

Statistique spatiale, Géostatistique, Régression et Interpolation

Florin Avram

18 décembre 2009

Objectif : L'interpolation déterministe et la régression sont parmi les méthodes les plus importantes dans la statistique et les mathématiques appliquées.

Leur but : est d'estimer la valeur d'une variable $Y(x)$ en un point x quelconque, en connaissant des valeurs observées $Y(x_i)$ aux points de mesure x_i , pour $i = 1, \dots, I$.

La motivation est la prédiction des valeurs inobservables des phénomènes spatio-temporaux (comme les valeurs futures des séries temporelles, ou moins accessibles physiquement, coûteuses, etc, en géostatistique) à partir des valeurs connues. On veut à la fois : a) enlever du bruit éventuel et b) "extrapoler" du connu à l'inconnu.

Domaines d'application :

- Prospection et exploitation pétrolières et minières
- Traitement du signal
- Imagerie médicale
- Sciences environnementales : océanographie, météorologie, hydrogéologie, ...
- Séries temporelles, appliquées en économétrie, économie, finances, météo, médecine, ...

Note : Pour maîtriser cette discipline, des rudiments d'interpolation (par splines et par RBF-fonctions/noyaux radiaux de base) et de modélisation des champs Gaussiens devront être acquis.

Références :

1. Notes de cours/TD distribuées en classe, incluant parties des notes de C. Rabut sur les splines.
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 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.
7. Reproducing Kernel Hilbert Spaces in Probability and Statistics, by Alain Berlinet, Christine Thomas-Agnan.
8. Hierarchical modeling by Bradley Carlin, S. Banerjee et A.E. Gelfand.
9. 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 : la démarche géostatistique	2
2	Interpolation et ajustement par splines	6
2.1	Le calcul des splines naturelles d'interpolation exacte	7
2.2	Splines d'ajustement	11
2.3	Exercices	12
2.4	(**) Interpolation par splines de degré $2m - 1, m \geq 1$	15
3	L'analyse des données spatiales	15
3.1	Premier aperçu : la visualisation des données	15
3.2	Rudiments de statistique bivariée	16
3.3	La distribution gaussienne/normale multivariée	16
3.4	Processus/champs gaussiens	17
3.5	Les distributions Gaussiennes conditionnelles, la régression linéaire et l'interpolation	19
3.6	Processus Gaussiens isotropiques	21
3.7	Le corrélogramme et le variogramme spatiaux empiriques	21
3.8	Le corrélogramme et variogramme théoriques	23
4	Exemples de krigeage	25
4.1	Le krigeage simple	25
4.2	Krigeage de la moyenne	25
4.3	Le krigeage ordinaire	26
4.4	Resultat auxiliaire : l'inverse d'une matrice partitionnée en blocs	28
4.5	Exercices	29
5	Interpolation optimale/regression à tendance paramétrique/krigeage universel	31
5.1	Les noyaux positifs (*)	31
5.2	Exercices	32
5.3	L'interpolation par noyaux de base à reproduction polynomiale	33
5.4	<u>Le système d'interpolation/krigeage dual</u>	34
5.5	Interpolation Hermite-Birkhoff	38
6	Examens d'entraînement	40
6.1	Examen 1	40

1 Introduction : la démarche géostatistique

On considère ici des champs des variables $Y(x)$, où $x \in \mathbb{R}^d$ dans la géostatistique, $x \in \mathbb{N}, \mathbb{Z}$ ou \mathbb{R}_+ pour les séries 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**.

On se propose d'estimer la valeur de la variable $Y(x)$ en un point x quelconque, en connaissant des valeurs observées

$$X_i = Y(x_i) + \epsilon_i, i = 1, \dots, I$$

dans quelques points de mesure x_i . L'information la plus importante est fournie par les observations X_i , $i = 1, \dots, I$.

Interpolation par noyaux radiales (RBF) La démarche habituelle dans la régression est d'estimer la valeur dans un point inconnu x par une moyenne pondérée des valeurs observées donc :

$$\hat{Y}(x) = \sum_i w_i(x) X_i \quad (1)$$

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 sera 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 (1) deviendra donc :

$$\hat{Y}(x) = \sum_{i=1}^N \frac{\rho(\|x - x_i\|)}{k(x)} Y(x_i) := \sum \alpha_i \varphi(\|x - x_i\|) \quad (2)$$

où $\alpha_i = k^{-1} Y(x_i)$.

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.
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 (2) devient dans ce cas exacte, i.e. on trouve que pour $x = x_i$ on a $w_j(x) = \delta_i(j)$ et donc $\hat{Y}(x_i) = Y(x_i)$. Par contre, les prévisions sont nonlisses, données par la fonction

$$\hat{Y}(x) = \sum_{i=1}^I X_i(x) \mathbf{1}_{x \in S_R(x_i)}$$

où $S_R(x_i)$ est la sphère de rayon R autour de x_i (et qui est incluse dans le polygone A_i qui entoure x_i dans le diagramme Voronoi de ces points).

¹L'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 champs, et aussi en statistique, si $Y(x)$ est un champs Gaussien isotropique (voir chapitre 3.4) avec covariance $\text{Cov}(X_i, Y) = \varphi(\|x_i - y\|)$ et **moyenne 0**.

4. Dans le cas $R \geq \frac{\min_{i,j} \|x_i - x_j\|}{2}$, la formule (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 $\hat{Y}(x_i) = Y(x_i), i = 1, \dots, I$ soit satisfait.

Définition 1.1 On appellera **interpolation exacte par une fonction de base radiale (RBF)** $\rho(r), r > 0$, une interpolation de la forme :

$$\hat{Y}(y) = \sum_{i=1}^I \alpha_i \rho(|y - x_i|) \quad (3)$$

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 (3) est facile à implémenter, car la demande qu'elle soit exacte dans les points observés nous fournit un système d'équations linéaires.

Rémarquez qu'on a deux formules pour la même prévision $\hat{Y}(x)$. Pour les distinguer, nous les appellerons (1) et (3) formules de régression et interpolation RBF, respectivement. Bien sûr, c'est la même formule (et la relation de "dualité" entre les deux formules sera encore discutée plus tard).

Une autre information importante pour la prévision de $\hat{Y}(x)$ est fournie par les coordonnées $x_i, i = 1, \dots, I$. Plus généralement, on pourrait chercher 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 champ par une combinaison linéaire :

$$\boxed{\hat{Y}(y) = \sum_{i=1}^I \alpha_i \varphi(|y - x_i|) + \sum_{j=1}^J \beta_j p_j(y)} \quad (4)$$

où $p_j(y), j = 1, \dots, J$ sont des prédicteurs exogènes (fonctions des coordonnées, par exemple polynômes), et $\varphi(|y - x_i|), i = 1, \dots, I$ font partie des poids de pondération à accorder aux observations X_i (qui seront incluses en α_i).

Cette formule d'interpolation peut-être justifiée dans un cadre déterministe, mais la démarche adoptée par les géostatisticiens et les météorologues est de supposer que le champ $Y(x)$ est une réalisation d'un processus stochastique (i.e. une famille des variables aléatoires), et de lier le noyau $\varphi(y - x)$ à la covariance du champ. Remarquons d'emblée que l'avantage de la modélisation stochastique sur l'approche déterministe n'est pas évident, surtout avec peu des données.

Le prochain pas décompose le processus en :

$$\boxed{Y(x) = m(x) + Z(x)} \quad \text{où :} \quad (5)$$

1. $m(x)$ représente la moyenne $\mathbb{E}Y(x)$, ou une première prévision basée sur la "tendance".

²La formule d'interpolation (3) apparaît souvent dans le contexte de l'interpolation déterministe. Par exemple, supposons que X_i sont les intensités des quelques sources d'un champ déterministe électro-magnétique, 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 $Y = \sum_i \alpha_i \varphi(y - x_i)$ indiqué en (21).

La fonction $m(x)$ est **inconnue** et nonisotropique, mais supposée d'appartenir à un espace vectoriel (des fonctions) de dimension finie, i.e. d'admettre une décomposition paramétrique

$$\boxed{m(x) = \sum_{j=1}^M \beta_j p_j(x) = \boldsymbol{\beta}^t \mathbf{p}(x)} \quad \text{où } p_1(x) = 1 \quad (6)$$

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).

On dénotera par \mathbf{P} 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); donc, on peut écrire le vecteur des moyennes dans les points d'observation comme

$$\mathbf{m} = \mathbf{P}\boldsymbol{\beta}$$

- $Z(x) = Y(x) - m(x)$ est souvent un processus stochastique (qui essaie de capter des corrélations 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). Sa distribution fournira des intervalles des confiance pour les prédictions, un premier pas vers la modélisation stochastique. Dans la présence des corrélations nonnulles, elles devront être prises en compte dans la prédiction.

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/Financi> <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 :

$$\tilde{\mathbf{C}}_{i,j} := \mathbb{E}[(Y(x_i) - m(x_i))(Y(x_j) - m(x_j))] = \sigma^2 \varphi(\|(x_i - x_j)/R\|)$$

où $\sigma^2 = \text{Var}(Z(x))$, R est appelé la "portée" ou "échelle de corrélation" et $\varphi(r)$ 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" $Y(x)$ 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 :

$$\boxed{X_i = Y(x_i) + \epsilon_i = m(x_i) + Z(x_i) + \epsilon_i}, i = 1, \dots, I \quad (7)$$

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 :

$$\boxed{X = P\beta + Z + \epsilon} \quad (8)$$

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 $Y(x)$ 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).

2 Interpolation et ajustement par splines

Nous aimerions approximer une "nuage" des points $(x, y_i) \in \mathbb{R}^2, i = 1, \dots, n$ par une courbe unidimensionnelle $\xi(x), x \in \mathbb{R}$. Comment la choisir ?

Rappelons que l'interpolation la plus simple est celle exacte par polynômes, qui fait passer un polynôme de degré $n - 1$ par n points, mais que cette approche est rarement satisfaisante (voir "exemple de Runge"), à cause de trop d'oscillations. Pour remédier ça, une première idée serait d'utiliser une interpolation polynomiale par morceaux ; pour des séries composées des morceaux à allure différente, cela serait incontournable. Pour des séries à allure plutôt homogène, il est raisonnable de demander aussi **lissité** dans les points de contact. Mais, peut-être que la polynomialité par morceaux n'est pas nécessaire ? De toute façon, nous aimerions que notre courbe :

1. minimise sa distance aux points, par exemple au sens des moindres carrés,
2. soit lisse, par exemple avec $\xi(x) \in \mathcal{C}^2$, et
3. oscille le moins possible.

Avec peu de données et très peu de bruit, on aimerait passer exactement par les points, en satisfaisant les deux derniers objectifs.

Le moyen pour atteindre ce but nous vient d'un principe physique qui détermine le positionnement d'une tige mince, flexible, obligée à passer par des points donnés (x_i, y_i) , qui sont : a) la continuité des premières deux dérivées et b) la minimisation de l'énergie de flexion

$$\int_{\mathbb{R}} (\xi''(x))^2 dx. \quad (9)$$

On arrive ainsi à un problème variationnel de minimiser (9) sous les contraintes de a) passer exactement par les points et b) de lissité.

Le nom mathématique de cette "tige d'interpolation exacte" qui résout le problème variationnel est **spline cubique naturelle d'interpolation**, et il se trouve qu'elle a une caractérisation très simple :

Définition 2.1 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 $]x_i, x_{i+1}[$, $i = 1, \dots, n - 1$, ξ est un polynôme de degré 3
- Sur $(-\infty, a]$ et sur $]b, +\infty$, ξ est un polynôme de degré 1 (i.e. $\xi''(a) = \xi''(b) = \xi'''(a) = \xi'''(b) = 0$)

Donc : une spline cubique naturelle "est un polynôme de degré 3 par morceaux", dont la dérivée seconde est continue, et nulle aux bords.

L'intérêt des splines naturelles vient du fait qu'ils sont la solution du problème variationnel ci-dessous :

Théorème 2.1 a) La spline cubique naturelle d'interpolation des points $(x_i, y_i)_{i \in [1:n]}$ réalise l'unique minimum de "l'énergie de flexion" :

$$\min_{\substack{f \in \mathcal{C}^2([a..b]) \\ \forall i \in [1:n], f(x_i) = y_i}} \int_{\mathbb{R}} (f''(x))^2 dx$$

b) **Réciproquement**, une fonction ξ réalisant le minimum contraint de "l'énergie de flexion" ci-dessus est la spline cubique naturelle d'interpolation des points $(x_i, y_i)_{i \in [1:n]}$

Pour la démonstration, voir Green et Silverman, Ch. 2.

Notes : 1) La spline cubique naturelle par $n = 2$ points est la droite. Par $n = 1$ points il y a une famille infinie. Pour $n \geq 3$, on verra que la spline cubique naturelle qui interpole n points donnés est unique. Comme une première justification, remarquez que avec $n + 1$ morceaux en \mathcal{P}_3 (polynômes de degré 3), lissité \mathcal{C}^2 et quatre conditions frontière, il restent $4(n + 1) - 3n - 4 = n$ paramètres à fixer. Donc, a priori on pourra interpoler n points, i.e. il y aura une spline cubique naturelle unique passant par n'importe quelles $n \geq 3$ points. On peut aussi vérifier l'unicité de la spline naturelle d'ordre m passant par n'importe quelles $n \geq m$ points.

2) A priori, on peut aussi interpoler n points par un polynôme de degré $n - 1$. Mais, avec les splines, on espère d'avoir :

- oscillations moins grandes
- un calcul plus facile (morceau par morceau).

3) Parfois, on utilise d'autres conditions frontière : par exemple, pour les splines périodiques,

$$\text{on impose : } \begin{cases} \xi(a) = \xi(b) \\ \xi'(a) = \xi'(b) \\ \xi''(a) = \xi''(b) \\ \xi'''(a) = \xi'''(b) \end{cases}$$

2.1 Le calcul des splines naturelles d'interpolation exacte

Idée : Tenant compte des problèmes avec l'utilisation des polynômes de grand degré, on propose de couper $[a, b] = [x_1, x_n]$ en petits morceaux, et s'imposer un petit degré dans chaque morceau + lissité dans les points de contact.

Idée : Une spline a une **l'écriture "locale"**, dans chaque intervalle où ξ est un polynôme :

$$\begin{aligned} \forall j \in [0 : n-1], \forall x \in [x_j \dots x_{j+1}], \\ \xi(x) &= \xi(x_j) + (x - x_j)\xi'(x_j) + \frac{(x-x_j)^2}{2}\xi''(x_j) + \frac{(x-x_j)^3}{6}\xi'''(x_j) \\ &= \xi_j + (x - x_j)\xi'_j + \frac{(x-x_j)^2}{2}\xi''_j + \frac{(x-x_j)^3}{6}\xi'''_j, \end{aligned}$$

$$\xi_i''' = \frac{\xi_{i+1}'' - \xi_i''}{h_i}$$

Note : En general, avec n noeuds, nous avons un total de $4(n-1) + 4 = 4n$ paramètres, donc $3n$ seront déterminées par les conditions de raccordement, de façon que ξ , ξ' , et ξ'' soient continues, et n par l'interpolation exacte.

Méthode générale : paramétrization 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. $d_i = \left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}\right), i \in \{2, \dots, n-1\}$
3. R la matrice tridiagonale **carré, indexée par $i = 1, \dots, n-2$** :

$$R = \begin{pmatrix} (h_1 + h_2)/3 & h_2/6 & 0 & \dots & \dots & \dots \\ h_2/6 & (h_2 + h_3)/3 & h_3/6 & 0 & \dots & \dots \\ 0 & h_3/6 & (h_3 + h_4)/3 & h_4/6 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & h_{n-2}/6 \\ \vdots & \ddots & \ddots & \ddots & h_{n-2}/6 & (h_{n-2} + h_{n-1})/3 \end{pmatrix}$$

définie par $R_{i,i} = (h_{i+1} + h_i)/3, i = 1, \dots, n-2, R_{i,i+1} = R_{i+1,i} = h_{i+1}/6, i = 1, \dots, n-1$

Théorème 2.2 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 $R\vec{\xi}'' = \vec{d} \iff$

$$\frac{h_{i-1}}{6}\xi_{i-1}'' + \frac{h_{i-1} + h_i}{3}\xi_i'' + \frac{h_i}{6}\xi_{i+1}'' = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{\xi_i - \xi_{i-1}}{h_{i-1}}, \quad i = 2, \dots, n-1 \quad (10)$$

e) L'énergie de flexion est :

$$\int_{\mathbb{R}} (\xi''(x))^2 dx = (\xi'')^t R \xi'' = \xi^t K \xi \quad (11)$$

b) La **continuité de ξ en x_{i+1}** détermine la première dérivée en x_i , par : $\xi_{i+1} = \xi_i + h_i \xi'_i + \frac{h_i^2}{2} \xi''_i + \frac{h_i^3}{6} \xi'''_i$

$$\Rightarrow \xi'_i = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{2} \xi''_i - \frac{h_i^2}{6} (\xi''_{i+1} - \xi''_i) \frac{1}{h_i^2}$$

$$\Rightarrow \boxed{\xi'_i = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{6} (\xi''_{i+1} + 2\xi''_i)}$$

c)- g) Exercices. *Indication* : Pour démontrer la formule (11), commencez par $n = 3, 4, 5, \dots$. Montrez d'abord que la formule pour l'intégral de 0 à h du carré d'une droite qui passe par les points $(0, y_1)$ et (h, y_2) est $\frac{h}{3}(y_1^2 + y_2^2 + y_1 y_2)$. ■

Corollaire 2.1 *Il existe une spline cubique naturelle unique qui passe par n'importe quelles $n \geq 2$ points.*

Démonstration : *Indication* : Montrez la positivité de la matrice R , en commençant par $n = 3, 4, 5, \dots$

Exercice 2.1 *Developper 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 $(-1, 0), (0, 2), (1, 0)$.*

La réponse est :

$$\begin{cases} \forall x \in]-\infty .. -1], & \xi(x) = 3(x+1) \\ \forall x \in [-1 .. 0], & \xi(x) = 2 - 3x^2 - x^3 = 3(x+1) - (x+1)^3 \\ \forall x \in [0 .. 1], & \xi(x) = 2 - 3x^2 + x^3 = (x-1)^3 - 3(x-1) \\ \forall x \in [1 .. \infty[, & \xi(x) = -3(x-1) \end{cases}$$

Exercice 2.2 a) *Déterminer la forme de la spline naturelle cubique $\xi(x)$ qui passe par les 3 points $(0, 0), (1, 1), (2, 0)$, en sachant que $\xi(x) \in C^2$ et que la fonction $\xi(x)$ est polynomiale d'ordre au plus 3 sur chaque interval déterminé par les "noeuds".*

b) *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.*

Avec trois points, on a deux morceaux, et il est convenable de choisir comme paramètres les valeurs communs des dérivés en $x_1 : \xi_1, \xi'_1, \xi''_1$, qui coïncident pour les deux morceaux.

Soit $h_i = x_{i+1} - x_i, i \in \{1, \dots, n-1\}$. La troisième dérivée est alors déterminée par le fait que la deuxième dérivée est linéaire :

$$\boxed{\xi'''_i = \frac{(\xi''_{i+1} - \xi''_i)}{h_i}} \text{ Les deux morceaux sont :}$$

$$\begin{aligned} \xi(x) &= \xi_1 + (x - x_1)\xi'_1 + \frac{(x - x_1)^2}{2} \xi''_1 + \frac{(x - x_1)^3}{6} \xi'''_1 \\ \tilde{\xi}(x) &= \xi_1 + (x - x_1)\xi'_1 + \frac{(x - x_1)^2}{2} \xi''_1 + \frac{(x - x_1)^3}{6} (-\xi'''_1) \end{aligned}$$

$$\begin{aligned} \xi_0 &= \xi_1 - \xi'_1 + \frac{1}{3} \xi''_1 \\ \xi_2 &= \xi_1 + \xi'_1 + \frac{1}{3} \xi''_1 \end{aligned}$$

$$\xi_1'' = \frac{3}{2} (\xi_0 - 2\xi_1 + \xi_2), \quad \xi_1' = \xi_1 - \xi_0 + \frac{1}{2} (\xi_0 - 2\xi_1 + \xi_2) = \frac{1}{2} (\xi_2 - \xi_0)$$

En conclusion : **Spline d'interpolation des 3 points**

(0,0) , (1,1) , (2,0)

Calcul des ξ_i'' : $\xi_0'' = 0 ; \xi_2'' = 0 \quad 4\xi_1'' = 6(\xi_2 - 2\xi_1 + \xi_0) = -12 \quad \text{alors} \quad \xi_1'' = -3$

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 :
$$\forall x \in \mathbb{R}, \xi(x) = 1,5x - \frac{x^3}{4} + \frac{(x-1)^3}{2} - \frac{(x-2)^3}{4}$$

Energie de flexion :
$$\int_0^2 (\xi''(x))^2 dx = 6$$

→ 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 :

$$\begin{array}{llll} \xi'(0) = \frac{3}{2} & g'(0) = \frac{\pi}{2} & \xi''(0) = 0 & g''(0) = 0 \\ \xi'(1) = 0 & g'(1) = 0 & \xi''(1) = -3 & g''(1) = -\pi^2/4 \simeq -2,5 \end{array}$$

2.2 Splines d'ajustement

Idée : Passer par les points, c'est peut-être unennecessary, si les points sont déjà incertaines ; 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 enleve ainsi du "bruit".

Ce but peut être atteint par la minimisation des objectives ³ :

$$\mu \min_{\xi(x) \in C^{2m-2}} \sum_i (y_i - \xi(x_i))^2 + \lambda \int_{\mathbb{R}} (\xi^{(m)}(x))^2 dx \quad (13)$$

Définition 2.2 ξ est dite spline d'ajustement (ou de lissage) d'ordre $2m-1$ des points $(x_i, y_i)_{i \in [1:m]}$, et de poids λ .

L'objectif (13) 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 (13) 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", i.e.

³En 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 evolution temporelle, les dernieres points sont les plus importants aux niveau de l'extrapolation vers le futur).

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

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

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

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

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

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

et la solution est la **droite des moindres carrés** ajustant $n > 2$ points.

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

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

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

et la solution est la **spline naturelle d'interpolation exacte** d'ordre $2m - 1$.

Théorème 2.3 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 $\xi \in C^{2m-2}(\mathbb{R})$ minimisant la somme des carrés ajustés (13).

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 $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n) = (I + \lambda K)^{-1} \mathbf{y}$. 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 (11).

2.3 Exercices

- Par analogie avec les splines cubiques, formulez et résolvez les cas limites $\lambda \rightarrow \infty, \mu \rightarrow \infty$ de l'interpolation ajusté d'ordre $m = 1$.
- 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 : $\xi_{i+1} =$
 $\xi_i + h_i \xi_i' + \frac{h_i^2}{2} \xi_i'' + \frac{h_i^3}{6} \xi_i''' \Rightarrow \xi_i' = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{2} \xi_i'' - \frac{h_i^2}{6} (\xi_{i+1}'' - \xi_i'') \frac{1}{h_i^2}$

$$\Rightarrow \xi'_i = \frac{\xi_{i+1} - \xi_i}{h_i} - \frac{h_i}{6} (\xi''_{i+1} + 2\xi''_i)$$

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''' .

3. Les deux exercices suivantes, des **représentations comme interpolations de noyaux/RBF**, 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 écrite comme :

$$\forall x \in \mathbb{R}, \xi(x) = ax + b + (1/2) \sum_{i \in [1:n]} \alpha_i |x - x_i|^3, \quad \text{ou bien } \xi(x) = ax + b + \sum_{i \in [1:n]} \alpha_i (x - x_i)_+^3$$

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 $x_i = x_0 + i$.

4. 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 ?
5. Déterminer la spline cubique s d'interpolation des 4 points $(0,0)$, $(1,1)$, $(2,0)$, $(3,-3/2)$.
6. 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 4 et 5 ci-dessus, $\xi - s$ est la spline d'interpolation des points $(0,0)$; $(1,0)$; $(2,0)$; $(3, \frac{3}{2})$

7. **Comparaison de divers interpolants.** On considère plusieurs interpolants des 4 points $(0,0)$, $(1,0)$, $(2,0)$, $(3,3/2)$:

- (a) 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.
- (b) Le polynôme p d'interpolation de degré le plus bas possible, interpolant ces quatre points. De quel degré est ce polynôme ?
- (c) 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 $f(3) = 3/2$.
- (d) 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 $f''(3) = 0$.
- (e) 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 $\int_0^3 (w''(x))^2 dx$ partir de ces résultats ? Eventuellement tracez ces courbes sur ordinateur.

Eventuellement calculez $\int_0^3 (w''(x))^2 dx$.

8. 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 .

9. Obtenez la formule du filtre correspondant à une spline d'ajustement correspondant à un opérateur différentiel arbitraire à coefficients constants.

Solutions :

3 a)

$$\left\{ \begin{array}{l} \xi(x) = a x + b + \sum_{i \in [1: n]} \alpha_i (x - x_i)_+^3 \\ \sum_{i \in [1: n]} \alpha_i = \sum_{i \in [1: n]} \alpha_i x_i = 0 \\ \forall j \in [1 : n], \xi(x_j) = y_j \end{array} \right\} \Rightarrow$$

$$\left\{ \begin{array}{l} \forall j \in [1 : n], y_j = a x_j + b + \sum_{i \in [1: n]} \alpha_i (x_j - x_i)_+^3 \\ \sum_{i \in [1: n]} \alpha_i = \sum_{i \in [1: n]} \alpha_i x_i = 0 \end{array} \right.$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \dots & \dots & \dots & 0 & 0 & 1 \\ 1^3 & 0 & 0 & 0 & \dots & \dots & \dots & 0 & 1 & 1 \\ 2^3 & 1^3 & 0 & 0 & \dots & \dots & \dots & 0 & 2 & 1 \\ 3^3 & 2^3 & 1^3 & 0 & \dots & \dots & \dots & 0 & 3 & 1 \\ \vdots & \vdots & \ddots & \ddots & \ddots & & & \vdots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots & \vdots & \vdots \\ (n-1)^3 & (n-2)^3 & \dots & \dots & 2^3 & 1^3 & 0 & 0 & n-1 & 1 \\ n^3 & (n-1)^3 & \dots & \dots & \dots & 2^3 & 1^3 & 0 & n & 1 \\ 0 & 1 & 2 & \dots & \dots & \dots & n-1 & n & 0 & 0 \\ 1 & 1 & 1 & \dots & \dots & \dots & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \\ a \\ b \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \\ 0 \\ 0 \end{pmatrix}$$

4 Spline d'interpolation des 4 points (0,0) , (1,1) , (2,0) , (3,0)

$$\left. \begin{array}{l} 4s_1'' + s_2'' = -12 \\ s_1'' + 4s_2'' = 6 \end{array} \right\} \Rightarrow \begin{cases} s_1'' = -3,6 \\ s_2'' = 2,4 \end{cases}$$

On en déduit : $\begin{cases} s_0''' = -3,6 & , & s_0' = 1,6 \\ s_1''' = 6 & , & s_1' = -0,2 & , & s_3' = 0,4 \\ s_2''' = -2,4 & , & s_2' = -0,8 \end{cases}$

D'où : $\begin{cases} \forall x \in]-\infty .. 0] , & \xi(x) = 1,6x \\ \forall x \in [0 .. 1] , & \xi(x) = 1,6x - 0,6x^3 \\ \forall x \in [1 .. 2] , & \xi(x) = 1 - 0,2(x-1) - 1,8(x-1)^2 + (x-1)^3 \\ \forall x \in \forall 23 , & \xi(x) = -0,8(x-2) + 1,2(x-2)^2 - 0,4(x-2)^3 \\ \forall x \in \forall 2+\infty , & \xi(x) = 0,4(x-3) \end{cases}$

2.4 (**) Interpolation par splines de degré $2m - 1$, $m \geq 1$

Définition 2.3 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} , i.e. satisfaisant :

$$\xi^{(m)}(x_0) = \dots = \xi^{(2m-1)}(x_0) = 0$$

$$\xi^{(m)}(x_n) = \dots = \xi^{(2m-1)}(x_n) = 0$$

Théorème 2.4 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 2.5 (***) Dans le cas equidistant, la solution de la récurrence (12) est donné, en posant $\lambda = \sqrt{3} - 2 \simeq -0,27$, par :

$$\xi_j'' = 6h^{-2}(1 - \lambda^{2n})^{-1} \sum_{i \in [0:n]} A_{ji} \xi_i, \text{ avec, pour tout } j \in [1 : n-1],$$

$$\forall i \in [1 : j-1], \quad \begin{aligned} A_{j0} &= -\lambda^j (1 - \lambda^{2n-2j}) \\ A_{ji} &= -\sqrt{3} \lambda^{j-i} (1 - \lambda^{2i}) (1 - \lambda^{2n-2j}) \\ A_{jj} &= 1 - \lambda^{2n} - \sqrt{3} (1 - \lambda^{2j}) (1 - \lambda^{2n-2j}) \end{aligned}$$

$$\forall i \in [j+1 : n-1], \quad \begin{aligned} A_{ji} &= -\sqrt{3} \lambda^{i-j} (1 - \lambda^{2j}) (1 - \lambda^{2n-2i}) \\ A_{jn} &= -\lambda^{n-j} (1 - \lambda^{2j}) \end{aligned}$$

Note : Cette "forme explicite" est plutôt d'intérêt théorique, car elle est moins aisée à utiliser que la résolution numérique du système tridiagonal.

3 L'analyse des données spatiales

Retenons du chapitre précédent la désirabilité des predictions lisses, pour decouvrir des "signaux" 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 ponderation λ des splines d'ajustement, on arrive à un point où le bruit disparaît totalement. Mais, pt-être que le bruit peut aider dans la prediction ? Pour arriver pt-être à un choix rationnel du λ ?

3.1 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, i.e. si il semble homogène dans toute la région ; si non, il faudra analyser séparément plusieurs régions ⁴.

En supposant ces manipulations préalables déjà faites, on arrive au dilemme de l'estimation de la covariance/sémivariance – voir prochain chapitre.

⁴Sans homogénéité, il est impossible d'estimer les paramètres globaux qui sont à la base de tous les modèles probabilistes.

3.2 Rudiments de statistique bivariée

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

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

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

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

Exercice 3.1 Montrez que le coefficient de corrélation est nécessairement compris entre -1 et $+1$.

Le coefficient de corrélation indique le degré de dépendance linéaire entre deux variables : - s'il est proche de ± 1 , il indique l'existence d'une relation linéaire entre variables ;

- s'il est proche de 0 , il indique que les deux variables ne sont pas linéairement dépendantes (l'indépendance complète entre les variables correspond à une hypothèse plus forte, soit en terme de probabilités : $f_{XY}(x, y) = f_X(x)f_Y(y)$, ce qui implique que la covariance et le coefficient de corrélation correspondants sont nuls - la réciproque n'étant pas nécessairement vraie).

Les géostatisticiens utilisent également la **sémivariance**

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

d'habitude pour des variables avec la même variance.

Exercice 3.2 Montrez que si $\text{Var}(X, X) = \text{Var}(Y, Y) = \sigma^2$, la covariance et sémivariance satisfont :

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

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

3.3 La distribution gaussienne/normale multivariée

Définition 3.1 Un vecteur colonne $\mathbf{X} = (X_1, \dots, X_I)$ a une distribution gaussienne/normale multivariée $N(\mathbf{m}, \mathbf{C})$, ou \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 $\mathbf{C}^{-1}, \mathbf{m}$).

Proposition 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 3.3 L'espérance et la matrice de covariance de la distribution gaussienne multivariée sont

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

i.e. on a $\text{Cov}(X_i, X_j) = \mathbf{C}_{i,j}$.

L'indépendance de deux v.a. 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 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 $X_1 \dots X_r$).

3.4 Processus/champs gaussiens

Hypothèse A : En géostatistique, on travaille souvent avec des "champs gaussiens" $Y(x)$, $x \in \mathbb{R}^d$.

Définition 3.2 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.

Proposition 3 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 prêtent 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 v.a. 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.

Note : Une difficulté dans la statistique spatiale c'est que les modèles théoriques, les processus stochastiques continus, spécialement en plusieurs dimensions, sont techniquement plus difficile à manipuler.

On ne peut plus considérer "le bruit blanc ponctuel". Au lieu de ça, on étudie le processus "Brownien", qui n'est pas indicé par points, mais par des sous-ensembles, et qui correspond conceptuellement à la somme/intégrale du bruit blanc dans des sous-ensembles.

Cette construction peut-être mise en oeuvre seulement dans le cas discret, quand on prend :

$$B(A) = \sum_{\{i \in A\}} \epsilon_i, \quad \forall A \subset \mathbb{N}$$

Exercice 3.4 Montrez que si ϵ_i est un bruit blanc Gaussien de variance $\sigma^2 = 1$, il suit que :

1. $B(A)$ est une variable aléatoire Gaussienne à distribution $N(0, m(A))$, où $m(A)$ denote la cardinalité de A
2. $B(A_1)$ et $B(A_2)$ sont v.a.i. pour chaque A_1, A_2 disjoints.
3. La fonction $B(A)$ est une mesure signée.

La prediction d'un champs dependra 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

$$\gamma(h, x) := \frac{\text{Var}(Y(x+h) - Y(x))}{2} \tag{14}$$

$$\varphi(h, x) := \mathbb{E}(Y(x+h) - m(x+h))(Y(x) - m(x)) \tag{15}$$

$$\rho(h, x) := \frac{\mathbb{E}(Y(x+h) - m(x+h))(Y(x) - m(x))}{\sigma(x+h)\sigma(x)} \tag{16}$$

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 $\sigma(x+h) = \sigma(x) := \sigma$.

Posons

$$\varphi(r) = \sigma^2 \rho(r),$$

où $\rho(r)$ denote la corrélation

Les fonctions ci-dessus, appelées variogramme, covariance et corrélogramme, expriment la variabilité/similitude entre deux valeurs du champs $Y(y), Y(x)$, 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

$$\gamma(h) = \sigma^2 - \varphi(h) = \sigma^2(1 - \rho(h)) \quad (17)$$

$$\varphi(h) = \gamma(\infty) - \gamma(h) \quad (18)$$

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 .

3.5 Les distributions Gaussiennes conditionnelles, la régression linéaire et l'interpolation

Il s'avère que pour la distribution gaussienne multivariée, les prédicteurs à variance minimale, données par des distributions conditionnelles, sont aussi Gaussiens. Cela explique l'intérêt d'utiliser cette distribution dans la modélisation (quand permis par les testes de normalité).

Révisons le resultat fondamental de la prediction Gaussienne, en vue de l'appliquer a un champ Gaussien $Y(x), x \in \mathbb{R}^d$.

Proposition 4 Soient $\mathbf{X} = (X_1 \dots X_I)$ (un vecteur colonne), et Y des v.a. réelles de moyennes nulles, définies sur (Ω, P) . Supposons la loi jointe de $X_1 \dots X_I, Y$ gaussienne. Posons

$\mathbf{C}_Y = \mathbf{C}_{X,Y} = (\text{Cov}(X_i, Y), i = 1, \dots, I)$ (un vecteur ligne), et soit $\mathbf{C} = \mathbf{C}_{X,X} = (\text{Cov}(X_i, X_j), i, j = 1, \dots, I)$ 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 :

$$\boxed{\hat{Y} = (\mathbf{C}_Y (\mathbf{C}_{XX})^{-1}) \mathbf{X} = \sum_{i=1}^I w_i X_i} \quad \text{Regression} \quad (19)$$

où le vecteur ligne $\mathbf{w} = (w_i, i = 1, \dots, I) = \mathbf{C}_Y (\mathbf{C}_{XX})^{-1}$ vérifie le système d'équations normales de régression :

$$\boxed{\mathbf{w} \mathbf{C}_{XX} = \mathbf{C}_Y} \quad (20)$$

Alternativement

$$\boxed{\hat{Y} = (\mathbf{C}_Y (\mathbf{C}_{XX})^{-1}) \mathbf{X} = \mathbf{C}_Y ((\mathbf{C}_{XX})^{-1}) \mathbf{X} = \sum_{i=1}^I \text{Cov}(X_i, Y) \alpha_i} \quad \text{Regression - Interpolation} \quad (21)$$

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 :

$$\boxed{\mathbf{C}_{X,X} \boldsymbol{\alpha} = \mathbf{X}} \quad \text{Systeme d'interpolation} \quad (22)$$

La variance de l'erreur de prediction de Y est :

$$\hat{\sigma}_{Y_Y}^2 = \mathbb{E}[(Y - \hat{Y})^2] = \sigma_{Y_Y}^2 - \mathbf{C}_{YX} (\mathbf{C}_{XX})^{-1} \mathbf{C}_{XY} = \sigma_Y^2 - \mathbf{w} \mathbf{C}_Y = \sigma_Y^2 - \sum_{i=1}^N w_i \text{Cov}(Y, X_i) \quad (23)$$

Notes :

1. La formule de prediction (19) peut être écrite comme un produit scalaire par rapport à l'inverse de la matrice de covariance $\mathbf{C} = \mathbf{C}_{XX}$:

$$\hat{Y}(x) = \langle \mathbf{C}_Y, \mathbf{X} \rangle_{\mathbf{C}^{-1}}$$

2. Le système d'"interpolation" (22) et le système d'"équations normales" (20) (dû à Kolmogorov et Wiener ⁵) 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 dépend 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).
3. L'extension de la formule (19) 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" $X_i - \mu_i, i = 1, \dots, I$:

$$\hat{Y} = \mu_y + \sum_{i=1}^I w_i (X_i - \mu_i) \quad (24)$$

4. La formule de la variance conditionnelle (23) permet de voir la contribution de chacune des variables dans la réduction de la variance; dans le cas univarié, ça donne : $\sigma_E^2 = \sigma_Y^2(1 - \rho^2)$.

Exemple 3.1 Supposons Y, X_1, X_2 sont des v.a. à espérance 0 et variance 1, tq $\text{Cov}(X_2, X_1) = \varphi$, $\text{Cov}(Y, X_1) = z$ et $\text{Cov}(Y, X_2) = 0$. a) Si $z = \varphi$, quel est le domaine de valeurs permises pour φ ?

b) Trouvez le meilleur prédicteur 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 la moyenne, en utilisant la section 4.2, et trouvez le meilleur prédicteur linéaire de Y en fonction de X_1, X_2 .

d) Quelles sont les domaines de valeurs permises pour φ, z , en general ?

Sol a) S : $|\varphi| \leq 2^{-1/2}$ c) $\hat{X}_0 = \frac{\varphi X_1 - \varphi^2 X_2}{1 - \varphi^2}$ d) $m = \frac{X_1 + X_2}{2}, \hat{X}_0 = m - \frac{\varphi}{1 - \varphi} X_2 = \frac{1}{2(1 - \varphi)} (X_1 + (1 - 2\varphi) X_2)$

⁵Ce 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.

Exercice 3.5 a) Soit $Y = (Y_1, Y_2)$ deux des coordonnées d'un vecteur Z à distribution gaussienne $N(\mathbf{0}, C)$, et soit X le reste. Montrez que la 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 $Y = X_i$ est une seule composante (dans quel cas on dénote les autres composantes par $Z = X_{-i}$), et ensuite quand $Y = (X_i, X_j)$ est un couple (dans quel cas on dénote les autres composantes par $Z = X_{-(i,j)}$), que :

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

3.6 Processus Gaussiens isotropiques

Pour le cas des observations spatiales, on suppose souvent que les covariance sont de la forme

$$\text{Cov}(X_i, X_j) = \varphi(|x_i - x_j|), i = 1, \dots, I \text{ Hypothèse Iso}$$

pour une certaine "fonction radiale" $\varphi(r)$ d'une variable. Dans ce cas, le champs s'appelle isotrope est la formule d'interpolation (21) devient :

$$\hat{Y}(y) = \sum_i \alpha_i \varphi(|y - x_i|) \quad \text{Interpolation RBF} \quad (25)$$

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 coincide avec celle de l'interpolation RBF –voir chapitre suivant.

Choisir la fonction $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ 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 noyaux $C(x, y) = \varphi(|x - y|)$ strictement positive défini. Une question plutôt difficile qui sera discuté dans la section ?? est de caractériser la classe des fonctions $SPD(d)$ donnant naissance à un noyaux strictement positive défini en d dimensions.

3.7 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 champs demoyenné et observé parfaitement, correspondant donc au cas $X_i = Z(x_i)$, 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.

Remarque 3.1 Implementation en GeoR : On aimerait bien 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)".

Les variogrammes ont 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 la quelle 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).

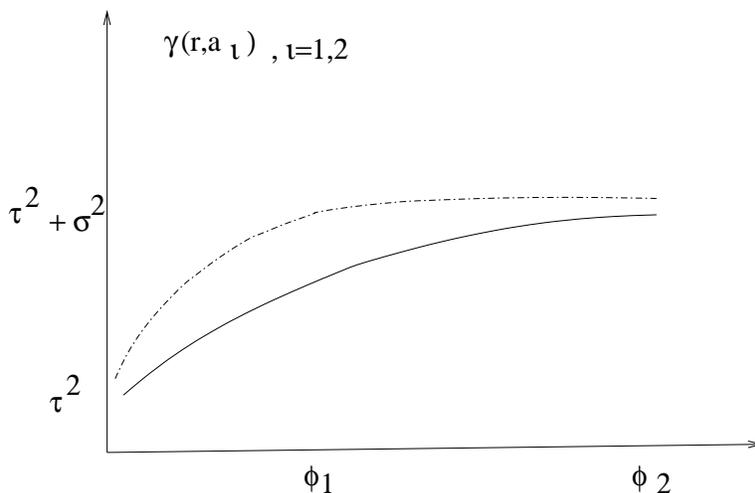


FIG. 1 – Variogrammes dans les directions d'anisotropie principales : $\gamma(r) = \tau^2 I_{\{r>0\}} + \sigma^2(1 - \rho(r)) \iff \varphi(r) = \gamma(\infty) - \gamma(r) = \tau^2 I_{\{r=0\}} + \sigma^2 \rho(r)$

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 di-

rection 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".

3.8 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 = 0.
2. moyennes mobiles, avec corrélation 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 ou de la corrélation 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/Matern, qui est Markovien. Mais, l'identification "à l'oeil" d'une classe convenable en partant de la corrélation 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 5.3). 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**.

- En pratique, pour tenir compte de la portée, pépité 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.

- La corrélation Gaussienne est rarement utilisée, car elle implique un champs "super-lisse" (infiniment différentiable), jamais rencontré en pratique.
- 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.
- Les corrélations "admissibles" sont définies par la propriété de positivité définie – voir section 5.1.
- 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 ??.
- 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 3.6 Supposons un champs déterministe $Y(x) = 4x^2$, 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 $\hat{Y}(x)$ basé sur le noyau Gaussien $\varphi(r) = e^{-r^2}$, i.e. 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 obtenu pour les α , pour une fonction $\varphi(r)$ arbitraire (appelé "système d'interpolation").

c) Calculez l'interpolation $\hat{Y}(x)$ basé sur le noyau $\varphi(r) = r^2$, i.e. 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".

Exercice 3.7 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 $\sup_{x \in [0,1]} |Y(x) - \hat{Y}(x)|$, et l'erreur intégrée $\int_x |Y(x) - \hat{Y}(x)|^2 dx$ 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.

4 Exemples de krigeage

Étant donnée une variable spatiale $Y(x)$ à increments stationnaires et à **moyenne fixe inconnue** m , et quelques valeurs observées

$$X_i = Y(x_i) + \epsilon_i, \quad i = 1, \dots, I$$

le krigeage consiste dans l'estimation de $Y(x), \forall x$. Quand on s'intéresse à une position particulière x , on denotera la variable correspondante par

$$Y = Y(x).$$

Nous présentons d'abord pour échauffement quelques cas particuliers de krigeage/régression spatiale.

4.1 Le krigeage simple

Le cas le plus simple est celui des champs sans tendance, i.e. $m(x) = 0$, appelé krigeage simple. En demandant une interpolation exacte, On utilise ici le système (22) :

$$C \alpha = X$$

qu'on appellera "système de krigeage simple"

Exemple 4.1 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** $R = 1/2, \tau^2 = 1/4, \sigma^2 = 1$.

4.2 Krigeage de la moyenne

Considerons l'estimation d'une moyenne constante. Une première possibilité serait de l'estimer par :

$$\hat{m} = \frac{\sum_{i=1}^N X_i}{N}$$

Mais, avec covariances nonnulles, une approche, plus raisonnable serait d'utiliser une moyenne pondérée :

$$\hat{m} = \sum_{i=1}^N w_i X_i = \mathbf{w} \mathbf{X}$$

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, en arrivant a un problème d'optimisation quadratique contrainte :

$$\begin{aligned} \min_{\mathbf{w}} \text{Var } \hat{m} &= \min_{\mathbf{w}} \mathbf{w} \mathbf{C} \mathbf{w}^t & (26) \\ \sum_{i=1}^N w_i &= 1 \end{aligned}$$

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 :

$$\begin{aligned} \mathbf{C} \mathbf{w} - d \mathbf{1} &= 0 & (27) \\ \mathbf{w} \mathbf{1} &= 1 \end{aligned}$$

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

$$\hat{m} = \mathbf{w} \mathbf{X} = d \mathbf{1}^t \mathbf{C}^{-1} \mathbf{X} = \frac{(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{X})}{(\mathbf{1}^t \mathbf{C}^{-1} \mathbf{1})} = \frac{\langle \mathbf{1}, \mathbf{X} \rangle_{\mathbf{C}^{-1}}}{\langle \mathbf{1}, \mathbf{1} \rangle_{\mathbf{C}^{-1}}} = \frac{\mathbf{v}^t \mathbf{X}}{\mathbf{v}^t \mathbf{1}} \quad (28)$$

où le produit scalaire est défini par rapport à la matrice \mathbf{C}^{-1} , et $\mathbf{v}^t = \mathbf{1}^t \mathbf{C}^{-1}$.

Rémarques : 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) Dans les cas des covariances croisées nulles, la moyenne pondérée (28) devient une moyenne arithmétique simple (en effet, on peut supposer que \mathbf{C} est la matrice identité (le $\sigma^2 + \tau^2$ se simplifie) et on vérifie que la formule donne alors de poids égaux).

4.3 Le krigeage ordinaire

On considère maintenant la prédiction de la variable spatiale $Y(x)$ à increments stationnaires, de moyenne **fixe et inconnue** m . On utilisera un estimateur :

$$\hat{Y}(x) = \sum_{i=1}^N w_i X_i = \mathbf{w} \mathbf{X}$$

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 :

$$\sigma_E^2 = \mathbb{E}(Y(x) - \hat{Y}(x))^2 = \mathbb{E}(Y - \mathbf{w}\mathbf{X})^2 = \sigma^2 + \mathbf{w}\mathbf{C}\mathbf{w}^t - 2\mathbf{w}\mathbf{C}_{X,Y} \quad (29)$$

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

$$\mathbf{C}\mathbf{w}^t - \mu\mathbf{1} = \mathbf{C}_{X,Y} \quad (30)$$

En forme matricielle, on a le "système normal étendu"

$$\begin{pmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w}^t \\ -\mu \end{pmatrix} = \tilde{\mathbf{C}} \begin{pmatrix} \mathbf{w}^t \\ -\mu \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{X,Y} \\ 1 \end{pmatrix}$$

En résolvant,

$$\begin{pmatrix} \mathbf{w}^t \\ -\mu \end{pmatrix} = \tilde{\mathbf{C}}^{-1} \begin{pmatrix} \mathbf{C}_{X,Y} \\ 1 \end{pmatrix}$$

où par le corollaire 4.1, la matrice $\tilde{\mathbf{C}}^{-1}$ est de la forme :

$$\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \mathbf{C}^{-1} - d\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1} & d\mathbf{C}^{-1}\mathbf{1} \\ d\mathbf{1}^t\mathbf{C}^{-1} & -d \end{pmatrix}$$

avec $d = (\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}$, ce qui coïncide avec le multiplicateur de Lagrange dans le krigeage de la moyenne. Donc,

$$\begin{aligned} \mathbf{w} &= (\mathbf{C}^{-1} - d\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1})\mathbf{C}_{X,Y} + d\mathbf{C}^{-1}\mathbf{1} \\ \hat{Y}(x) &= \mathbf{X}^t(\mathbf{C}^{-1}\mathbf{C}_{X,Y} - d\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} + d\mathbf{C}^{-1}\mathbf{1}) \\ &= \mathbf{X}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} - d(\mathbf{X}^t\mathbf{C}^{-1}\mathbf{1})\mathbf{1}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} + d(\mathbf{X}^t\mathbf{C}^{-1}\mathbf{1}) \\ &= \mathbf{X}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} - \hat{m}\mathbf{1}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} + \hat{m} \end{aligned}$$

Finalement, on trouve que

$$\hat{Y}(x) - \hat{m} = (\mathbf{X}^t - \hat{m}\mathbf{1}^t)\mathbf{C}^{-1}\mathbf{C}_{X,Y} = \mathbf{w}(\mathbf{X} - \hat{m}\mathbf{1})$$

et on retrouve la formule classique de régression avec une moyenne zero ou constante, avec la moyenne (inconnue) estimée par (28).

Exercice 4.1 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", i.e. 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 .

Exercice 4.2 Etudier la relation entre les multiplicateurs de Lagrange pour les krigeage ordinaires et de la moyenne.

4.4 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 $\mathbf{C} = (\text{Cov}(X_i, Y(x_j)) : i, j = 1, \dots, I)$. 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.

Lemma 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 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 :

$$\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} & \mathbf{C}^{-1} \mathbf{P} \mathbf{D} \\ \mathbf{D} \mathbf{P}^t \mathbf{C}^{-1} & -\mathbf{D} \end{pmatrix} \quad (31)$$

où $\mathbf{D} = (\mathbf{P}^t \mathbf{C}^{-1} \mathbf{P})^{-1}$.

Exemple : Pour le krigeage ordinaire, $\mathbf{P} = \mathbf{1}$, et

$$\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \mathbf{C}^{-1} - d\mathbf{C}^{-1}\mathbf{1}\mathbf{1}^t\mathbf{C}^{-1} & d\mathbf{C}^{-1}\mathbf{1} \\ d\mathbf{1}^t\mathbf{C}^{-1} & -d \end{pmatrix} \quad (32)$$

où $d = (\mathbf{1}^t\mathbf{C}^{-1}\mathbf{1})^{-1}$.

4.5 Exercices

1. Le champs déterministe $y(x) = x^2$ sera "estimé" à partir des observations en $x_1 = 0, x_2 = 1/2, x_3 = 1$, et en supposant (à tort) un noyau de corrélation sphérique $\rho(x, y) = (1 - \frac{|x-y|}{R})_+$, avec $R = 1$.
 - (a) Donnez la relation entre la corrélation spatiale $\rho(r)$ d'un champs stationnaire et son variogramme $\gamma(r)$.
 - (b) Calculez les semivariances empiriques $\gamma(r_0), \gamma(r_1)$ et $\gamma(r_2)$ (i.e. le variogramme empirique) pour $r_0 = |x - y| = 0, r_1 = |x - y| = 1/2$ et $r_2 = |x - y| = 1$.
 - (c) Formulez le problème de choisir le palier partiel σ^2 et la pépite τ^2 tq le variogramme théorique Gaussien $\tau^2 + \sigma^2(1 - (1 - r)_+) = \tau^2 + \sigma^2 r$ 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 .
Résolvez ce problème dans le cas particulier ci-dessus.
 - (d) Calculez (à partir du noyau) la matrice \mathbf{C} des corrélations théoriques entre les observations, et son inverse.
 - (e) Calculez la prédiction $\hat{y}(x)$ obtenue par le krigeage simple. Quelle est la valeur obtenue pour $x = 2/3$?
 - (f) 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.

- (g) Calculez les valeurs du champs demoyenné $Z(x) = Y(x) - m$, dans les points d'observation.
- (h) Calculez la prédiction $\hat{y}(x)$ obtenue par le krigeage ordinaire, i.e. 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 obtenues en supposant $m(x) = \beta_0 + \beta_1 x$ et $m(x) = \beta_0 + \beta_1 x + \beta_2 x^2$?

Solutions :

- (a) i. Utilisons le variogramme $\gamma(r) = \tau^2 + \sigma^2(1 - \rho(r)) = \tau^2 + \sigma^2 \min[r, 1]$, et les covariances pour la prediction.

ii.

$$\gamma(0) = 0, \gamma(1/2) = (1/16 + 9/16)/4 = 5/32, \gamma(1) = 1/2$$

iii. La regression 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 krigage directement avec la matrice des corrélations (i.e. comme σ^2 se simplifie, on peut le prendre comme 1).

iv.

$$C = \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & 1/2 \\ 0 & 1/2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a & 0 \\ a & 1 & a \\ 0 & a & 1 \end{pmatrix}$$

avec $a = 1/2$.

$$C^{-1} = \begin{pmatrix} \frac{1-a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{a^2}{1-2a^2} \\ -\frac{a}{1-2a^2} & \frac{1}{1-2a^2} & -\frac{a}{1-2a^2} \\ \frac{a^2}{1-2a^2} & -\frac{a}{1-2a^2} & \frac{1-a^2}{1-2a^2} \end{pmatrix} = \begin{pmatrix} 3/2 & -1 & 1/2 \\ -1 & 2 & -1 \\ 1/2 & -1 & 3/2 \end{pmatrix}$$

v. Soit $\alpha = \mathbf{X}C^{-1} = (1/4, -1/2, 5/4)$. Par le krigage 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 lineaire par morceaux interpolant les points. En particulier, pour la prédiction en $x = 2/3$ on trouve $\hat{y}(2/3) = \frac{1}{2}$.

vi. $\mathbf{v}^t = \mathbf{1}^t C^{-1} = (1, 0, 1)$, $\frac{\mathbf{v}^t}{\mathbf{v}^t \mathbf{1}} = (1/2, 1/3, 1/2)$ et la moyenne du champs est

$$m = \frac{\mathbf{v}^t \mathbf{X}}{\mathbf{v}^t \mathbf{1}} = 1/2$$

pendant que la moyenne arithmétique est $5/12$.

vii. le champs demoyenné est $z = (-1/2, -1/4, 1/2)$.

viii. Soit $\alpha = \mathbf{z}C^{-1} = (-1/4, -1/2, 3/4)$. Par le krigage 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 lineaire par morceaux interpolant les points! On obtient encore le même resultat 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.

ix. 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 $\beta = (P^t C^{-1} P)^{-1} (P^t C^{-1} X) = \begin{pmatrix} \frac{16}{17} & \frac{8}{17} \\ \frac{8}{17} & \frac{25}{34} \end{pmatrix}^{-1} \begin{pmatrix} \frac{2}{17} \\ \frac{19}{34} \end{pmatrix} = \begin{pmatrix} -\frac{3}{8} \\ 1 \end{pmatrix}$. La moyenne du champs est

$$m(x) = -\frac{3}{8} + x$$

Le champs demoyenné est $z = (3/8, 1/8, 3/8)$.

Soit $\alpha = z C^{-1} = (-1/2, 1, -1/2)$. Par le krigeage ordinaire, on trouve $\hat{y}(x) = m + \langle \alpha, C_{XY} \rangle = -\frac{3}{8} + x - \frac{1}{2}(1-x^3) + (1 - |x - 1/2|^3) - \frac{1}{2}(1 - (1-x)^3) = \begin{cases} y/4 + y^3, & \text{si } x \leq 1/2 \\ y/4 + y^3 - 2(y - 1/2)^3 & \text{si } x \geq 1/2 \end{cases}$

On a obtenu la spline cubique qui interpole les points! Les valeurs communes des dérivés en $1/2$ sont $f'(1/2) = 1, f''(1/2) = 3$.

5 Interpolation optimale/regression à tendance paramétrique/krigeage universel

5.1 Les noyaux positifs (*)

Définition 5.1 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 $C_{\mathcal{T}} = C(x_i, x_j), i, j = 1, \dots, n$, où $x_i, x_j \in \mathcal{T}, i, j = 1, \dots, n$, sont (strictement) positives définies.

Exemple 5.1 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 verifie que pour n'importe quel n et n'importe quel ensemble de n points \mathcal{T} , la matrice $C_{\mathcal{T}} = Id(n)$ est strictement positive définie.

Corollaire 5.1 Si $C(t_1, t_2)$ est un noyau strictement positif, alors les matrices $C = C(x_i, x_j), i = 1, \dots, n, j = 1, \dots, n$ 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 5.1 Montrez que si $X(x)$ est un champs Gaussien, alors pour n'importe quelles positions $\mathcal{T} = (x_1 \dots x_I)$, la matrice $C_{\mathcal{T}} = (\text{Cov}(X(x_i), X(x_j)), i, j = 1, \dots, I)$ est positive définie, i.e. pour n'importe quel vecteur $w = (w_1, \dots, w_I)$ on a :

$$w^t C w \geq 0$$

L'inverse est aussi vraie :

Proposition 5 *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 noyau positif est simplement la condition de positivité nécessaire pour pouvoir définir un champs Gaussien ayant la covariance donnée par ce noyau, cf Proposition 5. 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.

Lemma 5.1 *Si C_1, C_2 sont des noyaux de covariance, alors les fonctions suivantes sont aussi des noyaux de covariance :*

- (a) $a_1 C_1(x, y) + w_2 C_2(x, y) \quad \forall w_1, w_2 \geq 0$
- (b) $C_1(x, y) C_2(x, y)$
- (c) $pol(C_1(x, y))$, où pol est un polynôme arbitraire à coefficients positives.
- (d) $e^{C_1(x, y)}$
- (e) $C_1(\psi(x), \psi(y))$ 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", coincide avec l'espérance $\mathbb{E} \sum_i (w_i X_i Y_i)^2$, 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 $\rho(r)$ 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.

5.2 Exercices

- (a) Montrez que la fonction $C(x, y) = f(x) f(y)$ est un noyau de covariance.
- (b) Montrez que la fonction $C(x, y) = e^{-|x-y|^2}$ est un noyau de covariance.
- (c) 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.
- (d) 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.

5.3 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 5.2 On appellera *interpolation par des noyaux de base, à reproduction polynomiale d'ordre M* , une interpolation de la forme :

$$\hat{Y}(x) := \hat{Z}(x) + m(x) = \sum_{i=1}^I \alpha_i \varphi(x, x_i) + \sum_{j=1}^M \beta_j p_j(x) \quad (33)$$

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

$$\sum_{i=1}^I \alpha_i p_j(x_i) = 0, \quad \forall j = 1, \dots, M \quad (34)$$

Le cas où le noyau $\varphi(x, y) = \rho(|x - y|)$ est une fonction de la distance $|x - y|$ seulement est appelé *interpolation par une fonction radiale de base (RBF)*.

Définition 5.3 Un vecteur $\alpha = (\alpha_1, \dots, \alpha_I)$ satisfaisant les équations (34) 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 :

- (a) L'interpolation "pure" par des fonctions de base radiales, sans une partie polynomiale, est à ordre 0. En géostatistique, elle s'appelle **krigeage simple**.
- (b) L'interpolation incluant une constante est à ordre 1 ; dans ce cas, (34) demande $\sum_i \alpha_i = 0$. En géostatistique, elle s'appelle **krigeage ordinaire**.
- (c) pour ordre 2, dans \mathbb{R}^1 , (34) demande $\sum_i \alpha_i = 0, \sum_i \alpha_i x_i = 0$.
- (d) La condition d'orthogonalité (34) a des justifications différentes dans la modélisation stochastique—voir chapitre 4 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.
- (e) Parfois, d'autres fonctions sont utilisées pour décomposer la tendance, au lieu des polynômes. Ces fonctions devraient donner une bonne approximation de la tendance/dérive globale $m(x)$.
- (f) Les premières équations du système dual de krigeage/interpolation

$$\boxed{\begin{pmatrix} C & P \\ P^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} X \\ \mathbf{0} \end{pmatrix}}$$

expriment dans le cas $\tau = 0$ une interpolation exacte des données.

- (g) 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.
- (h) La formule d'interpolation (33) fournit un cadre général pour une grande classe de méthodes d'interpolation. Notons qu'elle contient :
 - i. des fonctions qui ne dépendent pas des données (i.e. points d'observation), comme les polynômes (pour les statisticiens, leur somme représente la tendance).
 - ii. 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).

Mais, la formule d'interpolation (33) laisse à l'utilisateur le dilemme de choisir une "bonne base" $\varphi(r)$ et une "bonne représentation" de la tendance.

5.4 Le système d'interpolation/krigeage dual

Dans ce chapitre, nous montrons que l'estimation des champs Gaussiennes avec tendance paramétrique, appelé aussi krigeage avec dérive, ramène précisément à la formule d'interpolation (33).

On déterminera l'estimateur "BLUE (Best linéar unbiased)" de la forme

$$\hat{Y}(x) = \sum_{i=1}^I w_i X_i, \forall x$$

pour le modèle additif à dérive paramétrique (7), (6) (les poids w_i dépendront de x et incluront déjà l'estimation de la moyenne $m(x)$).

Théorème 5.1 Soit $Y(x)$ un processus Gaussien, avec espérance $m(x)$ appartenant à un espace de dimension finie, engendré par les fonctions $p_1(x) = 1, p_2(x), \dots, p_M(x)$, i.e.

$$m(x) = \sum_{j=1}^M \beta_j p_j(x),$$

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

$$X_i = Y(x_i) + \epsilon_i, i = 1, \dots, I$$

les observations. Soit

$$\hat{Y}(x) = \sum_{i=1}^I w_i X_i = \mathbf{X}^t \mathbf{w} \quad (35)$$

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 $\mathbf{C} = (\text{Cov}(X_i, Y(x_j)) + \tau^2 \delta(i - j) : i, j = 1, \dots, I)$ 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 (35) "(BLUE)" qui est nonbiaisé et à variance minimale, donné par :

$$\begin{aligned} \hat{Y}(x) &= \mathbf{X}^t \mathbf{w} = \boxed{(\mathbf{X}^t \quad \mathbf{0}^t) \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix}} = (\boldsymbol{\alpha} \quad \boldsymbol{\beta})^t \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix} \quad (36) \\ &= \sum_{i=1}^I \alpha_i \mathbf{C}_{X,Y}(i) + \sum_{j=1}^M \beta_j p_j(x) \end{aligned}$$

où $\mathbf{C}_{X,Y}$ est un vecteur colonne défini par $(\text{Cov}(X_i, Y(x)) : i = 1, \dots, I)$, $\mathbf{p}(x)$ 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.

$$\boxed{\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix}} \quad (37)$$

(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 (36) :

$$\boxed{\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ -\boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{X,Y} \\ \mathbf{p}(x) \end{pmatrix}} \quad (38)$$

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 :

$$\sigma^2 - \mathbf{C}_{X,Y}^t(\mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{P}\mathbf{D}\mathbf{P}^t\mathbf{C}^{-1})\mathbf{C}_{X,Y} - \mathbf{p}(x)^t\mathbf{D}\mathbf{p}(x) \quad (39)$$

(d) Finalement, l'estimation est

$$\hat{Y}(x) = [(\mathbf{X}^t - \mathbf{m}^t)\mathbf{C}^{-1}\mathbf{C}_{X,Y} + \beta^t\mathbf{p}(x)] = (\mathbf{X}^t - \mathbf{m}^t)\mathbf{C}^{-1}\mathbf{C}_{X,Y} + m(x) \quad (40)$$

où $m(x) = \mathbf{p}(x)^t\boldsymbol{\beta} = \mathbf{p}(x)^t(\mathbf{P}^t\mathbf{C}^{-1}\mathbf{P})^{-1}\mathbf{P}^t\mathbf{C}^{-1}\mathbf{X}$ represente l'estimation de la moyenne en x .

Par rapport au processus avec la moyenne enlevé $Z(x) = Y(x) - m(x)$, la formule devient simplement la formule de régression à bruit corrélé de moyenne 0 :

$$\boxed{\hat{Z}(x) = \mathbf{z}^t\mathbf{C}^{-1}\mathbf{C}_{X,Y} = \langle \mathbf{z}, \mathbf{C}_{X,Y} \rangle_{\mathbf{C}^{-1}}} \quad (41)$$

Notes :

(a) Les formules (40), (41) offrent une généralisation vectorielle du krigeage ordinaire. On calculera d'abord une matrice $\mathbf{v} = \mathbf{P}^t\mathbf{C}^{-1}$, ensuite

$$\boldsymbol{\beta} = (\mathbf{v}\mathbf{P})^{-1}(\mathbf{v}\mathbf{X}),$$

et ensuite on calcule le champs demoyenné, en enlevant la moyenne $m(x) = \mathbf{p}^t(x)\boldsymbol{\beta}$.

(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 cotés doivent être égaux, et on trouve que $\forall j = 1, \dots, M$

$$p_j(x) = \sum_{i=1}^I w_i p_j(x_i) = \mathbf{p}_{(j)}^t \mathbf{w} \quad (42)$$

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 cotés 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 :

$$\mathbf{P}^t \mathbf{w} = \mathbf{p}(x) \quad (43)$$

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 (38) :

$$\mathbf{P}^t \mathbf{w} = \mathbf{p}_x \quad (44)$$

Pour obtenir la première équation du système (38), 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 (38).

1) suit immédiatement, en résolvant $(\mathbf{w} - \mathbf{c})$ du système (38) 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 $\tilde{\mathbf{w}}^t = (\mathbf{X}^t \quad \mathbf{0}^t) \begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}^{-1}$ et $\tilde{\mathbf{w}} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}$ satisfait $\begin{pmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix}$.

Note : Dans la formule de prédiction (36), 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 :

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \|\mathbf{Q} \mathbf{Y} \mathbf{Y}^t \mathbf{Q} - \mathbf{Q} \mathbf{C}_\theta \mathbf{Q}\| \quad \text{avec } \|(w_{i,j})\| := \sum_{i,j} w_{i,j}^2$$

5.5 Interpolation Hermite-Birkhoff

Une légère modification permet dans le contexte déterministe d'interpoler en sachant les valeurs des dérivées de u .

Exemple 5.2 Interpolation Hermite-Birkhoff. *Supposons un champs déterministe $y(x) = 4x^2$, pour le quel on observe la fonction et les premières deux dérivés 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 $\varphi(x, y) = e^{-(x-y)^2}$, i.e. 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 5.1 (*) *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 5.2 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 :

$$\begin{aligned}\alpha_1 &= \frac{1}{1 - 3e^{-2}} = \frac{e^2}{e^2 - 3} \\ \alpha_2 &= \frac{e^{-1}}{1 - 3e^{-2}} = \frac{e}{e^2 - 3} \\ \alpha_3 &= \frac{2e^{-1}}{1 - 3e^{-2}} = \frac{2e}{e^2 - 3}\end{aligned}$$

6 Examens d'entraînement

6.1 Examen 1

1. Le champs 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, i.e. calculez 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 demoyenné $\mathbf{z} = \mathbf{u} - \mathbf{P}\boldsymbol{\beta}$.
 (d) Obtenez la prédiction $\hat{u}(x)$ obtenue par le krigeage avec dérive (en appliquant un krigeage simple au champs demoyenné par la moyenne $m(x)$ obtenu ci-dessus).
 (e) Est-ce que les résultats obtenues pour les dérivées $m(x) = \beta_0$ et $m(x) = \beta_0 + \beta_1x + \beta_2x^2$ sont différents ?
 (f) Quel est l'effet de changer le noyaux à $\varphi(x, y) = \exp(-\frac{x-y}{R})^2$?
 (g) Quel est l'effet de changer la portée R ?
 (h) Quel est l'effet d'ajouter un effet de pepite τ^2 ?
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 fonctionelles :

$$\lambda_1(u) = u'(1), \quad \lambda_2(u) = u'(0), \quad \lambda_3(u) = u(0)$$

- 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 fonctionelles :

$$\lambda_1(u) = u'(1), \quad \lambda_2(u) = u(1), \quad \lambda_3(u) = u(0)$$

3. Supposons $X_0, X_1, X_2, \dots, X_n$ sont des v.a. à variance 1, et $\text{Cov}(X_i, X_j) = a$ si $i \neq j$.
- a) Quel est le domaine de valeurs permmissibles pour ρ ?
 b) Trouvez le meilleur predicteur linéaire de X_0 en fonction de X_1, X_2, \dots, X_n , et la variance de l'erreur de prediction, si :
- i) l'espérance m est 0.
 ii) l'espérance m est constante et **inconnue**. Comparez avec le cas d'une espérance connue
 iii) les espérance est $\mathbb{E}X_i = \beta_i$, où β est **inconnue**.
4. Trouver les coefficients $\mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ et la prediction $\hat{U}(x)$ 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 connues, 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 $\mathbf{P}^t \mathbf{C}^{-1} = \begin{pmatrix} -\frac{4}{17} & \frac{24}{17} & -\frac{4}{17} \\ -\frac{24}{17} & \frac{24}{17} & \frac{13}{17} \end{pmatrix}$ 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

$$m(x) = \frac{3}{2}, \quad \forall x$$

(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 $\tilde{\mathbf{C}} = \begin{pmatrix} \sigma^2 \mathbf{I} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix}$ et donc $\tilde{\mathbf{C}}^{-1} = \begin{pmatrix} \sigma^{-2}(\mathbf{I} - \mathbf{P}(\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t) & \mathbf{P}(\mathbf{P}^t \mathbf{P})^{-1} \\ (\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t & -\tau^2 (\mathbf{P}^t \mathbf{P})^{-1} \end{pmatrix}$

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).