

Université Montpellier II
Université Montpellier I
Montpellier SupAgro

Master Biostatistique

Statistiques spatiales : introduction à la géostatistique

Denis Allard

Table des matières

1	Introduction	3
2	Champs aléatoires stationnaires d'ordre deux	5
2.1	Champs aléatoires	5
2.2	Stationnarité	6
2.3	Fonctions de covariance	7
2.4	Stationnarité d'ordre 2	8
2.5	Ergodicité	10
2.6	Continuité et Différentiabilité	11
2.7	Allure générale d'une fonction de covariance	14
2.8	Représentation spectrale	15
2.9	Exemples de fonctions de covariances	18
2.10	Fonctions de covariance isotropes	21
3	Variogramme – Stationnarité intrinsèque	22
3.1	Le variogramme d'un champ aléatoire stationnaire d'ordre 2	22
3.2	Hypothèses intrinsèques	23
3.3	Fonctions conditionnellement définies négatives	24
3.4	Représentation spectrale	26
3.5	Propriétés spécifiques du variogramme	27
3.6	Les variogrammes puissances	28
3.7	Allure générale d'un variogramme	30
4	Estimation du variogramme	32
4.1	Pourquoi estimer le variogramme ?	32
4.2	Le variogramme expérimental	32
4.3	Estimation paramétrique d'un modèle de variogramme	35
4.4	Estimation paramétrique d'un variogramme anisotrope	37
4.5	Validation d'un modèle de variogramme	39

1 Introduction

En statistiques spatiales, les données sont collectées en des lieux dont on a relevé la position géographique dans le but d'utiliser cette information spatiale dans la modélisation statistique. En particulier, on cherche à modéliser ce que l'expérience courante nous enseigne : deux données proches géographiquement tendent à être similaires en valeur. Cette modélisation nous permettra de réaliser des prédictions spatiales ou de tester des hypothèses en intégrant explicitement cette dépendance spatiale dans les calculs.

On classe habituellement les statistiques spatiales en trois grands domaines, qui correspondent à des types de données distincts. Lorsque les données sont échantillonnées irrégulièrement mais peuvent en principe être mesurées en tout point d'un domaine continu (p.ex. les teneurs en matière organique dans un champ agricole) on utilise le formalisme des champs aléatoires continus ; c'est le domaine habituel de la **géostatistique**. Lorsque, par leur nature même, les données sont liées à un réseau (pour des images ou des données récoltées sur des entités administratives, etc.), il est certes possible d'utiliser le formalisme de la géostatistique, mais celui des **champs de Markov** est souvent plus adapté. Enfin, lorsque ce sont les coordonnées qui portent l'information principale (positions des arbres dans une forêt, points d'impact la foudre dans une région, etc.), on parle de **processus de points**.

Bien entendu, les frontières entre ces trois domaines ne sont pas toujours aussi nettes. Ainsi, par exemple, le formalisme géostatistique apporte des outils pertinents pour l'analyse et la modélisation de l'hétérogénéité spatiale sur des images issues de la télédétection. Lorsque des processus de points portent une variable attachée au point (par exemple, le diamètre d'un tronc d'arbre), le formalisme adopté pour étudier ce type de données utilise à la fois le formalisme des processus de points et le formalisme de la géostatistique.

Le champ des statistiques spatiales est trop vaste pour pouvoir être couvert, même superficiellement, en un cours de 20 heures. A sa parution, le livre *Statistics for Spatial Data* de Noël Cressie (1993, 900p.) couvrait quasiment l'ensemble des connaissances dans ce domaine. Refaire un tel livre serait aujourd'hui impossible.

J'ai choisi de restreindre ce cours à la théorie permettant d'analyser des données irrégulièrement échantillonnées dans un espace continu. Considérons l'exemple suivant : la température est mesurée à Montpellier ; cela nous permet d'avoir une bonne idée de la température à Nîmes, car par expérience nous savons qu'entre ces deux villes proches les températures varient peu. Mais cela nous renseigne nettement moins sur la température à Paris, et pas du tout sur la température à Kiev ou Bucarest. Cela traduit le fait que les températures mesurées en des lieux géographiquement proches sont moins variables que des températures mesurées en des lieux éloi-

gnés. Plus généralement, dans nombre de phénomènes naturels, les corrélations diminuent et la variabilité augmente avec la distance.

La **théorie des champs aléatoires** est l'outil probabiliste permettant de modéliser ce fait que nous expérimentons tous les jours. La **géostatistique** est l'application de cette théorie pour répondre à des questions de prédiction spatiale. L'objectif de ce cours est de fournir une introduction à la théorie des champs aléatoires et à la géostatistique en tant que domaine des statistiques spatiales.

La géostatistique est née dans le contexte très particulier de l'estimation des réserves des gisements miniers. Georges Matheron (1930-2000) en a posé les fondations dans l'ouvrage *Les variables régionalisées et leur estimation* paru en 1965. Il s'agissait d'une contribution pionnière et très originale en statistiques appliquées (je me permets un léger anachronisme, car ce concept, et donc cette expression, n'existaient pas à cette époque) visant à poser un cadre théorique et méthodologique répondant à deux difficultés majeures : *i*) comment définir un cadre statistique cohérent pour étudier un phénomène *sans répétition* ; *ii*) comment prendre en compte les corrélations spatiales existant entre des données récoltées sur un événement unique. Les années 1965 à 1985 ont essentiellement été des années de consolidation et de diffusion auprès des utilisateurs de ces concepts (principalement les géologues, mais également en halieutique, en sciences du sol et en sciences de l'environnement), tout en restant en marge de la communauté statistique. De ce contexte minier est resté un vocabulaire assez spécifique : krigeage, effet de pépite, effet de support, etc. En 1989 est paru *Estimating and Choosing*, un livre de réflexion épistémologique sur la place de la modélisation probabiliste en géoscience. A partir des années 1980, la géostatistique a été re-découverte par les statisticiens, principalement américains. Le livre *Statistics for Spatial Data* de Noël Cressie a marqué l'ancrage définitif de la géostatistique dans le domaine plus vaste des "statistiques spatiales" et a entraîné sa pleine reconnaissance par la communauté statistique. Les recherches les plus récentes dans ce domaine concernent la pratiques des tests, la modélisation spatio-temporelle, la géostatistique multivariée et fonctionnelle, la géostatistique sur les réseaux et la modélisation géostatistique dans les modèles hiérarchiques.

A la suite de cette introduction, ce cours est organisé en 3 chapitres. Le chapitre 2 est consacré à la définition des champs aléatoires stationnaires d'ordre 2 et aux propriétés théoriques des fonctions de covariance, qui caractérisent leur variabilité spatiale. Dans le chapitre 3, on affaiblit les hypothèses de stationnarités et on présente le variogramme, qui est l'outil de référence en géostatistique ; on montre les liens entre variogramme et fonction de covariance. Le chapitre 4 est consacré au calcul du variogramme expérimental à l'estimation des paramètres du variogramme théorique.

2 Champs aléatoires stationnaires d'ordre deux

2.1 Champs aléatoires

Définition 1 Soit (Ω, \mathcal{A}, P) un espace probabilisé et D un domaine de \mathbf{R}^d . Un champ aléatoire (ou une fonction aléatoire) sur D à valeur dans E , est une fonction de deux variables, notée $Z(x; \omega)$, telle que pour chaque $x_0 \in D$, $Z(x_0; \omega)$ est une variable aléatoire sur (Ω, \mathcal{A}, P) et pour chaque $\omega_0 \in \Omega$, $Z(x, \omega_0)$ est une fonction de $D \rightarrow E$.

L'ensemble E est l'ensemble des valeurs de Z . On ne considère ici que le cas $Z(\cdot) \in \mathbf{R}$ ou $Z(\cdot) \in \mathbf{C}$, mais on pourrait aussi envisager d'autres ensembles comme \mathbf{N} (ou une partie de \mathbf{N}), \mathbf{R}^p , etc. De même on pourrait considérer d'autres ensembles supports que \mathbf{R}^d , comme \mathbf{Z}^d , une $d - 1$ variété de \mathbf{R}^d , etc.

Lorsque le contexte le permet, on note un champ aléatoire $Z(\cdot)$ sans référence aux réalisations ω ni à l'espace probabilisé (Ω, \mathcal{A}, P) . On parle généralement de processus stochastique lorsque $d = 1$ et de champ aléatoire lorsque $d \geq 2$.

En géostatistique on distingue la variable étudiée, $z(x)$, que l'on appelle **variable régionalisée** (lettre z minuscule) et sa modélisation par un champ aléatoire $Z(x)$ (lettre Z majuscule). La variable régionalisée $z(x)$ est **une** réalisation du champ aléatoire $Z(x)$.

Un champ aléatoire est totalement caractérisé par les distributions fini-dimensionnelles pour toute valeur de n et tout ensemble de n -uplets $x_i \in \mathbf{R}^d$ $i = 1, \dots, n$:

$$F_{x_1, \dots, x_n}(B_1 \times \dots \times B_n) = P((Z(x_1), \dots, Z(x_n)) \in B_1 \times \dots \times B_n),$$

où B_1, \dots, B_n sont des boréliens de \mathbf{R} .

Les lois de distribution ne peuvent pas être quelconques. Elles doivent vérifier des conditions minimales de cohérence, appelées conditions de Kolmogorov.

Définition 2 (Conditions de Kolmogorov) On appelle conditions de Kolmogorov les deux conditions suivantes.

(i) condition de symétrie : la loi de distribution fini-dimensionnelle doit être constante par permutation :

$$F_{x_1, \dots, x_n}(B_1 \times \dots \times B_n) = F_{x_{i_1}, \dots, x_{i_n}}(B_{i_1} \times \dots \times B_{i_n})$$

où (i_1, \dots, i_n) est une permutation quelconque de $1, \dots, n$.

(ii) condition de cohérence par les marginales :

$$F_{x_1, \dots, x_{n-1}}(B_1 \times \dots \times B_{n-1}) = F_{x_1, \dots, x_n}(B_1 \times \dots \times B_{n-1} \times \mathbf{R}).$$

Si la famille des lois $F_{x_1, \dots, x_{n-1}}$ ne vérifient pas les conditions de Kolmogorov, alors il ne peut pas exister de champ aléatoire sur \mathbf{R}^d ayant les lois $F_{x_1, \dots, x_{n-1}}$ comme lois fini-dimensionnelles.

A l'inverse, si les lois $F_{x_1, \dots, x_{n-1}}$ vérifient les conditions de Kolmogorov, alors il existe un champ aléatoire ayant $F_{x_1, \dots, x_{n-1}}$ comme lois fini-dimensionnelles.

Dans la suite, on considérera que ces conditions sont toujours vérifiées.

2.2 Stationnarité

En statistique classique, l'inférence des paramètres est rendue possible par la répétition indépendante des données. En statistiques spatiales, on observe très souvent une réalisation **unique** des données, par exemple **un** épisode de pollution à l'ozone, **une** région agricole particulière, **une** épidémie végétale, etc... Pour pouvoir réaliser l'inférence statistique pour un événement unique, il faut donc en quelques sortes remplacer l'hypothèse sur les répétitions indépendantes par une hypothèse sur le champ aléatoire qui considère d'une part que certaines de ses caractéristiques sont identiques d'un point à l'autre de l'espace, et d'autre part que l'espérance de certaines grandeurs sont accessibles par des intégrales sur l'espace. On pose donc des hypothèses de **stationnarité** et d'ergodicité.

Définition 3 (Stationnarité stricte) *Un champ aléatoire $Z(\cdot)$ sur \mathbf{R}^d est dit strictement stationnaire si toutes les lois fini-dimensionnelles sont invariantes pour toute translation h , $h \in \mathbf{R}^d$, c'est à dire si on a*

$$F_{x_1, \dots, x_n}(B) = F_{x_1+h, \dots, x_n+h}(B),$$

pour tout n , tous boréliens $B = B_1 \times \dots \times B_n$ de \mathbf{R}^d , tout n -uplet x_1, \dots, x_n et tout vecteur $h \in \mathbf{R}^d$.

Cette hypothèse présente deux défauts majeurs, qui font qu'elle est rarement utilisée : d'une part elle ne dit rien sur l'existence ou non des moments de $Z(\cdot)$, alors que l'espérance et la variance sont des grandeurs nécessaire pour l'analyse statistique de ces données ; d'autre part, cette hypothèse est totalement invérifiable puisqu'elle est faite pour tout n , donc y compris pour des n supérieurs au nombre de données disponibles. On lui préfère l'hypothèse plus faible de stationnarité d'ordre 2, présentée à la section 2.4.

Définition 4 (Champ à l'ordre 2) *Un champ aléatoire est dit être à l'ordre 2 si*

$$Var\{Z(x)\} < \infty, \quad \forall x \in D.$$

Pour un champ à l'ordre 2, la fonction

$$C(x, y) = \text{Cov}\{Z(x), Z(y)\}.$$

s'appelle la fonction de covariance de $Z(\cdot)$. Notons qu'elle est bien définie puisque $C(x, y) < (\text{Var}\{Z(x)\}\text{Var}\{Z(y)\})^{1/2} < \infty$. Enfin, c'est une fonction symétrique et telle que $C(x, x) = \text{Var}\{Z(x)\}$.

2.3 Fonctions de covariance

Définie positivité

Une fonction de covariance $C(x, y)$ d'un champ aléatoire à l'ordre 2, $Z(\cdot)$, doit vérifier :

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j) \geq 0, \quad (1)$$

pour tout n , tout n -uplet (x_1, \dots, x_n) de \mathbf{R}^d et tout n -uplet $(\lambda_1, \dots, \lambda_n)$. Cette propriété découle immédiatement de ce que

$$\text{Var} \left\{ \sum_{i=1}^n \lambda_i Z(x_i) \right\} = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j) \geq 0 \quad (2)$$

doit nécessairement être positif ou nul, puisqu'une variance ne peut être négative. Une fonction qui vérifie la condition (1) est dite **définie positive** sur $\mathbf{R}^d \times \mathbf{R}^d$. En toute rigueur, la condition (1) est une condition de semi-définie positivité, le terme "semi" signifiant qu'on autorise la quantité à être nulle. La définie positivité stricte impose $\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j) > 0$. Afin d'alléger l'écriture de ce cours, je commettrai un léger abus de langage et j'utiliserai le terme de définie positivité pour (1).¹

Nous venons donc de montrer que la condition (1) est une condition **nécessaire** pour qu'une fonction $C(\cdot, \cdot)$ soit une fonction de covariance d'un champ $Z(\cdot)$. C'est également une condition **suffisante dans le cas gaussien**. En effet, considérons que le vecteur $(Z(x_1), \dots, Z(x_n))$ est distribué selon une loi multivariée normale ayant pour espérance le vecteur (μ_1, \dots, μ_n) et pour matrice de covariance la matrice dont l'élément (i, j) est $C(x_i, x_j)$. Alors l'ensemble des lois fini-dimensionnelles définies de cette façon vérifient les conditions de Kolmogorov. D'après des théorèmes classiques en probabilité (Billingsley, 1995 ; Gihman and Skorohod, 1974), il existe donc un champ aléatoire Gaussien admettant ces lois fini-dimensionnelles.

1. Attention, la définie positivité ne doit pas être confondue avec la positivité ! Notons au passage qu'une fonction de covariance peut présenter des valeurs négatives, comme nous le verrons plus loin.

Écriture matricielle

Considérons un champ aléatoire d'ordre 2 possédant une fonction de covariance $C(x, y)$. Soit un vecteur (x_1, \dots, x_n) de points de \mathbf{R}^d et notons \mathbf{C} la matrice $n \times n$ dont l'élément C_{ij} est $C(x_i, x_j)$. En écriture matricielle, (1) s'écrit

$$\text{Var}\{\lambda^t \mathbf{Z}\} = \lambda^t \mathbf{C} \lambda \geq 0,$$

pour tout vecteur $\lambda = (\lambda_1, \dots, \lambda_n)^t$, où λ^t est le vecteur transposé de λ . La matrice \mathbf{C} est donc une matrice semi-définie positive, ce qui entraîne que toutes ses valeurs propres sont non négatives.

Une fonction définie positive $C(\cdot, \cdot)$ est donc une fonction telle que toute matrice construite par $C_{ij} = C(x_i, x_j)$, à partir d'un vecteur (x_1, \dots, x_n) de points de \mathbf{R}^d est une matrice définie positive.

Propriétés élémentaires des fonctions définies positives

1. Soient C_1 et C_2 deux fonctions définies positives. Alors $a_1 C_1 + a_2 C_2$ est une fonction définie positive $\forall a_1, a_2 \geq 0$.
2. Si C_1 et C_2 sont deux fonctions définies positives, alors $C = C_1.C_2$ est une fonction définie positive.
3. Soit C_1, C_2, \dots est une famille de fonctions définies positives, et $\lim_{n \rightarrow \infty} C_n(x, y) = C(x, y)$, $\forall x, y$. Alors $C(x, y)$ est une fonction définie positive.

Les propriétés 1 et 3 se démontrent de façon directe en revenant à la définition (équation (1)). La propriété 2 se démontre par construction : soit $Z_1(\cdot)$ et $Z_2(\cdot)$ deux champs gaussiens indépendants ayant respectivement C_1 et C_2 comme fonctions de covariance. Alors, le champ $Z(x) = Z_1(x).Z_2(x)$ a comme fonction de covariance $C(x, y) = C_1(x, y).C_2(x, y)$.

Ces propriétés sont fort utiles pour construire des fonctions définies positives.

2.4 Stationnarité d'ordre 2

Définition 5 (Stationnarité d'ordre 2) *Un champ aléatoire à l'ordre 2, $Z(\cdot)$, sur \mathbf{R}^d est dit stationnaire d'ordre 2 si les deux premiers moments existent et sont invariants par translation² :*

$$E[Z(x)] = E[Z(x + \tau)] \tag{3}$$

$$\text{Cov}(Z(x), Z(y)) = C(x, y) = \text{Cov}(Z(x + \tau), Z(y + \tau)), \tag{4}$$

2. Certains auteurs préfèrent utiliser le terme de **stationnarité faible**, qui lui est équivalent.

quelques soient x, y et $\tau \in \mathbf{R}^d$.

Notons que la première condition est équivalente à dire que l'espérance est constante et que la seconde condition est équivalente à dire que la covariance entre $Z(x)$ et $Z(y)$ ne dépend que du vecteur $x - y = h$. Dans ce cas, la fonction de covariance ne dépend plus que d'un seul argument. Au prix d'un léger abus de langage, on écrit alors que la fonction $C(h), h \in \mathbf{R}^d$, est la fonction de covariance du champ aléatoire $Z(\cdot)$. La stationnarité d'ordre deux se reformule donc de la façon suivante :

Définition 6 (Stationnarité d'ordre 2 : définition usuelle) *Un champ aléatoire $Z(\cdot)$ sur \mathbf{R}^d est dit stationnaire d'ordre 2 si les deux premiers moments existent et vérifient :*

$$E[Z(x)] = \mu \quad (5)$$

$$\text{Cov}(Z(x), Z(x+h)) = C(h), \quad (6)$$

quelques soient x et $h \in \mathbf{R}^d$.

Propriétés d'une fonction de covariance d'un champ stationnaire d'ordre 2

Soit $Z(\cdot)$ un champ aléatoire **stationnaire d'ordre 2**. Sa fonction de covariance, $C(h)$, possède les propriétés immédiates suivantes :

i) $C(0) = \text{Var}(Z(x)) > 0, \forall x \in \mathbf{R}^d$.

Notons qu'on suppose ici que la variance de $Z(x)$ est strictement positive. En effet, une variance nulle est un cas limite dégénéré (champ aléatoire constant sur \mathbf{R}^d), qui ne présente aucun intérêt ici.

ii) $C(-h) = C(h)$: la fonction de covariance est une fonction paire.

iii) $|C(h)| \leq C(0)$ Notons qu'une fonction de covariance peut prendre des valeurs négatives.

iv) $\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j) \geq 0$.

v) Si $C_\theta(h)$ est une fonction définie positive pour toute valeur de $\theta \in \Theta$. Alors

$$\int_{\Theta} C_\theta(h) \mu(d\theta)$$

est définie positive si μ est une mesure de masse finie telle que $\int_{\Theta} C_\theta(0) \mu(d\theta) < \infty$.

Les deux premières propriétés sont évidentes, et leur démonstration est laissée à titre d'exercice. La troisième propriété est une conséquence de l'inégalité :

$$|\text{Cov}(Z(x), Z(x+h))| \leq \{\text{Var}(Z(x))\text{Var}(Z(x+h))\}^{1/2} \quad (7)$$

$$\Rightarrow |C(h)| \leq C(0). \quad (8)$$

La quatrième n'est qu'un rappel. La propriété v) se démontre à l'aide des propriétés 2 et 3 de la subsection précédente.

2.5 Ergodicité

L'ergodicité est un concept en relation avec la capacité de pouvoir en inférer les paramètres d'un champ aléatoire $Z(\cdot)$ à partir d'une seule réalisation. Il est donc nécessaire, dans ce paragraphe, d'utiliser la notation complète $Z(x; \omega)$ afin de pouvoir raisonner à réalisation fixée.

L'ergodicité est une propriété qui renforce la notion de stationnarité et qui permet d'obtenir la convergence presque sûre d'une moyenne empirique spatiale (d'une réalisation unique) lorsque le domaine "tend vers l'infini".

Définition 7 *Un champ aléatoire $Z(x; \omega)$ sur \mathbf{R}^d est dite **ergodique pour la moyenne** (ou **ergodique au premier ordre**) si la moyenne spatiale sur un domaine D de $Z(x; \omega)$ converge vers son espérance mathématique lorsque $D \rightarrow \mathbf{R}^d$.*

$$\bar{Z}_D = \frac{1}{|D|} \int_D Z(x; \omega) dx \xrightarrow{D \rightarrow \mathbf{R}^d} E[Z(x; \omega)] = \mu, \quad (9)$$

où $|D|$ désigne le d -volumue de D (sa mesure de Lebesgue) et où la convergence s'entend en moyenne quadratique.

Un contre-exemple trivial est le champ aléatoire $Z(x, \omega) = A(\omega)$. Il s'agit d'un champ spatialement constant, dont la valeur est une variable aléatoire. Une réalisation unique d'un tel champ ne permet pas d'inférer la loi de la variable aléatoire A : en effet, il est clair que la moyenne spatiale sur un domaine D sera toujours égale à $A(\omega)$, et donc ne convergera pas vers l'espérance.

L'ergodicité est un concept lié à celui de la mélangeance, concept familier de la théorie des processus. Une condition suffisante qui assure l'ergodicité est le théorème de Slutsky.

Théorème 1 (Théorème de Slutsky) *Soit $Z(x)$ un champ aléatoire stationnaire d'ordre 2, de fonction de covariance $C(h)$. Si*

$$\frac{1}{|D|} \int_D C(h) dh \xrightarrow{D \rightarrow \mathbf{R}^d} 0,$$

alors $Z(x)$ est ergodique pour la moyenne.

Le démonstration est directe en observant que

$$\text{Var}\{\bar{Z}_D\} = \frac{1}{|D|^2} \int_D \int_D C(x - x') dx dx' = \frac{1}{|D|^2} \int_D \int_{D_x} C(h) dh dx$$

où D_x est le domaine dans lequel peut varier h , tel que $x + h$ est dans D .

Notons que si $\frac{1}{|D|} \int_D C(h)dh \rightarrow_{D \rightarrow \mathbf{R}^d} 0$, alors quelque soit $\epsilon > 0$, il existe une distance a telle que lorsque $\|h\| \geq a$, $C(h) < \epsilon$. Dans ce cas, lorsque D est suffisamment grand,

$$\int_{D_x} C(h)dh = \int_{\mathbf{R}^2} C(h)dh - \int_{\mathbf{R}^2 \setminus D_x} C(h)dh = A - \int_{\mathbf{R}^2 \setminus D_x} C(h)dh,$$

et

$$\frac{1}{|D|^2} \int_D \int_{D_x} C(h)dhdx = \frac{A}{|D|} - \frac{1}{|D|} \int_{\mathbf{R}^2 \setminus D_x} C(h)dh.$$

Lorsque $D \rightarrow \mathbf{R}^2$, le dernier terme tend vers 0. Ainsi lorsque la condition de Slutsky est vérifiée, il est immédiat de constater que $\text{Var}\{\bar{Z}_D\} \rightarrow 0$ lorsque $D \rightarrow \mathbf{R}^d$. L'application du théorème de Chebychev mène au résultat. \square

Il existe un théorème analogue pour la covariance. Une fonction aléatoire est ergodique en covariance (ou ergodique d'ordre 2) si

$$\lim_{D \rightarrow \mathbf{R}^d} \frac{1}{|D|} \int_D C(h)^2 dh = 0.$$

2.6 Continuité et Différentiabilité

Continuité

Dans ce paragraphe, on établit les relations entre les propriétés de continuité et différentiabilité d'un champ aléatoire et des caractéristiques de la fonction de covariance. On débute par un rappel.

Rappel : Soit X_1, X_2, \dots une famille de variable aléatoire. Si $E[(X_n - X)^2] \rightarrow 0$ lorsque $n \rightarrow \infty$ et $\text{Var}(X) < \infty$, on dit que la famille $\{X_n\}$ converge en moyenne quadratique vers X .

Définition 8 *Un champ aléatoire stationnaire d'ordre 2, $Z(\cdot)$ sur \mathbf{R}^d est dit continu en moyenne quadratique si*

$$\lim_{y \rightarrow x} E \left[\{Z(y) - Z(x)\}^2 \right] = 0.$$

Il est important de rappeler que la continuité en moyenne quadratique n'entraîne pas la continuité presque sûre des réalisations. Les réalisations d'un champ aléatoire continu en moyenne quadratique peuvent donc présenter des discontinuités, mais celles-ci devront être "rares", c'est à dire que leur support devra être un ensemble de mesure nulle sur Ω et donc, par ergodicité, également de mesure nulle sur D .

Théorème 2 *Un champ aléatoire stationnaire d'ordre 2, $Z(\cdot)$, de covariance $C(h)$ est continu en moyenne quadratique si et seulement si sa fonction de covariance est continue à l'origine.*

Preuve : En effet,

$$\begin{aligned} E \left[\{Z(y) - Z(x)\}^2 \right] &= \text{Var}(Z(y)) + \text{Var}(Z(x)) - 2\text{Cov}(Z(x), Z(y)) \\ &= 2\{C(0) - C(y - x)\}. \end{aligned}$$

Donc, en posant $h = y - x$,

$$\lim_{y \rightarrow x} E \left[\{Z(y) - Z(x)\}^2 \right] = 0 \iff \lim_{h \rightarrow 0} C(h) = C(0).$$

□

Pour un champ aléatoire qui n'est pas stationnaire d'ordre deux, la condition de continuité en moyenne quadratique devient : $C(x, y)$ est continu sur la diagonale : $\lim_{y \rightarrow x} C(x, y) = C(x, x)$.

Proposition 1 *Soit un champ aléatoire stationnaire d'ordre 2 de covariance $C(h)$. Si $C(h)$ est continu à l'origine, elle est continue partout.*

Preuve : En effet,

$$\begin{aligned} |C(h) - C(h')| &= |\text{Cov}\{Z(h), Z(0)\} - \text{Cov}\{Z(h'), Z(0)\}| \\ &= |\text{Cov}\{Z(h) - Z(h'), Z(0)\}| \\ &\leq [\text{Var}\{Z(h) - Z(h')\} \text{Var}\{Z(0)\}]^{1/2} \\ &= [\{2C(0) - 2C(h - h')\}C(0)]^{1/2} \end{aligned}$$

La dernière égalité provient d'un développement similaire à celui rencontré dans la preuve du théorème précédent. Ainsi donc,

$$\lim_{h' \rightarrow h} |C(h) - C(h')| \leq \lim_{\epsilon \rightarrow 0} [\{2C(0) - 2C(\epsilon)\}C(0)]^{1/2},$$

on posant $\epsilon = h' - h$. La continuité en tout h est donc contrôlée par la continuité à l'origine. □

Corrolaire : *Une conséquence importante de ce théorème est qu'une fonction de covariance ne peut-être discontinue qu'à l'origine, et non ailleurs.*

Preuve : En effet, si la fonction de covariance est discontinue à l'origine, on la décompose en la somme d'une distribution de Dirac à l'origine, et d'une fonction continue :

$$C(h) = c\delta_0(h) + C_a(h).$$

Comme $C_a(h)$ est continue à l'origine, elle est continue partout et $C(h)$ n'est donc discontinue qu'à l'origine. □

Différentiabilité

Soit $Z(x)$ un champ aléatoire sur \mathbf{R}^d . Introduisons le champ aléatoire

$$Z_{\epsilon e}(x) = \frac{Z(x + \epsilon e) - Z(x)}{\epsilon},$$

où e est un vecteur unité de \mathbf{R}^d . Considérons une famille de valeur ϵ_n , avec $\epsilon_n \rightarrow 0$ lorsque $n \rightarrow \infty$.

On dit que $Z(x)$ est différentiable en moyenne quadratique si la famille de champs $Z_{\epsilon_n e}(x)$ converge vers une limite notée $Z'_e(x)$, avec $\text{Var}\{Z'_e(x)\} < \infty$. Dans ce cas, le champ

$$Z'_e(x) = \lim_{\epsilon \rightarrow 0} Z_{\epsilon e}(x) \tag{10}$$

est le champ dérivé de $Z(x)$ dans la direction e .

Calculons la fonction de covariance de $Z_{\epsilon e}(x)$:

$$\begin{aligned} C_{\epsilon e}(h) &= \text{Cov}\{Z_{\epsilon e}(x+h), Z_{\epsilon e}(x)\} \\ &= \frac{1}{\epsilon^2} \text{Cov}\{Z(x+h+\epsilon e) - Z(x+h), Z(x+\epsilon e) - Z(x)\} \\ &= \frac{1}{\epsilon^2} \{2C(h) - C(h+\epsilon e) - C(h-\epsilon e)\}. \end{aligned}$$

Le champ aléatoire $Z(\cdot)$ admet une dérivée si et seulement si la limite (10) existe, avec $\text{Var}\{Z'_e(x)\} < \infty$. On voit ainsi que $Z(\cdot)$ est dérivable ssi la covariance $C(h)$ admet une dérivée seconde dans la direction e , c'est-à-dire si

$$\lim_{\epsilon \rightarrow 0} C_{\epsilon e}(h) = -C''_e(h),$$

pour $h = 0$. En effet, lorsque cette condition est remplie, la dérivée seconde existe également pour tout h . On a ainsi établi la proposition suivante :

Proposition 2 (Différentiabilité en moyenne quadratique) *Un champ aléatoire stationnaire d'ordre 2 est différentiable en moyenne quadratique si et seulement si $C''_e(0)$ existe et est finie pour toute direction e ; dans ce cas, $-C''_e(h)$ est la fonction de covariance du champ dérivé dans la direction e .*

Un champ aléatoire stationnaire d'ordre 2 différentiable en moyenne quadratique dans toutes les directions e est dit différentiable en moyenne quadratique.

Ce résultat peut se généraliser : un champ aléatoire stationnaire d'ordre 2 est p fois **différentiable en moyenne quadratique** ssi $C_e^{(2p)}(0)$ existe et est finie pour toute direction e ; dans ce cas, $(-1)^p C_e^{(2p)}(h)$ est la fonction de covariance du champ dérivé p fois dans la direction e .

2.7 Allure générale d'une fonction de covariance

Nous avons vu dans les sections précédentes les caractéristiques les plus importantes des fonctions de covariance, $C(h)$.

- La valeur en de la fonction de covariance pour $h = 0$ est égale à la variance du champ aléatoire : $C(0) = \text{Var}\{Z(\cdot)\} = \sigma^2$.
- La continuité en $h = 0$ de $C(h)$ est une quantité liée à la régularité du champ aléatoire. La fonction de covariance est continue à l'origine ou non ; elle est continue partout ailleurs. Si elle est discontinue à l'origine, le saut de discontinuité correspond à la variance d'un champ aléatoire $Z_0(\cdot)$, tel que $Z_0(x)$ et $Z_0(y)$ sont non corrélés lorsque $x \neq y$. On peut alors écrire

$$Z(x) = Z_0(x) + Z_1(x),$$

où $Z_0(x)$ et $Z_1(x)$ sont deux champs indépendants tels que

$$\begin{aligned} \text{Cov}\{Z(x), Z(x+h)\} &= \text{Cov}\{Z_0(x) + Z_1(x), Z_0(x+h) + Z_1(x+h)\} \\ &= \text{Cov}\{Z_0(x), Z_0(x+h)\} + \text{Cov}\{Z_1(x), Z_1(x+h)\} \\ &= \sigma_0^2 \delta_0(h) + C_1(h), \end{aligned}$$

où $\delta_0(h) = 1$ lorsque $h = 0$ et 0 sinon, et où $C_1(h)$ est une fonction de covariance continue. Le saut de discontinuité est appelé **effet de pépité** en géostatistique.

- La condition de Slutsky vue plus haut impose que $C(h) \rightarrow 0$ lorsque $\|h\| \rightarrow \infty$. Deux cas peuvent se produire.
 1. Soit il existe une distance $a > 0$, telle que lorsque $\|h\| \geq a$, $C(h) = 0$. La plus petite valeur de a qui vérifie cette condition s'appelle la **portée** de la fonction de covariance.
 2. Soit il n'existe pas une telle distance parce que $C(h)$ tend asymptotiquement vers 0. Dans ce cas, par analogie, on appelle **portée pratique** de $C(h)$ la plus petite valeur a qui est telle que $C(h) \leq 0.05C(0)$ lorsque $\|h\| \geq a$. Ainsi par exemple, pour une covariance exponentielle $C(h) = \exp\{-\|h\|/a\}$, on considère généralement que la portée pratique vaut $3a$, car $e^{-3} = 0.0498$.

La figure 2.7 représente à gauche une fonction de covariance sans effet de pépité, avec une variance égale à 2, une portée $a = 10$ et à droite une fonction de covariance avec effet de pépité (égal à 0.5) avec une portée pratique $a = 10$.

FIGURE 1 – Deux exemples de fonctions de covariance.

2.8 Représentation spectrale

Théorème de Bochner

Pour introduire les représentations spectrales des champs aléatoires, il est plus simple de travailler sur des champ complexes $Z(x) = U(x) + iV(x)$ où $U(x)$ et $V(x)$ sont des champs aléatoires réels sur \mathbf{R}^d . Pour se ramener au cas réel, il suffit de poser $V(x) = 0$.

Soit $U(x)$ et $V(x)$ sont des champs stationnaires d'ordre 2 et soit $\text{Cov}\{U(x), V(x)\} = g(x-y)$. Dans ce cas, $Z(x)$ est champ aléatoire complexe stationnaire d'ordre 2.

On définit la fonction de covariance de $Z(x) : K(h) : \mathbf{R}^d \rightarrow \mathbf{C}$:

$$\begin{aligned} K(h) &= \text{Cov}\{Z(x+h), \overline{Z(x)}\} \\ &= \text{Cov}\{U(x+h), U(x)\} + \text{Cov}\{V(x+h), V(x)\} \\ &\quad + i[\text{cov}\{V(x+h), U(x)\} - \text{Cov}\{U(x+h), V(x)\}], \end{aligned}$$

pour laquelle on a $K(-h) = \overline{K(h)}$. Cette fonction doit être définie positive, c'est à dire que l'on a pour tout n et tout ensemble de pondérateurs complexes $\lambda_1, \dots, \lambda_n$:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \overline{\lambda_j} K(x_i, x_j) \geq 0.$$

Théorème 3 (Théorème de Bochner) *Une fonction complexe $K(h)$ sur \mathbf{R}^d est la fonction de covariance d'un champ aléatoire stationnaire d'ordre 2 complexe continu en moyenne quadratique sur \mathbf{R}^d ssi*

$$K(h) = \int_{\mathbf{R}^d} \exp\{i \langle \omega, h \rangle\} F(d\omega) \quad (11)$$

où $\langle a, b \rangle$ désigne le produits scalaire de a et $b \in \mathbf{R}^d$, et où $F(d\omega)$ est une mesure positive finie sur \mathbf{R}^d , c'est-à-dire telle que $\int_{\mathbf{R}^d} F(d\omega) = K(0) < \infty$.

Eléments de preuve : Considérons une famille de valeurs $\omega_1, \omega_2, \dots$ dans \mathbf{R}^d et une famille de variables aléatoires Z_1, Z_2, \dots complexes, i.i.d., et d'espérance nulle (i.e. $E[Z_j] = 0$ et $E[Z_j \overline{Z_k}] = 0, \forall j$ et $\forall k \neq j$). A chaque valeur de ω_k , qui doit être vue comme une fréquence, est associée une valeur Z_k qui doit être vue comme une amplitude. Notons $f_k = E[Z_k \overline{Z_k}] < \infty$,

l'amplitude de Z_k , et définissons

$$Z(x) = \sum_{k=1}^n Z_k \exp\{i \langle \omega_k, x \rangle\}. \quad (12)$$

$Z(x)$ est un champ aléatoire obtenu comme la somme de fonctions de bases $\exp\{i \langle \omega_k, x \rangle\}$ pondérées par des amplitudes aléatoires Z_k .

Alors, la fonction de covariance de $Z(x)$ vaut :

$$\begin{aligned} K(h) &= \text{Cov}(Z(x+h), \overline{Z(x)}) \\ &= \text{Cov} \left(\sum_{k=1}^n Z_k \exp\{i \langle \omega_k, x+h \rangle\}, \sum_{l=1}^n \overline{Z_l} \exp\{-i \langle \omega_l, x \rangle\} \right) \\ &= \sum_{k=1}^n f_k \exp\{i \langle \omega_k, x+h \rangle\} \exp\{-i \langle \omega_k, x \rangle\} \end{aligned} \quad (13)$$

$$= \sum_{k=1}^n f_k \exp\{i \langle \omega_k, h \rangle\}, \quad (14)$$

ce qui montre que le champ $Z(x)$ ainsi défini est (i) stationnaire d'ordre 2, car sa fonction de covariance ne dépend que de h ; (ii) continu en moyenne quadratique, car les dérivées de sa fonction de covariance sont continues en 0 (par les propriétés de dérivation de la fonction exponentielle).

En considérant les limites des sommes (12) lorsque $n \rightarrow \infty$, on montre que tous les champs stationnaires d'ordre 2, continus en moyenne quadratique peuvent s'écrire comme une somme continue identique à (12) :

$$Z(x) = \int_{\mathbf{R}^d} \exp\{i \langle \omega, x \rangle\} M(d\omega) \quad (15)$$

où $M(d\omega)$ est une mesure complexe, aléatoire, orthogonale sur \mathbf{R}^d , c'est à dire vérifiant $E[M(B)] = 0$, $E[M(B)\overline{M(B)}] = F(B) \geq 0$ et $E[M(B)\overline{M(B')}] = 0$ lorsque $B \cap B' = \emptyset$.

$M(d\omega)$ s'appelle la représentation spectrale du champ faiblement stationnaire $Z(x)$. La covariance du champ défini par (15) vaut alors

$$K(h) = \int_{\mathbf{R}^d} \exp\{i \langle \omega, h \rangle\} F(d\omega),$$

où $F(d\omega)$ s'interprète comme le module de la mesure aléatoire $M(d\omega)$, et doit vérifier $\int_{\mathbf{R}^d} F(d\omega) < \infty$. \square

Pour $h = 0$, on obtient

$$K(0) = \text{Var}\{Z(x)\} = \int_{\mathbf{R}^d} F(d\omega).$$

Si la mesure $F(d\omega)$ possède une densité, on note $F(d\omega) = f(\omega)d\omega$, et

$$K(h) = \int_{\mathbf{R}^d} \exp\{i \langle \omega, h \rangle\} f(\omega) d\omega.$$

On reconnaît la formule de la transformation de Fourier. Ainsi le théorème de Bochner peut se formuler ainsi : *une fonction de covariance d'un champ aléatoire stationnaire d'ordre 2 est la transformée de Fourier d'une mesure positive finie.*

La formule d'inversion des transformées de Fourier nous donne

$$f(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} \exp\{-i \langle \omega, h \rangle\} K(h) dh.$$

Si le champ $Z(x)$ est réel, la fonction de covariance est également réelle et la transformée de Fourier peut s'écrire comme une série de cosinus. Le théorème de Bochner devient alors

$$C(h) = \int_{\mathbf{R}^d} \cos(\langle \omega, h \rangle) F(d\omega).$$

Lien avec la régularité du champ aléatoire

J'établis les résultats de cette section de façon peu rigoureuse, en me limitant à $d = 1$, pour éviter d'alourdir les notations avec les dimensions de l'espace. Pour une présentation plus complète et plus rigoureuse, je renvoie à Yaglom (1986) et Stein (1999).

On part de l'écriture spectrale du champ (15). Supposons que la dérivée du champ existe. Alors

$$Z'(x) = \int_{\mathbf{R}} i\omega \exp\{i \langle \omega, x \rangle\} M(d\omega).$$

La covariance de $Z'(x)$ – si elle existe – sera donc

$$\begin{aligned} \text{Cov}\{Z'(x), \overline{Z'(x+h)}\} &= E[Z'(x) \overline{Z'(x+h)}] \\ &= E \left[\int_{\mathbf{R}} i\omega \exp\{i \langle \omega, x \rangle\} M(d\omega) (-i\omega) \exp\{-i \langle \omega, x+h \rangle\} \overline{M(d\omega)} \right] \end{aligned}$$

qui n'existera que si

$$\int_{\mathbf{R}} \omega^2 E[M(d\omega) \overline{M(d\omega)}] = \int_{\mathbf{R}} \omega^2 F(d\omega) < \infty.$$

Pour résumé, un champ $Z(x)$ est différentiable en moyenne quadratique ssi

$$C''(0) < \infty \text{ existe} \iff \int_{\mathbf{R}^d} \|\omega\|^2 dF(\omega) < \infty.$$

Une condition nécessaire (mais non suffisante) sur la densité spectrale $f(\omega)$ (lorsque celle-ci existe) est que la densité décroisse plus vite que $\|\omega\|^{-2}$. En d'autres termes, les hautes fréquences doivent être peu chargées. Ce résultat mathématique correspond en effet à l'intuition que l'on se fait d'un phénomène suffisamment régulier pour être différentiable : les variations rapides dans l'espace (i.e. les hautes fréquences) ne peuvent pas être trop importantes.

Ce résultat peut s'itérer à l'ordre m : un champ $Z(x)$ est différentiable en moyenne quadratique m fois ssi

$$C^{(2m)}(0) < \infty \text{ existe} \iff \int_{\mathbf{R}^d} \|\omega\|^{2m} dF(\omega) < \infty.$$

2.9 Exemples de fonctions de covariances

1. **Delta de Dirac** : Comme nous l'avons vu à la section 2.7, la fonction $C(h) = \sigma^2 \delta_0(h)$ (où $\delta_0(h) = 1$ lorsque $h = 0$ et $\delta_0(h) = 0$ sinon) correspond à un champ aléatoire sans corrélation spatiale. Utilisée avec d'autres fonctions de covariance, celle-ci sert à modéliser un effet de petite observé sur les données.
2. **Fonction triangulaire sur \mathbf{R}** . Considérons la fonction triangulaire $C(h) = \sigma^2(1 - |h|/a)$, si $|h| < a$, et $C(h) = 0$ sinon. Sa densité spectrale associée est $f(\omega) \propto \sin(a\omega)^2/\omega^2 > 0$, avec $\int_{\mathbf{R}} f(\omega) d\omega = \sigma^2$. En vertu du théorème de Bochner, la fonction triangulaire est donc bien une fonction de covariance.
En revanche, $\int_{\mathbf{R}} \omega^2 f(\omega) d\omega = \int_{\mathbf{R}} \sin(a\omega)^2 d\omega$ n'est pas finie : la fonction triangulaire correspond à un champ qui n'est pas différentiable en moyenne quadratique.
Le même résultat s'obtient en observant le comportement à l'origine de la fonction de covariance. En $h = 0$, la dérivée à gauche (h négatif) est $1/a$, tandis que la dérivée à droite (h positif) est $-1/a$. Il n'y a donc pas de dérivée en $h = 0$.
3. **Fonction conique sur \mathbf{R}^d** La fonction $C(h) = 1 - \|h\|/a$, si $\|h\| < a$, et $C(h) = 0$ sinon, n'est pas définie positive sur \mathbf{R}^d , $d > 1$ car sa transformée de Fourier contient des termes négatifs. Elle ne peut donc pas être une fonction de covariance pour $d > 1$.

Cet exemple illustre le résultat mathématique général suivant. Si une fonction est définie positive sur un ensemble E , elle est également définie positive dans tout ensemble $E' \subset E$. En revanche, elle n'est pas nécessairement définie positive sur un ensemble $F \supset E$ quelconque. En particulier, une fonction définie positive sur \mathbf{R}^d le sera également sur $\mathbf{R}^{d'}$, avec $d' \leq d$. Mais elle ne le sera pas nécessairement sur $\mathbf{R}^{d''}$, avec $d'' > d$.

4. **Fonctions auto-convoluées** Toute fonction $g(h)$ qui serait l'auto-convolution d'une fonction $u(x)$:

$$g(h) = \int_{\mathbf{R}^d} u(x)u(x+h)dx,$$

est une fonction de covariance. En effet, la transformation de Fourier échange convolution (dans l'espace direct) et multiplication (dans l'espace spectral). Notons $\phi(\omega)$ la transformée de Fourier de $u(x)$. Alors, en utilisant un résultat classique de la théorie de Fourier, la transformée de Fourier de $g(h)$ est $\phi^2(\omega)$, et est donc toujours positive. Ainsi, $g(h)$ est une fonction définie positive, donc une fonction de covariance admissible.

5. **Covariance sphérique** Il est possible d'étendre la fonction triangulaire aux dimensions supérieures d'une façon qui la rend définie positive pour toutes les dimensions de l'espace. Pour cela on observe que la fonction triangulaire vue au point 1, est l'auto-convolution de la fonction indicatrice de l'intervalle $I_a = [-a/2, a/2]$: $\mathbf{1}_{[-a/2, a/2]}(x) = 1$ si $x \in I_a$ et 0 sinon. En utilisant le résultat général du point précédent, il est immédiat que la fonction triangulaire est une fonction de covariance admissible.

On généralise ce résultat de la façon suivante. Dans l'espace \mathbf{R}^d , on considère la fonction indicatrice de la sphère à d dimension, de rayon $a/2$, notée S_a^d . Il est évident que $S_a^1 = I_a$. L'auto-convolution de cette fonction indicatrice sera une fonction définie positive sur \mathbf{R}^d . Ainsi par exemple, pour $d = 3$, on obtient la fonction de covariance

$$\begin{aligned} C(h) &= \int_{\mathbf{R}^3} I_{S_a^d}(x)I_{S_a^d}(x+h)dx \\ &= |S_a^d \cap S_a^d \oplus h| \\ &= 1 - 1.5\|h\|/a + 0.5(\|h\|/a)^3, \quad h \leq a, \end{aligned}$$

et $C(h) = 0$ si $h > a$. La seconde ligne est la mesure de Lebesgue (ici, le volume) de l'intersection de la sphère S_a^d avec la même sphère translatée du vecteur h . Cette fonction est définie positive sur \mathbf{R}^d , avec $d \leq 3$. Le même procédé peut s'appliquer à des sphères à $d > 3$ dimensions pour obtenir des fonctions définies positives dans des espaces de plus grande dimension. Notons cependant que les calculs sont beaucoup plus simples sur les espaces de dimensions impaires, ce qui explique pourquoi on ne présente pas le résultat pour $d = 2$.

Cette fonction de covariance est usuellement appelée fonction de covariance sphérique. Elle est souvent utilisée en modélisation géostatistique, car (i) elle possède une valeur a , au-delà de laquelle la covariance est strictement nulle ; (ii) elle possède une interprétation physique simple.

Notons qu'elle est continue à l'origine, sans être différentiable.

6. **Covariances exponentielles** Sur \mathbf{R} , à la densité spectrale $f(\omega) = \phi^2(1/a^2 + \omega^2)^{-1}$, est associée la fonction de covariance $C(h) = \pi\phi a \exp\{|h|/a\}$. De la même façon, à la densité spectrale $f(\omega) = \phi^2(1/a^2 + \omega^2)^{-2}$, est associée la fonction de covariance $C(h) = 0.5\pi\phi a^3 \exp\{-|h|/a\}(1 + |h|/a)$.

Ces résultats se généralisent de deux façons : (i) d'une part, on peut montrer que la transformée de Fourier de la fonction exponentielle $C(h) = \pi\phi a \exp\{-||h||/a\}$ sur \mathbf{R}^d est proportionnelle à $(a^{-2} + \omega^2)^{-(d+1)/2} > 0$. Elle donc définie positive pour toute dimension d . (ii) Les fonctions de type $C(h) = k||h||^\nu \exp\{-||h||/a\}$, ont une transformée de Fourier proportionnelle à $(a^{-2} + \omega^2)^{-(\nu+d/2)} > 0$. Elles sont donc également définies positives pour $\nu > 0$.

Enfin, comme le somme de fonction définies positives est une fonction définie positive, il s'ensuit que les fonctions du type $C(h) = Q_p(||h||) \exp\{-||h||/a\}$, où $Q_p(||h||)$ est un polynôme à coefficients positifs sont des fonctions de covariance sur \mathbf{R}^d , $\forall d$.

7. **Covariance gaussienne** Sur \mathbf{R}^d , la fonction $C(h) = \sigma^2 \exp\{-(||h||/a)^2\}$ est parfois appelée "fonction de covariance gaussienne" en raison de sa forme analytique. Sa transformée de Fourier est $f(\omega) = 0.5a\sigma^2\pi^{-1/2} \exp\{-4a^2||\omega||^2\} \geq 0$, ce qui montre que la fonction $C(h)$ est bien une fonction de covariance.

Comme la TF ne fait pas intervenir la dimension de l'espace, cette fonction est définie positive sur \mathbf{R}^d , $\forall d \geq 1$.

Cette fonction de covariance est indéfiniment dérivable à l'origine, (c'est un petit peu long à établir). Il est plus facile d'observer que $\int_{\mathbf{R}^d} ||\omega||^{2m} \exp\{-4a^2||\omega||^2\} d\omega < \infty$ pour tout m . Le champ aléatoire est donc indéfiniment dérivable en moyenne quadratique.

8. La **classe de Matern** Sur \mathbf{R}^d (Matern, 1960) est une classe de fonctions de covariance fort importante pour la modélisation des variables régionalisées s'obtient à partir d'une généralisation des cas vus au point 6. Partant des densités spectrales

$$f(\omega) = \phi(a^{-2} + ||\omega||^2)^{-\nu-d/2}, \quad \nu, \phi, a > 0,$$

on obtient la classe de Matern de fonctions de covariance

$$C(h) = \frac{\pi^{1/2}\phi a^{2\nu}}{2^{\nu-1}\Gamma(\nu + 1/2)} (||h||/a)^\nu \mathcal{K}_\nu(||h||/a),$$

où \mathcal{K}_ν est la fonction de Bessel modifiée (Abramwitz and Stegun, 1965, pp. 374-379). Lorsque $\nu = 1/2$, on retrouve la fonction exponentielle; lorsque $\nu = 3/2$, on retrouve le second exemple du point 4. Lorsque $\nu \rightarrow \infty$, on retrouve la covariance gaussienne vue au point 3.

Le paramètre ν est essentiellement un paramètre de régularité. En effet, l'intégrale

$$\int_{\mathbf{R}^d} \|\omega\|^{2m} f(\omega) d\omega = \int_{\mathbf{R}^d} \|\omega\|^{2m} \phi(a^{-2} + \|\omega\|^2)^{-\nu-d/2} d\omega$$

est finie si $\nu > m$. En d'autres termes, étant donné un coefficient ν , le champ aléatoire est m fois différentiable en moyenne quadratique, où m est le plus grand entier positif ou nul, inférieur à ν . Si $\nu < 1/2$, le processus n'est pas continu en moyenne quadratique.

Le cas $\nu = 1/2$ correspond à la fonction de covariance exponentielle. Lorsque $\nu = m + 1/2$, on obtient une fonction de covariance de type $Q_m(\|h\|) \exp\{-\|h\|/a\}$. Enfin, lorsque $\nu \rightarrow \infty$ on retrouve la covariance gaussienne.

9. **Classe des covariances stables** Les fonctions $C(h) = \exp\{-(\|h\|/a)^\alpha\}$, avec $a > 0$ et $0 < \alpha \leq 2$ sont définies positives sur \mathbf{R}^d pour toutes les dimensions d . Cela peut se montrer à partir de la classe de Matern, mais les développements sont assez techniques. On peut les trouver dans (Stein, 1999, pp. 32–33).

2.10 Fonctions de covariance isotropes

Une fonction de covariance, $C(h)$, est dite isotrope si elle ne dépend que de la distance $r = \|h\|$ et non de la direction du vecteur h . Par abus de notation, on notera dans ce cas $C(h) = C(\|h\|) = C(r)$.

Il existe un théorème de caractérisation des fonctions de covariances isotropes pour toutes les dimensions d de l'espace dû à Schoenberg (1938). La démonstration de ce théorème se trouve dans (Yaglom, 1987; pp. 348-390).

Théorème 4 (Schoenberg, 1938) *Une fonction $C(r)$ est une fonction de covariance isotrope dans \mathbf{R}^d pour tout d , ssi $C(r)$ est de la forme*

$$C(r) = \int_0^\infty \exp\{-r^2/t^2\} \nu(dt) \tag{16}$$

où ν est une mesure positive bornée sur \mathbf{R}^+ .

Il faut remarquer que (16) génère toutes les formes de régularité à l'origine, y compris les faibles degrés de régularité, car l'opération de randomisation (par la mesure ν) produit des modèles qui peuvent être *moins* réguliers que le modèle de base, mais jamais plus réguliers. Par exemple, randomiser (16) par une fonction Gaussienne définie sur \mathbf{R}^+ , $\nu(dt) = 1/(a\sqrt{\pi}) \exp\{-t^2/4a^2\}$ mène à la fonction exponentielle :

$$\int_0^\infty \exp\{-r^2/t^2\} \frac{1}{a\sqrt{\pi}} \exp\{-t^2/4a^2\} dt = \frac{1}{a\sqrt{\pi}} \int_0^\infty \exp\{-t^2/4a^2 - r^2/t^2\} dt = \exp\{-r/a\},$$

pour $a > 0$. La fonction exponentielle décroissante est donc bien une fonction de covariance sur \mathbf{R}^d , $\forall d$, ce que nous avons déjà vu à la section 2.8.

Enfin, il est intéressant de noter que la fonction exponentielle négative a la propriété d'être complètement monotone³. Une extension du théorème de Schoenberg stipule que toute fonction de covariance complètement monotone peut s'écrire comme une randomisation de la fonction exponentielle :

Théorème 5 *Une fonction $C(r)$ est une fonction de covariance isotrope et complètement monotone dans \mathbf{R}^d pour tout d , ssi $C(r)$ est de la forme*

$$C(r) = \int_0^\infty \exp\{-r/t\} \nu(dt) \quad (17)$$

où ν est une mesure positive bornée sur \mathbf{R}^+ .

3 Variogramme – Stationnarité intrinsèque

3.1 Le variogramme d'un champ aléatoire stationnaire d'ordre 2

Définition 9 (Une première définition) *Soit $Z(x)$ un champ aléatoire stationnaire d'ordre 2 sur \mathbf{R}^d . Le variogramme de $Z(x)$ est la fonction*

$$\gamma(h) = \frac{1}{2} E[(Z(x+h) - Z(x))^2] = \frac{1}{2} \text{Var}\{Z(x+h) - Z(x)\}$$

Propriétés élémentaires

- $\gamma(0) = 0$
- $\gamma(h) \geq 0$
- $\gamma(-h) = \gamma(h)$
- $\gamma(h) = C(0) - C(h)$. En effet,

$$\begin{aligned} \gamma(h) &= 0.5 E[(Z(x+h) - Z(x))^2] \\ &= 0.5 E[Z(x+h)^2 - 2Z(x+h)Z(x) + Z(x)^2] \\ &= 0.5 [\text{Var}(Z(x+h)) - 2\text{Cov}(Z(x+h), Z(x)) + \text{Var}(Z(x))] \\ &= 0.5 [2C(0) - 2C(h)] \end{aligned}$$

- $\lim_{\|h\| \rightarrow \infty} \gamma(h) = C(0) - \lim_{\|h\| \rightarrow \infty} C(h) = C(0)$

3. Une fonction $f(r)$, $r \in \mathbf{R}$ est complètement monotone si elle est continue en 0 et si les dérivées successives de f sont monotones et de signes alternés, i.e., $(-1)^n f^{(n)}(r) \geq 0, r \in \mathbf{R}, n = 0, 1, 2, \dots$

Ces deux dernières propriétés sont importantes : pour un champ aléatoire stationnaire d'ordre 2, le variogramme atteint une valeur asymptotique égale à la variance.

Les propriétés de régularité que nous avons vues au sujet de la fonction de covariance se traduisent directement sur le variogramme.

1. Si $\lim_{\|h\| \rightarrow 0} \gamma(h) \rightarrow 0$, i.e. si $\gamma(h)$ est continu à l'origine, alors le champ $Z(x)$ est continu en moyenne quadratique.
2. Si $\gamma(h)$ est dérivable deux fois à l'origine, alors le champ $Z(x)$ est différentiable.
3. Si $\gamma(h)$ est dérivable $2p$ fois à l'origine, alors le champ $Z(x)$ est différentiable p .

3.2 Hypothèses intrinsèques

Comme le montre la définition 9, le variogramme ne se définit pas directement à partir de valeurs de $Z(x)$ mais plutôt à partir de différences $Z(x+h) - Z(x)$. L'hypothèse de stationnarité d'ordre 2, qui suppose l'existence des deux premiers moments de $Z(x)$, est donc inutilement restrictive si l'on souhaite n'utiliser que le variogramme et non la fonction de covariance.

Il est donc utile de définir des hypothèses de stationnarité plus faibles, appelées en géostatistique les **hypothèses de stationnarité intrinsèque** :

Définition 10 (Hypothèses intrinsèques et définition usuelle du variogramme) Soit $Z(x)$ un champ aléatoire sur \mathbf{R}^d . On dit que $Z(x)$ vérifie les hypothèses intrinsèques si

$$E[Z(x) - Z(x+h)] = \langle a, h \rangle \quad \text{et} \quad \frac{1}{2} \text{Var}\{Z(x+h) - Z(x)\} = \gamma(h),$$

$\forall x, \forall h \in \mathbf{R}^d$. Le terme $\langle a, h \rangle$ est une tendance linéaire de $Z(x)$. Si la tendance linéaire est nulle ($a = 0$), alors les hypothèses intrinsèques se reformulent selon leur formulation la plus habituelle :

$$\begin{cases} E[Z(x) - Z(x+h)] &= 0 \\ 0.5E[(Z(x+h) - Z(x))^2] &= \gamma(h) \end{cases}$$

Il est important de souligner que l'hypothèse de stationnarité d'ordre 2 est plus forte que les hypothèses intrinsèques et les entraîne. L'inverse, en revanche, n'est pas vrai, car les hypothèses intrinsèques ne font pas d'hypothèses sur l'existence des moments de $Z(\cdot)$, mais seulement sur l'existence des moments pour les incréments. Chez certains auteurs (p. ex. Yaglom, 1987), les hypothèses intrinsèques sont appelées hypothèses de **stationnarité des incréments**.

Ainsi, la classe des champs aléatoires qui vérifient les hypothèses intrinsèques est plus vaste que la classe des champs aléatoires qui vérifient l'hypothèse de stationnarité d'ordre 2.

Enfin il est important de noter que si toute fonction de covariance peut s'écrire sous la forme d'un variogramme à l'aide de $\gamma(h) = C(0) - C(h)$, l'inverse n'est pas vrai en toute généralité. Le passage

variogramme \rightarrow covariance

n'est possible que sous les hypothèses de stationnarité d'ordre 2.

3.3 Fonctions conditionnellement définies négatives

Nous avons vu à la section 2.3 que la fonction de covariance est une fonction définie positive. Il existe une propriété similaire pour les variogrammes. Pour présenter cette propriété, il faut tout d'abord définir la notion de **combinaison linéaire autorisée**.

Proposition 3 *Soit $Z(x)$ un champ aléatoire vérifiant les hypothèses intrinsèques. Une combinaison linéaire $\sum_{i=1}^n \lambda_i Z(x_i)$ possède un premier et un second moment ssi $\sum_{i=1}^n \lambda_i = 0$.*

Preuve :

- (i) C'est une condition nécessaire : soit une C.L. $S = \sum_{i=1}^n \lambda_i Z(x_i)$ telle que $\sum_{i=1}^n \lambda_i = 0$. Alors, $\sum_{i=1}^n \lambda_i Z(x_i) = \sum_{i=1}^n \lambda_i (Z(x_i) - Z(x_0))$ car $\sum_{i=1}^n \lambda_i Z(x_0) = 0$. Comme S peut s'écrire comme une somme d'incrément, S possède une espérance et une variance.
- (ii) C'est une condition suffisante : sous les hypothèses intrinsèques, toute différence $Z(x_i) - Z(x_0)$ possède en effet une espérance et une variance. Alors, $\sum_{i=1}^n \lambda_i (Z(x_i) - Z(x_0))$ est une combinaison linéaire qui possède une espérance et une variance, et dont la somme des pondérateurs vaut $\sum_{i=1}^n \lambda_i - \sum_{i=1}^n \lambda_i = 0$. \square

Ce résultat est très important, car il nous indique que sous hypothèse intrinsèque, on ne peut calculer des espérances et des variances que pour des combinaisons linéaires dont le poids total est nul. On appelle **Combinaison Linéaire Autorisée (CLA)** un ensemble de pondérateurs $(\lambda_i)_{i=1, \dots, n}$ ayant la propriété $\sum_i \lambda_i = 0$.

Proposition 4 *Soit $Z(x)$ un champ aléatoire vérifiant les hypothèses intrinsèques et soit $(\lambda_i)_{i=1, \dots, n}$ une CLA. Alors,*

$$E\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] = 0$$

et

$$\text{Var}\left\{\sum_{i=1}^n \lambda_i Z(x_i)\right\} = -\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j). \quad (18)$$

Preuve : La preuve est directe en développant les calculs des espérances, et des variances. Mais **attention !** de se rappeler que $E[Z(x_i)]$ et $\text{Var}\{Z(x_i)\}$ n'existent pas.

i)

$$\begin{aligned}
E\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] &= E\left[\sum_{i=1}^n \lambda_i (Z(x_i) - Z(x_0))\right] \\
&= \sum_{i=1}^n \lambda_i E[Z(x_i) - Z(x_0)] \\
&= 0
\end{aligned}$$

La première égalité est vraie pour tout $Z(x_0)$, car $\sum_{i=1}^n \lambda_i = 0$; la troisième égalité résulte de l'hypothèse intrinsèque.

ii)

$$\begin{aligned}
\text{Var}\left\{\sum_{i=1}^n \lambda_i Z(x_i)\right\} &= \text{Var}\left\{\sum_{i=1}^n \lambda_i (Z(x_i) - Z(x_0))\right\} \\
&= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}\{Z(x_i) - Z(x_0), Z(x_j) - Z(x_0)\}
\end{aligned}$$

Comme

$$(Z(x_i) - Z(x_j))^2 = (Z(x_i) - Z(x_0))^2 + (Z(x_j) - Z(x_0))^2 - 2(Z(x_i) - Z(x_0))(Z(x_j) - Z(x_0)),$$

on obtient

$$\gamma(x_i - x_j) = \gamma(x_i - x_0) + \gamma(x_j - x_0) - \text{Cov}\{Z(x_i) - Z(x_0), Z(x_j) - Z(x_0)\}$$

et donc

$$\text{Cov}\{Z(x_i) - Z(x_0), Z(x_j) - Z(x_0)\} = \gamma(x_i - x_0) + \gamma(x_j - x_0) - \gamma(x_i - x_j).$$

Et donc, finalement,

$$\begin{aligned}
\text{Var}\left\{\sum_{i=1}^n \lambda_i Z(x_i)\right\} &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \{\gamma(x_i - x_0) + \gamma(x_j - x_0) - \gamma(x_i - x_j)\} \\
&= - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j) \\
&\geq 0
\end{aligned}$$

$$\text{car } \sum_{j=1}^n \lambda_j \gamma(x_i - x_0) = \sum_{i=1}^n \lambda_i \gamma(x_j - x_0) = 0.$$

□

Remarques :

1. Cette formule est une forme quadratique formellement identique à (2) sous hypothèse stationnaire d'ordre 2, mais dans laquelle $C(h)$ est remplacée par $-\gamma(h)$.

Notons que sous hypothèse stationnaire d'ordre 2, en appliquant (2) à $C(h) = C(0) - \gamma(h)$ pour une CLA, on retrouve (18) car la constante $C(0)$ est éliminée par le fait que $\sum_{i=1}^n \lambda_i = 0$.

2. Une variance est une quantité positive ou nulle. Il faut donc que la fonction $\gamma(h)$ vérifie (18) pour toutes CLA. Une fonction qui vérifie cette propriété est appelée fonction **conditionnellement définie négative**.
3. Comme toute fonction de covariance est définie positive, elle est en particulier conditionnellement définie positive ; mais l'inverse n'est pas nécessairement vrai. Il en résulte que la classe des variogrammes est plus large que la classe des fonctions de covariance.

Ces différents éléments se résument au théorème suivant.

Théorème 6 *Une fonction γ sur \mathbf{R}^d est un variogramme ssi γ est conditionnellement définie négative (ou, ce qui est équivalent ssi $-\gamma$ est conditionnellement définie positive).*

3.4 Représentation spectrale

Les variogrammes ont également une représentation spectrale, et à ce titre, il existe un théorème qui est une généralisation du théorème de Bochner.

Théorème 7 *Soit $\gamma(h)$ une fonction sur \mathbf{R}^d telle que $\gamma(0) = 0$. Les trois conditions sont équivalentes :*

1. $\gamma(h)$ est le variogramme d'un champ aléatoire intrinsèque vérifiant la définition 10.
2. la fonction $\psi(h) = e^{-t\gamma(h)}$ est une fonction de covariance $\forall t > 0$.
3. $\gamma(h)$ est de la forme

$$\gamma(h) = \int \frac{1 - \cos(2\pi \langle u, h \rangle)}{4\pi^2 \|u\|^2} \chi(du) \quad (19)$$

où χ est une mesure positive symétrique sans atome en 0 vérifiant

$$\int \frac{\chi(du)}{1 + 4\pi^2 \|u\|^2} < \infty.$$

Preuve :

i) 1. entraîne 2. Soit $Z(x)$ un champ aléatoire intrinsèque admettant le variogramme $\gamma(h)$. Alors, le champ aléatoire $Y(x) = Z(x) - Z(0)$ admet la covariance (non stationnaire) $C(x, y) = \gamma(x) + \gamma(y) - \gamma(x + y)$. Désignons par $Y_n(x)$ une suite de champs aléatoires indépendantes admettant la même fonction de covariance $C(x, y)$ et par N une variable de Poisson d'espérance $E[N] = t$, indépendante des $Y_n(x)$. Le champ aléatoire définie par $Y_0(x) = 1$ si $N = 0$ et $Y_0(x) = \sum_{n=1}^N Y_n(x)$ si $N > 0$ admet la covariance

$$C_0(x, y) = \sum_{n=1}^{\infty} \frac{t^n}{n!} e^{-t} C^n(x, y) = e^{t\{C(x,y)-1\}}.$$

Le champ aléatoire $Y_0(x)/E[Y_0(x)^2]$ admet alors la covariance

$$\begin{aligned} E[Y_0(x)Y_0(y)] / (E[Y_0(x)^2]E[Y_0(y)^2]) &= \exp\{t[\gamma(x) + \gamma(y) - \gamma(x - y) - \gamma(x) - \gamma(y)]\} \\ &= \exp\{-t\gamma(x - y)\}, \end{aligned}$$

qui est une fonction de $x - y$. Donc $e^{-t\gamma(h)}$ est bien une fonction de covariance.

ii) 2. entraîne 3. La fonction $\psi(h) = e^{-t\gamma(h)}$ est la fonction caractéristique d'une variable aléatoire infiniment divisible (ce qui se montre facilement car la classe des distributions infiniment divisible coïncide avec la classe des limites de mélanges de distributions de Poisson ; et donc, il en va de même de leur fonction caractéristique). Selon le théorème de Lévy, la fonction caractéristique d'une distribution indéfiniment divisible (cf. Feller, 1971, ch. XVII) a pour forme générale (dans le cas réel) l'expression recherchée (19).

iii) 3. entraîne 1. Il est immédiat de montrer que lorsque $\gamma(h)$ est de la forme (19), alors $-\gamma(h)$ est conditionnellement définie positive.

3.5 Propriétés spécifiques du variogramme

Proposition 5 *Si un champ $Z(x)$ possède un variogramme $\gamma(h)$ borné, c'est à dire tel que*

$$\lim_{\|h\| \rightarrow \infty} \gamma(h) = \gamma(\infty) < \infty,$$

alors $Z(x)$ est un champ aléatoire stationnaire d'ordre 2.

Preuve : Ce résultat reprend ce que nous avons déjà vu. En effet, si le variogramme est borné, il existe une valeur σ^2 telle que $\sigma^2 = \lim_{\|h\| \rightarrow \infty} \gamma(h)$. Alors la fonction $C(h) = \sigma^2 - \gamma(h)$ tend vers 0 lorsque $\|h\| \rightarrow \infty$, et $C(h)$ est une fonction de covariance car elle est définie positive (puisque $\gamma(h)$ est conditionnellement définie négative). \square

Les variogrammes bornés héritent de toutes les propriétés des fonctions de covariance (vues au chapitre 2) à travers la transformation $C(h) = \gamma(\infty) - \gamma(h)$.

De manière plus général, **tous** les variogrammes vérifient les propriétés suivantes, vraies sur $\mathbf{R}^d, \forall d$.

Proposition 6 *Si $\gamma(h)$ est le variogramme d'un champ aléatoire intrinsèque continu en moyenne quadratique, alors*

$$\lim_{\|h\| \rightarrow \infty} \gamma(h) < A\|h\|^2.$$

Preuve : D'après la représentation spectrale du variogramme du théorème 7, on a

$$\frac{\gamma(h)}{\|h\|^2} = \int \frac{1 - \cos(2\pi \langle u, h \rangle)}{4\pi^2 \|h\|^2 \|u\|^2} \chi(du).$$

Or,

$$\left| \frac{1 - \cos(2\pi \langle u, h \rangle)}{4\pi^2 \|h\|^2 \|u\|^2} \right| \leq \frac{(uh)^2}{\|h\|^2 \|u\|^2} \leq 1.$$

Par ailleurs, la fonction à intégrer tend vers 0 lorsque $\|h\| \rightarrow \infty$, sauf en $u = 0$. D'après le théorème de convergence dominée, on a donc

$$\int \frac{1 - \cos(2\pi \langle u, h \rangle)}{4\pi^2 \|h\|^2 \|u\|^2} \chi(du) \rightarrow 0$$

lorsque $\|h\| \rightarrow \infty$. Par conséquent $\gamma(h)/\|h\|^2 \rightarrow 0$ lorsque $\|h\| \rightarrow \infty$. □

Ce résultat nous indique que les variogrammes des champs aléatoires intrinsèques peuvent être non bornés, mais qu'ils ne peuvent pas croître trop vite, i.e., pas plus vite qu'une fonction quadratique.

3.6 Les variogrammes puissances

En appliquant le théorème 7 à la famille de fonctions de covariance $c(h) = \exp -a\|h\|^\alpha$, avec $a > 0$ et $0 < \alpha \leq 2$ (cf section 2.3) on voit immédiatement que les fonctions $\gamma(h) = a\|h\|^\alpha$, avec $a > 0$ et $0 < \alpha \leq 2$, sont des variogrammes valides (i.e. des fonctions conditionnellement définies négatives). Nous allons voir ci-dessous que c'est effectivement le cas pour $0 < \alpha < 2$. Le cas $\alpha = 2$ est particulier et méritera une discussion spécifique.

Considérons

$$\gamma(h + h') = \frac{1}{2} E [(Z(x + h + h') - Z(x))^2]$$

$$\begin{aligned}
&= \frac{1}{2} E [(Z(x+h+h') - Z(x+h')) + (Z(x+h') - Z(x))]^2 \\
&= \gamma(h) + \gamma(h') + \text{Cov}(Z(x+h+h') - Z(x+h'), Z(x+h') - Z(x)) \quad (20) \\
&\leq \gamma(h) + \gamma(h') + 2(\gamma(h)\gamma(h'))^{1/2}.
\end{aligned}$$

En appliquant ce résultat à $\gamma(h) = a||h||^\alpha$, on obtient

$$||h+h'||^\alpha \leq ||h||^\alpha + ||h'||^\alpha + 2||hh'||^{\alpha/2}.$$

Poser $||h'|| = ||h||$ mène à

$$2^\alpha ||h||^\alpha \leq 4||h||^\alpha,$$

et donc

$$2^\alpha \leq 4,$$

c'est à dire finalement à $\alpha \leq 2$. La condition $\alpha > 0$ vient de ce que l'on a imposé que le champ est continu en m.q. Lorsque $\alpha = 2$ il est immédiat de montrer que le champ est différentiable en moyenne quadratique. Lorsque $\alpha < 2$, le champ n'est que continu en moyenne quadratique, mais non différentiable en m.q.

Allons un peu plus loin, et repartons de (20) en posant $h = h'$. Alors,

$$\text{Cov}(Z(x+2h) - Z(x+h), Z(x+h) - Z(x)) = a||h||^\alpha(2^\alpha - 2),$$

ou encore

$$\text{Cov}(W_h(x+h), W_h(x)) = a||h||^\alpha(2^\alpha - 2),$$

en notant $W_h(x) = Z(x+h) - Z(x)$ l'accroissement de longueur h localisé en x . Ce résultat montre que le signe de la covariance d'accroissements successifs dépend de la valeur α :

- Lorsque $0 < \alpha < 1$, nous avons

$$\text{Cov}(W_h(x+h), W_h(x)) < 0.$$

Les accroissements successifs sont anti-corrélés : après une hausse, une baisse est plus probable qu'une nouvelle hausse. Bien qu'ayant une variance *a priori* infinie, ils correspondent à des processus aléatoires présentant une certaine force de rappel vers l'origine $Z(0)$.

- Lorsque $\alpha = 1$, c'est à dire $\gamma(h) = a||h||$, le processus est dit à **accroissements indépendants**. C'est la généralisation à $d \geq 1$ dimensions de l'espace du mouvement brownien. Ces processus s'écartent indéfiniment de l'origine $Z(0)$.

- Lorsque $1 < \alpha < 2$, nous avons

$$\text{Cov}(W_h(x+h), W_h(x)) > 0.$$

Ces processus ont des accroissements successifs corrélés : après une hausse, une nouvelle hausse est probable. Ils correspondent à des processus qui explosent.

- Lorsque $\alpha = 2$, on a

$$\text{Cov}(W_h(x+h), W_h(x)) = 2a||h||^2$$

et donc

$$\begin{aligned} \text{Cor}(W_h(x+h), W_h(x)) &= \frac{2a||h||^2}{(\text{Var}\{Z(x+2h) - Z(x+h)\}\text{Var}\{Z(x+h) - Z(h)\})^{1/2}} \\ &= \frac{2a||h||^2}{(2a||h||^2 2a||h||^2)^{1/2}} \\ &= 1. \end{aligned}$$

Ainsi, lorsque $\alpha = 2$, les accroissements successifs de longueur identiques sont parfaitement corrélés. Le variogramme $\gamma(h) = a||h||^2$ correspond en fait à un champ $Z(x) = \langle A, x \rangle$, où A est une variable aléatoire centrée. En effet, pour un tel champ, on a bien $E[Z(x+h) - Z(x)] = 0$ et $\text{Var}\{Z(x+h) - Z(x)\} = \sum_{i=1}^d A_i^2 ||h||$. Ce modèle, très particulier, n'est pas ergodique.

3.7 Allure générale d'un variogramme

Il faut distinguer deux cas : les variogrammes bornés et les variogrammes non bornés.

Variogrammes bornés

Nous avons vu à la proposition 5 que lorsque le variogramme est borné, il correspond à une fonction de covariance $C(h) = \gamma(\infty) - \gamma(h)$. On retrouve donc sur les variogrammes les grandes caractéristiques des fonctions de covariance, vues à la section 2.7.

- La valeur $\gamma(\infty) = \lim_{||h|| \rightarrow \infty} \gamma(h)$ est égale à la variance du champ. On l'appelle pallier du variogramme.
- La continuité en $h = 0$ de $\gamma(h)$ est liée à la régularité du champ aléatoire. Le saut de discontinuité à l'origine du variogramme est l'effet de pépite.
- La portée est la distance à laquelle le variogramme atteint son pallier, soit de façon exacte (portée vraie), soit asymptotiquement (portée pratique)

La figure 3.7 représente à gauche un variogramme sans effet de pépite, avec une portée $a = 10$ et à droite un variogramme avec effet de pépite et une portée pratique $a = 10$.

FIGURE 2 – Deux exemples de variogrammes bornés.

FIGURE 3 – Deux exemples de variogrammes puissances.

Variogrammes non bornés

Dans ce cas, il n'y a pas de pallier, ni de portée. Les caractéristiques importantes sont la régularité du variogramme à l'origine et sa pente.

La figure 3.7 représente à gauche un variogramme puissance sans effet de pépité, avec un exposant $\alpha = 1.5$ et à droite un variogramme puissance avec effet de pépité et $\alpha = 0.5$.

4 Estimation du variogramme

4.1 Pourquoi estimer le variogramme ?

Deux raisons justifient une estimation du variogramme plutôt que de la fonction de covariance, et ce même dans le cas où les hypothèses stationnaires d'ordre 2 sont vérifiées.

1. Tout d'abord, nous avons vu tout au long du chapitre 3 que la classe des fonctions variogrammes est plus vaste et contient la classe des fonctions de covariance. Ainsi, lorsque nous estimons un variogramme nous faisons des hypothèses moins restrictives. En particulier, nous avons vu (point 2 du paragraphe 3.5) que si le variogramme présente un pallier (i.e. une valeur asymptotique pour $\|h\| \rightarrow \infty$), alors le champ aléatoire est stationnaire d'ordre 2.

Dès lors, la présence d'un pallier lors de l'estimation du variogramme est un diagnostic de stationnarité⁴.

2. Même si l'hypothèse de stationnarité d'ordre 2 est vérifiée, l'estimation de la fonction de covariance sera biaisée. En effet, dans la définition de la fonction de covariance,

$$C(h) = \text{Cov}(Z(x+h), Z(x)) = E[(Z(x+h) - m)(Z(x) - m)]$$

intervient l'espérance mathématique, m , qui n'est pas connue. Remplacer m dans l'estimation de la fonction de covariance par une estimation \hat{m} entraîne un biais qu'il n'est pas

4. J'utilise le mot diagnostic plutôt que le mot test ; en effet, cela ne constitue pas un test de stationnarité au sens statistique du terme.

possible de corriger, comme nous le verrons plus tard.

Pour ces deux raisons, il est préférable d'estimer le variogramme que la fonction de covariance

4.2 Le variogramme expérimental

Le variogramme expérimental est au variogramme théorique ce que l'histogramme est à la fonction de densité : une représentation graphique, nécessairement imparfaite, construite à partir des données.

On dispose d'échantillons géoréférencés (x_1, \dots, x_n) dans un domaine D de \mathbf{R}^d . Par la suite, les indices α et β , avec $1 \leq \alpha, \beta \leq n$, seront utilisés pour parcourir l'ensemble des échantillons. Pour alléger les notations, on notera $Z_\alpha = Z(x_\alpha)$ et $h_{\alpha\beta} = x_\beta - x_\alpha$ le vecteur séparant les points x_α et x_β .

On définit $\gamma_{\alpha\beta}^* = (Z_\alpha - Z_\beta)^2/2$ la dissimilarité entre les échantillons $Z(x_\alpha)$ et $Z(x_\beta)$. On a donc $\gamma_{\alpha\beta}^* = \gamma_{\beta\alpha}^*$. Par ailleurs, on peut vérifier que

$$E[\gamma_{\alpha\beta}^*] = \frac{1}{2}E[(Z_\alpha - Z_\beta)^2] = \gamma(h_{\alpha\beta}).$$

Ainsi, on voit que $\gamma_{\alpha\beta}^*$ est un échantillon sans biais de $\gamma(h_{\alpha\beta})$.

Sous hypothèse de stationnarité intrinsèque, les incréments pour un vecteur h donné sont stationnaires d'ordre 2, i.e. ont mêmes premier et second moments. Des couples de points (x_α, x_β) séparés par des vecteurs $h_{\alpha\beta}$ proches de h ont donc des premiers moments semblables et peuvent être mélangés pour construire un estimateur $\gamma^*(h)$ du variogramme $\gamma(h)$.

Définition 11 (Variogramme expérimental) *Soit des échantillons $Z(x_1), \dots, Z(x_n)$, le variogramme expérimental est la fonction*

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{\alpha, \beta: h_{\alpha\beta} \simeq h} (Z(x_\alpha) - Z(x_\beta))^2, \quad (21)$$

où $N(h)$ est le nombre de paires (x_α, x_β) telles que $h_{\alpha\beta} \simeq h$.

Pour calculer effectivement le variogramme expérimental, il faut préciser ce qu'on entend par $h_{\alpha\beta} \simeq h$. On va distinguer successivement deux cas. Dans le premier cas, on pense qu'il n'y a pas d'effet directionnel et on suppose que le variogramme est isotropique (cf section ?? pour une définition de ce terme). Dans ce cas, le variogramme $\gamma^*(h)$ ne dépend en réalité que de la distance $r = \|h\|$. Dans le second cas, on modélise des effets directionnels, et le variogramme $\gamma^*(h)$ dépend à la fois de la distance $r = \|h\|$ et de la direction du vecteur h .

Variogramme isotropique

On note $r = \|h\|$ la norme de h . Le variogramme expérimental $\gamma^*(r)$ est calculé à la distance r en se définissant une tolérance Δr autour de r :

$$\gamma^*(r) = \frac{1}{2N(r)} \sum_{\alpha, \beta: \|h_{\alpha\beta}\| \in [r - \Delta r; r + \Delta r]} \{Z(x_\alpha) - Z(x_\beta)\}^2, \quad (22)$$

où $N(r)$ est le nombre de paires de points (x_α, x_β) qui vérifient $\|h_{\alpha\beta}\| \in [r - \Delta r; r + \Delta r]$. Afin d'obtenir une représentation graphique commode du variogramme expérimental, on le calcule pour une série de distances r_1, \dots, r_K , puis on représente $\gamma^*(r_k)$ en fonction de r_k , $k = 1, \dots, K$.

La construction du variogramme expérimental ressemble donc à la construction de l'histogramme, pour lequel on doit se donner une série de classes puis représenter le nombre de données dans chacune de ces classes. On sait qu'en statistique exploratoire une question très importante (et pas vraiment résolue) est de définir le nombre et l'amplitude des classes. Il en va de même en géostatistique, où le nombre et l'amplitude des classes de distances doivent être données pour le calcul du variogramme expérimental. En règle générale, on adopte une amplitude constante, i.e., on pose que $r_k - r_{k-1} = a$, $\forall k$. Par ailleurs, afin de ne négliger aucune données, on pose également $\Delta r = a/2$. En définitive, les deux paramètres qui définissent le variogramme expérimental sont a , le **pas du variogramme**, et K , le **nombre de pas**. L'amplitude a est choisie de sorte à ce que les effectifs pour chaque classe de distance soient suffisamment nombreux (p. ex. ≥ 30). Le nombre de classe K est choisi afin que la distance maximale, aK soit environ égale à $d/2$, où d est le diamètre minimal du domaine D .

Avec ces notations, le variogramme expérimental isotropique s'écrit

$$\gamma^*(r_k) = \frac{1}{2N(r_k)} \sum_{\alpha, \beta: \|h_{\alpha\beta}\| \in [r_k - a/2; r_k + a/2]} \{Z(x_\alpha) - Z(x_\beta)\}^2, \quad k = 1, \dots, K$$

où $r_k = ka$ et $\gamma(0) = 0$.

Le choix du pas du variogramme est crucial. En effet, si le pas est trop petit, le nombre de paires impliquées dans le calcul de la moyenne sera faible, ce qui entraîne de grandes fluctuations statistiques. A l'inverse, si le pas est trop important, des dissimilarités calculées à des distances assez différentes seront mélangées dans la moyenne. Le pas optimal résulte donc d'un compromis entre biais et variance de l'estimation.

Variogramme anisotropique

Certains phénomènes environnementaux ou géologiques présentent des directions de continuité privilégiées. Lorsque c'est le cas on calcule le variogramme expérimental selon des directions.

Aux classes de distance, on peut ajouter des classes de direction. Pour le vecteur $h = r(\cos \omega, \sin \omega)$ de direction ω dans \mathbf{R}^d , on prendra

$$\begin{aligned} r &\in [r_k - a/2; r_k + a/2] \\ \omega &\in [\omega_j - \delta/2; \omega_j + \delta/2] \end{aligned}$$

où r_k est un ensemble de distances et ω_j est un ensemble de directions. Les directions sont en général choisies de façon régulière et en nombre pas trop important. La connaissance *a priori* du phénomène peut aider à choisir des distances préférentielles. Sinon, on choisit généralement les directions 0, 45, 90 et 135.

Une anisotropie se traduit sur les variogrammes expérimentaux par des portées qui varient selon les directions. On parle alors de variogramme expérimentaux anisotropes. Le plus souvent, on observe que les directions de plus longue et de plus courte portées sont orthogonales.

4.3 Estimation paramétrique d'un modèle de variogramme

Le variogramme expérimental est utile pour explorer les données, mais il n'est pas adéquat pour répondre aux questions de prédiction que nous verrons au chapitre suivant. Pour cela nous devons disposer d'un **modèle de variogramme**, pour les deux raisons suivantes :

- Pour réaliser la prédiction, il faudra connaître le variogramme à des distances pour lequel il n'a pas été calculé empiriquement selon les équations ci-dessus. Il nous faut donc un modèle pour ces distances.
- Par ailleurs, comme nous l'avons vu à la section 3.3, le variogramme doit être une fonction conditionnellement définie négative ; il faut donc faire l'estimation à l'intérieure d'une classe paramétrique de fonctions vérifiant cette propriété.

Les modèles de variogrammes $\gamma(\cdot; \theta)$ présentés à la section 3 dépendent d'un vecteur de paramètre $\theta \in \Theta \subset \mathbf{R}^p$ qu'il faut estimer. Nous allons présenter deux méthodes d'estimation de θ : les moindres carrés et le maximum de vraisemblance.

Estimation par moindre carré

L'estimateur empirique naturel du variogramme théorique $\gamma(h)$ est l'estimateur des moments (21)

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha, \beta: h_{\alpha\beta} \simeq h} (Z(x_\alpha) - Z(x_\beta))^2, \quad h \in D. \quad (23)$$

Proposition 7 *La quantité $\hat{\gamma}(h)$ est un estimateur sans biais de $\gamma(h)$.*

Cela découle directement du fait que, comme nous l'avons déjà vu, $E[\{Z(x_\alpha) - Z(x_\beta)\}^2] = 2\gamma(h_{\alpha\beta})$.

Pour une famille paramétrique $\gamma(h; \theta)$ l'estimation de θ par moindres carrés ordinaires (MCO) est

$$\hat{\theta}_{MCO} = \arg \min_{\eta \in \Theta} \sum_{k=1}^K \{\hat{\gamma}(h_k) - \gamma(h_k; \eta)\}^2,$$

où K est le nombre de classes retenues pour le calcul du variogramme expérimental. Dans cette formule, la somme est réalisée sur des classes de distance (et éventuellement des classes de direction), comme nous cela a été discuté à la section précédente.

La méthode des MCO n'est en général pas la plus efficace car les $\hat{\gamma}(h_k)$ ont en général des variances très différentes, à la fois parce que la variance théorique des termes $\{Z(x_\alpha) - Z(x_\beta)\}^2$ dépend de $h_{\alpha\beta}$, et par ce que le nombre de termes $N(h)$ varie fortement d'une classe de distance à une autre.

L'estimation de θ par moindres carrés pondérés (MCP) consiste à introduire un poids w_k pour chaque classe, inversement proportionnel à une approximation de la variance de $\hat{\gamma}(h_k)$. Pour un champ $Z(\cdot)$ Gaussien d'ordre 2, $0.5\{Z(x_\alpha) - Z(x_\beta)\}^2$ est distribué comme une χ_1^2 , dont la variance est $\gamma^2(h_{\alpha\beta})$. En négligeant les corrélations intervenant entre les incréments d'une même classe, on peut écrire $Var\{\hat{\gamma}(h_k)\} \simeq \gamma^2(h_k; \alpha)/N(h_k)$. L'estimation par MCP revient à résoudre

$$\hat{\theta}_{MCP} = \arg \min_{\eta \in \Theta} \sum_{k=1}^K \frac{N(h_k)}{\gamma^2(h_k; \eta)} \{\hat{\gamma}(h_k) - \gamma(h_k; \eta)\}^2, \quad (24)$$

De nombreuses études ont montré que les performances de l'estimation par MCP sont généralement largement meilleures que par MCO.

Estimation par maximum de vraisemblance

L'estimation par maximum de vraisemblance nécessite de faire une hypothèse distributionnelle complète pour le vecteur d'observations $Z = (Z(x_1), \dots, Z(x_n))$. Soit Z un vecteur gaussien de moyenne μ et de covariance $Var\{Z\} = \Sigma(\theta)$, où θ est le vecteur de paramètre de la fonction de covariance $C(h; \theta)$. L'élément (i, j) de $\Sigma(\theta)$ vaut $\Sigma_{[i,j]}(\theta) = C(x_j - x_i; \theta)$. La log-vraisemblance vaut alors :

$$l_n(\theta) = -\frac{1}{2} \{\ln \det(\Sigma(\theta)) + (Z - \mu)^t \Sigma^{-1}(\theta) (Z - \mu)\}. \quad (25)$$

L'estimateur de maximum de vraisemblance est la valeur $\hat{\theta}$ qui maximise $l_n(\theta)$ et $C(h; \hat{\theta})$ est l'estimateur par MV de $C(h; \theta)$.

Remarques

1. L'estimation par MV selon (25) suppose l'existence d'une espérance μ et d'une fonction de covariance $C(h; \theta)$. En d'autre terme, elle n'est possible que sous hypothèse stationnaire d'ordre 2, ce qui exclut d'emblée les variogrammes non bornés. Les paramètres de ceux-ci doivent être estimés par MCP ou par vraisemblance restreinte (REML), voir Cressie (1993).
2. L'estimateur de MV repose fortement sur l'hypothèse gaussienne ; des exercices de simulation ont montré que cet estimateur n'est pas très robuste par rapport à des écarts à cette hypothèse. Sur certaines données, cet estimateur donne des résultats assez instables ; il suffit parfois de retirer une seule donnée pour voir les paramètres varier fortement. Il est donc plutôt recommandé d'utiliser l'estimateur par MCP.
3. Pour maximiser (25) il faut itérer le calcul de $\det(\Sigma(\theta))$ et de $\Sigma(\theta)^{-1}$, ce qui exige $O^3(n)$ opérations. Ce n'est donc envisageable que pour de relativement petits échantillons (tout au plus quelques centaines de valeurs).

4.4 Estimation paramétrique d'un variogramme anisotrope

Lorsque le variogramme expérimentale présente des anisotropies, on va modéliser les différents variogrammes expérimentaux par un variogramme unique $\gamma(h; \theta)$, valable dans toutes les directions et pour toutes les distances à la fois. Il est important de noter qu'il ne s'agit pas de modéliser un variogramme par direction, mais de modéliser une fonction variogramme $\gamma(h; \theta)$, qui ne dépend pas uniquement du module $\|h\|$ de h , mais également de sa direction.

L'idée générale est que l'on peut transformer les coordonnées $x = (x_1, x_2)$ qui décrivent D en un nouveau système de coordonnées, $x' = (x'_1, x'_2)$ dans lequel le variogramme est isotropique. Les paramètres du variogramme isotrope sont alors estimés (par MCP par exemple) dans ce nouveau système de coordonnées.

Anisotropie géométrique

Un modèle particulièrement facile à mettre en œuvre estimer et à ajuster est le modèle d'*anisotropie géométrique*. Ce modèle consiste à faire l'hypothèse que la transformation est la succession d'une rotation et d'une dilatation/contraction.

Dans \mathbf{R}^2 , une rotation s'exprime l'aide d'une matrice de rotation

$$Q = \begin{pmatrix} \cos \tau & \sin \tau \\ -\sin \tau & \cos \tau \end{pmatrix},$$

qui transforme le système de coordonnées d'origine en un système de coordonnées parallèle aux axes principaux. Dans \mathbf{R}^3 , la matrice prend une forme un peu plus compliquée et s'écrit comme

une composition de 3 rotations élémentaires autour des trois axes principaux. Voir Wackernagel (2003) pour une description détaillée. On se limite dorénavant au cas $d = 2$.

Une second matrice, diagonale, opère une dilatation ou une contraction de chaque axe :

$$\sqrt{D} = \begin{pmatrix} \sqrt{d_1} & 0 \\ 0 & \sqrt{d_2} \end{pmatrix}.$$

Ainsi, les coordonnées x' s'expriment comme la composition $x' = \sqrt{D}Qx$. Le variogramme est supposé isotropique dans le nouveau système de coordonnées. Il ne dépend donc que de la distance $r = \|h'\|$, c'est à dire :

$$r = \|h'\| = \sqrt{(h')^t h'} = \sqrt{h^t Q^t D Q h}.$$

En d'autre termes, les lignes d'iso-variogramme, qui sont circulaires dans le nouveau système de coordonnées, décrivent une ellipse d'équation

$$r^2 = h^t B h = h^t Q^t D Q h,$$

ayant pour direction la direction τ et ayant pour diamètre $2r/\sqrt{d_p}$ dans la direction $p = 1, 2$.

Finalement, on a donc un modèle de variogramme $\gamma(h; \theta, B)$, dans lequel θ sont les paramètres d'un variogramme anisotropique $\gamma_{iso}(r; \theta)$ et où

$$\gamma(h; \theta, B) = \gamma_{iso}(\sqrt{h^t B h}; \theta)$$

avec $B = Q^t D Q$. Ce modèle d'anisotropie comporte trois paramètres : la direction principale, τ et les paramètres d_1 et d_2 . En réalité le variogramme isotropique γ_{iso} ayant lui même un paramètre d'échelle (la portée), ce modèle est surparamétré. Il est préférable de poser $d_1 = 1$ et $d_2 > 1$ afin d'éviter tout problème de surparamétrisation.

Notons que dans le modèle d'anisotropie géométrique, le pallier est identique dans toutes les directions. Seule la portée varie.

L'anisotropie zonale

Il peut arriver que les variogrammes expérimentaux présentent des palliers différents selon les directions. Cela est par exemple le cas lorsque le domaine présente des objets géomorphologiques parallèles de très grandes dimensions qui traversent le domaine. Dans ce cas, le domaine apparaît comme divisé en zones plus ou moins parallèles. Cela explique une plus grande variabilité dans

la direction qui traverse cette zonation que dans la direction parallèle à ces objets. Posons que x_2 est la direction de plus grande variabilité.

Une stratégie efficace pour modéliser une telle anisotropie consiste à ajuster un modèle de variogramme $\gamma_1(h)$ le long de la direction x_1 puis à le poser comme un variogramme isotrope. On ajoute ensuite un variogramme $\gamma_2(h)$ avec une anisotropie géométrique, pour lequel τ est dans la direction x_2 et d_2 présente une valeur très petite, qui le rend sans effet visible dans la direction x_1 . Le variogramme total est alors

$$\gamma(h) = \gamma_1(h) + \gamma_2(h).$$

Déformation générale de l'espace

On se rappelle que l'anisotropie géométrique consiste à déformer le système de coordonnées par une transformation paramétrique simple en un nouveau système de coordonnées dans lequel le variogramme est isotrope. On généralise cette approche en considérant qu'il existe une transformation générale, f , des coordonnées x en un nouveau système de coordonnées x' ($f(x) = x'$), telle que $\gamma(\|x'_i - x'_j\|) = \gamma(\|f(x_i) - f(x_j)\|)$ est isotropique. Pour être capable d'estimer la fonction f , il faut *i*) imposer des conditions de régularité sur f (en général, imposer que f est un difféomorphisme de D dans \mathbf{R}^d suffit) et *ii*) disposer de N répétitions (par exemple temporelle) du phénomène afin de pouvoir disposer d'un estimateur empirique $\gamma^*(x_\alpha, x_\beta) = 1/2N \sum_{i=1}^N (Z(x_\alpha^i) - Z(x_\beta^i))^2$.

4.5 Validation d'un modèle de variogramme

Les techniques d'estimation vues ci-dessus permettent d'estimer les paramètres θ d'une covariance ou d'un modèle de variogramme à l'intérieur d'une famille paramétrique, selon certains critères d'optimalité (moindres carrés pondérés, maximum de vraisemblance). Il reste à valider cette estimation, ce qui revient à choisir la famille paramétrique. Il peut également s'agir de sélectionner le meilleur couple (famille paramétrique ; paramètres estimés) parmi un ensemble donné. Nous allons décrire deux procédures alternatives.

Le bootstrap paramétrique

On se place dans le cas le plus général, dans lequel on a estimé les paramètres $\hat{\theta}$ d'un variogramme paramétrique $\gamma(h; \theta)$, à partir d'un variogramme empirique $\hat{\gamma}^{(0)}(h_k)$ calculés pour des classes de distance $(h_k)_{k=1, \dots, K}$ à partir d'un échantillon $Z = (Z(x_1), \dots, Z(x_n))$ de valeurs mesurées aux points $x_1, \dots, x_n \in D$. On réalise alors N simulations⁵ $Z^{(j)}$, $j = 1, \dots, N$ d'un

5. Pour les techniques de simulation, voir Section ??

vecteur gaussien (éventuellement sous hypothèse intrinsèque) aux points de données x_1, \dots, x_n avec le variogramme $\gamma(h; \hat{\theta})$. Pour chaque simulation j , on calcule les estimations empiriques $\{\hat{\gamma}^{(j)}(h_k), k = 1, \dots, K\}$ pour les mêmes classes de distance $(h_k)_{k=1, \dots, K}$. On en déduit les enveloppes inférieures $\hat{\gamma}_{inf}(h_k) = \min_{j=1, \dots, N} \hat{\gamma}^{(j)}(h_k)$ et supérieures $\hat{\gamma}_{sup}(h_k) = \max_{j=1, \dots, N} \hat{\gamma}^{(j)}(h_k)$. Ces deux enveloppes définissent un intervalle de confiance approximatif empirique de niveau $1 - 2/(N + 1)$, car

$$P(\hat{\gamma}^{(j)}(h_k) < \hat{\gamma}_{inf}(h_k)) = P(\hat{\gamma}^{(j)}(h_k) > \hat{\gamma}_{sup}(h_k)) \leq 1/(N + 1).$$

On représente alors graphiquement les fonctions $\hat{\gamma}_{inf}(h_k)$, $\hat{\gamma}_{sup}(h_k)$ et $\hat{\gamma}^{(0)}(h_k)$. Si le variogramme empirique initial $\hat{\gamma}^{(0)}(h_k)$ sort de l'enveloppe $[\hat{\gamma}_{inf}(h_k), \hat{\gamma}_{sup}(h_k)]$ pour l'une des distances h_k , on rejette le modèle $\gamma(h; \theta)$. Il faut bien noter qu'on ne rejette pas seulement l'estimateur $\gamma(h; \hat{\theta})$, mais bien toute la famille paramétrique. En effet, $\gamma(h; \hat{\theta})$ étant le meilleur estimateur dans cette famille, si celui-ci est rejeté, alors toute la famille doit l'être également. Si le variogramme empirique initial reste toujours dans l'intervalle de confiance, on ne rejette pas le modèle $\gamma(h; \theta)$ et $\gamma(h; \hat{\theta})$ en est son estimateur.

Cette procédure, bien que simple et élégante, souffre cependant de deux défauts :

- En réalité, on ne sait simuler que des vecteurs distribués selon une loi multigaussienne. Cette procédure n'est donc en théorie valide que pour des échantillons qui peuvent raisonnablement être modélisés par des vecteurs gaussiens. Dans le cas où cette hypothèse n'est pas raisonnable, il peut être nécessaire de ne pas travailler directement sur les données, mais sur une transformation de celles-ci. Ce point sort du cadre de ce cours ; pour plus de détails on peut se référer par exemple à Chilès et Delfiner (1999).
- Telle que décrite ci-dessus, cette procédure fait l'impasse sur deux points majeurs : d'une part, il y a un problème de tests multiples, puisqu'on teste pour chaque distance h_k , $k = 1 \dots, K$ si $\hat{\gamma}^{(0)}(h_k)$ est ou non dans l'intervalle $[\hat{\gamma}_{inf}(h_k), \hat{\gamma}_{sup}(h_k)]$. Il faudrait donc ajuster le niveau réel pour prendre en compte ce phénomène. Mais d'autre part, les valeurs du variogramme empirique initial à des distances k et k' ne sont pas indépendantes. Cette dépendance est très difficile à prendre en compte et est en générale négligée. Elle vient en partie contrebalancer l'effet des tests multiples. Au final, cette procédure est plutôt conservatrice ; elle tend à maintenir (à ne pas rejeter) l'hypothèse validée.

Cette procédure permet de valider **un** modèle de variogramme, mais ne permet pas de sélectionner parmi un ensemble de modèles, ce que permet la procédure suivante.

La validation croisée

Le principe de la validation croisée est d'éliminer à tour de rôle chaque observation et de la prédire à l'aide des autres observations en utilisant le krigeage⁶, sans réestimer le modèle de variogramme⁷. En chaque site x_i , $i = 1, \dots, n$ on dispose donc de la valeur observée, $Z_i = Z(x_i)$, d'une valeur prédite par krigeage $Z_i^K = Z_i^k(x_i)$ et d'une variance de krigeage associée, $\sigma_{K,i}^2$. A partir de ces valeurs, on peut construire trois indicateurs :

$$\begin{aligned} BIAIS &= 1/n \sum_{i=1}^n (Z_i^K - Z_i), \\ EQM &= 1/n \sum_{i=1}^n (Z_i^K - Z_i)^2, \\ EQNM &= 1/n \sum_{i=1}^n (Z_i^K - Z_i)^2 / \sigma_{K,i}^2. \end{aligned}$$

Le premier critère est un biais, et le second l'erreur quadratique moyenne. Le dernier critère est la moyenne de l'erreur quadratique normalisée (i.e. divisée par la variance de krigeage qui n'est autre que l'espérance de l'erreur quadratique). Si le modèle de variogramme est valide, EQNM doit être proche de 1. Si en première approximation, on suppose que les erreurs normalisées $(Z_i^K - Z_i)/\sigma_{K,i}$ sont des normales centrées réduites indépendantes, on validera le modèle si

$$q(n, \alpha/2) \leq nEQNM \leq q(n, 1 - \alpha/2) \quad (26)$$

où $q(n, \beta)$ désigne le β -quantile d'une χ_n^2 . Ce dernier test permet de rejeter ou non un modèle $\gamma(h; \theta)$. L'observation des critères $BIAIS$, EQM et $EQNM$ permet de sélectionner le meilleur modèle parmi un ensemble de modèles. Dans un objectif de prédiction, on va chercher l'EQM minimum. Pour autant, les variances de krigeage annoncées doivent en adéquation avec les erreurs quadratiques constatées. On ne peut donc accepter un modèle ayant un EQNM hors de l'intervalle (26). Il peut être préférable de sélectionner un modèle présentant un EQM légèrement moins bon, si son EQNM est proche de 1.

6. Voir la Section ??

7. Dans la littérature sur la validation croisée, cette procédure s'appelle "Leave One out". Une forme plus générale consiste à éliminer p observations et à les prédire à l'aide des $n - p$ autres observations ; cette procédure s'appelle "Leave p Out". Déterminer le p optimal est un sujet de recherche actif, et il n'y a pas de réponse définitive sur la meilleure valeur de p . Mais on sait que $p = 1$ est un choix qui n'est jamais mauvais.

References

- Billingsley, P. (1995) *Probability and Measure*, third ed. Wiley, New-York.
- Chilès, J.-P. and Delfiner, P. (1999) *Geostatistics : modeling spatial uncertainty*, Wiley.
- Cressie, N. (2003) *Statistics for Spatial Data*, revised ed., Wiley.
- Gihman, I.I. and Skorohod, A.V. (1974) *The theory of Stochastic Processes*, vol. 1, Springer Verlag.
- Matern, B. (1960) *Spatial Variation*, Springer-Verlag.
- Stein, M. (1999) *Interpolation for Spatial Data : Some Theory for Kriging*, Springer-Verlag.
- Wackernagel, H. (2003) *Multivariate Geostatistics : An Introduction with Application*, Third Edition, Springer.
- Yaglom, A.M. (1987) *Correlation Theory of Stationary and Related Random Functions I. Basic Results*, Springer-Verlag.