbf{C}_{XX} = \mathbf{C}_{YX} \quad (12.2)$$

Alternativement

$$\hat{Y} = (\mathbf{C}_{YX} (\mathbf{C}_{XX})^{-1}) \mathbf{X} = \mathbf{C}_{YX} ((\mathbf{C}_{XX})^{-1}) \mathbf{X} = \sum_{i=1}^I \text{Cov}(Y_i, Y) \alpha_i \quad \text{Regression - Interpolation} \quad (12.3)$$

où le vecteur colonne $\boldsymbol{\alpha} = (\alpha_i, i = 1, \dots, I) = (\mathbf{C}_{XX})^{-1} \mathbf{X}$ vérifie le système d'équations d'interpolation :

$$\mathbf{C}_{X,X} \boldsymbol{\alpha} = \mathbf{X} \quad \text{Systeme d'interpolation} \quad (12.4)$$

La variance de l'erreur de prediction de Y est :

$$\hat{\sigma}_{YY}^2 = \mathbb{E}[(Y - \hat{Y})^2] = \sigma_{YY}^2 - \mathbf{C}_{YX} (\mathbf{C}_{XX})^{-1} \mathbf{C}_{XY} = \sigma_Y^2 - \mathbf{w} \mathbf{C}_{YX} = \sigma_Y^2 - \sum_{i=1}^I w_i \text{Cov}(Y, Y_i) \quad (12.5)$$

Notes :

1. La formule de prediction (12.1) peut être écrite comme un produit scalaire par rapport à l'inverse de la matrice de covariance $\mathbf{C} = \mathbf{C}_{X,X}$:

$$\hat{Y}(x) = \langle \mathbf{C}_{YX}, \mathbf{X} \rangle_{\mathbf{C}^{-1}}$$

- Le système d' "interpolation" (12.4) et le système d' "équations normales" (12.2) (dû à Kolmogorov et Wiener²) utilisent la même matrice \mathbf{C} , contenant les covariances entre les variables explicatives, dans le premier membre. Le deuxième membre par contre est indépendant de Y pour l'interpolation, et dépendent de Y pour le système normal (car faisant intervenir les covariances entre les variables explicatives et la variable estimée). Il suit que pour applications à plusieurs Y , le système d'interpolation est le plus convenable (car il suffit de le résoudre une fois).
- L'extension de la formule (12.1) au cas des moyennes non nulles, mais **connues**, est immédiate; on arrive ainsi à la fameuse formule de régression, qui prédit une variable Y par une pondération linéaire de sa moyenne et d'autres variables "explicatives moyennées" $Y_i - \mu_i, i = 1, \dots, I$:

$$\hat{Y} = \mu_y + \sum_{i=1}^I w_i (Y_i - \mu_i) \quad (12.6)$$

- La formule de la variance conditionnelle (12.5) permet de voir la contribution de chacune des variables dans la réduction de la variance.

Exercice 12.4.1 Soient X_1, \dots, X_n des variables Gaussiennes, indépendantes centrées, des variances $\sigma_1^2, \dots, \sigma_n^2$. Calculez la loi de $X_1 | X_2, \dots, X_n$.

Exercice 12.4.2 a) Soit $Y = (Y_1, Y_2)$ deux des coordonnées d'un vecteur \mathbf{Z} à distribution gaussienne $N(\mathbf{0}, \mathbf{C})$, et soit X le reste. Montrez que la corrélation conditionnelle de Y_1, Y_2 en sachant X , noté par $\hat{C}or_{Y_1 Y_2}$, est :

$$\hat{C}or_{Y_1 Y_2} = -\frac{Q_{Y_1 Y_2}}{Q_{Y_1 Y_1} Q_{Y_2 Y_2}}$$

b) Montrez, en utilisant les formules de la distribution Gaussienne conditionnelle d'abord quand $Y = Y_i$ est une seule composante (dans quel cas on dénote les autres composantes par $Z = X_{-i}$), et ensuite quand $Y = (Y_i, X_j)$ est un couple (dans quel cas on dénote les autres composantes par $Z = X_{-(i,j)}$), que :

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

12.5 Processus/champs gaussiens

Hypothèse A : En géostatistique, on travaille souvent avec des "champs gaussiens" $g(x), x \in \mathbb{R}^d$.

Définition 12.5.1 Un processus aléatoire/champs indexé par T est dit gaussien si pour toute famille finie $x_1 \dots x_I \in T$, la loi jointe de $(X_{x_1} \dots X_{x_I})$ est gaussienne.

2. Ce système apparaît souvent en statistique; en particulier, il généralise la formule de prévision univariée $\hat{Y} = \mu_Y + a(Z - \mu_Z)$, où $w = \frac{\text{Cov}(Y,Z)}{\text{Cov}(Z,Z)} = \frac{\sigma_{YX}}{\sqrt{\sigma_Z^2 \sigma_Y^2}}$, ainsi que le système de Yule-Walker pour les séries temporelles autorégressives.

Proposition 12.5.1 *Un processus gaussien X indexé par \mathbb{R}^d ou \mathbb{Z}^d est stationnaire strict si et seulement si il est stationnaire à l'ordre 2.*

Preuve : Si X est gaussien, $(X_{t_1} \dots X_{t_r})$ et $(X_{t_1+s} \dots X_{t_r+s})$ sont des vecteurs gaussiens. Leurs lois jointes coïncident si et seulement si elles ont même moyenne et même matrice de covariance. \square

Un champ gaussien $Y(t)$ est complètement déterminé par sa fonction d'espérance $m(t)$ et son "noyau" de covariance des deux variables $C(t_1, t_2) = \text{Cov}(Y(t_1), Y(t_2))$; espérance $m(t)$ est arbitraire, mais afin que le noyau de covariance $C(t_1, t_2)$ soit "admissible", il faut qu'il soit positive définie.

Les processus gaussiens se pretent facilement au calcul. En effet, non corrélation et indépendance coïncident, les moments d'ordre 2 déterminent complètement la loi globale du processus, stationnarité stricte et stationnarité à l'ordre 2 sont équivalentes, et conditionner par une famille de variable aléatoire revient à faire une projection orthogonale sur l'espace vectoriel engendré par cette famille. Ainsi, pour les processus gaussiens, les techniques de prédiction linéaire optimale - les seules d'emploi facile - sont en fait optimales dans un sens plus fort, plus précisément au sens des moindres carrés. Pour toutes ces raisons, le modèle gaussien est extrêmement utilisé dans la pratique. De plus, le théorème limite central justifie, dans certains cas, son utilisation d'un point de vue théorique. Il faut toutefois être prudent aussi bien d'un point de vue pratique que théorique. Suivant les vitesses de convergence dans le T.L.C., l'asymptotique n'est pas forcément atteinte. D'autre part, si le calcul d'un estimateur des moindres carrés est généralement aisé, la solution obtenue peut être très instable.

La prédiction d'un champs dépendra essentiellement de l'estimation de sa fonction de covariance/variogramme. On s'attend a priori que ces mesures de similitude/dissimilitude vont dépendre de la distance entre deux points observés x, y , et peut être aussi du vecteur $h = y - x$ (donc de son magnitude *et direction*). **Hypothèse de stationarité** : Pour simplifier, supposons que les fonctions

$$\gamma(h, x) := \frac{\text{Var}(Y(x+h) - g(x))}{2} \quad (12.7)$$

$$\varphi(h, x) := \mathbb{E}(Y(x+h) - m(x+h))(g(x) - m(x)) \quad (12.8)$$

$$\rho(h, x) := \frac{\mathbb{E}(Y(x+h) - m(x+h))(g(x) - m(x))}{\sigma(x+h)\sigma(x)} \quad (12.9)$$

ne dépendent pas du point de départ x , qui serait donc omis par la suite. On suppose en particulier que l'écart type est indépendant du point d'observation, et donc $\sigma(x+h) = \sigma(x) := \sigma$.

Les fonctions ci-dessus, appelées variogramme, covariance et corrélogramme, expriment la variabilité/similitude entre deux valeurs du champs $Y(y), g(x)$, par rapport au vecteur $h = y - x$ donnant leur position relative. Elles forment la base de l'approche géostatistique, qui postule que le meilleur choix de base pour l'interpolation RBF est la fonction $\rho(h)$ (ou $\gamma(h)$). Remarquons maintenant que les relations

$$\gamma(h) = \sigma^2 - \varphi(h) = \sigma^2(1 - \rho(h)) \quad (12.10)$$

$$\varphi(h) = \gamma(\infty) - \gamma(h) \quad (12.11)$$

rendent les deux analyses équivalentes en principe. En effet, pour chaque test de la covariance il existe un test analog de la sémivariance. Le fait que ces paires de tests donnent des résultats parfois différents, et qu'il n'y a pas d'avantages clairs de l'un par rapport à l'autre, est l'un des nombreux mystères de la statistique spatiale.

12.6 Processus Gaussiens stationnaires et isotropiques

L'analyse la plus simple des observations spatiales se fait dans le cas où on peut supposer que le champ observé est stationnaire et isotrope. Dans ce cas, il suit que les covariances sont de la forme

$$\text{Cov}(Y_i, X_j) = \varphi(|x_i - x_j|), i = 1, \dots, I \text{ Hypothèse Isotropie}$$

pour une certaine "fonction radiale" $\varphi(r)$ d'une variable. Pour les champs isotropes, la formule d'interpolation (12.3) devient :

$$\hat{Y}(y) = \sum_i \alpha_i \varphi(|y - x_i|) \quad \text{Interpolation RBF} \quad (12.12)$$

On arrive ainsi à une interpolation par des translatées d'une **fonction radiale de base (RBF)** $\varphi(r)$ et la formule de prévision des champs Gaussiens isotropiques coïncide avec celle de l'interpolation RBF. Posons désormais

$$\varphi(r) = \sigma^2 \rho(r),$$

où $\rho(r)$ denote la fonction de corrélation. Choisir la fonction $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}$ est la question fondamentale. Comme l'estimation d'une fonction (même d'une seule variable) demande beaucoup de données, on se contente normalement à supposer une famille paramétrique déterminée par deux-trois paramètres. Quand-même, il est naturel (en statistique ainsi qu'en interpolation) de travailler avec des fonctions $\varphi(r)$ qui donne naissance à un noyau $C(x, y) = \varphi(|x - y|)$ strictement positif défini. Une question plutôt difficile qui sera discutée dans la section ?? est de caractériser la classe des fonctions d'une variable $SPD(d)$ donnant naissance à un noyau strictement positif défini en d dimensions.

12.7 Le corrélogramme et le variogramme spatiaux empiriques

Un des aspects saillants de la géostatistique est la préoccupation de donner une bonne estimation (et utilisation) de la fonction de covariance. Considerons donc cet aspect, en supposant pour l'instant, pour simplifier, un champ moyenné **isotrope**, observé "sans pepite", c.-à-d. le cas $Y_i = Z(x_i)$, où Z est un processus Gaussien de moyenne 0 et covariance/variogramme $\varphi(h)/\gamma(h), h \in \mathbb{R}^d$, qu'on aimerait à estimer.

1. Soient $r_i, i = 1, \dots, I$ les distances entre les points observés, et supposons que pour chaque distance un nombre suffisant de paires a été observé, tq. l'estimation empirique de $\gamma(r_i)$ est possible (cela est toujours réalisable, en groupant si besoin les paires observés

avec des distances proches dans un seul group, c.-à-d. en réalisant un groupement (histogram) de $\gamma(r)$. On calcule alors

$$\gamma(r_i) = N(r_i)^{-1} \sum_{\{(x_i, x_j): |x_i - x_j| \approx r_i\}} \frac{(Y_i - X_j)^2}{2}, \quad (12.13)$$

où $N(r_i)$ est le nombre des paires à distance r_i .

2. Soit maintenant $\gamma(r)$ un ajustement (polynomiale, ou par splines, ou par un noyau, ...) du variogram empirique.

Les variogrammes $\gamma(\rho)$ obtenues ainsi auront typiquement trois caractéristiques saillantes :

1. la **pépite**="nugget" τ^2 ; c'est une discontinuité (saut) en l'origine, qui peut être attribuée à l'impossibilité d'estimer les détails microscopiques.
2. le **palier total**="sill" $\tau^2 + \sigma^2$; c'est la limite de la sémivariance quand la distance des paires augmente.
3. la **portée**="range parameter" R , dénoté par φ en geoR : c'est la distance au delà de laquelle la corrélation devient pratiquement 0 et la sémivariance devient constante (égale au palier total), indépendamment de la distance. Autrement dit, au delà de la portée, la structure disparaît, et la sémivariance devient égale à la variance des observations (l'inclusion des paires au delà de la portée dans le variogramme ne ramène donc aucune information).

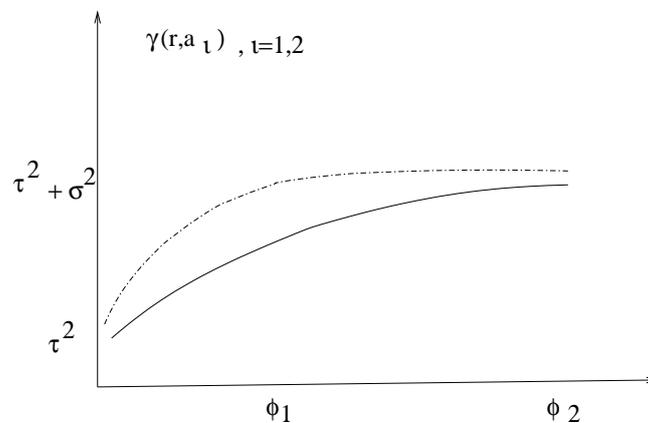


FIGURE 12.1 – Variogrammes dans les directions d’anisotropie principales : $\gamma(r) = \tau^2 I_{\{r>0\}} + \sigma^2(1 - \rho(r)) \iff \varphi(r) = \gamma(\infty) - \gamma(r) = \tau^2 I_{\{r=0\}} + \sigma^2 \rho(r)$

La figure ci-dessus expose les variogrammes dans les deux directions d’anisotropie principales, qui sont les vecteurs propres de l’approximation quadratique de $\gamma(h)$ autour de l’origine. La direction à portée plus grande φ_2 est la direction du gisement ; celle à portée plus courte φ_1 est perpendiculaire à la direction du gisement. Si on trouve des allures différentes pour les corrélogrammes/variogrammes directionnels, on appelle le champs **anisotropique**. Si par contre l’allure est similaire, on appelle le champs **isotropique** et on peut grouper toutes les paires d’observations ensemble. La première estimation ponctuelle des angles d’isotropie, des paramètres ϕ, σ, τ , d’une tendance linéaire, etc. se fait en geoR avec la commande

"likfit" suivi par "summary". L'estimation de la distribution des paramètres se fait avec "profilik".

Remarque 12.7.1 Implementation en GeoR : *Au cas non-isotrope, on aimerait quand-même réduire l'étude des fonctions $\varphi(h), \gamma(h)$ au cas uni-dimensionnel (comme dans le cas isotropique). Pour cela, en fixant d'abord une direction, on pourrait examiner le corrélogramme ou le **variogramme** pour chaque direction, ou plutôt à l'intérieur d'un "angle de tolérance" autour des quelques directions principales (la valeur défaut de l'angle de tolérance en geoR est 22.5 def). En pratique, on examine deux-quatre directions principales (geoR offre la commande spécialisé `variog4`). Si le champs est approximativement isotropique, on peut même grouper toutes les corrélations observées ensemble ; c'est en effet ce qui est fait par `variog` si on ne spécifie pas la "direction".*

Après la restriction à un angle donné, on a à faire avec un corrélogramme/variogramme unidimensionnel, où on s'intéresse à la décroissance/croissance de la covariance/sémivariance par rapport à la distance r entre deux points.

*Quoi-que le "nuage=cloud" initial des paires (distance, sémivariance) a d'habitude un aspect chaotique, on rencontre souvent, après le groupage des données habituel en histogrammes, des **variogrammes empiriques** qui suggèrent une variation structurée par rapport à la distance. En geoR, la commande `variog` a deux options principales : "cloud" et "bin", la deuxième étant le défaut. Les multiples sorties peuvent être listées par "names(variogramme)".*

12.8 Le corrélogramme et variogramme théoriques

La première question dans une analyse spatiale c'est de décider si le champs est isotropique ou pas. Supposons maintenant un champs isotropique –c'est d'ailleurs principalement le seul cas pour le quel les avantages de la modélisation théorique sont manifestes. Comme dans le cas des séries temporelles, on aimerait avoir la possibilité de choisir, à partir du corrélogramme/variogramme, un modèle théorique convenable. Rappelons qu'en séries temporelles, les modèles théoriques les plus simples pour le bruit sont :

1. bruit blanc, avec corrélation $\rho = 0$.
2. moyennes mobiles, avec corrélation ρ de **portée finie**.
3. processus AR(p), de portée infinie, entre les quelles le AR(1) satisfait aussi la propriété de Markov.

Une examination de la corrélation ρ ou de la corrélation ρ partielle permet un choix entre ces modèles. Des analogues existent dans la statistique spatiale : bruit blanc (invariants par rapport aux permutations des valeurs sur les positions), le modèle sphérique et le modèle Sobolev/Bessel/Matérn, qui est Markovien. Mais, l'identification "à l'oeil" d'une classe convenable en partant de la corrélation ρ est plutôt impossible. On recourt donc à des essais, en traversant une liste des modèles définis par leur fonctions de covariance (qui sont aussi les **fonctions radiales positives définies** utilisées dans l'interpolation déterministe—voir chapitre 14.1). Quelques unes des corrélations/noyaux radiales les plus utilisées, sans pépite et **normalisées** (à portée 1 pour le cas de support compact), sont :

Nom	$\hat{\rho}(r)$	CPD : m=
Sobolev/Bessel/Matern	$\frac{2\pi^{d/2}}{\Gamma(\nu+d/2)} K_\nu(r)(r/2)^\nu, \nu > 0$	0
exponentiel, $\nu = 1/2$	e^{-r}	0
$\nu = 3/2$	$(1+r)e^{-r}$	0
$\nu = 5/2$	$(1+r+r^2/3)e^{-r}$	0
Whittle, $\nu = 1$	$rK_1(r)$	0
Gaussien	e^{-r^2}	0
sphérique, $d = 1$ (support compact)	$(1-r)_+$	0
sphérique, $d = 2$ (support compact)	$(1 - \frac{2}{\pi}(r\sqrt{1-r^2} + \arcsin(r)))_+$	0
sphérique, $d = 3$ (support compact)	$(1 - 3/2 r + 1/2 r^3)_+$	0
linéaire	r	1
cubique	r^3	2

Notes :

1. Le Sobolev/Bessel/Matern, où $K_\nu(r)$ est la fonction modifiée de Bessel d'index ν , est un "super-modèle" qui contient en dehors des trois paramètres usuels un paramètre $\nu = k - d/2$ de lissité. Quand $\nu = k - d/2$ est multiple de $1/2$, on obtient plusieurs cas particuliers remarquables : le modèle exponentiel, avec $\nu = 1/2$ et covariance $\hat{\rho}(r) = \exp(-r)$; le modèle de Whittle, avec $\nu = 1$. Populaires aussi sont le modèle Gaussien, avec $\hat{\rho}(r) = -\exp(-r^2)$, obtenu par la limite $\nu \rightarrow \infty$ et le modèle de de Wijs, avec $\hat{\rho}(r) = -\log(r)$, obtenu par la limite $\nu \rightarrow 0$; ce dernier modèle est rencontré assez souvent dans l'agriculture pour mériter le nom : **la loi du terroir**.

2. En pratique, pour tenir compte de la portée, pépite et du palier, on utilisera la fonction de base

$$\varphi(r) = \tau^2 I_{\{r=0\}} + \sigma^2 \hat{\rho}\left(\frac{r}{R}\right),$$

où $\hat{\rho}(r)$ est la fonction de corrélation standardisée (tq $\hat{\rho}(1)$ est "suffisamment petit"). Par exemple, dans le cas de support compact, on normalise tq $\hat{\rho}(1) = 0$, la portée R coïncide avec le rayon du domaine. Pour les corrélations ρ de support non-compact, il s'agit plutôt des choix traditionnels.

3. La corrélation Gaussienne est rarement utilisée, car elle implique un champs "super-lisse" (infiniment différentiable), jamais rencontré en pratique.
4. La corrélation sphérique représente le volume de l'intersection des deux sphères de rayon $1/2$, situées à distance $r \leq 1$ l'une de l'autre.
5. Les corrélations "admissibles" sont définies par la propriété de positivité définie – voir section 12.9.
6. On utilise parfois des "corrélations généralisées", comme $\rho(r) = r^3$, qui ne convergent pas vers 0, et sont positives définies seulement sur un espace des fonctions orthogonales aux polynômes d'ordre m . Cela est indiqué dans la dernière colonne, par un "CPD" m qui n'est pas 0 – voir chapitre ??.
7. Le choix du $\rho(r)$ porte sur la lissité : par exemple, en $d = 1$, $\rho(r) = r$ ramène à une interpolation linéaire par morceaux, tandis que $\rho(r) = e^{-r^2}$ ramène à une interpolation très lisse.

Exercice 12.8.1 En utilisant un logiciel, calculez l'interpolation $\hat{Y}(x)$ basé sur le noyau Gaussien $\varphi(r) = e^{-\left(\frac{r}{R}\right)^2}$, pour $R = 1, R = .5, R = 2$ et $R = 10$. Calculez aussi l'erreur maximale $\sup_{x \in [0,1]} |g(x) - \hat{Y}(x)|$, et l'erreur intégrée $\int_x |g(x) - \hat{Y}(x)|^2 dx$ dans chaque cas.

Le choix du meilleur R est un dylème.

Sommaire : En pratique, on observe donc le corrélogramme ou plutôt le variogramme empiriques, en identifiant la portée, le palier et l'effet de pepite, et on choisit plutôt arbitrairement l'une des fonctions radiales ci-dessus, avec les trois paramètres ci-dessus qui resorted des valeurs empiriques. Des tests adéquats devront être faits sur les résidus, pour confirmer le modèle théorique choisi. Après le choix du modèle, on utilise le système linéaire de krigeage/interpolation pour donner une prévision du champs.

12.9 Les noyaux positifs (*)

Définition 12.9.1 Soit $n \in \mathbb{N}$ et soit $\mathcal{T} \in \mathbb{R}^d$ un ensemble arbitraire de n points. Une fonction $C(t_1, t_2)$ sera appelée noyau (strictement) positif ssi pour n'importe quel n et n'importe quel ensemble de n points \mathcal{T} , les matrices $\mathbf{C}_{\mathcal{T}} = C(x_i, x_j), i = 1, \dots, n, j = 1, \dots, n$, où $x_i, x_j \in \mathcal{T}, j = 1, \dots, n$, sont (strictement) positives définies.

Exemple 12.9.1 La fonction $C(t_1, t_2) = \delta(t_1 - t_2)$, où δ est le symbol de Kronecker, est un noyau strictement positif. Pour voir cela, on vérifie que pour n'importe quel n et n'importe quel ensemble de n points \mathcal{T} , la matrice $\mathbf{C}_{\mathcal{T}} = \mathbf{Id}(n)$ est strictement positive définie.

Corollaire 12.9.1 Si $C(t_1, t_2)$ est un noyau strictement positif, alors les matrices $\mathbf{C} = C(x_i, x_j), i = 1, \dots, n, j = 1, \dots, n$ sont **inversables**, pour n'importe quel n et n'importe quels points $x_i \in \mathbb{R}^d, i = 1, \dots, n$.

Ce corollaire demontre l'avantage d'interpoler avec un noyau strictement positif : c'est qu'on est assuré d'avoir toujours des solutions uniques pour les systèmes d'interpolation/krigeage qu'on aura besoin de résoudre.

Exercice 12.9.1 Montrez que si $X(x)$ est un champs Gaussien, alors pour n'importe quelles positions $\mathcal{T} = x_1 \dots x_I$, la matrice $\mathbf{C}_{\mathcal{T}} = (\text{Cov}(X(x_i), X(x_j))), i, j = 1, \dots, I$ est positive définie, c.-à-d. pour n'importe quel vecteur $\mathbf{w} = (w_1, \dots, w_I)$ on a :

$$\mathbf{w}^t \mathbf{C} \mathbf{w} \geq 0$$

L'inverse est aussi vraie :

Proposition 12.9.1 Il existe un champs Gaussien ayant la fonction de covariance $C(x, y)$ et l'espérance $m(x)$ ssi les matrices $C(x_i, x_j), i = 1, \dots, n, j = 1, \dots, n$ sont positives définies pour n'importe quel n et n'importe quels points $x_i, i = 1, \dots, n$.

Donc, la définition d'un noyaux positif est simplement la condition de positivité nécessaire pour pouvoir définir un champs Gaussien ayant la covariance donnée par ce noyau, cf Proposition 12.9.1. Pour cela, on appelle les noyaux positifs aussi noyaux de covariance. Les noyaux de covariance ont plusieurs propriétés qui permettent de construire des noyaux plus compliqués à partir des noyaux simples.

Lemme 12.9.1 Si C_1, C_2 sont des noyaux de covariance, alors les fonctions suivantes sont aussi des noyaux de covariance :

1. $a_1 C_1(x, y) + w_2 C_2(x, y) \quad \forall w_1, w_2 \geq 0$
2. $C_1(x, y) C_2(x, y)$
3. $pol(C_1(x, y))$, où pol est un polynôme arbitraire à coefficients positives.
4. $e^{C_1(x, y)}$
5. $C_1(\psi(x), \psi(y))$ où ψ est une fonction arbitraire.

Pour démontrer la deuxième propriété algébriquement, fixons d'abord un ensemble $\mathcal{T} = x_1 \dots x_I$. En suite, on construit le produit tensoriel des deux matrices $(C_1)_{\mathcal{T}}, (C_2)_{\mathcal{T}}$, qui est positivement défini parce que ses valeurs propres sont les produits des valeurs propres des ces deux matrices. Alternativement, on associe des champs gaussiens \mathbf{X}, \mathbf{Y} aux deux noyaux, et on note que $\mathbf{w}^t[(C_1)_{\mathcal{T}} \cdot (C_2)_{\mathcal{T}}]\mathbf{w}$, où \cdot denote le produit "Schur/Hadamard", coïncide avec l'espérance $\mathbb{E} \sum_i (w_i Y_i)^2$, étant donc positive pour chaque vecteur \mathbf{w} . Les noyaux les plus usuelles sont générés par une formule de type $C(x, y) = \varphi(|x - y|)$, où $\varphi(r)$ est une fonction d'une variable tq le noyau est positif défini. Ils correspondent aux champs gaussiens isotropiques, et comme les deux concepts sont "isomorphes", on peut se poser des questions sur l'utilité des noyaux. Mais, ce concept s'avère utile; par exemple, on aura besoin d'utiliser aussi des "noyaux de covariance généralisé", qui sont seulement positives définies sur certaines sous-espaces, à être définis. On reconnaît ces noyaux dans le tableau des fonctions radiales de base pour avoir un "CPD(m)" strictement positif : par exemple, les noyaux engendrés par les fonctions linéaire, quadratique et cubique.

12.10 Exercices

1. Montrez que la fonction $C(x, y) = f(x) f(y)$ est un noyau de covariance.
2. Montrez que la fonction $C(x, y) = e^{-|x-y|^2}$ est un noyau de covariance.
3. Montrez que la fonction $C(x, y) = h(x+y) - h(x-y)$, où $h(x)$ est une fonction positive, avec minimum en 0, est un noyau de covariance.
4. Calculez la corrélation sphérique standardisée (qui représente la mesure de l'intersection des deux sphères de rayon 1/2, situées à distance $r \leq 1$ une de l'autre), en dimensions 1 et 2.
5. Filtrage par krigeage. Soit le modèle de régression habituel où les y_i sont typiquement des mesures prises chacune en un point x_i du plan, c.-à-d. un champ spatial (intensité lumineuse en un point d'une image, mesure de pollution en un endroit, etc.). On se propose d'exprimer la corrélation comme une fonction paramétrisée des localisations, par exemple :

$$Y_i = m_i \beta + u_i, i = 1, \dots, I$$

$$E[u_i u_j] = \sigma^2 \delta_{ij} + c \exp(-b \|x_i - x_j\|^a)$$

Chapitre 13

Exemples de krigeage

Étant donné une variable spatiale $g(x)$ à incréments stationnaires et à **moyenne fixe inconnue** m , et quelques valeurs observées

$$Y_i = Y(x_i) + \epsilon_i, \quad i = 1, \dots, I$$

le krigeage consiste dans l'estimation de $g(x)$, $\forall x$. Quand on s'intéresse à une position particulière x , on dénotera la variable correspondante par

$$Y = g(x).$$

Nous présentons d'abord pour échauffement quelques cas particuliers de krigeage/régression spatiale.

13.1 Le krigeage simple

Le cas le plus simple est celui des champs avec tendance connue, appelée krigeage simple, qui est une application directe de la formule de régression (12.2) :

$$\mathbf{w} \mathbf{C}_{XX} = \mathbf{C}_{YX} \implies \mathbf{w} = \mathbf{w}_s = \mathbf{C}_{YX} \mathbf{C}_{XX}^{-1}$$

(on a donné ici le cas sans tendance, c.-à-d. avec $m(x) = 0$. Rappelons aussi le **”système de krigeage simple dual”** (12.4) :

$$\mathbf{C}_{XX} \boldsymbol{\alpha} = \mathbf{X}$$

(associé à une interpolation exacte).

Exemple 13.1.1 *Ecrivez le système de krigeage simple pour $n + 1$ points équidistants $x_i = i h, i = 0, \dots, n$ en \mathbb{R}^1 , avec $n = 3$, et avec distance entre les points de $h = .3$, en supposant une covariance sphérique de portée $R = 1/2$, $\tau^2 = 1/4$, $\sigma^2 = 1$.*

13.2 Krigeage de la moyenne

Considerons l'estimation d'une moyenne constante. Une première possibilité serait de l'estimer par :

$$\hat{m} = \frac{\sum_{i=1}^N Y_i}{N}$$

Mais, avec covariances nonnulles, une approche, plus raisonnable serait d'utiliser une moyenne pondérée :

$$\hat{m} = \sum_{i=1}^N w_i Y_i = \mathbf{w} \mathbf{X}$$

Afin que l'estimateur \hat{m} soit sans biais, on doit demander que

$$\mathbb{E}\hat{m} = m \iff \sum_{i=1}^N w_i = 1$$

Pour déterminer finalement les poids w_i , on demandera que l'estimateur \hat{m} a une variance minimale, ou plutôt le deuxième moment minimal (comme $\text{Var } \hat{m} = E[\hat{m}^2] - m^2$). On arrive ainsi à un problème d'optimisation quadratique contrainte :

$$\begin{aligned} \min_{\mathbf{w}} E[\hat{m}^2] &= \min_{\mathbf{w}} \mathbf{w} \mathbf{C} \mathbf{w}^t & (13.1) \\ \sum_{i=1}^N w_i &= 1 \end{aligned}$$

Par la méthode de Lagrange, on converti le problème au problème d'optimisation quadratique sans contraintes, en dénotant le multiplicateur de Lagrange par $2d$:

$$\min \mathbf{w} \mathbf{C} \mathbf{w}^t - 2d(\mathbf{w} \mathbf{1} - 1)$$

Après différentiation par rapport aux poids \mathbf{w} , ça rend le système :

$$\begin{aligned} \mathbf{C} \mathbf{w} - d \mathbf{1} &= 0 \\ \mathbf{w} \mathbf{1} &= 1 \end{aligned} \quad (13.2)$$

En résolvant,

$$\mathbf{w} = d \mathbf{1}^t \mathbf{C}^{-1} \quad \text{où } d = (\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}$$

et

$$\hat{m} = \mathbf{w} \mathbf{X} = d \mathbf{1}^t \mathbf{C}^{-1} \mathbf{X} = \frac{(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{X})}{(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})} = \frac{\langle \mathbf{1}, \mathbf{X} \rangle_{\mathbf{C}^{-1}}}{\langle \mathbf{1}, \mathbf{1} \rangle_{\mathbf{C}^{-1}}} = \frac{\mathbf{v}^t \mathbf{X}}{\mathbf{v}^t \mathbf{1}}, \quad \mathbf{v}^t = \mathbf{1}^t \mathbf{C}^{-1} \quad (13.3)$$

où le produit scalaire est défini par rapport à la matrice \mathbf{C}^{-1} .

Remarque 13.2.1 1) La variance de cet estimateur est : $\text{Var } \hat{m} = \mathbf{w} \mathbf{C} \mathbf{w}^t = d \mathbf{w} \mathbf{1} = d$, et donc le paramètre de Lagrange d peut être interprété comme la variance de l'estimateur de la moyenne. 2) La moyenne pondérée (13.3) devient une moyenne arithmétique simple ssi \mathbf{v} est proportionnel à $\mathbf{1}$, c.-à-d. si les sommes de lignes de la matrice \mathbf{C} sont égales. Cela arrive par exemple dans le cas des deux observations, et aussi dans le cas des covariances croisées nulles, c.-à-d. quand \mathbf{C} est la matrice identité multipliée par $\sigma^2 + \tau^2$.

13.3 Le krigeage ordinaire

On considère maintenant la prédiction de la variable spatiale $g(x)$ à increments stationnaires, de moyenne **fixe et inconnue** m . On utilisera un estimateur :

$$\hat{Y}(x) = \sum_{i=1}^N w_i Y_i = \mathbf{w}\mathbf{X}$$

où les poids a_i (qui dépendront de x), incluent déjà l'estimation de la moyenne. Pour que l'estimateur soit sans biais, il faut encore que la somme des poids soit égale à 1. On minimise en suite la variance d'estimation :

$$\sigma_E^2 = \mathbb{E}(g(x) - \hat{Y}(x))^2 = \mathbb{E}(Y - \mathbf{w}\mathbf{X})^2 = \sigma^2 + \mathbf{w}\mathbf{C}\mathbf{w}^t - 2\mathbf{w}\mathbf{C}_{X,Y} \quad (13.4)$$

sous cette contrainte, où $\mathbf{C}_{X,Y}$ est un vecteur colonne. Par la méthode de Lagrange, on minimisera :

$$\sigma_E^2 = \sigma^2 + \mathbf{w}\mathbf{C}\mathbf{w}^t - 2\mathbf{w}\mathbf{C}_{X,Y} - 2\mu(\mathbf{w}\mathbf{1} - 1)$$

En annulant les dérivées de cette quantité par rapport aux poids, on arrive au système

$$\mathbf{C}\mathbf{w}^t - \mu\mathbf{1} = \mathbf{C}_{X,Y} \quad (13.5)$$

En forme matricielle, on a le "système normal étendu"

$$\begin{pmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^t & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w}^t \\ -\mu \end{pmatrix} = \tilde{\mathbf{C}} \begin{pmatrix} \mathbf{w}^t \\ -\mu \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{X,Y} \\ 1 \end{pmatrix}$$

En résolvant directement (exercice! §)

$$\begin{pmatrix} \mathbf{w}^t \\ -\mu \end{pmatrix} = \tilde{\mathbf{C}}^{-1} \begin{pmatrix} \mathbf{C}_{X,Y} \\ 1 \end{pmatrix}$$

où en utilisant la forme de la matrice $\tilde{\mathbf{C}}^{-1}$, comme donnée par le Corollaire 13.4.1 :

$$\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \mathbf{C}^{-1} - d\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1} & d\mathbf{C}^{-1}\mathbf{1} \\ d\mathbf{1}^t\mathbf{C}^{-1} & -d \end{pmatrix}$$

avec $d = (\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}$ (qui coïncide avec le multiplicateur de Lagrange dans le krigeage de la moyenne), on trouve :

$$\begin{aligned} \mathbf{w}^t &= (\mathbf{C}^{-1} - d\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1})\mathbf{C}_{XY} + d\mathbf{C}^{-1}\mathbf{1} = \mathbf{C}^{-1}\mathbf{C}_{XY} + d\mathbf{C}^{-1}\mathbf{1}(1 - \mathbf{1}^t\mathbf{C}^{-1}\mathbf{C}_{XY}) \\ \hat{Y}(x) &= (\mathbf{C}_{YX}\mathbf{C}^{-1} - d\mathbf{C}_{YX}\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1} + d\mathbf{1}^t\mathbf{C}^{-1})\mathbf{X} \\ &= \mathbf{w}_s\mathbf{X} - \hat{m}\mathbf{w}_s\mathbf{1} + \hat{m} = \mathbf{w}_s\mathbf{X} + \hat{m}(1 - \mathbf{w}_s\mathbf{1}), \end{aligned}$$

où \hat{m} est la moyenne inconnue estimée par (13.3).

Remarque 13.3.1 Comme les poids \mathbf{w}_s du krigeage simple ont typiquement une somme différente de 1, le deuxième terme ci-dessus fourni une correction du krigeage simple !

§. Il suffit de calculer μ , en vérifiant que $\mu = d(1 - \mathbf{1}^t\mathbf{C}^{-1}\mathbf{C}_{XY})$

Alternativement, après que la moyenne \hat{m} a été estimée par (13.3), on peut voir ça comme la formule classique de régression avec une moyenne connue

$$\hat{Y}(x) - \hat{m} = (\mathbf{X}^t - \hat{m}\mathbf{1}^t)\mathbf{C}^{-1}\mathbf{C}_{XY} = \mathbf{w}(\mathbf{X} - \hat{m}\mathbf{1})$$

Exercice 13.3.1 Comparaison des krigeages simple et ordinaire. Supposons Y, X_1, X_2 sont des variable aléatoire à espérance 0 et variance 1, tq $\text{Cov}(X_2, X_1) = \rho$, $\text{Cov}(Y, X_1) = z$ et $\text{Cov}(Y, X_2) = 0$. a) Si $z = \rho$, quel est le domaine de valeurs permmissibles pour φ ? b) Trouvez le meilleur predicteur linéaire de Y en fonction de X_1, X_2 , et la variance de l'erreur de prediction. Est-ce que X_2 influence la prediction de Y ? c) Supposons cette fois que les variables ont une moyenne inconnue, constante. Estimez d'abord la moyenne, en utilisant la formule

$$\hat{m} = \frac{(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{X})}{(\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})} = \frac{\mathbf{v}^t\mathbf{X}}{\mathbf{v}^t\mathbf{1}}, \quad \mathbf{v}^t = \mathbf{1}^t\mathbf{C}^{-1}$$

de la section 13.2. En suite, trouvez le meilleur predicteur linéaire de Y en fonction de X_1, X_2 . d) Quelles sont les domaines de valeurs permmissibles pour φ, z , en general? **Sol** a) $\mathbf{S} : |\rho| \leq 2^{-1/2}$ c) $\hat{Y} = \frac{z}{1-\varphi^2}(X_1 - \rho X_2)$ d) $m = \frac{X_1+X_2}{2}, \hat{Y} = m - \frac{\rho}{1-\rho}X_2 = \frac{1}{2(1-\rho)}(X_1 + (1-2\rho)X_2)$

Exercice 13.3.2 Obtenez, avec le logiciel de votre choix, les résultats de l'exemple sur les pages 3-6 <http://www2.imm.dtu.dk/aa/krexample.pdf> concernant un champs avec une semivariance théorique $\gamma(h) = C_0 + C_1(1 - \rho(h))$ où

$$1 - \rho(h) = \frac{3}{2}\frac{h}{R} - \frac{1}{2}\left(\frac{h}{R}\right)^3, h \in (0, R).$$

Indication : Calculer le noyau de covariance en utilisant $\varphi(r) = \gamma(\infty) - \gamma(r) = C_0I_{\{r=0\}} + C_1\rho(r)$.

Exercice 13.3.3 Montrez que la variance d'estimation du krigeage ordinaire admet une décomposition similaire comme la somme de la variance d'estimation du krigeage simple et celle due à la méconnaissance de la moyenne. (Ainsi, la variance d'estimation est supérieure à celle obtenue dans le cas du krigeage simple où la moyenne est supposée connue).

Sol : La variance du krigeage est

$$\sigma_E^2 = \sigma^2 + \mathbf{w}^t\mathbf{C}\mathbf{w} - 2\mathbf{w}^t\mathbf{C}_{X,Y} = \sigma^2 - \mathbf{w}^t\mathbf{C}_{X,Y} + \mu\mathbf{w}^t\mathbf{1} = \sigma^2 - \mathbf{w}^t\mathbf{C}_{X,Y} + \mu$$

En conclusion, le problème du krigeage ordinaire est en effet "la somme des deux problèmes", c.-à-d. la solution est la somme de deux termes, un correspondant à un krigeage simple à moyenne 0, et l'autre correspondant à un krigeage d'une moyenne constante m .

13.4 Resultat auxiliaire : l'inverse d'une matrice partitionnée en blocs

Dans la regression multivarié à moyennes 0/"krigeage simple", le rôle principal dans la prévision est joué par la matrice des covariances $\mathbf{C} = (\text{Cov}(Y_i, Y(x_j)) : i, j = 1, \dots, I)$. Dans

la regression multivarié à moyennes paramétriques/ "krigeage universel", ce rôle sera pris par la matrice $\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}$ Il sera donc utile d'avoir une formule pour l'inverse d'une matrice partitionnée de ce genre.

Lemme 13.4.1 Soit $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{D} & \mathbf{C} \end{pmatrix}$ une matrice, et supposons que \mathbf{A} , et la matrice $\mathbf{S}_C := \mathbf{C} - \mathbf{D}\mathbf{A}^{-1}\mathbf{B}$, appelé en algèbre **complement de Shur** de \mathbf{C} et en probabilité variance conditionnelle, sont inversibles. Supposons aussi que \mathbf{C} , et la matrice $\mathbf{S}_A := \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{D}$ sont inversibles. Alors,

1.

$$\mathbf{S}_A \mathbf{A}^{-1} \mathbf{B} = \mathbf{B} \mathbf{C}^{-1} \mathbf{S}_C, \quad \mathbf{S}_C \mathbf{C}^{-1} \mathbf{D} = \mathbf{D} \mathbf{A}^{-1} \mathbf{S}_A$$

2.

$$\boxed{\mathbf{S}_A^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} (\mathbf{S}_C^{-1} \mathbf{D} \mathbf{A}^{-1})}$$

3.

$$\det(\mathbf{M}) = \det(\mathbf{A}) \det(\mathbf{S}_C) = \det(\mathbf{C}) \det(\mathbf{S}_A)$$

4. \mathbf{M} est inversible et son inverse est :

$$\boxed{\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{S}_A^{-1} & -\mathbf{A}^{-1} \mathbf{B} \mathbf{S}_C^{-1} \\ -\mathbf{C}^{-1} \mathbf{D} \mathbf{S}_A^{-1} & \mathbf{S}_C^{-1} \end{pmatrix}}$$

Note : Pour faciliter la verification, ces formules sont écrites tel que les seuls produits des matrices intervenant sont dans l'ordre "montant cyclique" des lettres A, B, C, D, A, \dots

Démonstration : 1. est immédiat, et 2. suit en remplaçant $\mathbf{S}_C^{-1} \mathbf{D} \mathbf{A}^{-1}$ par $\mathbf{C}^{-1} \mathbf{D} \mathbf{S}_A^{-1}$. 4.

Paramétrisant la matrice inverse par $\mathbf{M}^{-1} = \begin{pmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{D}} & \tilde{\mathbf{C}} \end{pmatrix}$, et en resolvant les inconnues dans l'équation $\mathbf{M} \mathbf{M}^{-1} = \mathbf{I}$, on trouve d'abord $\tilde{\mathbf{B}} = -\mathbf{A}^{-1} \mathbf{B} \tilde{\mathbf{C}}$, $\tilde{\mathbf{D}} = -\mathbf{C}^{-1} \mathbf{D} \tilde{\mathbf{A}}$, et ensuite $\tilde{\mathbf{A}} = \mathbf{S}_A^{-1}$, $\tilde{\mathbf{C}} = \mathbf{S}_C^{-1}$. En utilisant $\mathbf{M}^{-1} \mathbf{M} = \mathbf{I}$ on arrive a des formules alternatives, qui sont vrais aussi par les identités (1). **Note :** Si on sait seulement que \mathbf{A} et \mathbf{S}_C (ou que \mathbf{C} et \mathbf{S}_A) sont inversibles, on peut encore, en utilisant 2., écrire des formules asymetriques :

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} \mathbf{S}_C^{-1} \mathbf{D} \mathbf{A}^{-1} & -\mathbf{A}^{-1} \mathbf{B} \mathbf{S}_C^{-1} \\ -\mathbf{S}_C^{-1} \mathbf{D} \mathbf{A}^{-1} & \mathbf{S}_C^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{E} \mathbf{S}_C^{-1} \mathbf{F} & -\mathbf{E} \mathbf{S}_C^{-1} \\ -\mathbf{S}_C^{-1} \mathbf{F} & \mathbf{S}_C^{-1} \end{pmatrix}$$

où $\mathbf{E} = \mathbf{A}^{-1} \mathbf{B}$, $\mathbf{F} = \mathbf{D} \mathbf{A}^{-1}$.

Corollaire 13.4.1 Supposons que \mathbf{C} et \mathbf{P} sont de rang maximal. Alors, la matrice $\tilde{\mathbf{C}} = \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}$ est inversible et :

$$\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} & \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \\ \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} & -\mathbf{D} \end{pmatrix} \quad (13.6)$$

où $\mathbf{D} = (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1}$.

Exemple : Pour le krigeage ordinaire, $\mathbf{P} = \mathbf{1}$, et

$$\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \mathbf{C}^{-1} - d \mathbf{C}^{-1} \mathbf{1} \mathbf{1}^t \mathbf{C}^{-1} & d \mathbf{C}^{-1} \mathbf{1} \\ d \mathbf{1}^t \mathbf{C}^{-1} & -d \end{pmatrix} \quad (13.7)$$

où $d = (\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})^{-1}$.

13.5 Exercices

1. Le champs déterministe $y(x) = x^2$ sera "estimé" à partir des n observations en $x_1 = 0, x_2 = 1/(n-1), \dots, x_i = (i-1)/(n-1), \dots, x_n = 1$, et en supposant (à tort) un noyau de corrélation cubique $\rho(x, y) = (1 - \frac{|x-y|}{R})_+$.

Part 0 Donnez la relation entre la corrélation spatiale $\rho(r)$ d'un champs stationnaire et son variogramme $\gamma(r)$. Montrez que le variogramme théorique sphérique est donné par la formule :

$$\gamma(r) = \tau^2 + \sigma^2(1 - (1 - r/R)_+) = \tau^2 + \sigma^2 \frac{r}{R}, \quad r \leq R. \quad (13.8)$$

Part 1 Pour $n = 3$, calculez les semivariances empiriques $\gamma(r_0), \gamma(r_1)$ et $\gamma(r_2)$ (c.-à-d. le variogramme empirique) pour $r_0 = |x-y| = 0, r_1 = |x-y| = 1/2$ et $r_2 = |x-y| = 1$, et les covariances empiriques $\varphi(r_0), \varphi(r_1)$ et $\varphi(r_2)$. En suite, donnez : la matrice $\mathbf{\Gamma}$ des semivariances empiriques des observations, la matrice \mathbf{C} des corrélations empiriques entre les observations, et préciser la relation entre les deux.

Part 2 Formulez et résolvez le problème de choisir le palier partiel σ^2 et la pepite τ^2 tq le variogramme théorique sphérique $\gamma(r) = \tau^2 + \sigma^2 r$ donne la meilleure approximation du variogramme empirique au sens des moindres carrés, dans le cas général d'un variogramme contenant n distances observées r_1, \dots, r_n . Quel est le résultat dans le cas particulier ci-dessus ? Repetez ce problème, a) en supposant $\tau = 0$. b) En supposant un variogramme théorique parabolique $\gamma(r) = \tau + \sigma r + \sigma_2 r^2$.

- (a) Calculez le vecteur \mathbf{c} des corrélations théoriques entre les observations, et une position non observé x .
- (b) Calculez la prédiction $\hat{y}(x)$ obtenue par le krigeage simple. Quelle est la valeur obtenue pour $x = 2/3$?
- (c) Estimer la moyenne du champs, à partir de la formule

$$m = \frac{\mathbf{v}^t \mathbf{X}}{\mathbf{v}^t \mathbf{1}}$$

où $\mathbf{v}^t = \mathbf{1}^t \mathbf{C}^{-1}$. Comparez avec la moyenne arithmétique.

- (d) Calculez les valeurs du champs demoyenné $Z(x) = g(x) - m$, dans les points d'observation.
- (e) Calculez la prédiction $\hat{y}(x)$ obtenue par le krigeage ordinaire, c.-à-d. calculez des approximations de la forme :

$$\hat{y}(x) = \alpha_1(1 - |x|)_+ + \alpha_2(1 - |x - 1/2|)_+ + \alpha_3(1 - |x - 1|)_+ + m$$

(Pratiquement, on applique un krigeage simple au champs demoyenné par la moyenne m obtenu ci-dessus). Quelles seront les résultats obtenus en supposant $m(x) = \beta_0 + \beta_1 x$ et $m(x) = \beta_0 + \beta_1 x + \beta_2 x^2$?

Solutions :

1. (a) Le variogramme satisfait $\gamma(r) = \tau^2 + \sigma^2(1 - \rho(r)) = \tau^2 + \sigma^2 \min[r, 1]$.

(b)

$$\gamma(0) = 0, \gamma(1/2) = (1/16 + 9/16)/4 = 5/32, \gamma(1) = 1/2$$

$$\varphi(0) = \frac{26}{144}, \varphi(1/2) = \frac{-2}{144}, \varphi(1) = \frac{-35}{144}$$

(c) La régression linéaire $\gamma(r) \approx \tau^2 + \sigma^2 r$ donne, en utilisant les formules :

$$\sigma^2 = \frac{\tilde{S}_{XY}}{\tilde{S}_{XX}}, \tau^2 = \bar{\gamma} - \sigma^2 \bar{r}$$

où $\bar{r} = 1/2, \bar{\gamma} = 7/32$, donne $\tau^2 = -1/32, \sigma^2 = 1/2$. Mais comme τ^2 doit être positive, nous imposons la contrainte $\tau^2 = 0$ et finalement

$$\sigma^2 = \frac{S_{XY}}{S_{XX}} = 37/80, \quad \varphi(r) = \sigma^2(1 - r).$$

Note : Comme $\tau^2 = 0$, on pourra faire le krigeage directement avec la matrice des corrélations (c.-à-d. comme σ^2 se simplifie, on peut le prendre comme 1).

(d)

$$C = \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a & 0 \\ a & 1 & a \\ 0 & a & 1 \end{pmatrix}$$

avec $a = 1/2$.

$$C^{-1} = \begin{pmatrix} \frac{1-a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{a^2}{1-2a^2} \\ -\frac{a}{1-2a^2} & \frac{1}{1-2a^2} & -\frac{a}{1-2a^2} \\ \frac{a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{1-a^2}{1-2a^2} \end{pmatrix} = \begin{pmatrix} 3/2 & -1 & 1/2 \\ -1 & 2 & -1 \\ 1/2 & -1 & 3/2 \end{pmatrix}$$

(e) Soit $\alpha = \mathbf{X}C^{-1} = (1/4, -1/2, 5/4)$. Par le krigeage simple, $\hat{y}(x) = \langle \alpha, \mathbf{C}_{XY} \rangle = \frac{1}{4}(1-x) - \frac{1}{2}(1-|x-1/2|) + \frac{5}{4}x = \frac{x}{2} + (x - \frac{1}{2})_+$, la fonction linéaire par morceaux interpolant les points. En particulier, pour la prédiction en $x = 2/3$ on trouve $\hat{y}(2/3) = \frac{1}{2}$.

(f) $\mathbf{v}^t = \mathbf{1}^t C^{-1} = (1, 0, 1)$, $\frac{\mathbf{v}^t}{\mathbf{v}^t \mathbf{1}} = (1/2, 1/3, 1/2)$ et la moyenne du champ est

$$m = \frac{\mathbf{v}^t \mathbf{X}}{\mathbf{v}^t \mathbf{1}} = 1/2$$

pendant que la moyenne arithmétique est $5/12$.

(g) le champ demoyonné est $z = (-1/2, -1/4, 1/2)$.

(h) Soit $\alpha = \mathbf{z}C^{-1} = (-1/4, -1/2, 3/4)$. Par le krigeage ordinaire, on trouve $\hat{y}(x) = m + \langle \alpha, \mathbf{C}_{XY} \rangle = \frac{1}{2} - \frac{1}{4}(1-x) - \frac{1}{2}(1-|x-1/2|) + \frac{3}{4}x = \frac{x}{2} + (x - \frac{1}{2})_+$, la fonction linéaire par morceaux interpolant les points! On obtient encore le même résultat en supposant $m(x) = \beta_0 + \beta_1 x$, mais avec $m(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ on obtiendra la parabole qui interpole les points.

(i) On a encore

$$C = \begin{pmatrix} 1 & a & 0 \\ a & 1 & a \\ 0 & a & 1 \end{pmatrix}$$

avec $a = 7/8$ et

$$C^{-1} = \begin{pmatrix} \frac{1-a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{a^2}{1-2a^2} \\ -\frac{a}{1-2a^2} & \frac{1}{1-2a^2} & -\frac{a}{1-2a^2} \\ \frac{a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{1-a^2}{1-2a^2} \end{pmatrix} = \frac{1}{34} \begin{pmatrix} -15 & 56 & -49 \\ 56 & -64 & 56 \\ -49 & 56 & -15 \end{pmatrix}$$

Soit $\beta = (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1} (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{X}) = \begin{pmatrix} 16 & 8 \\ 17 & 17 \\ 17 & 34 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 17 \\ 19 \\ 34 \end{pmatrix} = \begin{pmatrix} -\frac{3}{8} \\ 1 \\ 1 \end{pmatrix}$. La moyenne du champ est

$$m(x) = -\frac{3}{8} + x$$

Le champ demoyenné est $z = (3/8, 1/8, 3/8)$. Soit $\alpha = \mathbf{z} \mathbf{C}^{-1} = (-1/2, 1, -1/2)$. Par le krigeage ordinaire, on trouve $\hat{y}(x) = m + \langle \alpha, \mathbf{C}_{XY} \rangle = -\frac{3}{8} + x - \frac{1}{2}(1 - x^3) + (1 - |x - 1/2|^3) - \frac{1}{2}(1 - (1 - x)^3) = \begin{cases} y/4 + y^3, & \text{si } x \leq 1/2 \\ y/4 + y^3 - 2(y - 1/2)^3 & \text{si } x \geq 1/2 \end{cases}$

On a obtenu la spline cubique qui interpole les points! Les valeurs communes des dérivés en $1/2$ sont $f'(1/2) = 1, f''(1/2) = 3$.

L'interpolation/ajustement des semivariances empiriques peut se faire par plusieurs méthodes, d'interpolation, c.-à-d. d'approximation d'un "nuage bruité" des points $(x_i, y_i) \in \mathbb{R}^2, i = 1, \dots, n$, par une "fonction simple".

Chapitre 14

Interpolation optimale/régression à tendance paramétrique/krigeage avec dérive

14.1 L'interpolation par noyaux de base à reproduction polynomiale

Soit \mathcal{P} un espace des fonctions de dimension finie M , avec base $p_1(x), \dots, p_M(x)$, qu'on espère une bonne approximation des données. Cet espace pourrait a priori contenir des fonctions périodiques, exponentielles, etc, mais pour l'instant on se contentera du cas le plus simple, quand \mathcal{P} est l'espace de tous les polynômes de degré plus petit que $M - 1$.

Définition 14.1.1 On appellera interpolation par des noyaux de base, à reproduction polynomiale d'ordre M , une interpolation de la forme :

$$\hat{Y}(x) := \hat{Z}(x) + m(x) = \sum_{i=1}^I \alpha_i \varphi(x, x_i) + \sum_{j=1}^M \beta_j p_j(x) \quad (14.1)$$

avec des coefficients α_i qui sont "orthogonaux" par rapport à l'espace de tous les polynômes de degré plus petit que $M - 1$, dans le sens que

$$\sum_{i=1}^I \alpha_i p_j(x_i) = 0, \quad \forall j = 1, \dots, M \quad (14.2)$$

Le cas où le noyau $\varphi(x, y) = \rho(|x - y|)$ est une fonction de la distance $|x - y|$ seulement est appelée interpolation par une fonction radiale de base (RBF).

Définition 14.1.2 Un vecteur $\alpha = (\alpha_1, \dots, \alpha_I)$ satisfaisant les équations(14.2) pour un ensemble des fonctions $p_j(x), j = 1, \dots, M$ s'appelle orthogonale par rapport à l'espace des fonctions \mathcal{P} engendré par $p_j(x), j = 1, \dots, M$.

Notes :

1. L'interpolation "pure" par des fonctions de base radiales, sans une partie polynomiale, est à ordre 0. En géostatistique, elle s'appelle **krigeage simple**.
2. L'interpolation incluant une constante est à ordre 1; dans ce cas (14.2) demande $\sum_i \alpha_i = 0$. En géostatistique, elle s'appelle **krigeage ordinaire**.
3. pour ordre 2, dans \mathbb{R}^1 , (14.2) demande $\sum_i \alpha_i = 0, \sum_i \alpha_i x_i = 0$.
4. La condition d'orthogonalité (14.2) a des justifications différentes dans la modélisation stochastique—voir chapitre 13 et dans la modélisation déterministe —voir chapitre ??, basé sur la théorie des distributions. On peut aussi vérifier qu'elle assure que l'interpolation des fonctions $p_j(x), x = 1, \dots, M$ soit exacte.
5. Parfois, d'autres fonctions sont utilisées pour décomposer la tendance, au lieu des polynômes. Ces fonctions devront donner une bonne approximation de la tendance/dérive globale $m(x)$.
6. Les premières équations du système dual de krigage/interpolation

$$\boxed{\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix}}$$

expriment dans le cas $\tau = 0$ une interpolation exacte des données.

7. La combinaison des **fonctions radiales de base** $\varphi(|x - x_i|)$ (où $\varphi(d)$ est en statistique la fonction de covariance à distance d), essaie de capturer, à travers les coefficients α_i , l'influence du point x_i sur un point x dans le reste du domaine.
8. La formule d'interpolation (14.1) fournit un cadre général pour une grande classe de méthodes d'interpolation. Notons qu'elle contient :
 - (a) des fonctions qui ne dépendent pas des données (c.-à-d. points d'observation), comme les polynômes (pour les statisticiens, leur somme représente la tendance).
 - (b) des fonctions qui dépendent des données : les translatées $\varphi(|x - x_i|)$ de la fonction $\rho(r)$, centrées dans les points observés x_i (pour les statisticiens, $\varphi(r)$ représente la covariance d'un processus isotropique).

ee Mais, la formule d'interpolation (14.1) laisse à l'utilisateur le dilemme de choisir une "bonne base" $\varphi(r)$ et une "bonne représentation" de la tendance.

14.2 Le système d'interpolation/krigeage dual

Dans ce chapitre, nous montrons que l'estimation des champs Gaussiennes avec tendance paramétrique, appelés aussi krigage avec dérive, ramène précisément à la formule d'interpolation (14.1). On déterminera l'estimateur "BLUE (Best linéar unbiased)" de la forme

$$\hat{Y}(x) = \sum_{i=1}^I w_i Y_i, \forall x$$

pour le modèle additif à dérive paramétrique (11.9), (11.8) (les poids w_i dépendront de x et incluront déjà l'estimation de la moyenne $m(x)$).

Théorème 14.2.1 Soit $g(x)$ un processus Gaussien, avec espérance $m(x)$ appartenant à un espace de dimension finie, engendré par les fonctions $p_1(x) = 1, p_2(x), \dots, p_M(x)$, c.-à-d.

$$m(x) = \sum_{j=1}^M \beta_j p_j(x),$$

et soit $\tilde{C}(x, y)$ sa covariance (par exemple, pour un processus isotropique, la covariance aura une forme radiale : $\tilde{C}(x, y) = \sigma^2 \varphi(|x - y|/R)$, où $\rho(r)$ est un noyau "admissible" (conditionnellement positif). Supposons que le processus est observé avec erreurs ϵ_i , i.i.d. de variance τ^2 , et soit

$$Y_i = Y(x_i) + \epsilon_i, i = 1, \dots, I$$

les observations. Soit

$$\hat{Y}(x) = \sum_{i=1}^I w_i Y_i = \mathbf{X}^t \mathbf{w} \quad (14.3)$$

un estimateur linéaire, où $\mathbf{X} = (U_1, \dots, U_I)$ et $\mathbf{w} = (w_i, i = 1, \dots, I)$ dénotent des vecteurs (colonnes) contenant les I observations/coefficients "de pondération". Alors, en supposant que $\mathbf{C} = (\text{Cov}(Y_i, Y(x_j) + \tau^2 \delta(i - j)) : i, j = 1, \dots, I)$ et \mathbf{P} sont de rang maximal, et donc $\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P}$ est inversible, il suit que :

(a) Il existe un estimateur unique (14.3) "BLUE" qui est nonbiaisé et à variance minimale, donnée par :

$$\begin{aligned} \hat{Y}(x) &= \mathbf{X}^t \mathbf{w} = \boxed{(\mathbf{X}^t \quad \mathbf{0}^t) \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix}} = (\boldsymbol{\alpha} \quad \boldsymbol{\beta})^t \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix} \quad (14.4) \\ &= \sum_{i=1}^I \alpha_i \mathbf{C}_{X,Y}(i) + \sum_{j=1}^M \beta_j p_j(x) \end{aligned}$$

où $\mathbf{C}_{X,Y}$ est un vecteur colonne défini par $(\text{Cov}(Y_i, g(x)) : i = 1, \dots, I)$, $\mathbf{p}(x)$ est un vecteur colonne défini par $(p_j(x) : j = 1, \dots, M)$ et $\boldsymbol{\alpha}, \boldsymbol{\beta}$ sont des vecteurs colonne vérifiant le système "de krigeage dual"/interpolation.

$$\boxed{\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix}} \quad (14.5)$$

(où dans le cas isotropique, $\mathbf{C}_{i,j} = \sigma^2 \varphi(|x_i - x_j|/R) + \tau^2 \delta(i, j)$).

(b) En posant $\mathbf{D} = (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1}$, les coefficients d'interpolation α_i, β_j sont donnés explicitement par :

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} &= \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} & \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \\ \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} & -\mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1}) \mathbf{X} \\ \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} \mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{-1} (\mathbf{X} - \mathbf{P} \boldsymbol{\beta}) \\ \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} \mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{-1} (\mathbf{X} - \hat{\mathbf{m}}) \\ (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1} \mathbf{P}^t \mathbf{C}^{-1} \mathbf{X} \end{pmatrix} \end{aligned}$$

où le vecteur $\hat{\mathbf{m}} = \mathbf{P} \boldsymbol{\beta}$ représente les moyennes dans les points observés.

- (c) Les coefficients "de pondération" $\mathbf{w} = (w_i, i = 1, \dots, I)$ du prédicteur BLUE satisfont le "système normal étendu" obtenu en groupant les deux derniers termes de (14.4) :

$$\boxed{\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ -\boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix}} \quad (14.6)$$

Explicitement,

$$\begin{pmatrix} \mathbf{w} \\ -\boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{P}\mathbf{D}\mathbf{P}^t\mathbf{C}^{-1} & -\mathbf{C}^{-1}\mathbf{P}\mathbf{D} \\ -\mathbf{D}\mathbf{P}^t\mathbf{C}^{-1} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix}$$

où $\mathbf{D}^{-1} = -\mathbf{P}^t\mathbf{C}^{-1}\mathbf{P}$. Finalement,

$$\boxed{\mathbf{w} = (\mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{P}\mathbf{D}\mathbf{P}^t\mathbf{C}^{-1})\mathbf{C}_{X,Y} - (\mathbf{C}^{-1}\mathbf{P}\mathbf{D})\mathbf{p}(x)}$$

La variance de l'estimateur BLUE est :

$$\sigma^2 - \mathbf{C}_{X,Y}^t(\mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{P}\mathbf{D}\mathbf{P}^t\mathbf{C}^{-1})\mathbf{C}_{X,Y} - \mathbf{p}(x)^t\mathbf{D}\mathbf{p}(x) \quad (14.7)$$

- (d) Finalement, l'estimation est

$$\hat{Y}(x) = [(\mathbf{X}^t - \mathbf{m}^t)\mathbf{C}^{-1}\mathbf{C}_{X,Y} + \boldsymbol{\beta}^t\mathbf{p}(x)] = (\mathbf{X}^t - \mathbf{m}^t)\mathbf{C}^{-1}\mathbf{C}_{X,Y} + m(x) \quad (14.8)$$

où $m(x) = \mathbf{p}(x)^t\boldsymbol{\beta} = \mathbf{p}(x)^t(\mathbf{P}^t\mathbf{C}^{-1}\mathbf{P})^{-1}\mathbf{P}^t\mathbf{C}^{-1}\mathbf{X}$ représente l'estimation de la moyenne en x . Par rapport au processus avec la moyenne enlevé $Z(x) = g(x) - m(x)$, la formule devient simplement la formule de régression à bruit corrélé de moyenne 0 :

$$\boxed{\hat{Z}(x) = \mathbf{z}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} = \langle \mathbf{z}, \mathbf{C}_{X,Y} \rangle_{\mathbf{C}^{-1}}} \quad (14.9)$$

Notes :

- (a) Les formules (14.8), (14.9) offrent une généralisation vectorielle du krigeage ordinaire. On calculera d'abord un vecteur $\mathbf{v} = \mathbf{P}^t\mathbf{C}^{-1}$, ensuite

$$\boldsymbol{\beta} = (\mathbf{v}\mathbf{P})^{-1}(\mathbf{v}\mathbf{X}),$$

et ensuite on calcule le champs demoyenné, en enlevant la moyenne $m(x) = \mathbf{p}^t(x)\mathbf{v}$.

- (b) Les deux formules explicites de base, celle du $\boldsymbol{\beta}$ et celle de l'estimation du processus avec la moyenne enlevé $Z(x)$ s'expriment les deux elegamment en fonction du produit scalaire $\langle, \rangle_{\mathbf{C}^{-1}}$.
- (c) On voit donc que le krigeage avec derive est équivalent, au niveau de l'estimation ponctuelle, à une interpolation déterministe, qui utilise la fonction de covariance du processus, qui, dans le cas des processus isotropes, devient "interpolation par des fonctions de base radiales".

Démonstration Començons par 3). Pour que l'estimateur soit sans biais, il faut qu'on ait, pour n'importe quel $b_j, j = 1, \dots, M$:

$$\sum_{j=1}^M b_j p_j(x) = \sum_{j=1, i=1}^{M, I} w_i (b_j p_j(x_i))$$

Il suit que les coefficients de b_j des deux côtés doivent être égaux, et on trouve que $\forall j = 1, \dots, M$

$$p_j(x) = \sum_{i=1}^I w_i p_j(x_i) = \mathbf{p}_{(j)}^t \mathbf{w} \quad (14.10)$$

où $\mathbf{p}_{(j)}$ dénote le vecteur colonne $\mathbf{p}_{(j)} = (p_j(x_i), i = 1, \dots, I)$, pour $j = 1, \dots, M$. Pour que l'estimateur soit sans biais, il faut qu'on ait, pour n'importe quel $b_j, j = 1, \dots, M$:

$$\sum_{j=1}^M b_j p_j(x) = \sum_{j=1, i=1}^{M, I} a_i (b_j p_j(x_i))$$

Il suit que les coefficients de b_j des deux côtés doivent être égaux, et on trouve que $\forall j = 1, \dots, M, p_j(x) = \sum_{i=1}^I w_i p_j(x_i) = \mathbf{p}^{(j)} \mathbf{w}$ où $\mathbf{p}^{(j)}$ dénote le vecteur ligne $\mathbf{p}^{(j)} = (p_j(x_i), i = 1, \dots, I)$, pour $j = 1, \dots, M$. **Note** : Par exemple, pour la fonction $p_1(x) = 1$, la contrainte est que la somme des poids soit égale à unité : $\sum_i w_i = 1$ et pour la fonction $p_2(x) = x$, la contrainte est que : $\sum_i w_i(x) = x$. En forme matricielle, on a :

$$\mathbf{P}^t \mathbf{w} = \mathbf{p}(x) \quad (14.11)$$

Note : Par exemple, pour la fonction $p_1(x)$, la contrainte est que la somme des poids soit égale à unité : $\sum_i w_i = 1$. Les contraintes de biais en notation matricielle donnent donc précisément la deuxième équation de (14.6) :

$$\mathbf{P}^t \mathbf{w} = \mathbf{p}_x \quad (14.12)$$

Pour obtenir la première équation du système (14.6), on minimise la variance d'estimation

$$\sigma_E^2 = \mathbb{E}(Z(x) - \hat{Z}(x))^2 = \sigma^2 + \mathbf{w}^t \mathbf{C} \mathbf{w} - 2\mathbf{w}^t \mathbf{C}_{X,Y}$$

sous les contraintes de biais. Par la méthode de Lagrange, on rajoute les contraintes, multipliées par un vecteur colonne $2\boldsymbol{\mu} = (\mu_j, j = 1, \dots, M)$, et on arrive à minimiser :

$$\sigma^2 + \mathbf{w}^t \mathbf{C} \mathbf{w} - 2\mathbf{w}^t \mathbf{C}_{X,Y} - 2(\mathbf{w}^t \mathbf{P} - \mathbf{p}^t(x)) \boldsymbol{\mu}$$

En annulant les dérivées de cette quantité par rapport aux poids \mathbf{w} , on arrive au système désiré, $\mathbf{C} \mathbf{w} = \mathbf{C}_{X,Y} + \mathbf{X} \boldsymbol{\mu}$. Donc, $(\mathbf{w}, -\mathbf{c})$ satisfait système (14.6). 1) suit immédiatement, en résolvant $(\mathbf{w} - \mathbf{c})$ du système (14.6) et en écrivant $\hat{Y}(x) = \mathbf{X}^t \mathbf{w} = (\mathbf{X}, \mathbf{0})^t (\mathbf{w}, -\mathbf{c})$. 2) Par la partie 1), on a :

$$\hat{Y}(x) = (\mathbf{X}^t \quad \mathbf{0}^t) \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix} = \tilde{\mathbf{w}}^t \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix}$$

ou $\tilde{w}^t = (\mathbf{X}^t \ \mathbf{0}^t) \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1}$ et $\tilde{w} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ satisfait $\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix}$.

Note : Dans la formule de prédiction (14.4), les difficultés principales sont de choisir une forme paramétrique convenable $\mathbf{C}_\theta(x, y)$, et ensuite d'estimer les paramètres θ . Cet exercice en suggère une méthode.

Soit $\mathbf{Y} = (Y(x_1), \dots, Y(x_I))$ et soit \mathbf{Q} la matrice

$$\mathbf{Q} = \mathbf{I}_d - \mathbf{X}^t (\mathbf{X} \mathbf{X})^{-1} \mathbf{X}$$

Montrez que $\mathbf{Q} \mathbf{X}^t = \mathbf{0}$, et $\mathbf{Q}^2 = \mathbf{Q} = \mathbf{Q}^t$. Déduisez que

$$\mathbb{E}(\mathbf{Q} \mathbf{Y}) = \mathbf{0} \text{ et}$$

$$\text{Var}(\mathbf{Q} \mathbf{Y}) = \mathbb{E}(\mathbf{Q} \mathbf{Y} \mathbf{Y}^t \mathbf{Q}) = \mathbf{Q} \mathbf{C} \mathbf{Q}$$

Cette formule suggère que après le choix d'une forme paramétrique \mathbf{C}_θ , une méthode pour choisir θ pourrait être de minimiser :

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \|\mathbf{Q} \mathbf{Y} \mathbf{Y}^t \mathbf{Q} - \mathbf{Q} \mathbf{C}_\theta \mathbf{Q}\| \quad \text{avec} \|(w_{i,j})\| := \sum_{i,j} w_{i,j}^2$$

14.3 Interpolation Hermite-Birkhoff

Une légère modification permet dans le contexte déterministe d'interpoler en sachant les valeurs des dérivées de u .

Exemple 14.3.1 Interpolation Hermite-Birkhoff. *Supposons un champs déterministe $y(x) = 4x^2$, pour le quel on observe la fonction et les premières deux dérivées en 0.*

Donc :

$$\lambda_1[y] = y''(0) = 8, \quad \lambda_2[y] = y'(0) = 0, \quad \lambda_3[y] = y(0) = 0$$

Calculons l'interpolation $\hat{Y}(x)$ basé sur le noyau Gaussien $\varphi(x, y) = e^{-(x-y)^2}$, c.-à-d. en utilisant la base :

$$(\lambda_i^{(y)}[e^{-(y-x)^2}], i = 1, 2, 3) = (4x^2 - 2) e^{-x^2}, 2x e^{-x^2}, e^{-x^2}$$

Nous cherchons donc une approximation de la forme :

$$\hat{Y}(x) = \sum_i^3 \alpha_i \lambda_i^{(y)}[e^{-(x-y)^2}] = \alpha_1 (4x^2 - 2) e^{-x^2} + \alpha_2 2x e^{-x^2} + \alpha_3 e^{-x^2}$$

Remarque 14.3.1 (*) *L'interpolation Hermite-Birkhoff en utilisant les dérivées, est assez facile à formaliser dans le contexte déterministe, en utilisant la théorie des distributions, mais beaucoup plus difficile à formaliser dans le contexte stochastique.*

Exercice 14.3.1 *Calculez l'interpolation généralisée de Hermite-Birkhoff $\hat{u}(x)$ basée sur le noyau Gaussien $\varphi(x, y) = e^{-(x-y)^2}$ pour la fonction $y(x) = x^2$, à partir des valeurs des fonctionnelles :*

$$\lambda_1(u) = u'(1), \quad \lambda_2(u) = u'(0), \quad \lambda_3(u) = y(0)$$

Sol :

$$\alpha_1 = \frac{1}{1 - 3e^{-2}} = \frac{e^2}{e^2 - 3}$$

$$\alpha_2 = \frac{e^{-1}}{1 - 3e^{-2}} = \frac{e}{e^2 - 3}$$

$$\alpha_3 = \frac{2e^{-1}}{1 - 3e^{-2}} = \frac{2e}{e^2 - 3}$$

14.4 Exercices : Krigeage et interpolation

1. Ecrivez le système de krigeage simple pour 9 points sur un maillage équidistant en \mathbb{R}^2 , en supposant une corrélation sphérique standardisée de portée 1, si la distance entre les points du maillage est $h = .4$.
2. **Interpolation Hermite-Birkhoff.** Supposons un champs déterministe $u(x) = 4x^2$, pour le quel on observe :

$$\lambda_1[u] = u'(1) = 8, \quad \lambda_2[u] = u'(0) = 0, \quad \lambda_3[u] = u(0) = 0$$

Calculez l'interpolation $n(x)$ basé sur le noyau Gaussien $C(x, y) = e^{-(x-y)^2}$, i.e. l'approximation de la forme :

$$n(x) = \sum_i^3 \alpha_i \lambda_i[e^{-(x-y)^2}] = \alpha_1(2(x-1)) e^{-(x-1)^2} + \alpha_2(2x) e^{-x^2} + \alpha_3 e^{-x^2}$$

3. Calculez l'interpolation généralisée $\bar{u}(x)$ basé sur le noyau Gaussien $C(x, y) = e^{-(x-y)^2}$, à reproduction polynômiale d'ordre 0, pour la fonction $u(x) = 1 + x$, à partir des valeurs des fonctionelles :

$$\lambda_1(u) = u'(1), \quad \lambda_2(u) = u(1), \quad \lambda_3(u) = u(0)$$

Solution : La base d'interpolation est :

$$(\lambda_1(e^{-(x-y)^2}) = 2(x-1)e^{-(x-1)^2}, \lambda_2(e^{-(x-y)^2}) = e^{-(x-1)^2}, \lambda_3(e^{-(x-y)^2}) = e^{-x^2})$$

donc on cherche $\bar{u}(x) = \alpha_1 2(x-1)e^{-(x-1)^2} + \alpha_2 e^{-(x-1)^2} + \alpha_3 e^{-x^2}$.

Le système $2\alpha_1 = 1 + 2\alpha_3 e^{-1}, \alpha_2 + \alpha_3 e^{-1} = 2, (\alpha_2 - 2\alpha_1)e^{-1} + \alpha_3 = 1$ donne : $(\alpha_1 = .89, \alpha_2 = 1.60, \alpha_3 = 1.06)$.

4. Supposons un champs déterministe $U(x) = 4x^2$, observé en $x_1 = 0, x_2 = 1/2, x_3 = 1$. On a :

$$\lambda_1[u] := u(0) = 0, \quad \lambda_2[u] := u(1/2) = 1, \quad \lambda_3[u] := u(1) = 4$$

- (a) Calculez le corrélogramme empirique. Formulez le problème de choisir R tq le noyau Gaussien $\rho(r) = e^{-(r/R)^2}$ donne la meilleure approximation du corrélogramme empirique au sens des moindres carrés et résolvez le!

- (b) Calculez le variogramme empirique si le champs est observé aussi en $x = 1/4$ et $x = 3/4$. Formulez le problème de choisir le palier partiel σ^2 , la pépite τ^2 et la portée R tq le variogramme théorique Gaussien $\tau^2 + \sigma^2(1 - e^{-(r/R)^2})$ donne la meilleure approximation du variogramme empirique au sens des moindres carrés. Indiquez une commande pour résoudre ce problème en R, pour un vecteur général de semivariances empiriques.
- (c) Donnez la forme générale du "système de krigeage à dérive paramétrique" obtenu pour une fonction $\rho(r)$ arbitraire. Donnez aussi son cas particulier obtenu pour une tendance constante (appelé "système de krigeage ordinaire").
- (d) En utilisant R, calculez les coefficients d'interpolation basé sur le noyau Gaussien $\rho(r) = e^{-(r)^2}$ quand on ajoute une dérive constante et linéaire, i.e. considérez des approximations de la forme :

$$\begin{aligned} n(x) &= \alpha_1 e^{-x^2} + \alpha_2 e^{-(x-1/2)^2} + \alpha_3 e^{-(x-1)^2} + \beta \\ \nu(x) &= \alpha_1 e^{-x^2} + \alpha_2 e^{-(x-1/2)^2} + \alpha_3 e^{-(x-1)^2} + \beta_0 + \beta_1 x \end{aligned}$$

- (e) Calculez l'erreur $\int_{x \in [0,1]} |u(x) - n(x)|^2 dx$ dans ces deux cas.
5. Montrez que la fonction $C(x, y) = f(x) f(y)$ est un noyau de covariance.
 6. Montrez que la fonction $C(x, y) = e^{-|x-y|^2}$ est un noyau de covariance.
 7. Montrez que la fonction $C(x, y) = h(x+y) - h(x-y)$, où $h(x)$ est une fonction positive, avec minimum en 0, est un noyau de covariance.
 8. Calculez la corrélation sphérique standardisée (qui représente la mesure de l'intersection des deux sphères de rayon $1/2$, situées à distance $r \leq 1$ une de l'autre), en dimensions 1 et 2.

Chapitre 15

Examens d'entraînement

15.1 Examen 1

1. Le champ déterministe $u(x) = x(2 - x)$ sera "estimé" à partir des observations en $x_1 = 0, x_2 = 1, x_3 = 2$, et en supposant (à tort) un noyau de corrélation sphérique $\varphi(x, y) = (1 - (\frac{|x-y|}{R})^3)_+$, avec $R = 2$ et une dérive linéaire, c.-à-d. calculer la meilleure approximation de la forme :

$$\hat{u}(x) = \alpha_1(1 - |x/2|^3)_+ + \alpha_2(1 - |(x-1)/2|^3)_+ + \alpha_3(1 - |(2-x)/2|^3)_+ + \beta_0 + \beta_1x$$

- (a) Calculez (à partir du noyau) la matrice \mathbf{C} des corrélations théoriques entre les observations, et son inverse.
- (b) Estimer la moyenne du champs, à partir de la formule

$$\boldsymbol{\beta} = (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1} \mathbf{P}^t \mathbf{C}^{-1} \mathbf{u}$$

- (c) Calculez les valeurs du champs demoyonné $\mathbf{z} = \mathbf{u} - \mathbf{P}\boldsymbol{\beta}$.
 - (d) Obtenez la prédiction $\hat{u}(x)$ obtenue par le krigeage avec dérive (pratiquement, on applique un krigeage simple au champs demoyonné par la moyenne $m(x)$ obtenu ci-dessus).
2. a) Calculez l'interpolation généralisée Hermite-Birkhoff $\hat{u}(x)$ basée sur le noyau Gaussien $\varphi(x, y) = e^{-(x-y)^2}$, à reproduction polynômiale d'ordre 0, pour la fonction $u(x) = x^2$, à partir des valeurs des fonctionnelles :

$$\lambda_1(u) = u'(1), \quad \lambda_2(u) = u'(0), \quad \lambda_3(u) = u(0)$$

- b) Calculez l'interpolation généralisée $\bar{u}(x)$ basé sur le noyau Gaussien $C(x, y) = e^{-(x-y)^2}$, à reproduction polynômiale d'ordre 0, pour la fonction $u(x) = 1 + x$, à partir des valeurs des fonctionnelles :

$$\lambda_1(u) = u'(1), \quad \lambda_2(u) = u(1), \quad \lambda_3(u) = u(0)$$

3. Supposons $X_0, X_1, X_2, \dots, X_n$ sont des variable aléatoire à variance 1, et $\text{Cov}(Y_i, X_j) = a$ si $i \neq j$. a) Quel est le domaine de valeurs permmissibles pour ρ ? b) Trouvez le meilleur

prédicteur linéaire de X_0 en fonction de X_1, X_2, \dots, X_n , et la variance de l'erreur de prédiction, si : i) l'espérance m est 0. ii) l'espérance m est constante et **inconnue**. Comparez avec le cas d'une espérance connue iii) l'espérance est $\mathbb{E}Y_i = \beta i$, où β est **inconnue**.

4. Trouver les coefficients $\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ et la prédiction $\hat{U}(x)$ pour les krigeage a) simple b) de la moyenne c) ordinaire pour chacun des deux cas suivants :
- (a) On dispose d'une seule donnée $U(x_1)$.
- (b) Toutes les distances entre les I données et le point estimé sont plus grandes que la portée.

Écrivez en chaque cas la matrice d'interpolation dans le cas des trois points connus, pour une fonction de base arbitraire $\rho(r)$.

Solution :

1. (a) On a encore

$$C = \begin{pmatrix} 1 & a & 0 \\ a & 1 & a \\ 0 & a & 1 \end{pmatrix}$$

avec $a = 7/8$ et

$$C^{-1} = \begin{pmatrix} \frac{1-a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{a^2}{1-2a^2} \\ -\frac{a}{1-2a^2} & \frac{1}{1-2a^2} & -\frac{a}{1-2a^2} \\ \frac{a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{1-a^2}{1-2a^2} \end{pmatrix} = \frac{1}{34} \begin{pmatrix} -15 & 56 & -49 \\ 56 & -64 & 56 \\ -49 & 56 & -15 \end{pmatrix}$$

- (b) On trouve $\mathbf{P}^t \mathbf{C}^{-1} = \begin{pmatrix} -\frac{4}{17} & \frac{24}{17} & -\frac{4}{17} \\ -\frac{21}{17} & \frac{24}{17} & \frac{13}{17} \end{pmatrix}$ et $\boldsymbol{\beta} = (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1} (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{u}) = \begin{pmatrix} \frac{16}{17} & \frac{16}{17} \\ \frac{16}{17} & \frac{50}{17} \end{pmatrix}^{-1} \begin{pmatrix} \frac{24}{17} \\ \frac{24}{17} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ 0 \end{pmatrix}$. La moyenne du champs est

$$m(x) = \frac{3}{2}, \forall x$$

- (c) Le champs demoyenné est $z = (-3/2, 1/2, -3/2)$.

- (d) Soit $\boldsymbol{\alpha} = \mathbf{z} \mathbf{C}^{-1} = (2, -4, 2)$. Par le krigeage simple appliqué au champs demoyenné, on trouve $\hat{u}(x) = m + \langle \boldsymbol{\alpha}, \mathbf{c}_x \rangle = \frac{3}{2} + 2(1 - (x/2)^3) - 4(1 - |x/2 - 1/2|^3) + 2(1 - (1 - x/2)^3) = \begin{cases} 3x/2 - x^3/2, & \text{si } x \leq 1 \\ (-2 + 9x - 6x^2 + x^3)/2 & \text{si } x \geq 1 \end{cases}$ On a obtenu la spline cubique qui interpole les points ! Les valeurs communes des dérivés en $1/2$ sont $f'(1) = 0, f''(1) = -3$.

2. La base d'interpolation est :

$$(\lambda_1(e^{-(x-y)^2}) = 2(x-1)e^{-(x-1)^2}, \lambda_2(e^{-(x-y)^2}) = e^{-(x-1)^2}, \lambda_3(e^{-(x-y)^2}) = e^{-x^2})$$

donc on cherche $\bar{u}(x) = \alpha_1 2(x-1)e^{-(x-1)^2} + \alpha_2 e^{-(x-1)^2} + \alpha_3 e^{-x^2}$. Le système $2\alpha_1 = 1 + 2\alpha_3 e^{-1}, \alpha_2 + \alpha_3 e^{-1} = 2, (\alpha_2 - 2\alpha_1)e^{-1} + \alpha_3 = 1$ donne : $(\alpha_1 = .89, \alpha_2 = 1.60, \alpha_3 = 1.06)$.

3. $\rho \in [-1/(n-1), 1]$ b) Pour $n = 3$, on trouve

$$C^{-1} = \begin{pmatrix} \frac{1+a}{1+a-2a^2} & -\frac{a}{1+a-2a^2} & -\frac{a}{1+a-2a^2} \\ -\frac{a}{1+a-2a^2} & \frac{1+a}{1+a-2a^2} & -\frac{a}{1+a-2a^2} \\ -\frac{a}{1+a-2a^2} & -\frac{a}{1+a-2a^2} & \frac{1+a}{1+a-2a^2} \end{pmatrix} = \begin{pmatrix} 3/2 & -1/2 & -1/2 \\ -1/2 & 3/2 & -1/2 \\ -1/2 & -1/2 & 3/2 \end{pmatrix}$$

i) $\hat{X}_0 = \frac{a}{1+(n-1)a}(X_1 + X_2 + \dots X_n)$, ii) $m = \frac{X_1+X_2+\dots X_n}{n} = \hat{X}_0$

4. b) On a $\tilde{C} = \begin{pmatrix} \sigma^2 \mathbf{I} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}$ et donc $\tilde{C}^{-1} = \begin{pmatrix} \sigma^{-2}(\mathbf{I} - \mathbf{P}(\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t) & \mathbf{P}(\mathbf{P}^t \mathbf{P})^{-1} \\ (\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t & -\tau^2(\mathbf{P}^t \mathbf{P})^{-1} \end{pmatrix}$

On trouve $\boldsymbol{\beta} = (\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t \mathbf{u}$, $\hat{U}(x) = 0$ et $\hat{U}(x) = m(x) + \hat{U}(x) = \mathbf{p}(x)^t \boldsymbol{\beta} = \mathbf{P}(\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t \mathbf{u}$, ce qui est la formule de prévision de la régression sur des variables exogènes. La RSS est $\mathbf{u}^t \mathbf{A} \mathbf{u}$. c) Pour l'estimation des locations non-observés, il en résulte : $w_i = c = I^{-1}, \forall i$ et $\sigma_E^2 = \sigma^2(1 + I^{-1})$ (donc dans le cas le plus défavorable d'une seule donnée, la variance d'estimation est égale à deux fois la variance a priori).

Chapitre 16

Rudiments de R

After starting an **R** session, load with the commands

```
install.packages("astsa") # install it ... choose the closest CRAN mirror
require(astsa)
data(jj) # load the data
jj      # print it to the screen
length(jj) # the number of elements
jj[1]    # the first element
jj[84]   # the last element
jj[1:4]  # the first 4 elements
jj[-(1:80)] # everything EXCEPT first 80 elements
plot(jj, ylab="Earnings per Share", main="J & J")

plot(diff(log(jj)), main="logged and diffed")

x = -5:5          # sequence of integers from -5 to 5
y = 5*cos(x)     # guess
par(mfrow=c(3,2)) # multifigure setup: 3 rows, 2 cols
#--- plot:
plot(x, main="plot(x)")
plot(x, y, main="plot(x,y)")
#--- plot.ts:
plot.ts(x, main="plot.ts(x)")
plot.ts(x, y, main="plot.ts(x,y)")
#--- ts.plot:
ts.plot(x, main="ts.plot(x)")
ts.plot(ts(x), ts(y), col=1:2, main="ts.plot(x,y)") # note- x and y are ts objects
#--- the help files [? and help() are the same]:
?plot.ts
help(ts.plot)
?par          # might as well skim the graphical parameters help file while you're here

noise = ts(rnorm(48), start=c(2163,6), frequency=12)
```

```

k = c(.5,1,1,1,.5)          # k is the vector of weights
(k = k/sum(k))
[1] 0.125 0.250 0.250 0.250 0.125
fjj = filter(jj, sides=2, k) # ?filter for help [but you knew that already]
plot(jj)
lines(fjj, col="red")       # adds a line to the existing plot
lines(lowess(jj), col="blue", lty="dashed")

Q = factor(rep(1:4,21))     # make (Q)uarter factors [that's repeat 1,2,3,4, 21 times]
reg = lm(log(jj)~0+trend+Q, na.action=NULL) # run the regression without an intercept
#-- the na.action statement is to retain time series attributes
summary(reg)

dljj = diff(log(jj))       # difference the logged data
plot(dljj)

par(mfrow=c(2,1))         # set up the graphics
hist(dljj, prob=TRUE, 12) # histogram
lines(density(dljj))      # smooth it - ?density for details
qqnorm(dljj)              # normal Q-Q plot
qqline(dljj)

x = cos(2*pi*1:100/4) + .1*rnorm(100)
plot.ts(x)
dev.new()
spectrum(x)
dev.new()
lag.plot(x, 4)
?lag.plot

x = cos(2*pi*1:100/4) + .01*rnorm(100)
plot.ts(x)
dev.new()
spectrum(x)

```

`library(geoR)` or `require(geoR)`. If the package is loaded correctly a message will be displayed.

Typically, default arguments are used for the function calls and the user is encouraged to inspect other arguments of the functions using the `args` and `help` functions. For instance, to see all the arguments for the function `variog` type `args(variog)` and/or `help(variog)`.

Typically, data are stored as an object (a list) of the class `geodata`. An object of this class contains two obligatory elements : the coordinates of data locations as first element

(\$coords) and the data values as second element (\$data), which can be a vector or a matrix ; autres elements possibles sont \$covariate, \$borders.

There are a few data-sets included in the package distribution, like data(s100) The list of all data-sets included in the package is given by data(package='geoR').

Pour convertir des données *.dat structurées par colonnes en un objet de type "geodata", utilisez : as.geodata(*.dat) ou read.geodata(*.dat, type=T).

A quick summary for the coordinates and data values of a geodata object can be obtained by summary(s100) . plot(s100) fourni les courbes de niveau, le histogram, ainsi que les graphes de la dependence des données pas rapport aux coordonnées.

16.1 Le Variogramme

Empirical variograms are calculated using the function variog, with options cloud, binned(default) or smoothed.

```
>cloud <- variog(s100, option ="cloud", max.dist=1)
>plot(cloud)
>bin <-variog(s100, uvec=seq(0,1,l=11))
>plot(bin)
```

Several results are returned by variog. The first three names(bin) are the more important ones and contain the distances, the estimated semivariance and the number of pairs for each bin.

The option "smooth" adds a theoretical variogram.

```
> smooth <- variog(s100, option = "smooth", max.dist = 1,
+ n.points = 100, kernel = "normal", band = 0.2)
> lines(smooth, type = "l", lty = 2)
> legend(0.4, 0.3, c("empirical", "exponential model",
+ "smoothed"), lty = c(1, 1, 2), lwd = c(1, + 3, 1))
```

The parameters of a theoretical variogram have to be estimated by some method :

1. by eye, using the function lines.variomodel, drawing curves of different theoretical variogram functions over an empirical variogram, changing the variogram model and/or its parameters and, at last, choosing one of them. The following commands show how to add a line with a variogram model to a variogram plot. Three different variogram models are used.

```
> plot(variog(s100, max.dist = 1))
> lines.variomodel(cov.model = "exp", cov.pars = c(1,
+ 0.3), nug = 0, max.dist = 1)
> lines.variomodel(cov.model = "mat", cov.pars = c(0.85,
+ 0.2), nug = 0.1, kappa = 1, max.dist = 1, + lty = 2)
> lines.variomodel(cov.model = "sph", cov.pars = c(0.8,
+ 0.8), nug = 0.1, max.dist = 1, lwd = 2)
```

2. by least squares fit of empirical variograms : with options for ordinary (OLS) and weighted (WLS) least squares, using the function variofit

3. by likelihood based methods : with options for maximum likelihood (ML) and restricted maximum likelihood (REML), using the function `likfit`, or Bayesian methods, using the function `krige.bayes`.

An example call to `likfit` is given below.

```
> ml <- likfit(s100, ini = c(1, 0.5))
> ml
likfit: estimated model parameters: beta tausq sigmasq phi "0.7766"
"0.0000" "0.7517" "0.1827" likfit: maximised log-likelihood = -83.57
```

When using these functions, the nugget effect parameter can either be estimated or set to a fixed value. The same applies for smoothness, anisotropy and transformation parameters. Options for taking trends into account are also included. Trends can be specified as polynomial functions of the coordinates and/or linear functions of given covariates. Fitting models with nugget fixed to zero :

```
> ml <- likfit(s100, ini = c(1, 0.5), fix.nugget = T)
> reml <- likfit(s100, ini = c(1, 0.5), fix.nugget = T,
+ method = "RML")
> ols <- variofit(variog1, ini = c(1, 0.5), fix.nugget = T,
+ weights = "equal")
> wls <- variofit(variog1, ini = c(1, 0.5), fix.nugget = T)
```

Fitting models with a fixed value for the nugget

```
> ml.fn <- likfit(s100, ini = c(1, 0.5), fix.nugget = T,
+ nugget = 0.15)
> reml.fn <- likfit(s100, ini = c(1, 0.5), fix.nugget = T,
+ nugget = 0.15, method = "RML")
> ols.fn <- variofit(variog1, ini = c(1, 0.5), fix.nugget = T,
+ nugget = 0.15, weights = "equal")
> wls.fn <- variofit(variog1, ini = c(1, 0.5), fix.nugget = T,
+ nugget = 0.15)
```

```
> ml.n <- likfit(s100, ini = c(1, 0.5), nug = 0.5)
> reml.n <- likfit(s100, ini = c(1, 0.5), nug = 0.5,
+ method = "RML")
> ols.n <- variofit(variog1, ini = c(1, 0.5), nugget = 0.5,
+ weights = "equal")
> wls.n <- variofit(variog1, ini = c(1, 0.5), nugget = 0.5)
```

Comands for plotting fitted models against empirical variograms :

```
> par(mfrow = c(1, 3))
> plot(variog1, main = expression(paste("fixed ", tau^2 ==
+ 0)))
> lines(ml, max.dist = 1)
> lines(reml, lwd = 2, max.dist = 1)
```

```

> lines(ols, lty = 2, max.dist = 1)
> lines(wls, lty = 2, lwd = 2, max.dist = 1)
> legend(0.5, 0.3, legend = c("ML", "REML", "OLS",
+ "WLS"), lty = c(1, 1, 2, 2), lwd = c(1, 2, + 1, 2), cex = 0.7)
> plot(variog1, main = expression(paste("fixed ", tau^2 ==
+ 0.15)))
> lines(ml.fn, max.dist = 1)
> lines(reml.fn, lwd = 2, max.dist = 1)
> lines(ols.fn, lty = 2, max.dist = 1)
> lines(wls.fn, lty = 2, lwd = 2, max.dist = 1)
> legend(0.5, 0.3, legend = c("ML", "REML", "OLS",
+ "WLS"), lty = c(1, 1, 2, 2), lwd = c(1, 2, + 1, 2), cex = 0.7)
> plot(variog1, main = expression(paste("estimated ",
+ tau^2)))
> lines(ml.n, max.dist = 1)
> lines(reml.n, lwd = 2, max.dist = 1)
> lines(ols.n, lty = 2, max.dist = 1)
> lines(wls.n, lty = 2, lwd = 2, max.dist = 1)
> legend(0.5, 0.3, legend = c("ML", "REML", "OLS",
+ "WLS"), lty = c(1, 1, 2, 2), lwd = c(1, 2, + 1, 2), cex = 0.7)
> prof <- proflik(ml, geodata = s100, sill.val = seq(0.48,
+ 2, l = 11), range.val = seq(0.1, 0.52, l = 11), + uni.only =
FALSE)
> par(mfrow = c(1, 3))
> plot(prof, nlevels = 16)

```

Directional variograms can also be computed by the function `variog` using the arguments *direction* and *tolerance*. For example, to compute a variogram for the direction 60 degrees with the default tolerance angle (22.5 degrees) the command would be : `vario60 <- variog(s100, max.dist = 1, direction=pi/3)`

For a quick computation in four directions we can use the function `variog4` which by default computes variogram for the direction angles 0°, 45°, 90° and 135°. `vario.4 <- variog4(s100, max.dist = 1)`

The command to perform ordinary kriging using the parameters estimated by weighted least squares with nugget fixed to zero :

```

> kc4 <- krige.conv(s100, locations = loci,
+ krige = krige.control(obj.m = wls))

```

list including the predicted values (`kc4$predict`) and the kriging variances (`kc4$krige.var`).

```

> plot(s100$coords, xlim = c(0, 1.2), ylim = c(0,
+ 1.2), xlab = "Coord X", ylab = "Coord Y")
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3,
+ 1, 1.1), ncol = 2)
> text(loci, as.character(1:4), col = "red")
> polygon(x = c(0, 1, 1, 0), y = c(0, 0, 1, 1),

```

```

+ lty = 2)
> pred.grid <- expand.grid(seq(0, 1, l = 51), seq(0,
+ 1, l = 51))
> kc <- krige.conv(s100, loc = pred.grid, krige = krige.control(obj.m = ml))

```

16.2 variog

package :geor Compute Empirical Variograms

1. *Usage* : `variog(geodata, coords = geodata$coords, data = geodata$data, uvec = "default", breaks = "default", trend = "cte", lambda = 1, option = c("bin", "cloud", "smooth"), estimator.type = c("classical", "modulus"), nugget.tolerance, max.dist, pairs.min = 2, direction = "omnidirectional", tolerance = pi/8, unit.angle = c("radians", "degrees"), messages, ...)`
2. *Arguments* :
 - (a) *geodata* : a list containing element 'coords' as described next. Typically an object of the class "'geodata'" - a 'geor' data-set. If not provided the arguments 'coords' must be
 - (b) *coords* : an $n \times 2$ matrix containing coordinates of the n data locations in each row. Defaults to 'geodata\$coords', if not provided.
 - (c) *data* : a vector or matrix with data values. If a matrix is provided, each column is regarded as one variable or realization. Defaults to 'geodata\$data', if not provided.
 - (d) *uvec* : a vector with values used to define the variogram binning. Only used when 'option = "bin"'. See 'DETAILS' below for more details on how to specify the bins.
 - (e) *breaks* : a vector with values to define the variogram binning. Only used when 'option = "bin"'. See 'DETAILS' below for more details on how to specify the bins.
 - (f) *trend* : specifies the mean part of the model. See documentation of 'trend.spatial' for further details. Defaults to "'cte"'.
 - (g) *lambda* : values of the Box-Cox transformation parameter. Defaults to 1 (no transformation). If another value is provided the variogram is computed after transforming the data. A case of particular interest is $\lambda = 0$ which corresponds to log-transformation.
 - (h) *option* : defines the output type : the options "'bin'" returns values of binned variogram, "'cloud'" returns the variogram cloud and "'smooth'" returns the kernel smoothed variogram. Defaults to "'bin'"
 - (i) *estimator.type* : "'classical'" computes the classical method of moments estimator. "'modulus'" returns the variogram estimator suggested by Hawkins and Cressie (see Cressie, 1993, pg 75). Defaults to "'classical"'.
 - (j) *nugget.tolerance* : a numeric value. Points which are separated by a distance less than this value are considered co-located. Defaults to zero.

- (k) *max.dist* : a numerical value defining the maximum distance for the variogram. Pairs of locations separated for distance larger than this value are ignored for the variogram calculation. If not provided defaults takes the maximum distance among all pairs of data locations.
- (l) *pairs.min* : a integer number defining the minimum numbers of pairs for the bins. For 'option = "bin"', bins with number of pairs smaller than this value are ignored. Defaults to 'NULL'.
- (m) *direction* : a numerical value for the directional (azimuth) angle. This used to specify directional variograms. Default defines the omnidirectional variogram. The value must be in the interval $[0, \pi]$ radians ($[0, 180]$ degrees).
- (n) *tolerance* : numerical value for the tolerance angle, when computing directional variograms. The value must be in the interval $[0, \pi/2]$ radians ($[0, 90]$ degrees). Defaults to $\pi/8$.
- (o) *unit.angle* : defines the unit for the specification of angles in the two previous arguments. Options are "radians" and "degrees", with default to "radians".

3. Description :

Computes sample (empirical) variograms with options for the -classical- or -robust- estimators. Output can be returned as a 'binned variogram', a 'variogram cloud' or a 'smoothed variogram'. Data transformation (Box-Cox) is allowed. "Trends" can be specified and are fitted by ordinary least squares in which case the variograms are computed using the residuals.

4. The defaults- If arguments 'breaks' and 'uvec' are not provided, the binning is defined as follows :

1. read the argument 'max.dist'. If not provided it is set to the maximum distance between the pairs of points.
2. the center of the bins are initially defined by the sequence 'u = seq(0, max.dist, l = 13)'. 3. the interval spanned by each bin is given by the mid-points between the centers of the bins.

If an vector is passed to the argument 'breaks' its elements are taken as the limits of the bins (classes of distance) and the argument 'uvec' is ignored.

-Variations on the default- The default definition of the bins is different for some particular cases. 1. if there are coincident data locations the bins follows the default above but one more bin is added at the origin (distance zero) for the collocated points. 2. if the argument 'nugget.tolerance' is provided the separation distance between all pairs in the interval $[0, \text{nugget.tolerance}]$ are set to zero. The first bin distance is set to zero ('u[1] = 0'). The remaining bins follows the default. 3. if a scalar is provided to the argument 'uvec' the default number of bins is defined by this number. 4. if a vector is provided to the argument 'uvec', its elements are taken as central points of the bins.

5. Examples :

```
# Loading data : data(s100)
```

```
# computing variograms :
```

```
(a) binned variogram
```

```
vario.b <- variog(s100, max.dist=1)
```

(b) variogram cloud

```
vario.c <- variog(s100, max.dist=1, op="cloud")
```

(c) smoothed variogram

```
vario.s <- variog(s100, max.dist=1, op="sm", band=0.2)
```

(d) plotting the variograms :

```
par(mfrow=c(2,2)) plot(vario.b, main="binned variogram") plot(vario.c, main="variogram cloud") plot(vario.bc, bin.cloud=TRUE, main="clouds for binned variogram") plot(vario.s, main="smoothed variogram")
```

(e) computing a directional variogram

```
vario.0 <- variog(s100, max.dist=1, dir=0, tol=pi/8) plot(vario.b, type="l", lty=2) lines(vario.0) legend(0, 1.2, legend=c("omnidirectional", expression(0 * degree)), lt y=c(2,1))
```

[http://www.ic.arizona.edu/ic/math574/class_notes/geoRintro\[1\].pdf](http://www.ic.arizona.edu/ic/math574/class_notes/geoRintro[1].pdf)

16.3 trend.spatial

Specifies the mean part of the model. The argument *trend* defines the form of the mean and the following options are allowed : a) "cte" (the default option) : the mean is assumed to be constant over the region. b) "1st" the mean is assumed to be a first order polynomial on the coordinates. c) "2nd" the mean is assumed to be a second order polynomial on the coordinates :

$$m(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1)^2 + \beta_4 (x_2)^2 + \beta_5 x_1 x_2$$

16.4 likfit

Likelihood Based Parameter Estimation for Gaussian Random Fields Description Maximum likelihood (ML) or restricted maximum likelihood (REML) parameter estimation for (transformed) Gaussian random fields.

16.5 cov.spatial

covariance function is one of : "matern", "powered.exponential", "cauchy" or "gneiting.matern".

16.6 krige.conv

Description This function performs spatial prediction for fixed covariance parameters using global neighbourhood. Options available implement the following types of kriging : SK (simple kriging), OK (ordinary kriging), KTE (external trend kriging) and UK (universal kriging).

16.7 image.kriging

Image or Perspective Plot with Kriging Results Description Plots image or perspective plots with results of the kriging calculations.

16.8 Données de type "geodata" disponibles

Pour voir les données de type "geodata" déjà disponibles, utilisez : `data(package='geoR')` qui nous donne la prochaine liste des données en format "geodata" :

1. *Ksat Saturated Hydraulic Conductivity*
2. *SIC Spatial Interpolation Comparison data*
3. *ca20 Calcium content in soil samples (*analysé en tutoriel 4*)*
4. *camg Calcium and magnesium content in soil samples*
5. *elevation Surface Elevations*
6. *gambia Gambia Malaria Data*
7. *hoef Data for spatial analysis of experiments*
8. *isaaks Data from Isaaks and Srisvastava's book*
9. *parana Rainfall Data from Parana State, Brasil*
10. *s100 Simulated Data-Sets which Illustrate the Usage of the Package geoR*
11. *s121 Simulated Data-Sets which Illustrate the Usage of the Package geoR, gstat*
12. *meuse Meuse river data set*
13. *meuse.all Meuse river data set – original, full data set*
14. *meuse.grid Prediction Grid for Meuse Data Set*
15. *oxford Oxford soil samples*

D'autres, comme `cruciani.dat`, se trouvent sur le WEB.

16.9 Projet : Krigeage d'un champs déterministe, avec symmetrie radiale par rapport a un certain centre

http://www.iag.asso.fr/pdf/krigeage_juillet2002.pdf

1. (a) *Nous allons simuler des valeurs d'un champs déterministe, $F(x) = F(r)$, $r = |x|$ avec symmetrie radiale par rapport a un certain centre (par exemple $F(x) = \frac{\text{Sin}(r)}{r}$), observés avec un petit bruit ϵ_i , dans des points aléatoires $x_i \in \mathbb{R}^2$, choisies en simulant une pair des variables aléatoires Gaussiennes (noncorrelées). En suite, on calculera les observations*

$$Y_i = \frac{\text{Sin}(|x_i|)}{|x_i|} + \epsilon_i, i = 1, \dots, I.$$

(b) On calculera alors le covariogramme empirique

$$\gamma(r_i) = N(r_i)^{-1} \sum_{\{(x_i, x_j) : |x_i - x_j| \approx r_i\}} \frac{(Y_i - X_j)^2}{2},$$

où $N(r_i)$ est le nombre des paires à distance r_i , en réalisant un groupement (histogram) de $\gamma(r)$.

(c) On calculera ensuite un ajustement $\gamma(r)$ (polynomiale, ou par splines, ou par un noyau, ...) du variogram empirique.

(d) Avec ces valeurs, on calculera les contours de niveau estimé du champs, en utilisant d'abord le krigeage ordinaire, et en suite un calcul par la formule

$$\hat{X}(x) = \sum_{i=1}^N \frac{\rho(\|x - x_i\|)}{k(x)} Y_i = \sum \alpha_i \rho(\|x - x_i\|)$$

où $\alpha_i = k^{-1} Y_i$, et ρ est un noyau fixe, comme $\rho(r) = r^{-1}$, ou $\rho(r) = r^{-2}$. Comparez les résultats.

2. Expliquez les objets suivants de *GeoR*, en précisant leur paramètres essentiels, et en donnant des exemples d'utilisation :

(a) `likfit`, `proflik`, `plot.proflik`, `cov.spatial`, `grf`, `image.grf`

(b) `variofit`, `trend.spatial`, `variog4`, `krige.conv`, `image.kriging`, `krweights`

(c) `lines.variomodel`, `variog.mc.env`, `variog.model.env`, `read.geodata`, `polygrid`, `krige.bayes`