

L'ESTIMATION GÉOSTATISTIQUE NON LINÉAIRE

par

Christian LAJAUNIE

Cours C-152

Novembre 1993

CENTRE DE GÉOSTATISTIQUE

35, RUE SAINT-HONORÉ, 77305 FONTAINEBLEAU (France)



ECOLE DES MINES
DE PARIS

TABLE DES MATIÈRES

Introduction	1
Chapitre 1 : Domaine d'application des méthodes non linéaires	3
1.1. — Introduction	3
1.2. — Besoin d'une estimation non linéaire	4
1.2.1. — Cas d'un vecteur gaussien	4
1.2.2. — Un contre exemple	5
1.3. — Un problème type : les dépassements de niveau	9
1.3.1. — Une application inappropriée du krigeage	9
1.3.2. — Quelques formes d'estimateurs	11
1.4. — Le problème de l'estimation des réserves	15
1.4.1. — L'effet de support	17
1.4.2. — L'effet d'information	17
Chapitre 2 : Estimation locale au support ponctuel par espérance conditionnelle	19
2.1. — Introduction	19
2.1.1. — L'espérance conditionnelle	19
2.2. — Espérance conditionnelle en modèle multinormal	20
2.2.1. — Estimation d'une indicatrice	22
2.2.2. — Utilisation du développement de h	22
2.3. — Applications aux anamorphoses de f.a. gaussiennes	23
2.3.1. — Le krigeage lognormal	23
2.4. — Construction d'une anamorphose gaussienne	24
2.4.1. — Anamorphose empirique	25
2.4.2. — Cas d'une loi non continue	27
2.4.3. — Estimation de quantité de métal	28
2.4.4. — Remarques	28
Chapitre 3 : Le krigeage disjonctif ponctuel	30
3.1. — Les équations du krigeage disjonctif	32
3.2. — Lois isofactorielles	33
3.3. — Le krigeage disjonctif isofactoriel	37
3.3.1. — Biais et variance	38
3.3.2. — Pseudo-densité k.d.	38
3.3.3. — Sur l'hypothèse " <i>H fermé</i> "	39

3.4. — Application à une FAST	39
3.4.1. — Remarques	42
3.4.2. — Importance du choix du modèle	43
3.4.3. — Modèles à résidus d'indicatrices autokrigeables	44
Chapitre 4 : Les modèles isofactoriels discrets symétriques .	51
4.1. — Processus de vie et de mort	51
4.2. — Application à des fonctions aléatoires	57
4.3. — Estimation des paramètres	59
4.3.1. — Inférence des coefficients de diffusion	62
Chapitre 5 : Estimation locale avec changement de support en modèle multigaussien discrétisé	65
5.1. — Rappels sur le modèle utilisé en estimation globale	65
5.2. — Passage aux lois multivariées	67
5.3. — L'espérance conditionnelle en modèle gaussien discrétisé	70
5.4. — Le krigeage disjonctif en modèle gaussien discrétisé	72
5.5. — Extension du krigeage disjonctif à des modèles non gaussiens	73
<i>Exercices</i>	
5.6. — Exercice 1 : Sur le modèle mosaïque	75
5.7. — Exercice 2	76
5.8. — Exercice 3 : Estimation par cokrigeage simple	77
5.9. — Exercice 4 : Effet d'information	77
5.10. — Exercice 5 : Présence de dérive	79
5.11. — Exercice 6 : Données sous forme d'intervalles	80
<i>Annexe A : Rappels élémentaires sur les projections dans des espaces de Hilbert</i>	81
<i>Annexe B : Rappels sur les lois gaussiennes</i>	88
B.1. — Loi normale à une dimension	88
B.2. — Loi normale vectorielle	89
<i>Références</i>	92

Introduction

Ce texte est constitué de notes relatives à la seconde partie du cours de géostatistique non linéaire, enseigné à Fontainebleau dans le cadre du DEA. Il fait suite au cours de sélectivité, dont il suppose connus les résultats essentiels. Toutefois, on trouvera en annexe, énoncés sans démonstration, les propriétés élémentaires utilisées. Cela concerne essentiellement les projections dans des sous espaces de Hilbert, et les lois gaussiennes.

En ce qui concerne les problèmes d'estimation, le cours de selectivité adopte le point de vue de l'estimation globale. Cela suppose que les quantités à estimer sont des moyennes sur des domaines suffisamment grands. Ces quantités peuvent alors être assimilées à des constantes, égales à des espérances, dans le cadre des modèles probabilistes ergodiques utilisés. La seconde partie du cours, développée ici, adopte le point de vue de l'estimation locale, dans lequel cette approximation n'est plus valable. Les grandeurs à estimer sont alors variables d'une réalisation à l'autre, ou tout aussi bien en translatant les domaines dans l'espace, à réalisation fixée. L'estimation se fera en conditionnant par les données de voisinage.

Le premier chapitre est destinée à motiver la suite du texte. On tente de dégager sur une base intuitive, les conditions qui rendent nécessaire le recours à des méthodes non linéaires. Le problème type utilisé par la suite, est celui de l'estimation des dépassements de niveaux. Il peut apparaître dans l'estimation d'un risque, ou dans l'estimation d'histogrammes locaux. Une attention particulière est portée au problème de l'estimation de réserves récupérables dans des panneaux miniers. En effet, outre le fait qu'il s'agit du domaine d'application dans lequel l'expérience est la plus riche, l'accumulation de difficultés présentée en fait un excellent problème prototype. Ainsi est rappelé le problème de la prise en compte de l'effet de support et de l'effet d'information.

Le second chapitre traite de l'estimation non linéaire au support ponctuel, par espérance conditionnelle. On insiste assez peu sur cette méthode, bien qu'elle fournisse en principe de la meilleure estimation. De fait, la méthode est tributaire de la bonne adéquation des modèles de fonctions aléatoires gaussiennes, et l'expérience montre que les données sont assez souvent incompatibles avec ces

modèles (par exemple en présence d'un fort effet zéro¹). Une place plus importante est faite au krigeage disjonctif, qui fait l'objet du chapitre suivant. Cela se justifie autant par l'originalité de la méthode que par la place que la pratique lui a faite. Sont ainsi proposés plusieurs modèles de lois bivariées permettant la mise en oeuvre du krigeage disjonctif. Le modèle bigaussien est utilisé comme exemple introductif, et des modèles plus récents sont rapidement décrits, comme le modèle à résidus d'indicatrices autokrigeable, et les modèles isofactoriels discrets. La diversité de modèles permettant la mise en oeuvre du krigeage disjonctif est un des intérêts essentiels de la méthode.

Le dernier chapitre traite de l'estimation avec effet de support. La transposition des techniques d'estimation du chapitre précédent suppose l'incorporation d'un modèle de changement de support. Cela entraîne une modélisation plus spécifique que celle nécessaire en estimation globale. On se limite dans l'exposé au modèle gaussien discret, bien que les autres modèles isofactoriels possèdent des versions non symétriques permettant l'estimation en présence d'effet de support. La bibliographie proposée en fin de texte permet un approfondissement de cet aspect.

1. proportion non négligeable de données nulles.

Chapitre 1

Domaine d'application des méthodes non linéaires

1.1 Introduction

Ce chapitre vise à préciser les domaines d'application des méthodes non-linéaires. Les estimateurs d'une grandeur T - qui peut être un scalaire ou une variable aléatoire - formés sur la base des informations Z_1, \dots, Z_n sont dans le cas le plus général des fonctions de n variables:

$$T^* = \Phi(Z_1, \dots, Z_n)$$

A été considéré dans le cours de géostatistique linéaire le cas particulier où cette dépendance est linéaire, ou affine. C'est le cas pour les krigeages intrinsèques :

$$Z^{k.o.} = \sum_i \lambda_i Z_i$$

et du krigeage à moyenne connue (krigeage simple):

$$Z^{k.s.} = m \left(1 - \sum_j \lambda_j \right) + \sum_i \lambda_i Z_i$$

Ces estimateurs ont le bon goût d'être simples et peu exigeants. Ainsi, dans le premier cas, le choix de l'ordre d'un modèle de fonction aléatoire intrinsèque et la spécification d'une covariance généralisée suffisent à déterminer l'estimateur.

Cependant, il est des circonstances où ces estimateurs sont insuffisants, et où le recours à des estimateurs non linéaires par rapport aux données est nécessaire. Ce recours a son prix, perceptible à deux niveaux. En terme d'exigence de modélisation d'abord, puisque sera fait appel aux propriétés distributionnelles, et non plus seulement aux moments d'ordre inférieur ou égal à deux. Enfin, en terme

de restriction sur la classe de modèles utilisables, puisque le plus souvent on se limitera à des modèles stationnaires.

Les principaux estimateurs que l'on considérera sont des projections dans des sous-espaces de Hilbert de $\mathcal{L}^2[\Omega, \mathcal{A}, P]$, qui contiennent les constantes. Il en résulte les propriétés suivantes:

- **absence de biais:** $E[T - T^*] = 0$.
- **variance minimale** (dans leur catégorie respective). Ainsi les poids de krigeage simple sont déterminés par la condition:

$$\min_{\lambda} \{Var(Z - \sum \lambda_i Z_i)\}$$

De même, les fonctions de pondérations f_i du krigeage disjonctif devront réaliser le minimum de:

$$\min_f \{Var(Z - \sum f_i(Z_i))\}$$

Enfin, la minimisation dans un ensemble des fonctions de n variables, $f(z_1, \dots, z_n)$, conduit à l'espérance conditionnelle.

- **Linéarité** par rapport à l'objet de l'estimation:

$$(\sum \alpha_u T_u)^* = \sum \alpha_u (T_u)^*$$

(La non linéarité porte sur les données elles mêmes)

1.2 Besoin d'une estimation non linéaire

Nous allons maintenant essayer de préciser les circonstances dans lesquelles les estimateurs linéaires échouent. Remarquons d'abord que, déterminés par les covariances, ils ne peuvent être efficaces que si ces covariances sont de bons indicateurs de liens entre variables. Il n'en est pas toujours ainsi.

1.2.1 Cas d'un vecteur gaussien

Commençons par un cas très favorable, celui d'un vecteur gaussien (ou d'une fonction gaussienne si nous considérons une fonction aléatoire). De fait nous savons que si le vecteur:

$$(X_1, \dots, X_n, Y_1, \dots, Y_m)$$

suit une loi de Gauss à $m + n$ dimensions, alors l'équivalence suivante a lieu:

$$\left\{ \begin{array}{l} \forall i = 1, \dots, n \\ \forall j = 1, \dots, m \end{array} \text{cov}(X_i, Y_j) = 0 \right\} \iff (X_1, \dots, X_n) \text{ ind. } (Y_1, \dots, Y_m)$$

(l'indépendance dans la proposition de droite porte sur les vecteurs de dimensions respectives n et m). En particulier, $\text{Cov}(X_1, X_2) = 0$ équivaut à l'indépendance de X_1 et X_2 . La corrélation est donc un très bon indicateur de liens dans ce cas. De fait, nous savons que le krigeage simple est alors identique à l'espérance conditionnelle, et épuise donc l'information apportée par les données.

1.2.2 Un contre exemple

Un exemple très défavorable est proposé dans la suite sous forme d'exercice. Cet exercice est en deux parties:

Exercice - Première partie - Soit X une variable aléatoire admettant une densité $f(x)$ paire, et donc telle que: $f(-x) = f(x)$. Soit $Y = \phi(X)$ une variable liée à X par une fonction ϕ également paire.

Un observateur (ignorant la relation fonctionnelle entre X et Y) note un grand nombre de réalisations indépendantes du couple (X, Y) , et on admet qu'il peut ainsi mesurer exactement les moyennes, variances et covariances des variables. Il cherche ensuite à ajuster un modèle de régression linéaire:

$$Y^* = m_Y + \lambda (X - m_X)$$

Que trouve t-il?

Réponse. Le pondérateur λ vaut:

$$\lambda = \frac{C_{XY}}{\sigma_X^2}$$

La covariance non centrée est:

$$E[XY] = \int_{-\infty}^{+\infty} x\phi(x)f(x)dx$$

La fonction à intégrer étant impaire, on a donc $E[XY] = 0$ (voir la figure 1.1). Comme $m_X = 0$ la covariance elle même est nulle, on a finalement $Y^* = m_Y$.

Conclusion La régression linéaire échoue à utiliser l'information portée par X , alors que la dépendance est fonctionnelle.

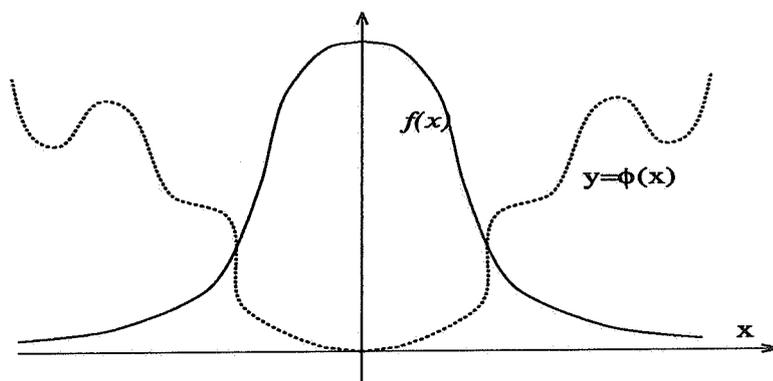


FIG. 1.1 - : Densité $f(x)$ et fonction lien ϕ

En pratique, si notre observateur avait commencé par regarder le nuage de corrélation, il aurait noté l'existence de la relation entre X et Y , ce qui lui aurait suggéré l'estimateur suivant:

$$Y^{**} = m_Y + \lambda' (\phi(X) - m_\phi)$$

Cet estimateur retrouve toute l'efficacité possible.

On notera cependant que dans un contexte multivariable, et en présence de bruit additionnel, il peut être beaucoup plus difficile de détecter ce genre de relation. La suite de l'exercice montre que cela peut dans une certaine mesure être réalisé automatiquement.

Commençons par le résultat élémentaire suivant, qui a le mérite de simplifier grandement la suite de la discussion.

Orthogonalisation de Schmidt. *En supposant que tous les moments de X existent, montrer qu'il est possible de constituer une famille orthonormée $\chi_0, \dots, \chi_n, \dots$, où les χ_n sont des polynômes de degré n en X :*

$$\chi_n = a_{nn}X^n + \dots + a_{n1}X + a_{n0}$$

à coefficients principaux $a_{nn} > 0$.

On pose $\chi_0 = 1$. puis $V_1 = X + b_{10}\chi_0$. L'orthogonalité de V_1 et χ_0 est assurée si $b_{10} = -\langle X, \chi_0 \rangle$. Il suffit ensuite de prendre:

$$\chi_1 = \frac{V_1}{\|V_1\|}$$

ce qui est possible puisque $V_1 \neq 0$. Alors (χ_0, χ_1) est une famille orthonormée.

L'orthogonalisation se fait ensuite de proche en proche: si $\chi_0, \dots, \chi_{n-1}$ est une famille orthonormée, on pose:

$$V_n = X^n + b_{n,n-1} \chi_{n-1} + \dots + b_{n0} \chi_0$$

V_n est alors orthogonal à $\chi_0, \dots, \chi_{n-1}$ si:

$$\langle V_n, \chi_i \rangle = \langle X^n, \chi_i \rangle + b_{ni} = 0 \text{ pour } i = 0, \dots, n-1$$

Comme $V_n \neq 0$ puisque $X^n \notin \mathcal{L}(X^0, \dots, X^{n-1})^1$, nous pouvons prendre:

$$\chi_n = \frac{V_n}{\|V_n\|} \quad \text{et} \quad a_{nn} = \frac{1}{\|V_n\|} > 0$$

Alors χ_0, \dots, χ_n est une famille orthonormée.

Nous sommes maintenant en mesure de traiter la suite de l'exercice.

Seconde partie On décide alors d'améliorer la régression linéaire vue dans l'exercice précédent en passant à une régression multiple sur les monômes de X :

$$Y_n^* = \lambda_0^n + \lambda_1^n X + \dots + \lambda_n^n X^n$$

Y_n^* admet une expression en terme des χ_n :

$$Y_n^* = c_0^n + c_1^n \chi_1 + \dots + c_n^n \chi_n$$

Donner l'expression des c_i^n , et la variance de l'erreur $Y - Y_n^*$. A quelles conditions a-t-on:

$$\lim_{n \rightarrow \infty} Y_n^* = Y ?$$

La régression multiple vérifie:

$$\forall i \leq n \quad \langle Y - Y_n^*, X^i \rangle = 0$$

ou de manière équivalente:

$$\forall i \leq n \quad \langle Y - Y_n^*, \chi_i \rangle = 0$$

Grace à l'orthogonalité des χ_i , cela implique:

$$c_i^n = \langle Y, \chi_i \rangle = \int \phi(x) \chi_i(x) f(x) dx = c_i$$

1. X étant une variable continue, une relation du type $\sum c_i X^i = 0$ implique que les c_i sont tous nuls.

Les coefficients c_i^n ne dépendent donc pas de n tant que $i \leq n$ (l'indice n sera donc omis par la suite). Du fait que $\chi_0 = 1$, l'estimateur est sans biais:

$$E[Y - Y_n^*] = \langle Y - Y_n^*, \chi_0 \rangle = 0$$

En ce qui concerne la variance, on a:

$$\begin{aligned} \text{Var}(Y - Y_n^*) &= \langle Y - Y_n^*, Y - Y_n^* \rangle = \langle Y - Y_n^*, Y \rangle \\ &= \|Y\|^2 - \sum_{i=0}^n c_i \langle \chi_i, Y \rangle \end{aligned}$$

Comme $c_0 = \langle \chi_0, Y \rangle = E[Y]$, la variance de Y vaut $\|Y\|^2 - c_0^2$. Il reste:

$$\text{Var}(Y - Y_n^*) = \text{Var}(Y) - \sum_{i=1}^n c_i^2$$

Finalement, $S_n = \sum_{i=1}^n c_i^2 \leq \text{Var}(Y)$ tend vers une limite S_∞ lorsque $n \rightarrow \infty$. Si $S_\infty = \text{Var}(Y)$ alors $Y^{*n} \rightarrow Y$ en moyenne quadratique.

Remarques:

- L'exemple choisi est extrême en ce que la transformation symétrique ϕ détruit complètement la corrélation. Plus généralement, on doit s'attendre à ce qu'une transformation non linéaire, sans aller jusqu'à la destruction de corrélation, entraîne une perte d'efficacité des estimateurs linéaires.
- L'enrichissement de l'espace d'approximation:

$$(1, X) \quad \longrightarrow \quad (1, X, X^2, \dots)$$

a permis de trouver le codage optimal $\phi(X)$.

- Les familles orthogonales telles que χ_0, χ_1, \dots simplifient les expressions des régressions multiples.
- Cette méthode de recherche de codage optimal de prédicteurs sera utilisable dans un cas multivariable, où $\phi(X_0)$ sera à estimer à partir d'un ensemble de m variables X_1, X_2, \dots, X_m .

Nous allons maintenant introduire un problème type, où ce genre de difficultés se retrouve.

1.3 Un problème type: les dépassements de niveau

De nombreux phénomènes présentent des seuils critiques. Citons par exemple:

- **En économie:** Il existe des seuils de rentabilité de projets. On accordera un intérêt particulier au problème de la sélection d'unités économiquement rentables dans un contexte minier.
- **En agriculture:** Des concentrations minimales en éléments nutritifs dans les sols, sont nécessaires pour assurer la viabilité des cultures. Des concentrations trop faibles demandent l'apport d'engrais spécifiques. Le même problème concerne aussi l'élevage.
- **En pollution:** Il existe des niveaux maximaux admissibles, au delà desquels une action de correction est nécessaire.

Dans ces exemples, les seuils critiques sont connus, et dépendent d'un contexte économique, de données biologiques, ou de la législation. Une grandeur d'intérêt peut être la proportion d'un domaine fixe D dans lequel le dépassement a lieu (pourcentage de forêt polluée par exemple):

$$R = \frac{|\{x \in D : Z(x) \geq z_c\}|}{|D|}$$

cette quantité peut encore s'écrire:

$$R = \frac{1}{|D|} \int_D 1_{Z(x) \geq z_c} dx$$

Le phénomène Z étant reconnu par un certain nombre de points de mesure $Z(x_1), \dots, Z(x_n)$, le problème posé au géostatisticien est celui de l'estimation de R sur la base des $Z(x_i)$. Nous allons d'abord envisager rapidement plusieurs possibilités. Les plus intéressantes seront reprises dans les chapitres suivants.

Commençons par une démarche naïve, qui échoue :

1.3.1 Une application inappropriée du krigeage.

Le krigeage ponctuel nous donne une interpolation optimale du phénomène étudié:

$$\begin{aligned} Z^{k.o.}(x) &= \sum \lambda_i^{k.o.}(x) Z(x_i) \\ Z^{k.s.}(x) &= \sum \lambda_i^{k.s.}(x) Z(x_i) + (1 - \sum \lambda_j^{k.s.}(x)) \end{aligned}$$

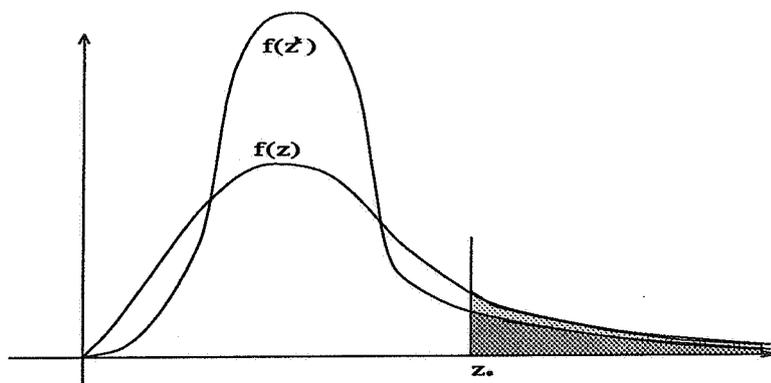


FIG. 1.2 - : Densité de la loi stationnaire de Z , et des valeurs Z^k prises par l'estimateur.

Nous pouvons alors calculer la proportion de dépassement de seuil pour cet interpolateur optimal, ce qui donne l'estimation:

$$R^* = \frac{1}{|D|} \int_D 1_{Z^K(x) \geq z_c} dx$$

Cette estimation est malheureusement inacceptable car entachée d'un biais considérable. La raison essentielle à cela est que le krigeage, conçu pour donner une estimation optimale ponctuelle, n'est pas apte à reproduire la variabilité du phénomène dans D . En d'autres termes l'histogramme des valeurs krigées dans D est très différent de l'histogramme réel. Pour voir cela, il suffit de considérer par exemple la relation de variance du krigeage simple, qui résulte de la décomposition suivante de Z :

$$Z = (Z - Z^k) + Z^k$$

en deux termes orthogonaux. D'où:

$$\sigma_Z^2 = \sigma_{Z-Z^k}^2 + \sigma_{Z^k}^2$$

Cela implique que la variance des valeurs krigées est sensiblement plus faible que la variance des valeurs réelles, et ce d'autant plus que l'erreur d'estimation est grande. La figure 1.2 illustre cette situation en termes d'histogrammes. Pour une coupure z_c élevée, par exemple, la proportion de valeurs krigées telles que $Z^k \geq z_c$ est donc beaucoup plus faible que la proportion de $Z(x) \geq z_c$, d'où résulte le biais mentionné.

Dans le cas limite d'une fonction aléatoire purement pépétique, le krigeage simple est constant et égal à la moyenne: $Z^K = m$, alors que la proportion au dessus d'un seuil est donnée par la loi marginale $R = 1 - F(z_c)$. L'estimation obtenue $R^* = 1_{m \geq z_c}$ est donc inacceptable.

En ce qui concerne l'allure des fonctions Z et Z^K , la figure 1.3 illustre une situation assez générale: le krigeage est beaucoup plus lisse que le phénomène

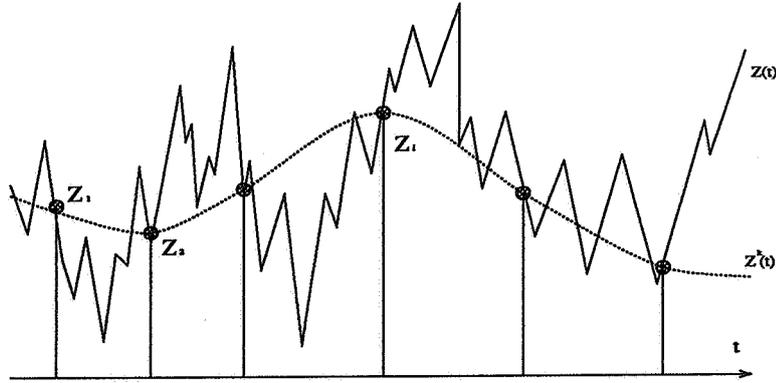


FIG. 1.3 - : L'estimateur $Z^k(x)$ présente un aspect plus régulier que le phénomène estimé Z

sous-jacent, et cela d'autant plus que l'information $Z(x_1), \dots, Z(x_n)$ est pauvre ou que le phénomène est peu structuré.

Plus généralement, nous pouvons formuler la remarque suivante:

Si la fonction ϕ est non linéaire, $\phi(Z^k)$ n'est pas, en général, un estimateur acceptable de $\phi(Z)$.

1.3.2 Quelques formes d'estimateurs

Il s'agira d'estimateurs projectifs, pour lesquels nous avons linéarité par rapport à l'objectif d'estimation (voir l'annexe A):

$$R^* = \frac{1}{|D|} \int_D (1_{Z(x) \geq z_c})^* dx$$

Il suffit donc d'estimer chaque indicatrice $1_{Z(x) \geq z_c}$ individuellement. Nous venons de voir que $1_{Z^k \geq z_c}$ n'est pas acceptable. D'autres choix sont envisageables. On notera $T(z) = E[1_{Z > z}]$ pour alléger un peu la notation, et m_Z l'espérance $E[Z(x)]$.

- Dans un cokrigage simple, la forme suivante est utilisée:

$$(1_{Z \geq z_c})^{c.k.s} = T(z_c) + \sum \lambda_i (Z(x_i) - m_Z)$$

Cette technique ne souffre pas de critiques aussi graves que l'estimateur naïf indiqué précédemment. Cependant, compte tenu de la mauvaise qualité de $Cov(1_{Z(x) \geq z_c}, Z(x_i))$ en tant que mesure de dépendance, l'utilisation

effective de l'information portée par les $Z(x_i)$ est mauvaise. Considérons par exemple le cas d'une fonction Z continue en moyenne quadratique, ainsi que son indicatrice $1_{Z \geq z_c}$, et estimée en x à partir du seul point $Z(y)$. Si le point de donnée y se rapproche du point à estimer x , on aura:

$$1_{Z(y) \geq z_c} \rightarrow 1_{Z(x) \geq z_c}$$

et donc convergence de l'estimateur formé par le terme de gauche: $1_{Z(y) \geq z_c}$. Cela n'est pas le cas de l'estimation par cokrigeage simple à partir de $Z(y)$:

$$(1_{Z(x) \geq z_c})^{c.k.s} = T(z_c) + \lambda(y)[Z(y) - m]$$

Ce qui suffit à disqualifier cette méthode. Cela suggère aussi qu'une estimation basée sur des variables de même nature que la variable à estimer (indicatrices pour estimer une indicatrice), est préférable à une estimation hétérogène (variables brutes pour estimer une indicatrice). D'où la procédure suivante:

- Dans la technique des **variables utiles**, $\phi(Z(x))$ sera estimé par combinaison linéaire des $\phi(Z(x_i))$:

$$[\phi(Z(x))]^* = m_\phi + \sum_i \lambda_i [\phi(Z(x_i)) - m_\phi]$$

Appliqué aux indicatrices à coupure z_c , cela donne le **krigeage des indicatrices**:

$$(1_{Z(x) \geq z_c})^{k.i.} = T(z_c) \cdot (1 - \sum \lambda_j) + \sum_i \lambda_i 1_{Z(x_i) \geq z_c}$$

Cette méthode est acceptable, contrairement aux deux méthodes précédentes. Elle présente en outre plusieurs avantages sur les méthodes plus élaborées que nous allons décrire ultérieurement. Sa mise en oeuvre est à priori peu exigeante, puisque la méthode est utilisable par quiconque dispose d'un programme de krigeage linéaire. Elle est conceptuellement simple et naturelle, ce qui a contribué dans une large mesure à sa large diffusion.

Cependant, elle entraîne une perte d'information regrettable, puisque savoir que $Z(x_i)$ est juste au dessus de la coupure z_c ne signifie en général pas la même chose, en ce qui concerne le comportement de $Z(x)$ dans un voisinage de x_i que, par exemple, $Z(x_i) = 10 \times z_c$. La méthode ignore cette différence, puisque l'indicatrice vaut 1 dans les deux cas.

- Dans le but de minimiser cette perte d'information, il est possible de considérer simultanément différentes coupures z_1, \dots, z_k . Soient donc $I_\alpha(x) = 1_{Z(x) \geq z_\alpha}$ les indicatrices associées aux coupures introduites, et $I_\alpha^i = I_\alpha(x_i)$

les valeurs aux points informés de ces indicatrices. L'estimateur correspondant est le **cokrigage des indicatrices**:

$$[1_{I_\alpha}(x)]^{c.k.i.} = T(z_\alpha) + \sum_{i,\beta} \lambda_{i,\beta} [I_\beta^i - T(z_\beta)]$$

Plus le nombre de coupures considéré simultanément est élevé, plus faible est la perte d'information due au codage par indicatrices.

Outre l'estimation des indicatrices aux coupures introduites, est possible par linéarité l'estimation de toute fonction h constante par morceaux, de la forme:

$$h(Z) = h_0 + \sum_{\alpha=1}^k h_\alpha [I_\alpha - I_{\alpha+1}]$$

(avec $z_{k+1} = +\infty$ et $I_{k+1} = 0$). Un tel estimateur utilise les covariances croisées de toutes les indicatrices:

$$\text{Cov}[I_\alpha(x+h), I_\beta(x)] = \langle I_\alpha(x+h), I_\beta(x) \rangle - T(z_\alpha)T(z_\beta)$$

et donc suppose leur estimation ainsi que leur modélisation. Outre sa lourdeur, une telle modélisation n'est pas sans soulever de délicats problèmes de cohérence, puisque la donnée de ces covariances équivaut à la spécification de la loi bivariable de l'indice de classe (fonction qui vaut α en x si $Y(x) \in]z_\alpha, z_{\alpha+1}]$). Des contraintes assez fortes existent donc sur ces covariances, et il est difficile de choisir une modélisation satisfaisante à moins justement de se baser sur une loi bivariable.

Remarquons que, en tant que fonction des valeurs $Z(x_i)$, le cokrigage des indicatrices est de la forme:

$$(\mathbf{1}_{Z(x) \geq z_\alpha})^{c.k.i.} = m(z_\alpha) + \sum_i \phi_i(Z(x_i))$$

avec des fonctions ϕ_i constantes par intervalles (figure 1.4):

$$\phi_i(z) = \sum_{\beta} c_{i,\beta} \mathbf{1}_{z \in [z_\beta, z_{\beta+1}[}$$

et d'espérance nulle. C'est dans un tel espace de fonctions ϕ_i que l'optimisation qui conduit au cokrigage des indicatrices est fait.

- Partant de l'estimateur défini précédemment, imaginons une augmentation indéfinie du nombre de coupures, de telle manière que la probabilité de chaque intervalle $P[Z(x) \in]z_\alpha, z_{\alpha+1}]$ tende vers zéro. Les fonctions de pondérations précédentes ϕ_i deviennent alors des fonctions Borel-mesurables d'espérance nulle et de moment d'ordre deux finis, (afin de ne pas sortir du cadre de travail choisi) mais à cela près quelconques.

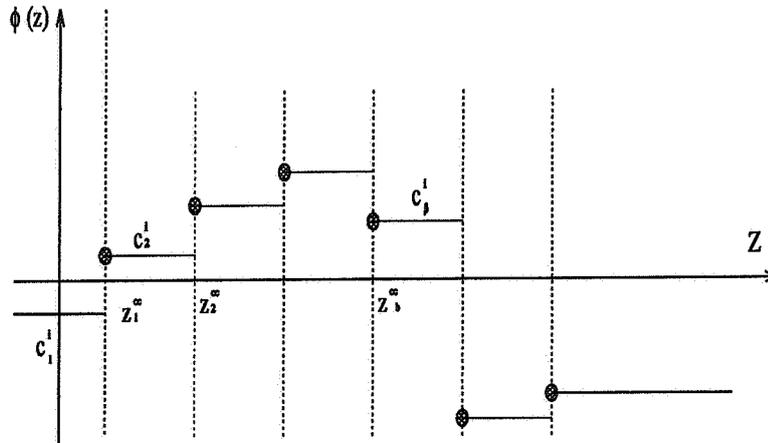


FIG. 1.4 - : Les fonctions de pondération du cokrigeage des indicatrices sont constantes sur les segments délimités par les coupures, et d'espérance nulle

La forme d'estimateur que l'on obtient par ce passage à la limite est le krigage disjonctif continu:

$$(1_{Z(x) \geq z_c})^{k.d.} = T(z_c) + \sum_i \phi_i(Z(x_i))$$

(Dans cette expression, le terme de moyenne peut tout aussi bien être incorporé aux fonctions ϕ_i , à condition de remplacer la condition $E[\phi_i(Z)] = 0$ par une condition de non biais sur l'estimateur.)

Pour une fonction aléatoire discrète ne prenant qu'un nombre fini de valeurs, le cokrigeage des indicatrices prenant en compte toutes les coupures et le krigage disjonctif coïncident donc.

- Enfin, la forme la plus générale d'estimateur admissible est une fonction de n variables, de variance finie. La minimisation de l'erreur quadratique pour un tel estimateur conduit à l'espérance conditionnelle:

$$E[1_{Z(x) \geq z_c} | Z(x_1), \dots, Z(x_n)] = \phi[Z(x_1), \dots, Z(x_n)]$$

C'est l'estimateur le plus efficace, en terme de variance d'erreur, que l'on puisse construire sur la base des $Z(x_i)$.

Nous reviendrons par le suite en détails sur ces deux dernières méthodes d'estimation. Nous allons maintenant considérer plus en détails le problème de l'estimation de réserves récupérables en mine, dans le cas d'une exploitation sélective.

1.4 Le problème de l'estimation des réserves

Il s'agit, ainsi qu'il a été dit dans l'introduction, d'un problème type, en ce qu'il permet d'illustrer des difficultés présentes dans d'autres applications. Il est en effet assez vraisemblable, que dans la plupart des applications nécessitant un calcul d'histogramme local, un effet de support existe, d'importance plus ou moins grande.

La terminologie établie étant d'origine minière, nous continuons à l'utiliser bien que le champ d'application déborde largement le cadre des applications minières.

Le problème est donc d'estimer la valeur économique de panneaux miniers. Ces panneaux sont des portions de gisement, de géométrie donnée. On se placera dans le cas de méthodes d'exploitation conformes au schéma suivant:

1. Le gisement entier est divisé en un certain nombre de blocs de taille constante. On note V le volume de l'un quelconque de ces blocs. Ce découpage est fixé une fois pour toute, et ne sera pas remis en question au moment de l'exploitation, même s'il apparaît à ce moment qu'une modification de positionnement ou de forme conduirait à une meilleure efficacité (Cela conduirait à une géométrie adaptative, non abordée ici).
2. Les blocs sont regroupés en panneaux, de tailles non nécessairement égales. Ces panneaux constituent des entités différentes pour l'exploitation, comme pour l'évaluation, et un premier niveau de sélection écarte les panneaux les moins intéressants. L'estimation géostatistique doit donc se faire panneau par panneau.
3. Les blocs de chaque panneau exploité ne sont pas tous envoyés en usine: une sélection tend à ne conserver que les plus riches d'entre eux. Cette sélection s'opère librement de toute contrainte géométrique.

On s'intéressera à l'estimation de la valeur économique d'un panneau particulier, \mathcal{P} , et on désignera par V_i l'ensemble des blocs qui le constituent. Pour un bloc de teneur $Z(V_i)$

- Le coût de traitement est proportionnel au tonnage traité, et est donc constant par bloc (soit c ce coût).
- La valeur économique est proportionnelle à la quantité de métal contenue (soit $\alpha.Z(V_i)$ cette valeur).

Soit \mathcal{S}_p le sous-ensemble de \mathcal{P} , constitué des blocs de ce panneau sélectionnés pour traitement. Le bénéfice réalisé à l'issue du traitement, et en ne tenant compte que

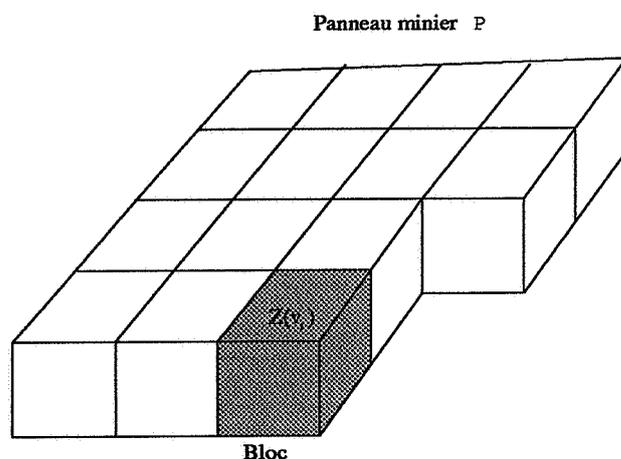


FIG. 1.5 - : *Panneau et blocs d'exploitation*

du seul coût de traitement, est donc:

$$B = \sum_{i \in \mathcal{S}_p} (\alpha \cdot Z(V_i) - c)$$

La maximisation du bénéfice est donc réalisée en ne conservant dans cette somme que les contributions positives. La règle de sélection qui se dégage est donc basée sur la teneur de coupure z_c :

$$\mathcal{S}_p = \left\{ i \in \mathcal{P} : Z(V_i) \geq z_c = \frac{c}{\alpha} \right\}$$

A laquelle est associé le bénéfice:

$$B = \sum_{Z(V_i) \geq z_c} (\alpha \cdot Z(V_i) - c)$$

Dans les conditions décrites ci dessus, la proportion de tonnage de minerai sélectionné est pour un panneau \mathcal{P} :

$$T_P(z_c) = \frac{1}{N} \sum_{i \in \mathcal{P}} 1_{Z(V_i) \geq z_c}$$

La sommation est ici effectuée sur l'ensemble des blocs du panneau. La quantité de métal récupérée à la même coupure est proportionnelle à la fonction:

$$Q_P(z_c) = \frac{1}{N} \sum_{i \in \mathcal{P}} Z(V_i) 1_{Z(V_i) \geq z_c}$$

puisque pour un bloc sélectionné, cette quantité est proportionnelle à la teneur (on obtient la quantité de métal en multipliant Q_P par le tonnage total du panneau).

Estimer les réserves récupérables c'est estimer les fonctions T_P et Q_P , pour l'ensemble des coupures économiques possibles, et pour l'ensemble des panneaux constituant le gisement. Cette estimation est effectuée sur la base des données de sondage disponibles à cette étape (il s'agit de données dont la maille est de l'ordre de la dimension des panneaux).

Deux effets dégradent de manière significative les résultats d'exploitation, et doivent donc impérativement être pris en compte par le géostatisticien. Nous introduisons ces deux effets dans les deux paragraphes suivants.

1.4.1 L'effet de support

L'estimation du tonnage T nous ramène au problème des dépassements de niveaux considéré précédemment, à ceci près que la sélection agit sur des teneurs de blocs, et non sur des teneurs de carottes de sondages. Sont donc disponibles pour former l'estimateur les indicatrices associées aux teneurs de carottes $1_{Z(v) \geq z_c}$, alors que l'on cherche à estimer des indicatrices portant sur des teneurs de blocs $1_{Z(V) \geq z_c}$. Or les teneurs de carottes présentent une dispersion beaucoup plus forte que les teneurs de blocs, comme les variances le montrent:

$$\begin{aligned}\sigma_v^2 &= \gamma(\infty) - \bar{\gamma}(v, v) \\ \sigma_V^2 &= \gamma(\infty) - \bar{\gamma}(V, V) < \sigma_v^2\end{aligned}$$

Le tonnage global récupérable pour une coupure donnée passe de

$$P[Z(v) \geq z_c] = 1 - F_v(z_c)$$

à

$$P[Z(V) \geq z_c] = 1 - F_V(z_c)$$

valeurs qui peuvent différer sensiblement.

Les autres fonctions de récupération ($Q(z)$, $Q(T)$, $B(z)$), sont également sensiblement affectées par le passage des distributions d'échantillons aux distributions de blocs. C'est l'effet de support. La prise en compte de cet effet dans l'estimation globale des réserves a été étudiée lors du cours de sélectivité. Nous verrons qu'elle est également possible en estimation locale, donc au niveau du panneau, par l'utilisation de modèles de changement de support.

1.4.2 L'effet d'information

On a supposé que le mineur était capable au moment de l'exploitation de distinguer les blocs économiques des autres, ce qui serait possible si les teneurs de blocs $Z(V_i)$ étaient connues exactement. En réalité, ces teneurs sont inconnues,

et doivent être estimées à partir de l'analyse plus ou moins précise des résidus de trous de tir. Notons Z_α les teneurs des échantillons constitués des trous de tir (connues au moment de l'exploitation, mais non au moment de l'estimation de réserves, où seules les données de sondages à grande maille sont connues). L'estimateur ultime de la teneur d'un bloc $Z(v_i)$, noté $Z(v_i)^*$, peut être par exemple la moyenne arithmétique des teneurs de tous de tirs intérieurs au bloc:

$$Z(V_i)^* = \frac{1}{n_\alpha} \sum_{\alpha} Z_\alpha$$

ou, ce qui est préférable mais demande une analyse séparée des trous de tirs, un krigeage. La sélection est alors effectuée sur la base de ces estimateurs. On a donc pour le tonnage minéral:

$$T_P(z_c) = \frac{1}{N} \sum_{i \in \mathcal{P}} 1_{Z(V_i)^* \geq z_c}$$

En ce qui concerne la quantité de métal, on sélectionne sur la base de la teneur estimée, mais on récupère une quantité proportionnelle à la teneur vraie, d'où une quantité de métal effective:

$$Q(z_c) = \frac{1}{N} \sum_{i \in \mathcal{P}} Z(V_i) 1_{Z(V_i)^* \geq z_c}$$

De cela résulte que des erreurs de sélection sont commises: des blocs estimés riches, mais en réalité de teneur $Z(v_i) < z_c$ sont envoyés au traitement, et vice versa. Il en résulte une dégradation des résultats d'exploitation. C'est l'effet d'information.

L'estimation des réserves récupérables nécessite donc l'estimation de ces nouvelles fonctions de récupération. Cela suppose la modélisation conjointe des vecteurs $(Z(V_i)$ et $Z(V_i)^*)$.

Chapitre 2

Estimation locale au support ponctuel par espérance conditionnelle

2.1 Introduction

On s'intéresse à l'estimation de quantités telles que:

$$\frac{1}{|D|} \int_D h(Z(x)) dx$$

où D est un certain domaine reconnu par un échantillonnage, et h une fonction non linéaire donnée. D'un intérêt particulier est le cas des fonctions de récupération en mine, que nous considérons ici sans effet de support ni d'information, laissant leur étude à un chapitre ultérieur:

$$\begin{aligned} \text{pour } h(y) = 1_{y \geq z} \quad \text{cela donne: } T(z_c) &= \frac{1}{|D|} \int_D 1_{Z(x) \geq z_c} dx \\ \text{pour } h(y) = y 1_{y \geq z} \quad Q(z_c) &= \frac{1}{|D|} \int_D Z(x) 1_{Z(x) \geq z_c} dx \end{aligned}$$

On cherchera à former l'espérance conditionnelle de telles quantités à $Z(x_1), \dots, Z(x_n)$ connus.

2.1.1 L'espérance conditionnelle.

Rappel (Voir annexe C) Soit (Y_0, \dots, Y_n) un vecteur aléatoire à $n + 1$ composantes. L'espérance conditionnelle de Y_0 à Y_1, \dots, Y_n fixés, que l'on note $E[Y_0 | Y_1, \dots, Y_n]$,

ou seulement $E_c[Y_0]$ s'il n'y a pas d'ambiguïté, est l'unique fonction des Y_i de variance finie vérifiant la condition d'orthogonalité de l'erreur:

$$\forall \psi \in \mathcal{H} \quad \langle Y_0 - E[Y_0 | Y_1, \dots, Y_n], \psi(Y_1, \dots, Y_n) \rangle = 0 \quad (2.1)$$

ou \mathcal{H} est l'ensemble des fonctions ψ de n variables telles que $E[\psi(Y_1, \dots, Y_n)^2] < \infty$. Cette équation met en jeu les lois à $n + 1$ variables de Y_0, Y_1, \dots, Y_n . La factorisation de ces lois:

$$F_{n+1}(dy_0, dy_1, \dots, dy_n) = F_{y_1, \dots, y_n}(dy_0) F_n(dy_1, \dots, dy_n)$$

avec $F_n = \int_{y_0} F_{n+1}$, permet de déterminer la loi conditionnelle $F_{y_1, \dots, y_n}(dy_0)$ de Y_0 à $Y_1 = y_1, \dots, Y_n = y_n$ fixés. L'espérance conditionnelle s'en déduit:

$$E[Y_0 | Y_1, \dots, Y_n] = \int u F_{Y_1, \dots, Y_n}(du)$$

Il est en effet immédiat de vérifier que cette quantité satisfait à l'équation de projection (2.1), qui caractérise $E[Y_0 | Y_1, \dots, Y_n]$. Cependant, une application dans un contexte de fonctions aléatoires suppose que nous disposions de modèles de lois multivariées satisfaisant à des exigences assez strictes. Il faut en effet:

- que les lois dépendent d'un petit nombre de paramètres pour être inférables pratiquement.
- qu'elles satisfassent à des contraintes de compatibilité, puisque les lois à n variables se déduisent des lois à $m > n$ variables par intégration.

Cela fait que seules sont utilisées en pratique les lois multigaussiennes, et nous nous plaçons dans ce contexte pour la suite de ce chapitre.

2.2 Espérance conditionnelle en modèle multi-normal

Nous supposons donc que Y est une fonction aléatoire gaussienne stationnaire, centrée et normée, et dont la covariance $C(h)$ est connue. Cette fonction est reconnue en un certain nombre de points de mesure $Y(x_1), \dots, Y(x_n)$, et nous cherchons à estimer en un point non informé x_0 , la transformée $h(Y(x_0))$. La technique d'estimation s'étend alors immédiatement par linéarité à toute transformée linéaire des valeurs prises par $h(Y(x))$ sur un domaine $\{x \in \mathcal{D}\}$, telle que $\int h(Y(x)) dx$, puisque:

$$E_c \left[\int h(Y(x)) dx \right] = \int E_c[h(Y(x))] dx$$

On se concentre donc sur l'estimation de $h(Y(x))$. Soient alors $Y^{k.s} = \lambda^i Y(x_i)$ le krigeage simple de $Y(x)$, et $\sigma_{k.s}$ l'écart-type d'estimation associé:

$$\begin{aligned} \forall i \quad \sum_j C_Y(x_i - x_j) \lambda_j &= C_Y(x_i - x_0) \\ \sigma_{k.s.}^2 &= 1 - \sum_i \lambda_i C_Y(x_i - x_0) \end{aligned}$$

Rappelons que le résidu normé:

$$U = \frac{Y(x_0) - Y^{k.s.}}{\sigma_{k.s.}}$$

est par construction orthogonal (et non corrélé) aux $Y(x_i)$:

$$\forall i = 1, \dots, n \quad \langle U, Y(x_i) \rangle = 0$$

Mais, le vecteur $(Y(x_0), Y(x_1), \dots, Y(x_n))$ étant gaussien, il en est de même de $(U, Y(x_1), \dots, Y(x_n))$. Les relations d'orthogonalité précédentes entraînent donc l'indépendance de U et $(Y(x_1), \dots, Y(x_n))$. La décomposition suivante:

$$Y(x_0) = Y^{k.s.} + \sigma_{k.s.} U$$

en un terme qui ne dépend que des $Y(x_i)$, et un terme qui leur est indépendant, donne donc la loi conditionnelle de $Y(x_0)$ à $Y(x_1), \dots, Y(x_n)$ connus: il s'agit d'une loi gaussienne d'espérance $Y^{k.s.}$, et d'écart type $\sigma_{k.s.}$. En particulier, l'espérance conditionnelle de $h(Y(x_0))$ est donc:

$$E[h(Y(x_0)) | Y(x_1), \dots, Y(x_n)] = \int h(Y^{k.s.} + \sigma^{k.s.} u) g(u) du$$

La précision de l'estimation est reflétée par la dispersion de la loi conditionnelle. Celle-ci traduit le niveau d'incertitude résiduel, compte tenu de la position spatiale des points de données, et de leur valeur. Cette dispersion peut être chiffrée par la variance conditionnelle:

$$Var[h(Y(x_0)) | Y(x_1), \dots, Y(x_n)] = E_c[h(Z(x_0))^2] - \{E_c[h(Z(x_0))]\}^2$$

Contrairement aux variances d'erreurs habituelles (qui sont non conditionnelles), les variances conditionnelles dépendent explicitement des valeurs $Y(x_i)$. Ainsi, le fait que dans une zone de faible teneur, les erreurs sont habituellement plus faibles que dans une zone riche, à configuration de krigeage donnée, peut-il être pris en compte.

Voyons les simplifications possibles dans quelques cas particuliers:

2.2.1 Estimation d'une indicatrice.

L'espérance conditionnelle d'une indicatrice à coupure y_c a une expression simple.

$$\begin{aligned} E[1_{Y(x_0) \geq y_c} | Y(x_1), \dots, Y(x_n)] &= \int 1_{\{Y^{k.s.} + \sigma^{k.s.} u \geq y_c\}} g(u) du \\ &= \int_{\frac{y_c - Y^{k.s.}}{\sigma^{k.s.}}}^{\infty} g(u) du = 1 - G\left(\frac{y_c - Y^{k.s.}}{\sigma^{k.s.}}\right) \end{aligned}$$

La variance conditionnelle est également facile à obtenir:

$$\begin{aligned} \text{Var}[1_{Y(x) \geq y_c}] &= E[1_{Y(x) \geq y_c}] - E[1_{Y(x) \geq y_c}]^2 \\ &= \left(1 - G\left(\frac{y_c - Y^{k.s.}}{\sigma^{k.s.}}\right)\right) G\left(\frac{y_c - Y^{k.s.}}{\sigma^{k.s.}}\right) \end{aligned}$$

2.2.2 Utilisation du développement de h .

On suppose ici connus les coefficients h_k du développement de la fonction h sur les polynômes d'Hermite H_k :

$$h(y) = \sum_k \frac{h_k}{k!} H_k(y).$$

On note en outre s^2 la variance des valeurs $Y^{k.s.}$, qui vaut:

$$s^2 = \sum_{ij} \lambda_i \lambda_j C_Y(x_i - x_j) = \sum_i \lambda_i C_Y(x_i - x_0)$$

En remarquant que, d'après la décomposition de $Y(x_0) = Y^{k.s.} + \sigma^{k.s.} U$, on a:

$$E[H_k(Y(x_0)) | Y(x_1), \dots, Y(x_n)] = E[H_k(Y(x_0)) | Y^{k.s.}]$$

et que $(Y(x_0), Y^{k.s.}/s)$ est bigaussien¹ de corrélation s , il vient:

$$E[H_k(Y(x_0)) | Y(x_1), \dots, Y(x_n)] = s^k H_k\left(\frac{Y^{k.s.}}{s}\right)$$

D'où résulte l'expression:

$$E[h(Y(x_0)) | Y(x_1), \dots, Y(x_n)] = \sum_k \frac{h_k}{k!} s^k H_k\left(\frac{Y^{k.s.}}{s}\right)$$

1. On se souvient que pour un vecteur gaussien (X, Y) de \mathbb{R}^2 , dont les lois marginales sont centrées, et normées, $E[H_k(X) | Y] = \text{Corr}(X, Y)^k H_k(Y)$.

2.3 Applications aux anamorphosés de f.a. gaussiennes

Les résultats d'estimation valables pour des fonctions aléatoires gaussiennes s'étendent immédiatement aux fonctions obtenues par transformation bijective de telles fonctions. Soit donc Y une fonction gaussienne, et Z la fonction stationnaire définie par $Z(x) = \phi(Y(x))$, pour une fonction ϕ donnée. Si Z est observable, et que les mesures collectées sont données par $Z(x_i)$, nous connaissons les valeurs normales associées $Y(x_i) = \phi^{-1}(Z(x_i))$. Alors il y a équivalence entre l'information portée par $(Z(x_1), \dots, Z(x_n))$ et celle portée par $(Y(x_1), \dots, Y(x_n))$. D'où résulte que

$$E[h(Z(x_0)) | Z(x_1), \dots, Z(x_n)] = E[h \circ \phi(Y(x_0)) | Y(x_1), \dots, Y(x_n)]$$

De même, en ce qui concerne les coupures:

$$Z(x) \geq z_c \iff Y(x) \geq y_c = \phi^{-1}(z_c)$$

Nous voyons donc que, dans la mesure où la fonction normalisante est supposée connue, les résultats précédents s'appliquent immédiatement. En résumé la démarche est constituée des étapes suivantes:

- Transformation des données $Y(x_i) = \phi^{-1}(Z(x_i))$.
- Analyse structurale de Y , et modélisation de la covariance C_Y .
- Krigeage simple de Y à partir des $Y(x_i)$, ce qui détermine $Y^{k.s.}$, $\sigma_{k.s.}$ et $s^2 = \text{Var}(Y^{k.s.})$.
- Décomposition de $h \circ \phi$ sur les polynômes d'Hermite. Soient c_n les coefficients de cette décomposition: $h \circ \phi = \sum (c_n/n!) \times H_n$.
- Alors $E_c[h(Z(x))] = \sum (c_n/n!) \times s^n H_n[Y^{k.s.}/s]$
- La décomposition de $(h \circ \phi)^2$ permet le calcul de la variance conditionnelle.

Ce schéma général se simplifie dans certains cas particuliers. En voici un exemple:

2.3.1 Le krigeage log-normal.

Une FAST Z est log-normale si elle est de la forme:

$$Z(x) = m e^{\sigma Y(x) - \frac{\sigma^2}{2}}$$

où Y est une fonction aléatoire gaussienne réduite. Avec cette paramétrisation, m est l'espérance de Z et σ^2 est dite variance logarithmique. Dans ces conditions, on a la décomposition:

$$Z(x) = m \exp\left\{\sigma Y^{k.s.} + \sigma\sigma_{k.s.} U - \frac{\sigma^2}{2}\right\}$$

où le résidu normalisé U , est une variable normale indépendante des $Y(x_i)$. L'espérance conditionnelle est donc:

$$E[Z(x_0) | Z(x_1), \dots, Z(x_n)] = m \exp\left\{\sigma Y^{k.s.} - \frac{\sigma^2}{2}(1 - \sigma_{k.s.}^2)\right\}$$

Par rapport à la démarche naïve qui consiste à prendre $Z^* = \phi\{[\phi^{-1}(Z(x))]^{k.s.}\}$ soit l'exponentielle du krigeage des logarithmes, nous voyons l'apparition d'un terme multiplicatif correcteur, sans lequel l'estimation est lourdement biaisée. Le tableau suivant permet de se faire une idée de l'ampleur de ce biais:

TAB. 2.1 - : Terme correcteur $\exp\left\{\frac{\sigma\sigma_{k.s.}^2}{2}\right\}$

σ	0.5	1	1.5	2
$\sigma_{k.s.} = 0.2$	1.005	1.02	1.05	1.08
$\sigma_{k.s.} = 0.5$	1.03	1.13	1.32	1.65
$\sigma_{k.s.} = 0.8$	1.08	1.38	2.05	3.6

Nous avons supposé la fonction normalisante ϕ connue. Dans la pratique cette fonction pourra être choisie dans une famille paramétrique simple. Ainsi $\phi(y) = m \exp\left\{\sigma y - \frac{\sigma^2}{2}\right\}$ dans le cas log-normal que nous venons de voir. Un procédé constructif plus général est donné dans le paragraphe suivant:

2.4 Construction d'une anamorphose gaussienne.

Commençons par poser le problème suivant:

Etant donnée une FAST $Z(x)$, de loi marginale (loi monovariante) F connue:

$$P[Z(x) < z] = F(z)$$

on se demande s'il est possible de trouver une transformation bijective ϕ et une FAST $Y(x)$ telles que:

i) $Z(x) = \phi(Y(x))$

ii) $P[Y(x) < y] = G(y)$.

Cela est toujours possible lorsque la fonction de répartition F est inversible. Par exemple, si F admet une densité f strictement positive dans un intervalle $I = [a, b]$ et nulle ailleurs, alors F^{-1} existe de $[0, 1]$ dans $[a, b]$. Alors $Y(x) = G^{-1} \circ F(Z(x))$ est une FAST de loi marginale normale:

$$P[Y(x) < y] = P[Z(x) < F^{-1} \circ G(y)] = G(y)$$

$\phi = F^{-1} \circ G$ est donc la fonction cherchée.

2.4.1 Anamorphose empirique

La construction précédente supposait connue la loi de la FAST. En pratique, cette loi n'est accessible qu'à travers un échantillonnage. Associé à cet échantillonnage, nous avons une loi empirique:

$$F_n(z) = \frac{1}{n} \sum_i 1_{Z(x_i) < z}$$

Cette loi empirique suggère la forme suivante pour la transformation ϕ : selon un usage habituel en statistique, on note $Z_{(i)}$ les données classées par ordre croissant ($Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$). L'anamorphose empirique est alors la fonction $\hat{\phi}$ suivante:

$$\hat{\phi}(y) = \sum_i Z_{(i)} 1_{y \in D_i}$$

où les D_i sont les segments:

$$D_i = \left] G^{-1}\left(\frac{i-1}{n}\right), G^{-1}\left(\frac{i}{n}\right) \right]$$

qui forment une partition de la droite en segments équiprobables, et de probabilité $1/n$. La figure 2.1 illustre cette construction.

Alors si Y est une variable normale, $\hat{\phi}(Y)$ suit la loi empirique, car la probabilité $P[\hat{\phi}(Y) = Z(x_i)]$ est de $1/n$ si les données sont deux à deux distinctes, et est de k/n si $Z(x_i)$ appartient à un paquet de k données égales.

Cependant, cette fonction n'est pas inversible, puisque:

- toute valeur $z \notin \{Z(x_1), \dots, Z(x_n)\}$ n'a pas d'image gaussienne.
- l'image réciproque de $Z_{(i)}$ est le segment D_i entier.

Il convient donc d'effectuer un lissage de $\hat{\phi}$ pour obtenir une fonction inversible. Il en résultera une modélisation de la loi de Z plus acceptable que la loi empirique. En effet, outre le fait que F_n ne charge que les valeurs $Z(x_i)$, sa variance est:

$$\text{Var}_{F_n} = \frac{1}{n} \sum (Z(x_i) - \bar{Z})^2$$

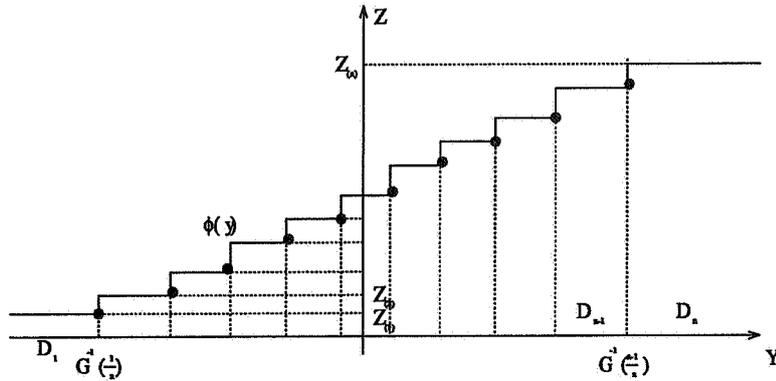


FIG. 2.1 - : *Anamorphose empirique*

qui est un estimateur biaisé de la variance de Z . (C'était déjà le cas pour des $Z(x_i)$ indépendants puisqu'il fallait alors diviser par $n - 1$ et non par n pour avoir un estimateur non biaisé, mais c'est bien d'avantage encore le cas lorsqu'une structure est présente puisque la variance empirique estime alors plutôt une variance de dispersion dans le champ d'étude qu'une variance stationnaire. Cette variance stationnaire est en fait obtenue sans biais comme palier du variogramme modélisé).

Il est en pratique facile d'obtenir une modélisation satisfaisante de la fonction d'anamorphose en prenant pour ϕ une fonction affine par morceaux. Le comportement aux extrémités du domaine (demi-droites D_0 et D_n) est alors fixé par des contraintes naturelles (telles que par exemple $\phi(-\infty) = 0$ si Z est une variable positive).

Un autre procédé de lissage utilisé dans la pratique courante consiste à ne conserver de $\hat{\phi}$ que son développement tronqué sur les polynômes d'Hermite. Le développement complet s'écrit:

$$\hat{\phi} = \sum_{k=0}^{\infty} \frac{\phi_k}{k!} H_k$$

avec des coefficients donnés par:

$$\begin{aligned} \phi_k &= \int \hat{\phi}(y) H_k(y) g(y) dy \\ &= \sum_i Z_{(i)} \int_{D_i} H_k(y) g(y) dy \\ &= \sum_i Z_{(i)} \Delta_i [H_{k-1}g] \end{aligned}$$

en ayant noté Δ_i les accroissements entre les valeurs $y_{i-1} = G^{-1}((i-1)/n)$ et $y_i = G^{-1}(i/n)$:

$$\Delta_i [H_{k-1}g] = H_{k-1}(y_i)g(y_i) - H_{k-1}(y_{i-1})g(y_{i-1})$$

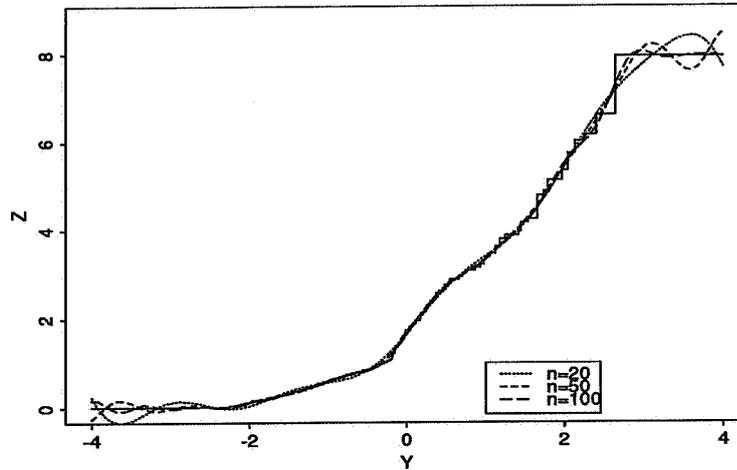


FIG. 2.2 - : Une anamorphose empirique, et trois approximations polynomiales, de degré respectivement 20, 50, et 100.

Le développement tronqué à un certain ordre p :

$$\phi^* = \sum_{k=0}^p \frac{\phi_k}{k!} H_k$$

donne la meilleure approximation en moyenne quadratique, avec pondération par la densité de Gauss, de $\hat{\phi}$ par un polynôme de degré p , tout en étant un modèle plus acceptable d'anamorphose. Des corrections demeurent cependant encore nécessaires, dans la mesure où la monotonie n'est pas assurée aux extrémités du domaine des valeurs. Un exemple d'approximation polynomiale d'une anamorphose empirique basée sur un échantillon constitué de 244 données est montré en figure 2.2. Les approximations considérées sont de degré 20, 50 et 100.

2.4.2 Cas d'une loi non continue

L'exemple précédent illustre le cas où la loi sous-jacente F est continue. Le cas d'une loi sous-jacente discrète ou plus généralement non continue pose des problèmes d'inversibilité. Il existe dans ce cas encore une fonction ϕ non décroissante telle que $\phi(Y)$ suive la loi non continue F , lorsque Y est une gaussienne centrée et normée, mais cette transformation présentera des paliers, comme illustré sur la figure 2.3.

L'image réciproque d'un point z tel que $P[Z = z] = F^+(z) - F(z) > 0$ est alors le segment $[G^{-1} \circ F(z), G^{-1} \circ F^+(z)]$. Deux options sont alors possibles. La plus simple, mais aussi la plus grossière, consiste à se ramener au cas précédent

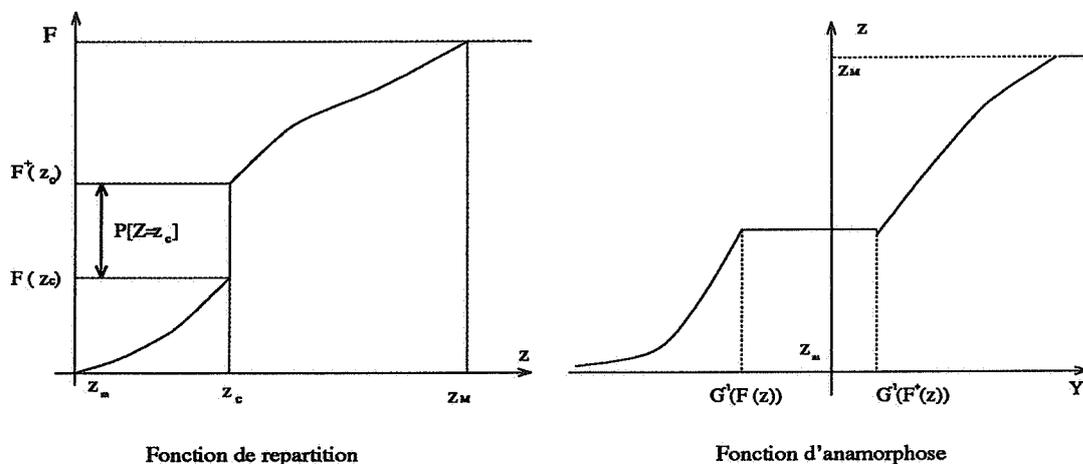


FIG. 2.3 - : Exemple de loi présentant un atome en z_0

en associant à une donnée $Z(x_i) = z$ une valeur $Y(x_i)$ plus ou moins arbitraire dans le segment image réciproque de z . La seconde option consiste à calculer la loi conditionnelle utilisant des données de type intervalles. Cela est possible en ayant recours à des simulations conditionnelles, au prix d'une certaine lourdeur algorithmique. Ces techniques ne seront pas décrites avec plus de détails ici.

2.4.3 Estimation de quantité de métal

Nous estimons la quantité:

$$Z(x_0) 1_{Z(x_0) \geq z_c} = \phi(Y(x_0)) 1_{Y(x_0) \geq y_c}$$

Sous forme intégrale, son espérance conditionnelle vaut:

$$Q_{z_c}(x_0) = \int_{\frac{y_c - Y^{k.s.}}{\sigma^{k.s.}}} \phi(Y^{k.s.} + \sigma^{k.s.}u) g(u) du$$

Cependant, il est plus simple d'utiliser le développement de ϕ lorsqu'il est connu. Il est en effet aisé d'en déduire celui de $\phi(y) 1_{y \geq y_c}$ (exercice), et d'utiliser le résultat du paragraphe 2.2.1, pour avoir le développement de l'espérance conditionnelle.

2.4.4 Remarques

- La FAST Y ainsi construite est à loi marginale normale:

$$\forall x \quad Y(x) \stackrel{\mathcal{L}}{\equiv} \mathcal{N}[0, 1]$$

Cela n'entraîne pas bien entendu, et loin s'en faut, que Y soit une fonction aléatoire gaussienne. Ce dernier cas demande en effet la propriété suivante, beaucoup plus exigeante:

$$\forall n; \forall (x_1, \dots, x_n) \quad (Y(x_1), \dots, Y(x_n)) \stackrel{\mathcal{L}}{\equiv} \text{Gauss}$$

et donc entraîne que toute transformée linéaire $Y(\lambda) = \int Y(x)\lambda(dx)$ est encore gaussienne.

- Le rôle joué ici par la loi de gauss peut être joué par n'importe quelle loi continue. En particulier pour toute FAST de loi marginale (monovariante) continue, il existe aussi, par exemple, une anamorphose gamma de paramètre b quelconque, c'est à dire qu'il existe une fonction stationnaire $Y^b(x)$ à loi monovariante gamma de paramètre b , et une transformation ψ inversible avec $Z(x) = \psi(Y^b(x))$.

Les calculs d'espérance conditionnelle développés dans les paragraphes précédents utilisent la multinormalité des vecteurs mis en jeu. Cela suppose que certains contrôles de modèle soient possibles. En pratique des tests multivariés sont difficiles à mettre en oeuvre, pour plusieurs raisons. Mentionnons d'abord la nécessité de pouvoir translater une même configuration $Z(x_0), \dots, Z(x_n)$ comportant $n + 1$ points dans l'espace, de façon à pouvoir constituer une population à soumettre au test. Ensuite, la puissance d'un test, et donc sa capacité à rejeter une hypothèse non valide, diminue très rapidement avec la dimension considérée ($n + 1$ ici). Cela fait que les contrôles utilisés dans la pratique se limitent en général à considérer les lois bivariées $(Z(x + h), Z(x))$.

Chapitre 3

Le krigeage disjonctif ponctuel

On se place donc, comme dans le chapitre précédent dans le cas de l'estimation locale au support ponctuel. Nous avons vu que l'espérance conditionnelle:

$$E[h(Z_0) | Z_1, \dots, Z_n]$$

était le meilleur estimateur -au sens de la norme quadratique- que l'on puisse former sur la base des informations disponibles, c'est à dire qu'il réalise le minimum de la norme de l'erreur, sur l'ensemble des fonctions quelconques des données:

$$\min_{h^* = \phi(Z_1, \dots, Z_n)} \|h(Z_0) - h^*\|$$

Le krigeage disjonctif, auquel on s'intéresse dans ce chapitre, réalise le minimum de cette même quantité, sur l'espace plus petit des estimateurs de la forme $\sum_i f_i(Z_i)$, où la variance de chaque terme $f_i(Z_i)$ est finie. Dans ces conditions, pourquoi se restreindre, au prix d'une perte d'efficacité évidente, à des estimateurs de cette forme?

Remarquons d'abord que l'optimalité de l'espérance conditionnelle est tributaire de la bonne adéquation des modèles de lois multivariées intervenant dans les estimations (lois de (Z_0, Z_1, \dots, Z_n)). Or cette adéquation ne peut être vérifiée valablement dès que le nombre de points intervenant dans l'estimation est de l'ordre de 4 ou 5.

En effet, dans le cas d'une estimation à partir de 4 variables, et les lois à tester sont alors de dimension 5, envisageons un test du χ^2 basé sur un découpage de chaque variable en 10 classes iso-effectif par exemple. La loi conjointe des 5 variables ainsi discrétisée comporte donc 10^5 modalités, à estimer à partir d'un jeu de données, en utilisant toutes les occurrences de la configuration spatiale présentée par les 5 points considérés (toutes les configurations de points de données égales à une translation près à (x_0, x_1, \dots, x_4)).

On mesure aisément le caractère hautement hypothétique d'un tel contrôle en

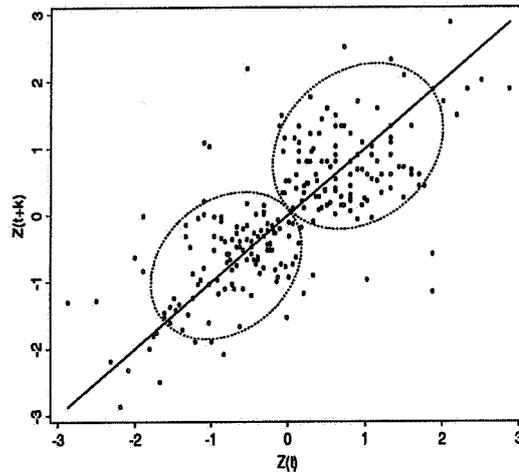


FIG. 3.1 - : *Un exemple de nuage de corrélation différé bimodal.*

pratique. Dans ces conditions, la supériorité de l'espérance conditionnelle sur des méthodes moins exigeantes du point de vue de la modélisation, risque fort de se révéler illusoire.

Une autre raison motivant la recherche d'autres estimateurs, est que les fonctions aléatoires gaussiennes (à une anamorphose près) sont les seuls modèles dans lesquels nous sommes capables de calculer effectivement cet estimateur. Or il est des circonstances où ce modèle se révèle inadapté:

- Les lois bivariées obtenues après anamorphose à une distance donnée h (lois des couples $(Y(x), Y(x+h))$) peuvent se révéler incompatibles avec un modèle gaussien, bien que la loi univariée (loi de $Y(x)$) le soit. C'est le cas, par exemple, pour des lois bivariées obtenues comme mélanges de deux gaussiennes, et admettant des densités de la forme:

$$p g_{\rho_1}(x, y) + (1 - p) g_{\rho_2}(x, y)$$

avec un coefficient p compris entre 0 et 1 (g_{ρ} désigne la loi de gauss bivariée de coefficient de corrélation ρ , et à lois marginales centrées et normées). Un exemple basé sur un jeu de données réel est montré dans la figure 2.1. Dans cet exemple, les distributions expérimentales étaient manifestement bimodales, et cette bimodalité se retrouve sur les nuages de corrélation différée des gaussiennes. Un modèle du type de celui indiqué ci-dessus est donc certainement mieux adapté ici qu'un modèle gaussien pur.

- Les lois univariées elles mêmes peuvent déjà être incompatibles avec la normalité (lois discrètes, ou lois présentant manifestement un atome, par exemple en zéro).

Il est souhaitable de disposer d'estimateurs qui ne soient pas tributaires des modèles gaussiens, et donc utilisables dans des situations plus générales. On va voir que, de fait, le krigeage disjonctif s'accommode d'une classe sensiblement plus vaste de modèles.

3.1 Les équations du krigeage disjonctif.

Le krigeage disjonctif est l'élément le plus proche de $h(Z_0)$ dans le sous-espace:

$$H = \{f_1(Z_1) + \dots + f_n(Z_n); \text{ avec } \forall i, E[f_i(Z_i)^2] < \infty\}$$

Nous pouvons appliquer à ce sous-espace de \mathbb{L}^2 que l'on supposera fermé (nous reviendrons par la suite sur cette hypothèse), le résultat sur les projections dans les espaces de Hilbert. Il existe donc un plus proche élément de $h(Z_0)$ dans H , qui est caractérisé comme le seul élément de H , noté $h^{k.d.}$, tel que $h(Z_0) - h^{k.d.} \in H^\perp$. Les équations du krigeage disjonctif sont donc:

$$\forall i; \forall \psi(Z_i) \text{ de variance finie} \quad \langle h(Z_0) - h^{k.d.}, \psi(Z_i) \rangle = 0$$

En se souvenant de la définition de l'espérance conditionnelle, nous voyons que ces équations expriment que:

$$\forall i \quad E[h(Z_0) - h^{k.d.} | Z_i] = 0$$

Ces équations demandent la spécification des lois bivariées $F_{ij}(du, dv)$, pour chacun des couples (Z_i, Z_j) et (Z_i, Z_0) . Elles se présentent sous la forme d'un système d'équations intégrales dont les fonctions de pondération f_i sont solutions. La première forme conduit à:

$$\left. \begin{array}{l} \forall j, \\ \forall \psi; E[\psi(Z_j)^2] < \infty \end{array} \right\} \sum_i \int f_i(u) \psi(v) F_{ij}(du, dv) = \int h(u) \psi(v) F_{0j}(du, dv)$$

tandis que la seconde donne le système suivant, d'apparence un peu plus simple:

$$\forall i \quad \sum_j \int f_i(u) F_i^j(du) = \int h(v) F_0^j(dv)$$

dans lequel on a noté F_i^j la loi conditionnelle de Z_i à Z_j fixé. Ce système ne pourrait en général être résolu que de manière approchée, à l'aide de méthodes numériques. Nous allons voir que dans le cas de lois bivariées isofactorielles, de grandes simplifications se produisent, qui rendent inutile le recours à ces méthodes numériques.

3.2 Lois isofactorielles.

On se limitera dans un premier temps au cas de lois symétriques, (lois bivariées symétriques par rapport à leurs arguments: $F(du, dv) = F(dv, du)$) seul nécessaire tant que l'on se limite à l'estimation ponctuelle. On note encore $F(du) = \int_v F(du, dv)$ la loi marginale, aucune confusion n'étant possible.

Définition: Une loi isofactorielle symétrique est une loi bivariable symétrique, $F(du, dv)$, pour laquelle il existe une famille dénombrable de fonctions réelles, notées χ_n , possédant les propriétés suivantes:

- i) Famille complète dans $\mathcal{L}^2(\mathbb{R}, F(du))$
- ii) Orthonormée dans ce même espace: $\int \chi_n(u) \chi_m(u) F(du) = \delta_{nm}$
- iii) Telle que l'orthogonalité s'étende aux lois bivariées: $\int \chi_n(u) \chi_m(v) F(du, dv) = \delta_{nm} T_n$

avec des coefficients T_n réels et dans $[-1, 1]$, puisqu'ils représentent les covariances (non centrées) des χ_n , et donc que $|T_n| \leq \|\chi_n\|^2 = 1$.

Propriétés: De cela résultent les remarques suivantes:

i) Développements sur les facteurs

Le premier point signifie que toute fonction de moment d'ordre deux fini admet une décomposition sur les χ_n :

$$\int f(u)^2 F(du) < \infty \implies f = \sum_n f_n \chi_n$$

avec, compte tenu de la seconde propriété, des coefficients f_n donnés par: (il suffit de multiplier les deux membres de l'égalité précédente par $\chi_m(u)F(du)$, puis d'intégrer)

$$f_n = \int f(u) \chi_n(u) F(du)$$

ii) Relations d'échange des facteurs

Pour un couple (U, V) suivant la loi isofactorielle $F(du, dv)$, on a pour tout n :

$$E[\chi_n(U) | V] = T_n \chi_n(V)$$

En effet, l'espérance conditionnelle $E[\chi_n(U) | V]$ étant elle même une fonction de $\mathcal{L}^2(\mathbb{R}, F)$, admet un développement sur les χ_m , dont on note les coefficients c_{nm} :

$$E[\chi_n(U) | V] = \sum_m c_{nm} \chi_m(V)$$

Ces coefficients s'obtiennent par les relations:

$$\begin{aligned} c_{nm} &= \int E[\chi_n(U) | V = v] \chi_m(v) F(dv) = E[E[\chi_n(U) | V] \chi_m(V)] \\ &= E[\chi_n(U) \chi_m(V)] = \delta_{nm} T_n \end{aligned}$$

ce qui fait que seul le terme $T_n \chi_n(V)$ subsiste du développement.

iii) Développement de la loi bivariable.

Enfin, notons que pour toute fonction $\phi \in \mathbb{L}^2$, on a:

$$E[\phi(U) | V] = \sum_n \phi_n T_n \chi_n(V)$$

en remplaçant les coefficients ϕ_n par leur expression, et en intervertissant intégration et sommation:

$$E[\phi(U) | V] = \int \phi(u) \sum_n T_n \chi_n(u) \chi_n(V) F(du)$$

D'où une identification de la loi conditionnelle:

$$F_v(du) = \sum_n T_n \chi_n(u) \chi_n(v) F(du)$$

et donc la forme suivante de la loi bivariable:

$$F(du, dv) = F(du) F(dv) \sum_n T_n \chi_n(u) \chi_n(v)$$

Ces calculs cessent évidemment d'être licites lorsque les développements $\sum T_n \chi_n \chi_n$ ne sont pas convergents, ce qui en pratique ne se produit que dans certains cas limites (nous en verrons un exemple dans un instant). Cette forme de loi bivariable isofactorielle est donc, de fait, assez générale.

Donnons maintenant quelques exemples:

• Loi de Gauss bivariable.

Cet exemple a déjà été rencontré dans le cours de sélectivité. Les facteurs en sont les polynômes d'Hermite normés:

$$\eta_k(y) = \frac{H_k(y)}{\sqrt{n!}}$$

Les coefficients T_n sont alors simplement les n-ièmes puissances des coefficients de corrélation:

$$\begin{aligned} E[\eta_n(X) \eta_n(Y)] &= \frac{1}{n!} E[E[H_n(X) | Y] H_n(Y)] \\ &= \frac{1}{n!} E[\rho^n H_n(Y)^2] = \rho^n \end{aligned}$$

Et la forme de la loi isofactorielle donne le développement isofactoriel connu de la loi de Gauss:

$$g_\rho(x, y) = g(x) g(y) \sum_n \rho^n \eta_n(x) \eta_n(y)$$

• **Lois hermitiennes.**

Ce sont des lois isofactorielles de lois marginales gaussiennes, dont les facteurs sont encore les polynômes d'Hermite normés. Dans ces conditions, la forme la plus générale des coefficients T_n est (résultat admis. Se reporter à [15] pour une démonstration):

$$T_n = \int_{-1}^1 r^n \omega(dr)$$

où ω est une loi de probabilité sur l'intervalle $[-1, +1]$. En d'autres termes, les coefficients T_n des lois hermitiennes sont les moments d'ordre n d'une variable aléatoire à valeurs dans $[-1, +1]$.

Cette forme est nécessaire et suffisante, en ce sens que les coefficients de toute loi hermitienne sont de cette forme, et que inversement, tout jeu de coefficients de cette forme définit une loi hermitienne.

La loi bivariable correspondante admet une densité lorsque ω ne possède pas d'atome en $+1$ ou en -1 ¹. Cette densité est alors un mélange de lois gaussiennes:

$$\begin{aligned} f(x, y) &= g(x)g(y) \sum_n \int r^n \omega(dr) \eta_n(x) \eta_n(y) \\ &= \int g_r(x, y) \omega(dr) \end{aligned}$$

Par exemple dans le cas particulier où la loi est dégénérée et ne charge que la valeur r_0 :

$$\omega(dr) = \delta_{r_0}(dr)$$

et donc que la variable associée R vaut r_0 presque sûrement, la loi correspondante est donc une gaussienne pure, de coefficient de corrélation r_0 .

Un autre cas particulier intéressant est celui où la variable associée ne prend que les valeurs 0 ou 1:

$$\omega(dr) = p \delta_1(dr) + (1 - p) \delta_0(dr)$$

Le mélange de loi gaussiennes ne comprend alors que deux composantes: dans la première, $X = Y$ presque sûrement, tandis que dans la seconde, X et Y sont indépendantes (voir l'exercice sur le modèle mosaïque)

Le modèle de fonction aléatoire associé à ce cas particulier de loi hermitienne, est le **modèle mosaïque**. Dans ce modèle l'espace est partitionné en compartiments dans lesquels la fonction est constante. En outre, les valeurs affectées aux différents compartiments sont indépendantes entre elles. Une telle fonction présente donc effectivement un aspect de mosaïque. Ainsi, deux points x et y peuvent appartenir à un même compartiment -ce qui a lieu avec la probabilité

1. En présence d'atome en $+1$ ou en -1 les développements utilisés cessent d'être convergents.

$p(x, y) = \omega_{(x,y)}(\{1\})$ - les valeurs $Y(x)$ et $Y(y)$ sont alors égales. Avec la probabilité complémentaire x et y se trouvent dans des compartiments différents. $Y(x)$ et $Y(y)$ sont alors deux gaussiennes indépendantes [12]. La fonction ainsi générée est à loi monovariante gaussienne. Le cas général s'obtient par anamorphose $Z = \phi(Y)$. La loi monovariante de Z peut alors être quelconque.

• **Lois bigamma.**

Les lois marginales de ces modèles sont des lois gamma, dont la densité dépend d'un paramètre de forme, noté ici α :

$$f_\alpha(x) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1}$$

Cette loi marginale prend des formes très différentes selon la valeur du paramètre, ainsi que le montre la figure. En particulier, lorsque le paramètre augmente indéfiniment $\alpha \rightarrow \infty$, on a la convergence en loi des variables centrées et normées:

$$\frac{X_\alpha - \alpha}{\sqrt{\alpha}} \rightarrow \mathcal{N}[0, 1]$$

Ce qui donne à ce modèle une plus grande généralité que le modèle gaussien.

Les facteurs de ces lois sont les polynômes de Laguerre normés, que l'on obtient par dérivation de la densité:

$$\chi_n(x) = l_n^\alpha(x) = \sqrt{\frac{\Gamma(\alpha + n)}{\Gamma(\alpha)n!}} \cdot \frac{f_\alpha^{(n)}(x)}{f_\alpha(x)}$$

Les coefficients T_n associés sont encore de la forme:

$$T_n = (\text{Corr}(X, Y))^n = \rho^n$$

et la loi bivariable admet la densité:

$$f_\rho(x, y) = f_\alpha(x) f_\alpha(y) \sum_n \rho^n l_n^\alpha(x) l_n^\alpha(y)$$

• **Lois de Laguerre.**

C'est la forme la plus générale des lois isofactorielles à lois marginales gamma, et à facteurs polynômiaux. Les facteurs sont donc les mêmes que ceux de la loi bigamma, et la forme générale des coefficients T_n est [7]:

$$T_n = \int_0^1 r^n \omega(dr)$$

ou ω est une loi de probabilité sur l'intervalle $[0, 1]$ cette fois. Les lois de Laguerre sont donc des mélanges de lois gamma de même paramètre α , et lorsque ω n'a pas d'atome en 1, elles ont une densité de la forme:

$$f(x, y) = \int_0^1 f_r(x, y) \omega(dr)$$

Voyons maintenant les simplifications qui résultent de cette forme de lois bivariées.

3.3 Le krigeage disjonctif isofactoriel

Nous supposons donc que les lois bivariées intervenant dans l'estimation sont toutes isofactorielles de la même famille:

$$\begin{aligned} (Y_i, Y_j) &\equiv \text{isofactorielle} & (\chi_n, T_n^{ij}) \\ (Y_0, Y_j) &\equiv \text{isofactorielle} & (\chi_n, T_n^{0j}) \end{aligned}$$

Alors, puisque les facteurs constituent une famille complète dans l'ensemble des fonctions de Y_0 admettant une variance finie, on a la décomposition:

$$h(Y_0) = \sum_n h_n \chi_n(Y_0)$$

et, par linéarité du krigeage disjonctif en tant que projecteur:

$$h(Y_0)^{k.d.} = \sum_n h_n [\chi_n(Y_0)]^{k.d.}$$

Le problème revient donc à estimer les termes $\chi_n(Y_0)$.

Une deuxième simplification résulte du caractère générateur de la famille χ_n : il suffit dans l'équation de projection, de vérifier l'orthogonalité de l'erreur aux $\chi_m(Y_i)$ (et non plus à terme quelconque $\psi(Y_i)$). On cherche donc l'élément $\chi_n(Y_0)^{k.d.}$, nécessairement unique, tel que:

$$\forall m, \forall i \quad \langle \chi_n(Y_0) - \chi_n(Y_0)^{k.d.}, \chi_m(Y_i) \rangle = 0$$

Il est alors facile de vérifier que le krigeage simple de $\chi_n(Y_0)$ à partir des $\chi_n(Y_i)$ vérifie ces conditions d'orthogonalité.

En effet, on a d'abord, en raison de l'orthogonalité des facteurs:

$$\langle \chi_n(Y_0) - \chi_n(Y_0)^{k.s.}, \chi_m(Y_j) \rangle = 0 \quad \text{pour } n \neq m$$

puisque le krigeage simple

$$\chi_n(Y_0)^{k.s.} = \sum_i \lambda_n^i \chi_n(Y_i)$$

ne comprend que des facteurs d'ordre n . Enfin, l'équation de projection pour $n = m$ donne le système de krigeage ordinaire:

$$\forall j \quad \sum_i T_n^{ij} \lambda_n^i = T_n^{0j}$$

On a donc bien $\chi_n^{k.d.} = \chi_n^{k.s.}$, ce qui achève de déterminer le krigeage disjonctif de $h(Y_0)$. Dans le système précédent, les termes diagonaux T_n^{ii} sont les normes des χ_n , et donc $T_n^{ii} = 1$. En résumé:

Lorsqu'il existe une famille génératrice orthogonale le krigeage disjonctif s'obtient en décomposant la fonction à estimer sur cette famille, et en krigeant séparément chaque terme de cette décomposition.

3.3.1 Biais et variance

La constante 1 appartenant à l'espace d'approximation, puisque de carré sommable, l'estimation est sans biais:

$$E[h(Y_0) - h(Y_0)^{k.d.}] = \langle h(Y_0) - h(Y_0)^{k.d.}, 1 \rangle = 0$$

En ce qui concerne la variance d'erreur, on a:

$$h(Y_0) - h(Y_0)^{k.d.} = \sum_n h_n [\chi_n(Y_0) - \chi_n(Y_0)^{k.d.}]$$

d'où des moments d'ordre deux:

$$\|h(Y_0) - h(Y_0)^{k.d.}\|^2 = \sum h_n^2 \|\chi_n(Y_0) - \chi_n(Y_0)^{k.d.}\|^2$$

qui sont égaux à la variance, puisqu'il n'y a pas de biais.

Très souvent, et c'est la cas pour tous les exemples de lois isofactorielles donnés précédemment, le premier facteur est constant: $\chi_0 = 1$. de ce fait il disparaît de la sommation précédente, et on a alors:

$$\text{Var} \{h(Y_0) - h(Y_0)^{k.d.}\} = \sum_{n \geq 1} h_n^2 \sigma_n^2$$

avec la variance d'estimation de chaque facteur donnée par $\sigma_n^2 = 1 - \sum_i \lambda_n^i T_n^{i0}$.

3.3.2 Pseudo-densité k.d.

Le krigeage disjonctif admet une représentation intégrale intéressante. Partons du développement:

$$h^{k.d.} = \sum_n h_n \chi_n^{k.d.}$$

et explicitons les coefficients $h_n = \int h(u) \chi_n(u) F^i(du)$, en intervertissant sommation et intégration, cela donne la représentation suivante:

$$h(Y_0)^{k.d.} = \int h(u) F^{k.d.}(du)$$

avec:

$$F^{k.d.}(du) = \sum_n \chi_n(u) \chi_n(Y_0)^{k.d.} F(du)$$

Nous voyons donc que cette mesure $F^{k.d.}$ agit dans le calcul d'un krigeage disjonctif, de manière analogue à une loi conditionnelle dans le calcul de l'espérance conditionnelle. Cependant, il faut bien voir que ce n'est pas une loi de probabilité, en ce qu'elle n'est pas positive. Il en résulte par exemple que:

$$F^{k.d.}(z) = \int^z F^{k.d.}(du)$$

peut très bien présenter tantôt des valeurs négatives, et tantôt des valeurs > 1 , même si:

$$\int F^{k.d.}(du) = [1]^{k.d.} = 1$$

3.3.3 Sur l'hypothèse " H fermé"

On a supposé dans le raisonnement l'espace d'approximation H fermé sans avoir donné de démonstration de la validité de cette propriété. Cela n'est pas nécessaire en réalité, comme le raisonnement suivant le montre:

On utilise comme espace d'approximation la fermeture \overline{H} de H , et nous utiliserons le projecteur de \overline{H} . L'approximant $h^{k.d.} = \Pi_{\overline{H}}(h(Z_0))$ est encore caractérisé par:

$$\forall i; \forall \psi(Z_i) \text{ de variance finie, } \langle h(Z_0) - h^{k.d.}, \psi(Z_i) \rangle = 0$$

mais il n'est pas en général de la forme $\sum f_i(Z_i)$. On peut seulement dire à ce stade qu'il s'agit d'une limite de variables de ce type. Ce n'est qu'à posteriori, c'est à dire après avoir exhibé une solution à cette équation, que nous pouvons conclure que l'estimateur est bien de la forme postulée. C'est ce qui a été fait lors des paragraphes précédents dans le cas isofactoriel.

3.4 Application à une FAST

On se place ici pour illustration dans le cas d'une application du modèle bi-gaussien.

On considère donc une fonction aléatoire stationnaire $Z(x)$, de loi monovariante continue. Nous avons vu comment associer à Z une autre fonction aléatoire Y , à loi monovariante gaussienne: $Z(x) = \phi(Y(x))$. Et on cherche donc à former le krigeage disjonctif de $\psi(Z(x_0)) = \psi \circ \phi(Y(x_0)) = h(Y(x_0))$.

Nous avons besoin de modèles pour les lois bivariées des couples $Y(x), Y(y)$. On fera donc à cette étape l'hypothèse que ces lois bivariées, dont les marges sont par construction gaussiennes, sont des lois gaussiennes à deux dimensions.

Il s'agit là d'une hypothèse supplémentaire, et non d'une propriété allant de soi, et comme telle, elle doit être soigneusement contrôlée. En pratique, on dispose des images gaussiennes des données: $Y(x_i) = \phi^{-1}(Z(x_i))$. Plusieurs possibilités existent. Citons en trois:

- L'examen de nuages de corrélation différée: $Y(x_i), Y(x_j)$ pour une classe de distance donnée: $h_i < |x_j - x_i| < h_s$ permet à la fois un contrôle complet, et un diagnostic des problèmes: le type d'écart à la normalité observé constitue un guide précieux dans le choix d'un modèle plus approprié. L'inconvénient de ce diagnostic est son caractère un peu fastidieux, puisqu'en principe il faut observer les nuages correspondant aux différentes classes de distance mises en jeu dans l'estimation.
- Un contrôle plus synthétique, et donc moins précis, est donné par la comparaison du variogramme d'ordre un de Y :

$$\gamma_1(h) = \frac{1}{2} E[|Y(x+h) - Y(x)|]$$

avec le variogramme d'ordre deux habituel. En effet, pour un modèle bigaussien, on a $\gamma_1(h) = \sqrt{\gamma(h)}$. On estimera donc simultanément ces deux quantités, et on réalisera un graphe portant en abscisse les estimations de $\gamma(h)$, pour différentes valeurs de h , et en ordonnée les estimations de γ_1 correspondantes.

- Enfin, il est possible d'estimer le variogramme de Y à partir de $Y(x_i)$, et de calculer le variogramme correspondant de Z selon le modèle bigaussien, puisque si $Z(x) = \sum (\phi_n/n!) H_n(Y(x))$:

$$C_Z(h) = \sum_{n \geq 1} \frac{\phi_n^2}{n!} [\rho_Y(h)]^n$$

On comparera ensuite la covariance ainsi obtenue avec celle des données brutes. C'est probablement le moins précis de ces trois contrôles, puisqu'il ne révèle rien si Z était déjà gaussienne au départ, et donc que l'anamorphose est affine, par exemple! Une amélioration possible consisterait à contrôler les variogrammes de quelques fonctions tests, tels que les facteurs eux mêmes, par exemple (corrélations en ρ_Y^n). Cependant, la procédure devient alors aussi lourde que l'examen des nuages de corrélations différés, tout en étant moins complète et moins informative.

Si cette phase révèle une inadéquation de la loi bivariable considérée, on s'orientera vers un autre modèle isofactoriel. La panoplie actuelle des modèles [15] permet de s'adapter à une grande diversité de situations pratiques.

Le reste de la méthode est évident. Après avoir estimé et modélisé la covariance de la transformée Y , on décompose la fonction à estimer sur les facteurs:

$$\psi \circ \phi(Y(x_0)) = h(Y(x_0)) = \sum h_k \chi_k(Y_0)$$

L'estimation d'une moyenne spatiale de $h(Y(x))$ se ramène alors à celle des moyennes de facteurs:

$$\left[\frac{1}{|v|} \int_v h(Y(x)) dx \right]^{k.d.} = \sum_k h_k \left[\frac{1}{|v|} \int_v \chi_k(Y(x)) dx \right]^{k.d.}$$

qui elle même est obtenue par krigeage simple:

$$\begin{aligned} \left[\frac{1}{|v|} \int_v \chi_0(Y(x)) dx \right]^{k.d.} &= 1 \\ \left[\frac{1}{|v|} \int_v \chi_k(Y(x)) dx \right]^{k.d.} &= \left[\frac{1}{|v|} \int_v \chi_k(Y(x)) dx \right]^{k.s.} \\ &= \sum_i \lambda_k^i \chi_k(Y(x_i)) \quad k \neq 0 \end{aligned}$$

avec des pondérateurs λ_n^i donnés par les systèmes linéaires:

$$\forall j \quad \sum_i \lambda_k^i T_k(x_i, x_j) = \frac{1}{|v|} \int_v T_k(x_j, x) dx$$

(on a noté $T_k(x_i, x_j)$ à la place de T_k^{ij} pour expliciter la dépendance spatiale). Pour le modèle bigaussien considéré, en notant $C_Y(h)$ la covariance de Y , ce système donne:

$$\forall j \quad \sum_i \lambda_k^i [C_Y(x_i - x_j)]^k = \frac{1}{|v|} \int_v [C_Y(x_j - x)]^k dx$$

La représentation intégrale reste bien entendu valable:

$$\left[\frac{1}{|v|} \int_v h(Y(x)) dx \right]^{k.d.} = \int h(u) F^{k.d.}(du)$$

De même que la décomposition de la variance d'estimation.

3.4.1 Remarques

Nous sommes partis de la forme cherchée de l'estimateur: $\sum_i f_i(Y(x_i))$ et de son optimalité en terme de variance d'estimation, pour caractériser le krigeage disjonctif. On aurait tout aussi bien pu l'introduire de l'une des deux manières suivantes, sans doutes plus intuitives:

- 1) Les facteurs χ_k sont très souvent des polynômes de degré k strictement. C'est le cas comme nous l'avons vu des modèles hermitiens, des modèles de Laguerre, mais aussi d'autres modèles non présentés ici, tels les modèles binômiaux, binômiaux négatifs. Dans ces conditions, l'espace engendré par les facteurs coïncide avec l'espace engendré par les monômes. Le krigeage disjonctif, qui est une régression linéaire multiple sur l'espace engendré par les facteurs:

$$\chi_0, \chi_1(Y_1), \chi_2(Y_1), \dots, \chi_1(Y_2), \chi_2(Y_2), \dots, \chi_1(Y_n), \chi_2(Y_n), \dots$$

est donc aussi un régression linéaire multiple sur tous les monômes obtenus à partir des données (Y_1, \dots, Y_n) :

$$1, Y_1, (Y_1)^2, (Y_1)^3, \dots, Y_2, (Y_2)^2, \dots, Y_n, (Y_n)^2, \dots$$

ou, plus explicitement:

$$h(Y_0)^{k.d.} = E[h(Y_0)] + \sum_{i,k} \lambda_k^i (Y(x_i)^k - \mu_k)$$

- 2) Le domaine des variables Y_i peut être découpé en classes, en introduisant les seuils:

$$y_0 = -\infty < y_1 < \dots < y_{n-1} < y_n = +\infty$$

et les indicatrices de classes:

$$J_i^p = 1_{\{y_{p-1} < Y_i \leq y_p\}}$$

On a ainsi un codage disjonctif complet des variables Y_i . Une indicatrice J_0^q peut alors être estimée par régression linéaire multiple sur l'ensemble des J_i^p pour $i = 1, \dots, N$ et $p = 1, \dots, n$. C'est le cokrigeage des indicatrices. Si maintenant le nombre de classes prises en compte augmente indéfiniment de telle manière que $E[J_i^p] \rightarrow 0$ l'estimateur obtenu coïncide avec le krigeage disjonctif.

L'estimation des indicatrices $1_{Y_0 < y}$ pour toutes les coupures y possibles permet à son tour l'estimation de toute fonction mesurable $h(Y_0)$ à variance finie. C'est la présentation adoptée dans le chapitre introductif.

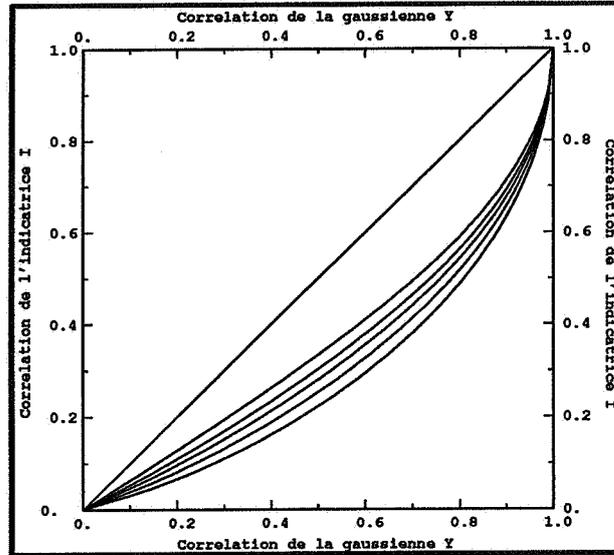


FIG. 3.2 - : *Destructuration des hautes teneurs dans le cas de lois bigaussiennes. En abscisse est porté le coefficient de corrélation de la gaussienne, et en ordonnée celui de l'indicatrice $1_{Y>y_c}$. Les coupures considérées sont $y_c = 0$, $y_c = 0.75$, $y_c = 1$, $y_c = 1.25$ et $y_c = 1.5$, valeurs qui correspondent aux courbes de plus en plus éloignées de la diagonale. La diagonale correspond au cas du modèle mosaïque, pour lequel toutes les indicatrices ont même corrélation que la variable Y elle même.*

3.4.2 Importance du choix du modèle

La qualité des résultats d'estimation dépend bien entendu de l'adéquation du modèle de loi bivariable utilisée. Un des caractères essentiels pris en charge dans cette modélisation est le niveau de destructuration des valeurs fortes [12]. Par là, nous entendons la chose suivante: à chaque valeur de coupure z_c correspond un variogramme d'indicatrice:

$$\gamma_{z_c}(h) = \frac{1}{2} E[(1_{Z(x+h) \geq z_c} - 1_{Z(x) \geq z_c})^2]$$

Ces variogrammes sont usuellement d'autant moins bien structurés que la coupure z_c est extrême. La figure 3.2 illustre cet effet dans le cas d'une fonction dont les lois bivariées sont gaussiennes centrées normées.

L'intensité de cette destructuration est un caractère essentiel de la loi bivariable, car c'est elle qui nous dira quelle importance doivent avoir quelques valeurs fortes trouvées dans l'échantillonnage, sur l'estimation:

- Si la destructuration est très intense, alors ces valeurs sont des valeurs isolées, qui n'auront que peu de poids dans l'estimation compte tenu de la courte portée des variogrammes associés aux indicatrices élevées.

- Dans le cas contraire, ces valeurs ont de bonnes chances de correspondre à des zones d'extension plus notable. Le poids accordé à ces valeurs fortes sera alors nécessairement plus élevé.

Le modèle mosaïque est un cas limite de modèle isofactoriel, dans lequel il n'y a pas de destructuration. Dans ce modèle, toutes les fonctions de la variable initiale Z sont autokrigeables (voir l'exercice sur le modèle mosaïque), en ce sens que leur krigeage disjonctif se réduit au krigeage simple:

$$h(Z(x_0))^{k.d.} = h(Z(x_0))^{k.s.} = m_h + \sum_i \lambda_i (h(Z(x_i)) - m_h)$$

avec des poids λ_i qui sont les mêmes quelle que soit la fonction à estimer h . En particulier, dans ce modèle, le krigeage disjonctif d'une indicatrice coïncide avec son krigeage simple. Il s'agit donc d'un modèle intéressant, en ce qu'il conduit à une mise en oeuvre d'une extrême simplicité, mais aussi d'un modèle très particulier.

3.4.3 Modèles à résidus d'indicatrices autokrigeables.

Le krigeage des indicatrices est très utilisé par de nombreux praticiens, beaucoup plus en raison de sa grande simplicité conceptuelle, que de la bonne adéquation du modèle qui lui est sous-jacent. Ici, est présenté un modèle proposé dans [18], qui tout en étant plus riche, reste d'une compréhension très intuitive. Commençons par introduire quelques notations spécifiques:

Notations Il s'agit d'un modèle de variables discrètes. Son application à des teneurs continues suppose donc une discrétisation préalable. On choisira n valeurs de coupures:

$$z_1 < z_2 < \dots < z_n$$

et on introduira les indicatrices associées:

$$I_0(x) = 1 \quad ; \quad I_1(x) = 1_{Z(x) \geq z_1} \quad \dots \quad I_n(x) = 1_{Z(x) \geq z_n}$$

On note T_i les espérances associées $T_i = E[I_i] = 1 - F(z_i)$, et les variances $\sigma_i^2 = T_i(1 - T_i)$. Les espérances de Z par classe de valeurs, seront notées m_i :

$$m_i = E[Z(x) \mid z_i \leq Z(x) < z_{i+1}]$$

Renonçant ensuite à la connaissance fine des valeurs, on ne retiendra par la suite que les informations données par les classes d'appartenance, et donc les valeurs des diverses indicatrices $I_i(x)$. Cela correspond à remplacer $Z(x)$ par la fonction:

$$\sum_i m_i (I_i(x) - I_{i+1}(x))$$

puisque $I_i - I_{i+1} = 1$ lorsque $Z(x)$ est comprise entre z_i et z_{i+1} , et seulement dans ce cas. La perte d'information ainsi consentie est minimale lorsque le découpage en classes est suffisamment fin.

On se place bien sûr dans un contexte stationnaire, et on note T_{ij} les covariances non centrées:

$$T_{ij}(h) = E[I_i(x+h) I_j(x)] = P[Z(x+h) \geq z_i \text{ et } Z(x) \geq z_j]$$

Notons la relation:

$$T_{ij}(0) = P[Z(x) \geq \max(z_i, z_j)] = T_{i \vee j}$$

avec la notation $u \vee v = \max(u, v)$.

Les résidus de la régression:

Les indicatrices ne prenant les valeurs 0 et 1, les régressions sont affines. Pour $i > j$, notons I_i^* la régression de $I_i(x)$, à $I_j(x)$ connu:

$$I_i^* = T_i + \lambda (I_j(x) - T_j)$$

le pondérateur λ vaut:

$$\lambda = \frac{\text{Cov}(I_i(x), I_j(x))}{\text{Var}(I_j(x))} = \frac{T_i - T_i T_j}{T_j(1 - T_j)} = \frac{T_i}{T_j}$$

Le résidu de cette régression $R_{ij}(x) = I_i(x) - I_i^*(x)$, est donc:

$$R_{ij}(x) = I_i(x) - \frac{T_i}{T_j} I_j(x)$$

et on a, par construction, toujours pour $i > j$, $R_{ij}(x) \perp I_j(x)$. Les questions que l'on va étudier maintenant sont les deux suivantes:

- Cette orthogonalité peut-elle être étendue aux fonctions aléatoires. Autrement dit, est-il possible d'avoir:

$$\forall (x, y); \forall i > j \quad R_{ij}(x) \perp I_j(y)$$

- Quel bénéfice est-il possible de tirer d'une telle propriété?

Condition d'orthogonalité:

Cette orthogonalité se traduit par:

$$\forall h \quad \text{Cov}(R_{ij}(x+h), I_j(x)) = T_{ij}(h) - \frac{T_i}{T_j} T_{jj}(h) = 0$$

D'où la propriété caractéristique:

$$\forall i > j, \forall h \quad T_{ij}(h) = \frac{T_i}{T_j} T_{jj}(h) \quad (3.1)$$

La même propriété se retrouve sur les covariances $C_{ij}(h) = T_{ij}(h) - T_i T_j$. On a donc:

La propriété d'orthogonalité demandée est assurée si et seulement si les covariances croisées de deux indicatrices sont proportionnelles à la covariance de l'indicatrice de la coupure la plus basse, avec un coefficient donné par le rapport T_i/T_j .

Cette propriété permet de tester le modèle à partir des analyses variographiques.

Interprétation géométrique:

L'interprétation de la condition précédente en terme de loi conditionnelle, compte tenu de:

$$\frac{T_i}{T_j} = \frac{P[Z(x) \geq z_i \text{ et } Z(x) \geq z_j]}{P[Z(x) \geq z_j]} = P[Z(x) \geq z_i | Z(x) \geq z_j]$$

et de:

$$\begin{aligned} \frac{T_{ij}(h)}{T_{jj}(h)} &= \frac{P[Z(x+h) \geq z_i \text{ et } Z(x) \geq z_j]}{P[Z(x+h) \geq z_j \text{ et } Z(x) \geq z_j]} \\ &= P[Z(x+h) \geq z_i | Z(x+h) \geq z_j \text{ et } Z(x) \geq z_j] \end{aligned}$$

donne:

$$P[Z(x) \geq z_i | Z(x) \text{ et } Z(x+h) \geq z_j] = P[Z(x) \geq z_i | Z(x) \geq z_j]$$

autrement dit, l'information portée par $I_j(x+h)$ n'apporte rien sur $I_i(x)$ quand $I_j(x)$ est connue (avec toujours bien sûr $i \geq j$). En termes d'ensembles de niveau:

$$B_j = \{x : Z(x) \geq z_j\}$$

cette propriété s'énonce:

$$P[Z(x) \geq z_i | x \text{ et } x+h \in B_j] = P[Z(x) \geq z_i | x \in B_j]$$

Cela est réalisé lorsque, à l'intérieur des B_j , les lois monovariabiles conditionnelles $P[Z(x) \geq z_i]$ sont stationnaires. Il n'y a donc pas d'effet de bord: la probabilité

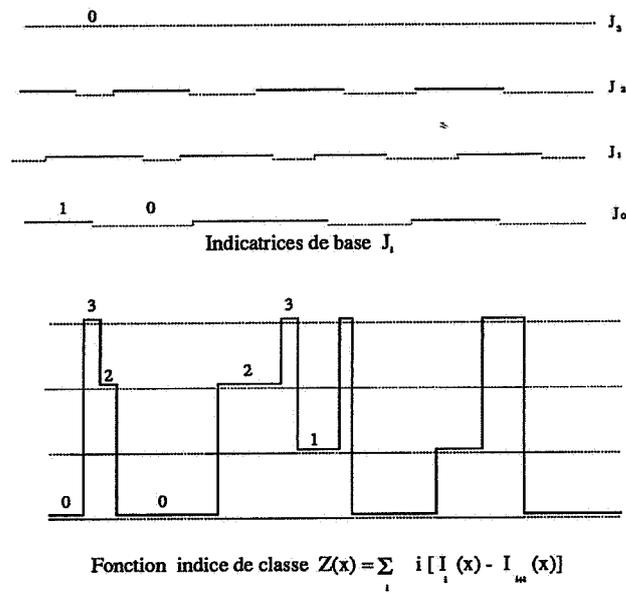


FIG. 3.3 - : Construction d'une fonction sans effets de bords à 4 niveaux dans \mathbb{R}

de trouver des valeurs élevées est la même près des frontières des B_j , et en leur centre.

Un modèle de fonction aléatoire:

On obtient ce type de comportement avec le modèle suivant: à chaque niveau i , nous définissons une fonction aléatoire $J_i(x)$, à valeur 0 ou 1. Au niveau le plus élevé est affecté conventionnellement la valeur $J_n = 0$. Les fonctions J_i sont choisies indépendantes d'un niveau à l'autre. Soit alors I_i la fonction produit:

$$I_i(x) = \prod_{j=0}^{i-1} J_j(x)$$

de telle sorte que la fonction $Z(x) = \sum m_i [I_i(x) - I_{i+1}(x)]$ vaut m_i dans le cas suivant:

$$J_0(x) = \dots = J_{i-1}(x) = 1 \quad \text{et} \quad J_i(x) = 0$$

la figure 3.3 illustre un exemple en dimension 1 d'une telle fonction dans le cas de 4 niveaux, et la figure 3.4 dans le cas de trois niveaux dans \mathbb{R}^2 .

Ce modèle est par construction sans effets de bords. De fait, les lois bivariées sont conformes à la relation caractéristique. On a d'abord:

$$T_i = \prod_{j=0}^{i-1} q_j$$

pour $q_j = E[J_j(x)]$. Alors:

$$T_{ij}(h) = E[I_i(x+h)I_j(x)] = E[J_{i-1}(x+h) \dots J_j(x+h) I_j(x+h) I_j(x)]$$

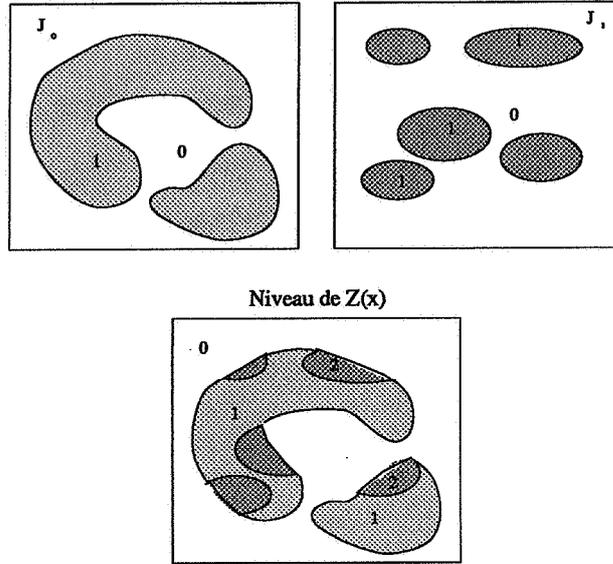


FIG. 3.4 - : Construction d'une F.a. suivant un modèle à résidus d'indicatrices autokrigeables dans \mathbb{R}^2

$$\begin{aligned}
 &= E[J_{i-1}(x+h) \cdots J_j(x+h)] E[I_j(x+h)I_j(x)] \\
 &= \frac{T_i}{T_j} T_{jj}(h)
 \end{aligned}$$

Les résidus D_i

Voyons maintenant les conséquences de cette propriété pour l'estimation non linéaire. Il est pour cela commode de considérer les résidus successifs avec la normalisation suivante:

$$D_i(x) = \frac{R_{i,i-1}(x)}{T_i} = \frac{I_i(x)}{T_i} - \frac{I_{i-1}(x)}{T_{i-1}}$$

et on notera $D_0(x) = 1$. On a alors la:

Propriété. Les résidus $D_i(x)$ constituent une famille génératrice et orthogonale.

- La famille est évidemment génératrice, puisque:

$$\begin{aligned}
 D_i(x) &= \frac{I_i(x)}{T_i} - \frac{I_{i-1}(x)}{T_{i-1}} \\
 \dots\dots &\dots\dots \\
 D_1(x) &= \frac{I_1(x)}{T_1} - \frac{I_0(x)}{T_0}
 \end{aligned}$$

$$D_0(x) = \frac{I_0(x)}{T_0} = 1$$

Les D_i engendrent donc les indicatrices I_i :

$$I_i(x) = T_i (D_0(x) + D_1(x) + \dots + D_i(x))$$

D'autre part, toute fonction $f(Z)$ s'écrit:

$$f(Z) = \sum_i f(m_i) (I_i(x) - I_{i+1}(x))$$

ce qui conduit à l'anamorphose:

$$f(Z) = \sum_i D_i \left\{ \sum_j f(m_j) [T_j 1_{\{j \geq i\}} - T_{j+1} 1_{\{j \geq i-1\}}] \right\}$$

• Et elle est orthogonale, puisque pour $i > j > 0$, on a:

$$\begin{aligned} E[D_i(x)D_j(y)] &= E \left[\left(\frac{I_i(x)}{T_i} - \frac{I_{i-1}(x)}{T_{i-1}} \right) \left(\frac{I_j(y)}{T_j} - \frac{I_{j-1}(y)}{T_{j-1}} \right) \right] \\ &= \frac{T_{ij}(x-y)}{T_i T_j} - \frac{T_{i-1,j}(x-y)}{T_{i-1} T_j} - \frac{T_{i,j-1}(x-y)}{T_i T_{j-1}} + \frac{T_{i-1,j-1}(x-y)}{T_{i-1} T_{j-1}} \\ &= \frac{T_{jj}(x-y)}{T_j^2} - \frac{T_{jj}(x-y)}{T_j^2} - \frac{T_{j-1,j-1}(x-y)}{T_{j-1}^2} + \frac{T_{j-1,j-1}(x-y)}{T_{j-1}^2} \\ &= 0 \end{aligned}$$

enfin, $E[D_i(x)D_0(y)] = E[D_i(x)] = 0$, ce qui achève de montrer l'orthogonalité.

La covariance des facteurs s'obtient facilement:

$$C_i(h) = E[D_i(x+h)D_i(x)] = \frac{T_{ii}(h)}{T_i^2} - \frac{T_{i-1,i-1}(h)}{T_{i-1}^2}$$

D'où la norme:

$$\|D_i\|^2 = \frac{1}{T_i} - \frac{1}{T_{i-1}}$$

Krigeage disjonctif dans ce modèle

Les seuils de classes z_i ayant été déterminés, les paramètres m_i (moyenne par classe) et T_i (probabilité de dépassement de z_i) suivent. Les résidus peuvent alors être calculés en tout point où $Z(x)$ est connu:

$$D_i(x_\alpha) = \frac{I_i(x_\alpha)}{T_i} - \frac{I_{i-1}(x_\alpha)}{T_{i-1}}$$

En vue de la mise en oeuvre du krigeage disjonctif, il faut disposer d'une modélisation des covariances de résidus, $C_i(h)$. Il est possible d'inférer ces covariances indépendamment les unes des autres, en se basant sur les variogrammes empiriques des résidus D_i . Cette modélisation présente l'inconvénient de ne pas respecter les contraintes de cohérence auxquelles ces covariances doivent satisfaire.

Une autre manière de procéder consiste à partir des fonctions J_i . Ces fonctions sont connues aux points x_α tels que $Z(x_\alpha) \geq z_i$, et donc $I_i(x_\alpha) = 1$. On a alors $J_i = I_{i+1}$. Elles restent inconnues si $Z(x_\alpha) < z_i$. A partir de ces valeurs expérimentales, il est possible d'estimer les covariances non centrées d'ensembles aléatoires² :

$$K_i(h) = \langle J_i(x+h), J_i(x) \rangle$$

On a ensuite:

$$T_{ii}(h) = K_{i-1}(h) \cdots K_0(h)$$

et:

$$C_i(h) = \frac{T_{ii}(h)}{T_i^2} - \frac{T_{i-1,i-1}(h)}{T_{i-1}^2}$$

L'estimation de $f(Z(x_0))$ est alors classique:

- i) Décomposition de f sur les D_i selon la formule d'anamorphose.
- ii) Estimation de $D_i(x_0)$ par krigeage simple à partir des $D_i(x_\alpha)$ et de $C_i(h)$.
- iii) Alors $f^{k.d.}(x_0) = \sum_i f_i D_i^{k.s.}$.

2. Les covariances d'ensembles aléatoires ne sont pas quelconques (voir par exemple [17]). Citons comme modèle admissible les covariances de la forme $C(h) = \exp\{-a.K(h)\}$ où $K(h)$ est un covariogramme géométrique.

Chapitre 4

Les modèles isofactoriels discrets symétriques

Ces modèles peuvent s'appliquer à la modélisation de variables par nature discrètes, telles que des comptages, ou tout aussi bien à des discrétisations en classes de variables continues, tout comme les modèles à résidus d'indicatrices orthogonaux vus précédemment.

A partir d'un jeu de données, il est possible d'estimer:

- les lois monovariées w_i ,
- et les lois bivariées $w_{i,j}(h) = \text{Prob}[I(x+h) = i, I(x) = j]$, pour un certain nombre de classes de distances $h = h_1, \dots, h_k$.

La méthodologie des modèles isofactoriels discrets proposée dans [16], et dont la mise en oeuvre est présentée dans [3], permet de construire des modèles respectant la loi monovariée w_i , et ajustant de manière fine les lois bivariées $w_{i,j}$, car s'appuyant sur une analyse des correspondances des lois expérimentales. Cette méthode présente donc un degré de généralité très appréciable, et qui justifie une plus grande complexité, dans des cas difficiles.

Le point de départ de cette construction est un processus de vie et de mort, qui permet de générer des lois bivariées symétriques.

4.1 Processus de vie et de mort.

Il s'agit d'une fonction aléatoire à valeurs discrètes, dans l'espace à une seule dimension, possédant un certain nombre de propriétés que nous allons décrire. On notera $I(t)$ ce processus, en on supposera pour simplifier le nombre d'états

fini: $I \in \{0, 1, \dots, N\}$. (le paramètre t sera interprété comme un temps par la suite).

i) Le processus est **markovien**:

$$P[I(t+t') = j | I(u); u \leq t] = P[I(t+t') = j | I(t)]$$

autrement dit, le futur du processus dépend du passé seulement à travers le présent, ou si l'on préfère, la connaissance de $I(t)$ fait écran aux événements antérieurs à t .

ii) Le processus est **homogène**:

$$P[I(t+t_0) = j | I(t_0) = i] = P_{ij}(t)$$

La probabilité de transition de i vers j entre les instants t_0 et $t+t_0$ dépend donc seulement de l'incrément t , et non de l'instant initial t_0 .

iii) Le processus est **diffusif**, c'est à dire que si le processus est dans un état i au temps t , seules les transitions vers des états voisins de i sont possibles aux instants immédiatement ultérieurs à t . On a ainsi:

$$\begin{aligned} P[I(t+dt) = i+1 | I(t) = i] &= a_i dt \\ P[I(t+dt) = i | I(t) = i] &= 1 - (a_i + b_i) dt \\ P[I(t+dt) = i-1 | I(t) = i] &= b_i dt \end{aligned}$$

à des infiniment petits d'ordre 2 en dt près. Les autres transitions sont donc impossibles. C'est ce qui explique la terminologie employée pour ce processus, le passage $i \rightarrow i+1$ s'interprétant comme une naissance, et la transition $i \rightarrow i-1$ comme une mort, lorsque la variable i est une population. On suppose par la suite $a_i > 0$ pour $N > i \geq 0$, et $b_i > 0$ pour $N \geq i > 0$ (et bien sûr, $a_N = b_0 = 0$).

Les coefficients a_i et b_i , qui définissent le comportement infinitésimal du processus, sont liés à la **destruction des hautes teneurs**. En effet, des valeurs fortes de ces coefficients pour un état i entraînent une faible persistance de cet état. Ainsi la destruction des hautes teneurs est d'autant plus intense que la croissance de a_i et b_i avec i l'est.

La **loi stationnaire** w_i d'un tel processus est telle que si $P[I(t_0) = i] = w_i$ alors $P[I(t+t_0) = i] = w_i$ également, ce qui se traduit à l'aide de la matrice de transition $P_{ij}(t) = P[I(t+t_0) = j | I(t_0) = i]$ par:

$$\forall j \quad \sum_i w_i P_{ij}(t) = w_j$$

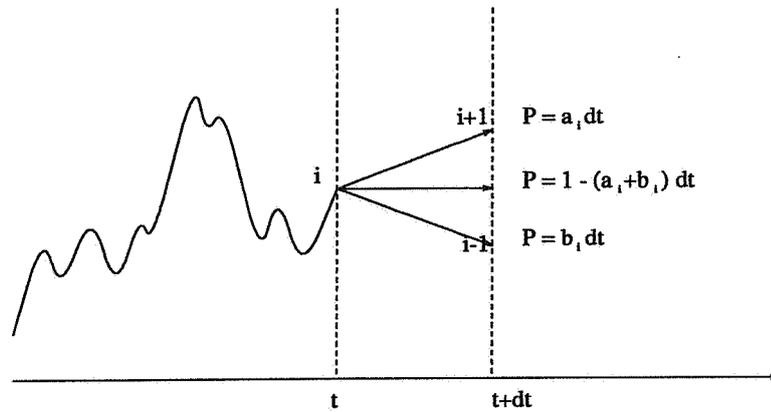


FIG. 4.1 - : Transitions d'un processus de vie et mort entre t et $t + dt$

En remplaçant t par dt , et en utilisant l'approximation d'ordre 1, cette condition donne:

$$w_j = w_{j+1} b_{j+1} dt + w_{j-1} a_{j-1} dt + w_j (1 - (a_j + b_j) dt)$$

puisque les états j en $t + dt$ ne peuvent provenir que des états $j + 1$, j , et $j - 1$ en t . On a donc:

$$b_{j+1} w_{j+1} - a_j w_j = b_j w_j - a_{j-1} w_{j-1}$$

relation valable pour $j > 0$. En $j = 0$, elle est remplacée par:

$$b_1 w_1 - a_0 w_0 = 0$$

De telle sorte que, de proche en proche, nous avons:

$$\forall i \quad a_i w_i = b_{i+1} w_{i+1}$$

en tenant compte de la relation $\sum w_i = 1$, nous constatons que la probabilité stationnaire, est nécessairement unique, et est déterminée par les a_i et b_i .

Il sera commode par la suite d'introduire l'espace des fonctions \mathcal{L}^2 , définies sur l'ensemble des états du processus, à valeurs dans \mathbb{R} , et de carré sommable pour la loi stationnaire (restriction triviale dans le cas où le nombre d'états est fini, mais qui prend tout son sens dans le cas contraire):

$$\mathcal{L}^2 = \left\{ f : \sum_i w_i f(i)^2 < \infty \right\}$$

espace que l'on muni du produit scalaire:

$$\langle f, g \rangle = \sum_i f(i) g(i) w_i$$

Le **générateur infinitésimal** du processus est défini de la manière suivante. On considère d'abord l'opérateur espérance conditionnelle:

$$f \rightarrow E_t[f]$$

qui, à $f \in \mathcal{L}^2$, fait correspondre:

$$E_t[f](i) = E[f(I(t)) | I(0) = i] = \sum_j P_{ij}(t) f(j)$$

Le générateur infinitésimal A du processus est alors défini comme la limite:

$$A(f) = \lim_{dt \rightarrow 0} \frac{E_{dt}[f] - f}{dt}$$

Comme $E_0 = I$, opérateur identité, A s'interprète comme une dérivée de E_t par rapport à t . En général, le domaine de définition de A est l'ensemble des f telles que cette limite existe. Ici, ce domaine est \mathcal{L}^2 entier. En effet:

$$E_{dt}[f](i) = a_i dt f(i+1) + [1 - (a_i + b_i) dt] f(i) + b_i dt f(i-1)$$

La limite existe donc toujours, et vaut:

$$Af(i) = a_i f(i+1) - (a_i + b_i) f(i) + b_i f(i-1)$$

ce qui donne la matrice tridiagonale suivante:

$$A = \begin{pmatrix} -a_0 & a_0 & 0 & \dots & 0 \\ b_1 & -(a_1 + b_1) & a_1 & \dots & 0 \\ 0 & b_2 & -(a_2 + b_2) & a_2 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & b_n & -b_n \end{pmatrix}$$

L'opérateur A est **autoadjoint** dans \mathcal{L}^2 . On a, en effet, pour tout couple de fonctions f et g de \mathcal{L}^2 :

$$\begin{aligned} \langle Af, g \rangle &= \sum_i w_i [Af](i) g(i) \\ &= \sum_i w_i g(i) [a_i f(i+1) - (a_i + b_i) f(i) + b_i f(i-1)] \end{aligned}$$

mais, d'après la relation sur la loi stationnaire:

$$\sum_i w_i g(i) f(i+1) a_i = \sum_i w_{i+1} g(i) f(i+1) b_{i+1} = \sum_i w_i g(i-1) f(i) b_i$$

et, de même:

$$\sum_i w_i g(i) f(i-1) b_i = \sum_i w_i g(i+1) f(i) a_i$$

finalement:

$$\langle Af, g \rangle = \sum_i w_i f(i) [a_i g(i+1) - (a_i + b_i) g(i) + b_i g(i-1)] = \langle f, Ag \rangle$$

Un tel opérateur a nécessairement ses valeurs propres réelles. Soient en effet $H_n \in \mathcal{L}^2$ et $\lambda_n \in \mathbb{C}$ tels que $AH_n = \lambda_n H_n$:

$$\left. \begin{aligned} \langle AH_n, H_n \rangle &= \lambda_n \|H_n\|^2 \\ \langle H_n, AH_n \rangle &= \bar{\lambda}_n \|H_n\|^2 \end{aligned} \right\} \lambda_n \in \mathbb{R}$$

Une autre conséquence du caractère autoadjoint de A est l'orthogonalité des facteurs associés à des valeurs propres différentes:

$$\left. \begin{aligned} \langle AH_n, H_m \rangle &= \lambda_n \langle H_n, H_m \rangle \\ \langle H_n, AH_m \rangle &= \lambda_m \langle H_n, H_m \rangle \end{aligned} \right\} \implies \langle H_n, H_m \rangle = 0 \text{ si } \lambda_n \neq \lambda_m$$

D'autre part, en raison du caractère tridiagonal de la matrice représentant A , chaque valeur propre est simple. En effet, si λ_n est une valeur propre, partant de $H_n(0) = 1$, fixé arbitrairement, la première relation:

$$-a_0 H_n(0) + a_0 H_n(1) = \lambda_n H_n(0)$$

détermine $H_n(1)$. La relation de récurrence:

$$b_i H_n(i-1) - (a_i + b_i) H_n(i) + a_i H_n(i+1) = \lambda_n H_n(i)$$

permet ensuite la détermination de proche en proche des $H_n(i)$. (La dernière relation est alors surabondante, et elle est automatiquement satisfaite, puisque $A - \lambda_n I$ est singulière). Ainsi, l'espace propre associé à λ_n est constitué des vecteurs proportionnels à H_n , et est donc de dimension 1.

Enfin, A est semi-définie négative, puisque:

$$\begin{aligned} \langle Af, f \rangle &= \sum_i w_i f(i) \{a_i (f(i+1) - f(i)) - b_i (f(i) - f(i-1))\} \\ &= \sum_i w_{i+1} b_{i+1} f(i) [f(i+1) - f(i)] - \sum_i w_i b_i f(i) [f(i) - f(i-1)] \\ &= -\sum_i w_i b_i [f(i) - f(i-1)]^2 \leq 0 \end{aligned}$$

Il s'ensuit que les valeurs propres de A sont négatives ou nulles:

$$\langle AH_n, H_n \rangle = \lambda_n \|H_n\|^2 \leq 0$$

On note de plus, que $\lambda_0 = 0$ est valeur propre de A , associée au vecteur propre $H_0(i) = 1$, puisque de manière triviale, $AH_0 = 0$.

En résumé, A a $N + 1$ valeurs propres distinctes, simples, réelles et ≤ 0 . A ces valeurs propres sont attachés des sous espaces de dimension 1, faciles à déterminer par la relation de récurrence. Les vecteurs propres normés, notés $\chi_n = H_n / \|H_n\|$ constituent un système orthonormé et générateur de \mathcal{L}^2 . Nous allons voir que ces éléments propres permettent le calcul des lois bivariées du processus.

A cause de la propriété markovienne, et de l'homogénéité du processus, l'opérateur E_t vérifie:

$$\begin{aligned} E_{t+t'}[f](i) &= \sum_k E[f(I(t+t')) | I(t) = k] P_{ik}(t) = \sum_k E_{t'}[f](k) P_{ik}(t) \\ &= E[E_{t'} f(I_t) | I_0 = i] = E_t[E_{t'}[f]](i) \end{aligned}$$

en terme d'opérateurs, c'est l'équation des semi-groupes:

$$E_{t+t'} = E_t \circ E_{t'}$$

Comme il a été vu, il existe une dérivée de E_t par rapport à t , ce qui nous autorise à dériver par rapport à t' l'identité précédente. En faisant ensuite $t' = 0$, on trouve l'équation de Kolmogorov suivante:

$$E_t' = E_t \circ A$$

En appliquant cette relation à un vecteur propre normé de A , cela donne:

$$E_t'[\chi_n] = \lambda_n E_t[\chi_n]$$

dont la solution générale est (pour $t \geq t_0$):

$$\ln[E_t(\chi_n)] - \ln[E_{t_0}(\chi_n)] = \lambda_n (t - t_0)$$

avec $E_0 = I$, on identifie:

$$E_t(\chi_n) = e^{\lambda_n t} \chi_n$$

On en déduit l'espérance conditionnelle d'un élément quelconque $f \in \mathcal{L}^2$. En effet, les vecteurs propres forment un système orthonormé et générateur de \mathcal{L}^2 , et donc:

$$f = \sum_n \langle f, \chi_n \rangle \chi_n$$

Puis, par linéarité:

$$E_t[f] = \sum_n e^{\lambda_n t} \langle f, \chi_n \rangle \chi_n$$

explicitement:

$$E_t[f](i) = \sum_{n,j} e^{\lambda_n t} w_j \chi_n(j) \chi_n(i) f(j)$$

Cela étant vrai pour toute fonction f , l'identification de la loi conditionnelle est possible:

$$P_{ij}(t) = w_j \sum_n e^{\lambda_n t} \chi_n(i) \chi_n(j)$$

Il est alors traditionnel d'utiliser un changement de signe sur les valeurs propres. On pose donc $\mu_n = -\lambda_n \geq 0$. D'autre part, on a considéré jusqu'ici seulement le cas $t \geq 0$. Plus généralement, par symétrie, on a l'expression:

$$w_{ij}(t) = w_i w_j \sum_n e^{-\mu_n |t|} \chi_n(i) \chi_n(j)$$

Cette loi est isofactorielle, avec les χ_n pour facteurs. On a vu, en effet, que ce système est générateur et orthonormé dans \mathcal{L}^2 . L'orthogonalité en $t \neq 0$ est immédiate:

$$\begin{aligned} \langle \chi_k(I(t_0 + t)), \chi_l(I(t_0)) \rangle &= \sum_n e^{-\mu_n |t|} \langle \chi_k, \chi_n \rangle \langle \chi_l, \chi_n \rangle \\ &= \delta_{kl} e^{-\mu_k |t|} \end{aligned}$$

Résumé: Les processus de vie et mort permettent donc la construction de familles de lois isofactorielles:

- i) admettant une loi marginale w_i donnée à l'avance,
- ii) ayant un niveau de destructuration choisi librement par l'utilisateur.

Oubliant maintenant cet aspect processus, on a une famille de lois indexée par un paramètre t , lequel reflète la plus ou moins grande corrélation entre les variables. Il reste donc à étudier les points suivants:

- i) Comment appliquer ces familles à des fonctions aléatoires dans des dimensions supérieures?
- ii) Comment réaliser l'inférence des paramètres du modèle?

La suite de ce chapitre aborde ces questions.

4.2 Application à des fonctions aléatoires

On a cette fois affaire à une fonction $Y(x)$, en dimension $n \geq 1$ et prenant ses valeurs dans l'ensemble $\{0, 1, \dots, N\}$. La fonction est supposée stationnaire, de loi marginale w_i , et de loi bivariable:

$$p_{ij}(x, y) = \text{Prob}[Y(x) = i, Y(y) = j] = p_{ij}(x - y)$$

ce qui ne suppose pas nécessairement isotropie. L'hypothèse de base est que ces lois bivariées appartiennent à la famille de lois générée par un processus de vie et de mort:

$$p_{ij}(x - y) = w_{ij}(t)$$

pour une certaine valeur du paramètre t , dépendant de $x - y$. On fait donc ainsi apparaître une fonction $t(h)$ mettant en correspondance les lois de la fonction Y et du processus $I(t)$. Ainsi:

$$p_{ij}(x - y) = w_i w_j \sum_n e^{-\mu_n t(x-y)} \chi_n(i) \chi_n(j)$$

Toute fonction $t(h)$ n'est bien sûr pas admissible pour modéliser cette dépendance. Par exemple, on a de manière évidente $t(-h) = t(h)$. Des conditions plus précises sont obtenues en considérant les covariances associées au modèle. En effet, en notant, pour une fonction f quelconque, $f_n = \langle f, \chi_n \rangle$ les coefficients de son développement sur les facteurs:

$$f = \sum_n f_n \chi_n$$

on a $E[f(Y(x))] = f_0$, puisque $\chi_0 = H_0 = 1$, et par suite:

$$\text{Cov}[f(Y(x+h)), f(Y(x))] = \sum_{n \geq 1} f_n^2 e^{-\mu_n t(h)}$$

en particulier, pour $f = \chi_n$, on voit que la covariance (spatiale) du facteur d'ordre n est $C_n(h) = e^{-\mu_n t(h)}$. Ces covariances doivent être de type positif, et nous allons choisir $t(h)$ de telle manière qu'il en soit bien ainsi. Commençons pour cela par rappeler un résultat du cours sur les fonctions aléatoires d'ordre deux [6].

Rappel: On a l'équivalence:

$$\forall \nu > 0 \quad e^{-\nu t(h)} \text{ est une covariance} \quad \iff \quad t(h) \text{ est un variogramme}$$

La modélisation de $t(h)$ se fera donc toujours en respectant cette condition, puisqu'il est vital d'avoir des matrices de covariances de facteurs positives dans le krigeage disjonctif.

On se contente dans la pratique du krigeage disjonctif de cette condition de modélisation, bien qu'elle ne garantisse pas l'existence d'une fonction aléatoire admettant $p_{ij}(h) = w_{ij}(t(h))$ pour loi bivariable. Notons cependant qu'une construction satisfaisant cette condition est possible en procédant par substitution [5]. On

se contente ici de donner une indication sur la manière de procéder, en renvoyant à la précédente référence pour plus de détails. Soit $Z(x)$ une fonction aléatoire intrinsèque, et:

$$Y(x) = I[Z(x)]^*$$

la fonction obtenue en substituant au paramètre t la valeur de la fonction Z au point x . Calculons alors la loi:

$$p_{ij}(x, y) = \text{Prob}[Y(x) = i; Y(y) = j]$$

en conditionnant par rapport à la réalisation de Z , on a:

$$\text{Prob}[Y(x) = i; Y(y) = j | Z] = w_i w_j \sum_n e^{-\mu_n |Z(y)-Z(x)|} \chi_n(i) \chi_n(j)$$

et donc, si ϕ_{y-x} est la transformée de Laplace de la valeur absolue de l'accroissement de Z :

$$\phi_{y-x}(s) = E[e^{-s |Z(y)-Z(x)|}]$$

on trouve pour loi bivariable:

$$p_{ij}(y-x) = w_i w_j \sum_n \phi_{y-x}(\mu_n) \chi_n(i) \chi_n(j)$$

Par exemple, si Z est à accroissements gaussiens, et à variogramme γ , on trouve:

$$\phi_h(s) = 2e^{(s\gamma(h))^2} [1 - G(2s\gamma(h))]$$

avec G pour la fonction de répartition de la loi de Gauss. Un tel processus existant toujours, la loi bivariable ainsi obtenue est toujours admissible, et nous savons simuler une telle fonction aléatoire Y .

4.3 Estimation des paramètres

On suppose donc que les opérations de découpage en classes et de détermination d'une loi marginale w_i ont été réalisées, soit que la loi empirique ait été utilisée telle quelle, soit qu'un modèle ait été ajusté à cette loi expérimentale. Il reste donc à déterminer:

1. Les coefficients de diffusion a_i
2. Et la fonction $t(h)$

Commençons par la première question. Celle-ci est grandement facilitée par l'utilisation de propriétés particulières du facteur d'ordre un, χ_1 , d'un processus de vie et mort.

Propriété 1. *Le facteur d'ordre un d'un processus de vie et de mort est monotone, d'espérance nulle, et d'autocorrélation maximale.*

Par cette dernière proposition, nous entendons que χ_1 réalise le maximum de la quantité:

$$\max_{\|f\|=1} \langle f(I(t+t_0)), f(I(t_0)) \rangle - E[f]^2$$

Il s'agit d'un maximum à t fixé, pris sur l'ensemble des fonctions de norme $[E[f^2]]^{1/2} = \|f\| = 1$. Il se trouve donc que, en fait, ce maximum ne dépend pas de t .

La démonstration de la première partie de cette proposition est assez lourde. Deux démonstrations différentes sont données dans [2] et dans [4], et ne sont pas reprises ici. La seconde résulte de l'orthogonalité des facteurs, en tenant compte de $\chi_0(i) = 1$. La troisième partie s'obtient immédiatement, en considérant la forme générale de la covariance d'une telle fonction:

$$\langle f(I(t+t_0)), f(I(t_0)) \rangle - E[f]^2 = \sum_{n \geq 1} f_n^2 e^{-\mu_n |t|}$$

quantité à maximiser sous la contrainte:

$$\|f\|^2 = \sum_n f_n^2 = 1$$

Compte tenu du fait que les valeurs propres sont ordonnées $0 < \mu_1 < \mu_2 < \dots < \mu_N$, le maximum est réalisé par $f_n = \delta_{n1}$, et donc par $f = \chi_1$.

Il sera plus commode par la suite de considérer le facteur H_1 , normalisé par la condition $H_1(0) = 1$, soit $H_1 = \chi_1/\chi_1(0)$. (χ_1 étant monotone et d'espérance 0, $\chi_1(0) \neq 0$). H_1 est alors décroissant en i .

Remarquons aussi que, d'après la structure de A , la transformation $a_i \rightarrow \nu a_i$ et $b_i \rightarrow \nu b_i$ transforme les valeurs propres selon $\lambda_n \rightarrow \nu \lambda_n$, sans modifier les vecteurs propres. La loi bivariable associée est alors transformée en:

$$w_{ij}(t) \rightarrow w_{ij}\left(\frac{t}{\nu}\right)$$

autrement dit, on a seulement effectué un changement de l'échelle des temps du processus, ce qui est du reste intuitivement évident. Il est donc possible, sans restreindre la généralité, de supposer par exemple $\lambda_1 = -1$.

Propriété 2. La donnée du facteur d'ordre un et de la loi marginale caractérise le processus de vie et de mort. De plus, à chaque couple w_i et H_1 tels que

i) $H_1(0) = 1$

ii) $E[H_1(I)] = \sum_i H_1(i) w_i = 0$

iii) H_1 est strictement décroissant en i

correspond un et un seul processus de vie et de mort admettant w_i pour loi stationnaire, H_1 pour facteur d'ordre un, et tel que $\lambda_1 = -1$.

Soit en effet w_i et H satisfaisant aux conditions de la proposition. Nous allons leur associer un processus de vie et mort de loi stationnaire w_i , et admettant H pour facteur d'ordre 1.

Les sommes partielles:

$$s_i = \sum_{j \leq i} w_j H(j)$$

vérifient: $s_0 = w_0$, $s_N = 0$. De plus, s_i étant croissante en i , tant que $H(i) > 0$, puis décroissante ensuite, il s'ensuit que $s_i \geq 0$. Les quantités a_i et b_i définies par les relations:

$$\begin{aligned} a_N &= b_0 = 0 \\ a_i &= -\frac{s_i}{w_i [H(i+1) - H(i)]} & i = 0, \dots, N-1 \\ b_i &= \frac{w_{i-1}}{w_i} a_{i-1} & i = 1, \dots, N \end{aligned}$$

définissent des coefficients positifs. Il leur est donc associé un processus de vie et mort à loi stationnaire w_i . Ces deux relations entraînent d'autre part la relation suivante, valable pour $i = 0, \dots, N-1$:

$$a_i [H(i+1) - H(i)] - b_i [H(i) - H(i-1)] = \frac{s_{i-1} - s_i}{w_i} = -H(i)$$

c'est à dire les $N-1$ premières relations de $AH = -H$. La dernière reste donc à montrer pour assurer que H est vecteur propre.

Définissons pour cela le polynôme $Q_i(\lambda)$, de degré i en λ , par la relation de récurrence (la récurrence est amorcée par $Q_{-1} = 0$):

$$\begin{cases} Q_0 = 1 \\ a_i (Q_{i+1} - Q_i) - b_i (Q_i - Q_{i-1}) = -\lambda Q_i & i = 0, \dots, N-1 \end{cases}$$

On a alors $Q_i(1) = H(i)$ et $Q_i(0) = 1$, et, plus généralement, si λ_n est une valeur propre de la matrice A associée aux coefficients a_i et b_i que nous venons de définir, $Q_i(\lambda_n)$ est un vecteur propre associé. L'orthogonalité des vecteurs propres montre que:

$$\sum_i Q_i(\lambda_n) Q_i(\lambda_m) w_i = 0 \quad \text{si } n \neq m$$

Soit alors $S(\lambda)$ le polynôme (de degré N en λ):

$$S(\lambda) = \sum_{i=0}^N w_i Q_i(\lambda)$$

La relation d'orthogonalité précédente avec $m = 0$ et $\lambda_m = 0$ montre que

$$S(\lambda_n) = \sum_i Q_i(\lambda_n) w_i = 0 \quad n = 1, \dots, N$$

et, compte tenu du degré de $S(\lambda)$, il s'agit là des seuls zéros de S . Autrement dit, on a l'équivalence:

$$S(\lambda) = 0 \quad \Longleftrightarrow \quad Q_i(\lambda) \text{ est vecteur propre de } A$$

Comme $S(1) = \sum_i H(i) w_i = 0$ par hypothèse, on en déduit que H est facteur, associé à la valeur propre -1 . D'autre part, les facteurs d'ordre supérieur à 1 du processus étant non monotones, on a bien $\lambda_1 = -1$ et $H_1 = H$.

Les propriétés 1 et 2 sont à la base de la méthode d'inférence des coefficients de diffusion décrite ici.

4.3.1 Inférence des coefficients de diffusion

Cette inférence utilise les lois bivariées empiriques, calculées pour une certaine classe de distance $h_i \leq h \leq h_s$:

$$\hat{w}_{ij}(h) = \frac{\#\{\text{données } (x, y) : Y(x) = i, Y(y) = j, h_i \leq |y - x| \leq h_s\}}{\#\{(x, y) : h_i \leq |y - x| \leq h_s\}}$$

on a noté $\#A$ le cardinal de l'ensemble A . Cette estimation est symétrique en i et j , mais sa loi marginale $\hat{w}_i = \sum_j \hat{w}_{ij}(h)$ est différente de l'estimation:

$$w_i^* = \frac{\#\{\text{données } x : Y(x) = i\}}{\#\{Y(x)\}}$$

puisque toutes les données interviennent dans cette dernière estimation, ce qui n'est en général pas le cas en ce qui concerne la précédente. La première étape

consiste à effectuer une correction pour qu'il en soit bien ainsi. Soit d'abord la matrice:

$$r_{ij} = \hat{w}_{ij} \cdot \frac{w_i^*}{\hat{w}_i}$$

On a $\sum_j r_{ij} = w_i^*$. Notons aussi $s_j = \sum_i r_{ij}$. On prendra alors pour estimation de loi bivariable la quantité:

$$w_{ij}^* = \sum_k \frac{r_{ik} r_{jk}}{s_k}$$

Il s'agit bien d'une loi bivariable symétrique, ayant w_i^* pour loi marginale, comme on le vérifie immédiatement. En outre, c'est une matrice semi définie positive, puisque:

$$\sum f_i f_j w_{ij}^* = \left\| \sum_i \frac{r_{ik}}{\sqrt{s_k}} \cdot f_i \right\|^2$$

Les valeurs propres de w_{ij}^* sont donc positives ou nulles. Notons que, si $\hat{w}_{ij}(h)$ était une estimation de la loi bivariable à la distance h , w_{ij}^* correspondrait à une estimation à la distance $2h$ dans le cas markovien.

On utilisera par la suite w_{ij}^* et w_i^* , sans le spécifier par l'utilisation du symbole *. Un facteur empirique est ensuite calculé sur la base de w_{ij} en maximisant la corrélation:

$$\max_f \frac{\sum f_i w_{ij} f_j - (\sum f_i w_i)^2}{\sum f_i^2 w_i}$$

Cela peut être réalisée de manière équivalent en maximisant:

$$\max_{\sum f_i w_i = 0} \frac{\sum f_i w_{ij} f_j}{\sum f_i^2 w_i}$$

puisque, partant d'une fonction f_i^1 centrée, et ajoutant une constante, on ne modifie pas le numérateur de l'expression dans la première maximisation, tandis que l'on augmente le dénominateur. La maximisation est donc nécessairement réalisée pour une fonction centrée.

Cette maximisation est obtenue en calculant les éléments propres $\lambda^{(n)}$ et $f^{(n)}$ associés au problème:

$$\forall i \quad \sum_j w_{ij} f^{(n)}(j) = \lambda^{(n)} w_i f^{(n)}(i)$$

Il est, en effet, élémentaire de vérifier que que $f^{(N-1)}$ maximise la corrélation, et est d'espérance $\sum w_i f^{(N-1)}(i) = 0$.

Remarque: $H(i) = F^{(N-1)}(i)/F^{(N-1)}(0)$ satisfait aux deux premières conditions de la propriété 2, mais non à la condition de monotonie. Dans la pratique, une modélisation graphique est effectuée, pour assurer le respect des trois conditions de la proposition 2.

Le générateur infinitésimal A est alors défini de la manière indiquée dans la démonstration associée. Les éléments propres de A permettent ensuite le calcul des lois bivariées:

$$w_{ij}(t) = w_i w_j \sum_n e^{-\mu_n |t|} \chi_n(i) \chi_n(j)$$

4.3.2 Estimation de $t(h)$:

Si les paires de données intervenant dans le calcul du comptage:

$$n_{ij}(h) = \#\{(x, y) : Y(x) = i, Y(y) = j, h_i \leq |y - x| \leq h_s\}$$

étaient indépendantes deux à deux, la vraisemblance attachée à une valeur t , et à une famille w_{ij} serait:

$$v(t) = \prod_{ij} [w_{ij}(t)]^{n_{ij}(h)}$$

et le logarithme de la vraisemblance donné par:

$$l(t) = \sum_{ij} n_{ij}(h) \left\{ \log \left\{ \sum e^{-\mu_n |t|} \chi_n(i) \chi_n(j) \right\} + \log(w_i w_j) \right\}$$

En ce cas, l'estimateur du maximum de vraisemblance \hat{t} serait obtenu par la maximisation de l . Comme les dérivées de l sont faciles à calculer, et que l'expérience montre que $l(t)$ est unimodale sur $t \geq 0$, le calcul effectif de \hat{t} est donc simple.

En réalité, la condition d'indépendance des paires est évidemment loin d'être réalisée. Cependant, il a été vérifié que l'estimation ainsi obtenue n'est pas entachée de biais.

Par utilisation de cette procédure pour différentes classes de distance, nous obtenons ainsi des estimations séparées $t(\hat{h}_1), \dots, t(\hat{h}_k)$. Un modèle de variogramme est ensuite ajusté à partir de ces différentes estimations, pour obtenir une fonction $t(h)$.

Remarque: Plus le paramètre t est élevé, et plus la loi $w_{ij}(t)$ tend vers l'indépendance $w_{ij}(\infty) = w_i w_j$. Comme cette convergence est gouvernée par des exponentielles décroissantes, il en résulte que le modèle $w_{ij}(t)$ est très peu sensible à des perturbations de t lorsque ce paramètre est élevé. L'estimation de $t(h)$ aux grandes distances est donc nécessairement très imprécise, sans que cela porte préjudice à la qualité de modélisation des lois bivariées.

Chapitre 5

Estimation locale avec changement de support en modèle multigaussien discrétisé

Le modèle Gaussien discrétisé, à la description duquel ce chapitre est consacré, permet l'estimation par espérance conditionnelle des fonctions de récupération locales, en présence d'effet de support. Une version un peu affaiblie de ce modèle, puisque ne spécifiant que les lois bivariées, permet l'estimation par krigeage disjonctif des mêmes quantités. Commençons par rappeler les hypothèses faites dans la spécification du modèle de changement de support.

5.1 Rappels sur le modèle utilisé en estimation globale

Le point de départ de ce modèle est une représentation gaussienne des valeurs d'échantillons:

$$Z(x) = \sum_n \frac{\psi_n}{n!} H_n(Y(x))$$

dans laquelle $Y(x)$ est une fonction stationnaire à loi monovariée gaussienne (on ne dit rien à ce niveau des lois multivariées des vecteurs obtenus en considérant n point d'appui $(Y(x_1), \dots, Y(x_n))$).

Aux moyennes par blocs $Z(v)$ sont également associées des représentations gaussiennes:

$$Z(v) = \sum_n \frac{\psi_n^v}{n!} H_n(Y(v))$$

La loi bivariée d'un couple constitué de $Z(v)$ et d'un échantillon $Z(\underline{x})$ de po-

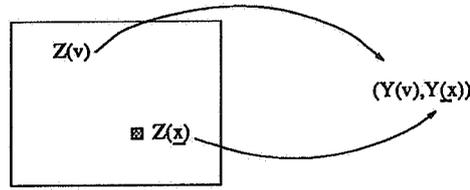


FIG. 5.1 - : Echantillon $Z(\underline{x})$ de position aléatoire et uniforme dans v

sition \underline{x} aléatoire et uniforme dans v , est spécifiée par celle des transformées gaussiennes, $(Y(\underline{x}), Y(v))$, L'hypothèse de base du modèle est que cette loi est gaussienne dans \mathbb{R}^2 , c'est à dire que le vecteur $(Y(\underline{x}), Y(v))$, est lui même gaussien. On note alors $r_{\underline{x},v}$ son coefficient de corrélation.

On utilise ensuite la relation de Cartier, à laquelle satisfait le couple $Z(v), Z(\underline{x})$ ¹:

$$E[Z(\underline{x})|Z(v)] = Z(v)$$

pour spécifier l'anamorphose de blocs, que l'on suppose bijective.

$$\begin{aligned} E[Z(\underline{x})|Z(v)] &= Z(v) = \sum_n \frac{\psi_n^v}{n!} H_n(Y(v)) \\ &= E[Z(\underline{x})|Y(v)] = \sum_n \frac{\psi_n}{n!} E[H_n(Y(\underline{x}))|Y(v)] \\ &= \sum_n \frac{\psi_n}{n!} r_{\underline{x},v}^n H_n(Y(v)) \end{aligned}$$

et donc, par unicité des développements sur les facteurs:

$$\psi_n^v = \psi_n r_{\underline{x},v}^n$$

Cette relation détermine complètement la loi monovariante des blocs.

Il reste à fixer le coefficient $r_{\underline{x},v}$. On utilise pour cela la variance de bloc, calculable à partir du variogramme:

$$\sigma_v^2 = \sigma_x^2 - \bar{\gamma}(v, v)$$

Or cette quantité est déterminée par les coefficients d'anamorphose:

$$\sigma_v^2 = \sum_{n \geq 1} \frac{\psi_n^2}{n!} r^{2n}$$

Comme la fonction $S(x)$:

$$S(x) = \sum_{n \geq 1} \frac{\psi_n^2}{n!} x^{2n}$$

est monotone croissante de $[0, 1]$ dans $[0, \sigma^2]$, il y a une solution unique $r_{\underline{x},v}$ à l'équation:

$$S(r) = \sigma_v^2$$

1. la variable Z est supposée additive, et se compose donc suivant des moyennes arithmétiques

5.2 Passage aux lois multivariées

Avant d'aller plus loin, soulignons que l'hypothèse introduite jusqu'ici concerne les lois $(Z(\underline{x}), Z(v))$ pour un échantillon de position \underline{x} aléatoire uniforme dans v , mais ne dit rien en ce qui concerne les lois bivariées avec des échantillons localisés: $(Z(x), Z(v))$. Il serait tentant de franchir une étape supplémentaire, en étendant à ces dernières lois, ou plutôt à leurs transformées $(Y(x), Y(v))$, l'hypothèse de binormalité. Il se trouve malheureusement qu'une telle hypothèse souffre d'un sérieux problème de cohérence. En effet, d'après un tel modèle, on aurait, en notant r_{xv} la corrélation associée:

$$\begin{aligned} E[Z(x)|Z(v)] &= \sum_n \frac{\psi_n}{n!} r_{xv}^n H_n(Y(v)) \\ E[Z(\underline{x})|Z(v)] &= E[E[Z(\underline{x})|Z(v), \underline{x} = x] | Z(v)] \\ &= \frac{1}{|v|} \int_v E[Z(x)|Z(v)] dx \\ &= \sum_n \frac{\psi_n}{n!} \frac{1}{|v|} \int_v r_{xv}^n dx H_n(Y(v)) \end{aligned}$$

l'unicité des développements entraîne alors les relations:

$$\forall n \quad r_{xv}^n = \frac{1}{|v|} \int_v r_{xv}^n dx$$

D'ou, en particulier:

$$\left[\frac{1}{|v|} \int_v r_{xv} dx \right]^2 = \frac{1}{|v|} \int_v r_{xv}^2 dx$$

et donc:

$$\frac{2}{|v|} \int_v \int_v r_{xv} r_{yv} dx dy = \frac{1}{|v|} \int_v \int_v (r_{xv}^2 + r_{yv}^2) dx dy$$

soit:

$$\frac{1}{|v|} \int_v \int_v (r_{xv} - r_{yv})^2 dx dy = 0$$

et donc $r_{xv} = k$ pour presque tout $x \in v$. Cela entraîne:

$$\forall x \in v \quad \text{Cov}(Z(x), Z(v)) = \sigma_v^2$$

Ce qui n'est compatible qu'avec des modèles très particuliers, et assez peu intéressants, tels l'effet de pépite pur, et pour lesquels $\sigma_v = 0$.

Il en résulte que les modèles bigaussiens ne sont plus applicables aux lois à échantillons localisés $Y(x), Y(v)$. On se limitera donc, si l'on tient à conserver la

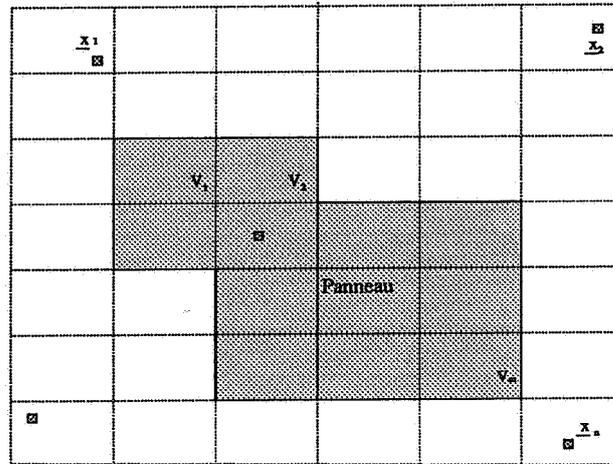


FIG. 5.2 - : L'estimation d'un panneau met en jeu la loi du vecteur $(Y(\underline{x}_1), \dots, Y(\underline{x}_n), Y(v_1), \dots, Y(v_m))$.

normalité, à la modélisation des échantillons de position "randomisée" dans des blocs.

Ceci nous conduit à une description du gisement en terme d'union de blocs disjoints et de même taille. les blocs sont ensuite munis d'un certain nombre d'échantillons dont les positions exactes ne sont pas prises en considération, seule l'appartenance à un bloc donné étant utilisée. Lorsque plusieurs échantillons appartiennent à un même bloc, leurs positions respectives sont supposées indépendantes à l'intérieur de ce bloc.

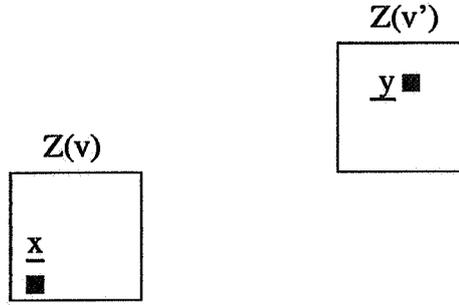
Au cours de l'estimation, on sera donc amené à utiliser des lois de dimension $n + m$, qui correspondent aux vecteurs constitués des gaussiennes associées aux données d'échantillons du voisinage de krigeage, et aux blocs constituant le panneau (figure 5.2):

$$(Y(\underline{x}_1), \dots, Y(\underline{x}_n), Y(v_1), \dots, Y(v_m))$$

Ces lois sont supposées gaussiennes centrées et donc déterminées par les différentes covariances. Nous connaissons la covariance associée à un couple $Y(\underline{x}), Y(v)$ pour $x \in v$. Il reste à déterminer les autres.

C'est ici qu'intervient l'hypothèse d'indépendance conditionnelle. Selon cette hypothèse, lorsque la teneur $Z(v)$ d'un bloc est fixée, les variables relatives à des grandeurs intérieures au bloc v , telles les teneurs d'échantillons intérieurs, sont indépendantes des grandeurs extérieures.

Ainsi si v et v' sont deux blocs distincts, et \underline{x} et \underline{y} deux échantillons aléatoires uniformes dans v et v' respectivement (voir la figure 5.2), les lois de $Y(\underline{x}), Y(v')$ et $Y(\underline{y}), Y(v)$ à $Y(v)$ fixé se factorisent.



Utilisons les décompositions suivantes:

$$\begin{aligned}
 Y(\underline{x}) &= r_{\underline{x},v} Y(v) + \sqrt{1 - r_{\underline{x},v}^2} U_1 \\
 Y(\underline{y}) &= r_{\underline{y},v} Y(v) + \sqrt{1 - r_{\underline{y},v}^2} U_2
 \end{aligned}$$

Dans lesquelles U_1 et U_2 sont des variables normales indépendantes de $Y(v)$, en raison du caractère bigaussien des lois de couples. L'indépendance conditionnelle implique alors que U_1 et U_2 sont indépendantes, puisque portant toute la variabilité de $Y(\underline{x})$ et $Y(\underline{y})$ à $Y(v)$ fixé. Cela entraîne:

$$\text{Cov}(U_1, U_2) \propto \text{Cov}(Y(\underline{x}) - r_{\underline{x},v} Y(v), Y(\underline{y}) - r_{\underline{y},v} Y(v)) = 0$$

et donc la relation:

$$r_{\underline{x},\underline{y}} = \text{Cov}(Y(\underline{x}), Y(\underline{y})) = r_{\underline{x},v} r_{\underline{y},v}$$

on a de même, en exploitant l'autre indépendance conditionnelle:

$$r_{\underline{x},v'} = r_{\underline{x},v} r_{vv'}$$

(la corrélation $r_{\underline{x},v}$ n'est autre que le coefficient r du changement de support, puisque $\underline{x} \in v$) La combinaison de ces relations donne, en définitive:

$$\begin{aligned}
 r_{\underline{x},\underline{y}} &= r^2 r_{vv'} \\
 r_{\underline{x},v'} &= r r_{vv'}
 \end{aligned}$$

de telle sorte que seules les corrélations associées à deux blocs différents v et v' restent à déterminer. La covariance des teneurs de blocs associées est:

$$\text{Cov}(Z(v), Z(v')) = \sum_{n \geq 1} \frac{[\psi_n^v]^2}{n!} r_{vv'}^n = \sum_{n \geq 1} \frac{\psi_n^2}{n!} [r^2 r_{vv'}]^n$$

D'où la valeur $r_{vv'}$:

$$r_{vv'} = \frac{1}{r^2} S^{-1}[\text{Cov}(Z(v), Z(v'))]$$

Le modèle est ainsi entièrement déterminé, puisque toutes les covariances le sont.

Remarque: Soulignons cependant, que l'application directe de la relation précédente n'a aucune raison, en général de conduire à des matrices de covariances r_{v_i, v_j} positives. Aussi est-il préférable de passer par une étape de modélisation des covariances de $Y(v)$:

$$C_v^Y(h) = \text{Cov}(Y(v_h), Y(v))$$

dans la mise en oeuvre de ce modèle. Ainsi, ayant obtenu les covariances en Z au support v , on détermine deux à deux les coefficients r_{v_i, v_i} par inversion de S . C'est sur la base de ces coefficients que $C_v^Y(h)$ est modélisée. Les coefficients utilisés en définitive dans l'estimation découlent de C_v^Y , les r_{v_i, v_j} n'ayant été que des intermédiaires de calcul.

5.3 L'espérance conditionnelle en modèle gaussien discrétisé

Le reste de la méthode est évident. L'estimation d'une fonction $\psi(Z(v_0))$ est réalisée en procédant selon les étapes suivantes:

1. Calcul de l'anamorphose ponctuelle $Z(x) = \phi(Y(x))$
2. Calcul de l'anamorphose de blocs ϕ^v , et donc du coefficient r
3. Calcul des covariances de blocs $C_{v, v'}$. Estimation, puis modélisation des $r_{v, v'}$, ce qui achève la détermination du modèle.
4. Décomposition de $\psi \circ \phi^v$ sur les polynômes H_n :

$$\psi \circ \phi^v = \sum_n \frac{c_n}{n!} H_n$$

5. Calcul des transformées des valeurs d'échantillons: $Y(x_i) = \phi^{-1}(Z(x_i))$, et affectation de ces valeurs à un bloc v_i . On note $Y(\underline{x}_i)$ ces variables
6. Estimation de $Y(v_0)$ par krigeage simple à partir des $Y(\underline{x}_i)$:

$$Y_{v_0}^k = \sum_i \lambda_i Y(\underline{x}_i)$$

avec des pondérateurs donnés par le système:

$$\forall i \quad \sum_j \lambda_j r_{\underline{x}_i, \underline{x}_j} = r_{\underline{x}_i, v_0}$$

La variance d'estimation associée est:

$$(\sigma^k)^2 = 1 - \sum_i \lambda_i r_{\underline{x}_i, v_0}$$

et la variance de $Y_{v_0}^k$:

$$\text{Var}(Y_{v_0}^k) = s^2 = \sum_i \lambda_i r_{\underline{x}_i v_0}$$

7. L'estimation de $\phi(Z(v_0))$ est alors donnée par:

$$[\psi(Z(v_0))]^* = \sum_n \frac{c_n}{n!} s^n H_n \left(\frac{Y_{v_0}^k}{s} \right)$$

Le cas de blocs contenant plusieurs données nécessite un commentaire: pour \underline{x} et \underline{y} situés dans le même bloc v , on a encore indépendance conditionnelle de $Y(\underline{x})$ et $Y(\underline{y})$ à $Y(v)$ fixé, puisque les implantations \underline{x} et \underline{y} sont indépendantes. Cela implique $r_{\underline{x}, \underline{y}} = r^2$.

Supposons maintenant que le bloc v_i contienne n données. Les pondérateurs associés à ces données sont identiques, et sont notés λ_i . On procédera ensuite par regroupement. Ainsi, dans l'équation relative à une donnée $k \notin v_i$, le pondérateur λ_i recevra un coefficient $n r_{\underline{x}_i, \underline{x}_k}$. Enfin, on sommerá toutes les équations associées aux données de v_i , ce qui donnerá en définitive:

- En colonne k le coefficient $n r_{\underline{x}_i, \underline{x}_k}$,
- pour le second membre $n r_{\underline{x}_i, \underline{x}_0}$,
- En colonne i , le terme $n(1 - (n - 1)r^2)$.

L'avantage de cette sommation de n équations identiques, est la préservation de la symétrie de la matrice de krigeage, qui est précieuse du point de vue de la résolution numérique. En ce qui concerne les variances, les formules sont inchangées.

Remarque: La méthode ainsi décrite peut sembler relativement lourde, puisque pour le tonnage minéral d'un panneau:

$$T_P(z) = \frac{1}{N} \sum_{u \in \mathcal{P}} 1_{Z(v_u) \geq z_c}$$

il faut estimer séparément chaque bloc v_u , et donc chaque gaussienne associée $Y(v_u)$ indépendamment. Cependant, dans ces différentes estimations, seul le second membre du système de krigeage change, ainsi que la variance d'estimation, si on a pris soin d'utiliser le même voisinage de krigeage pour tous les blocs du panneau. Dans ce cas, le coût de la méthode en terme de temps de calcul est tout à fait acceptable.

5.4 Le krigeage disjonctif en modèle gaussien discrétisé

Les hypothèses de modélisation sont un peu affaiblies ici, puisque seule la binormalité est utilisée. Le calcul des différentes covariances est inchangé par rapport au cas précédent.

Les étapes du calcul sont un peu différentes, et l'estimation d'une moyenne par zone se fait ici en une seule étape, contrairement au cas précédent. Soit en effet l'estimation de:

$$\Psi = \frac{1}{N} \sum_{u \in \mathcal{P}} \psi(Z(v_u))$$

La fonction $\psi \circ \phi$ est, comme précédemment décomposée sur les facteurs:

$$\Psi = \sum_n \frac{c_n}{n!} \frac{1}{N} \sum_{u \in \mathcal{P}} H_n(Y(v_u))$$

et les quantités:

$$R_n = \frac{1}{N} \sum_{u \in \mathcal{P}} H_n(Y(v_u))$$

sont tour à tour estimées par krigeage, puisque avec la binormalité, on a toujours:

$$\begin{aligned} \langle H_n(Y(\underline{x}_i)), H_m(Y(\underline{x}_j)) \rangle &= \delta_{nm} n! (r_{\underline{x}_i, \underline{x}_j})^n \\ \langle H_n(Y(\underline{x}_i)), H_m(Y(v_j)) \rangle &= \delta_{nm} n! (r_{\underline{x}_i, v_j})^n \end{aligned}$$

et donc:

$$R_n^{k.d.} = \sum_i \lambda_i^n H_n(Y(\underline{x}_i))$$

avec des pondérateurs λ_i^n donnés par:

$$\forall i \quad \sum_j (r_{\underline{x}_i, \underline{x}_j})^n \lambda_i^n = \frac{1}{N} \sum_{u \in \mathcal{P}} (r_{\underline{x}_i, v_u})^n$$

et la variance d'estimation suivante:

$$(\sigma_{k.d.}^n)^2 = \frac{n!}{N^2} \sum_{u, u'} r_{u, u'}^n - \frac{n!}{N} \sum_{u, i} \lambda_i^n r_{\underline{x}_i, u}^n$$

On a finalement l'estimation:

$$\Psi^{kd} = \sum_n \frac{c_n}{n!} R_n^{k.d.}$$

et la variance

$$\text{Var}(\Psi - \Psi^{kd}) = \sum_{n \geq 1} \left(\frac{c_n}{n!} \right)^2 (\sigma_{kd}^n)^2$$

5.5 Extension du krigeage disjonctif à des modèles non gaussiens

On se borne ici à donner quelques indications générales. On se reportera aux références citées pour plus de détails. Cette extension fait appel le plus souvent à des lois isofactorielles non symétriques, et nous commençons par en donner la forme générale. Une loi bivariable $F(du, dv)$, de lois marginales:

$$\begin{aligned} F^1(du) &= \int_v F(du, dv) \\ F^2(dv) &= \int_u F(du, dv) \end{aligned}$$

est isofactorielle si il existe deux familles de fonctions réelles, χ_n^1 et χ_n^2 vérifiant:

i) chaque famille χ_n^i est complète dans $\mathcal{L}^2(\mathbb{R}, F^i)$.

ii) Chaque famille est orthonormée:

$$\int \chi_n^i(x) \chi_m^i(x) F^i(dx) = \delta_{nm}$$

iii) L'orthogonalité s'étend aux lois bivariées:

$$\int \chi_n^1(u) \chi_m^2(v) F(du, dv) = \delta_{nm} T_n$$

Mentionnons à titre d'exemple le modèle bigamma (voir [1] pour une description détaillée des modèles à lois marginales gamma, et pour des cas d'étude convaincants). Dans ce modèle F^1 et F^2 sont des lois gamma de paramètres respectifs α et α' , et les facteurs associés sont les polynômes de Laguerre déjà rencontrés l_n^α et $l_n^{\alpha'}$. La densité bivariable admet le développement:

$$f_\rho^{\alpha, \alpha'}(u, v) = g_\alpha(u) g_{\alpha'}(v) \sum_n \rho^n \sqrt{\frac{\Gamma(\alpha + n) \Gamma(\alpha')}{\Gamma(\alpha' + n) \Gamma(\alpha)}} l_n^\alpha(u) l_n^{\alpha'}(v)$$

avec un coefficient $\rho \in [0, 1]$. On obtient un vecteur (U, V) suivant cette loi à partir d'un vecteur (U, X) suivant lui même une loi bigamma symétrique, de paramètre α (si $\alpha \leq \alpha'$), en prenant $V = X + W$, si W suit une loi gamma de paramètre $\alpha' - \alpha$, et est indépendant de (U, X) .

Le paramètre α est associé au petit support, tandis que α' est associé au support de bloc, puisque la densité associée est plus proche d'une loi normale.

Ainsi, l'une des lois F^1 représente à une anamorphose près les valeurs $Z(\underline{x}_i)$, l'autre les valeurs de blocs. On écrira alors le développement isofactoriel sous la forme:

$$F^{x,v}(u, v) = F^x(du) F^v(dv) \sum_n T_n^{x,v} \chi_n^x(u) \chi_n^v(v)$$

Ces modèles posent l'embarassant problème du choix de la loi destinée à modéliser -toujours à une anamorphose près- la distribution des blocs (paramètre α' dans le cas de la loi bigamma).

L'estimation locale met en jeu les lois bivariées suivantes:

$$\begin{aligned} (Y(\underline{x}_i), Y(\underline{x}_j)) &\equiv \text{isofactorielle symétrique } (\chi_n^x, T_n^{x,x}(\underline{x}_i, \underline{x}_j)) \\ (Y(v_i), Y(v_j)) &\equiv \text{isofactorielle symétrique } (\chi_n^v, T_n^{v,v}(v_i, v_j)) \\ (Y(\underline{x}_i), Y(v_j)) &\equiv \text{isofactorielle non symétrique } (\chi_n^x, \chi_n^v, T_n^{x,v}(\underline{x}_i, v_j)) \end{aligned}$$

La propriété de découplage des estimations de facteurs subsiste, puisque des facteurs d'ordre différent sont toujours orthogonaux dans toutes les lois mises en jeu.

Enfin, l'inférence des paramètres T_n est simplifiée par une factorisation des lois conditionnelles analogue à celle rencontrée dans le modèle gaussien discrétisé. En reprenant le cas de figure utilisé alors: $\underline{x} \in v$, $\underline{x}' \in v'$ on suppose que la loi conditionnelle de $Y(\underline{x})$ à $Y(v')$ fixé s'écrit²

$$F_{y_{v'}}(d(y_x)) = \int_{y_v} F_{y_v}(d(y_x)) F_{y_{v'}}(d(y_v))$$

En utilisant les développements isofactoriels:

$$\begin{aligned} F_{y_v}(d(y_x)) &= F(d(y_x)) \sum T_n^{x,v}(x, v) \chi_n^x(y_x) \chi_n^v(y_v) \\ F_{y_{v'}}(d(y_x)) &= F(d(y_x)) \sum T_n^{x,v}(x, v') \chi_n^x(y_x) \chi_n^v(y_{v'}) \\ F_{y_{v'}}(d(y_v)) &= F(d(y_v)) \sum T_n^{v,v}(v, v') \chi_n^v(y_v) \chi_n^v(y_{v'}) \end{aligned}$$

On en déduit les relations:

$$T_n^{x,v}(x, v') = T_n^{x,v}(x, v) T_n^{v,v}(v, v')$$

et de même:

$$T_n^{x,x'}(x, x') = (T_n^{x,v}(x, v))^2 T_n^{v,v}(v, v')$$

Ainsi, les paramètres principaux sont les coefficients de changement de support, $T_n^{x,v}(x, v)$ ($x \in v$), et les $T_n^{v,v}(v, v')$ associés à toutes les paires de blocs du voisinage.

2. L'expression générale serait:

$$F_{y_{v'}}(d(y_x)) = \int_{y_v} F_{y_v, y_{v'}}(d(y_x)) F_{y_{v'}}(d(y_v))$$

La simplification utilisée est donc vraie dans le cas de l'indépendance conditionnelle vue plus haut.

Exercices

5.6 Exercice 1. Sur le modèle mosaïque.

Rappels. Une fonction aléatoire stationnaire $Z(x)$ est dite hermitienne si elle est de loi monovariante gaussienne, et si les polynômes d'hermite normés $\eta_n = H_n/\sqrt{n!}$ sont orthogonaux:

$$\begin{array}{ll} i) \forall x & P[Z(x) < z] = G(z) \\ ii) \forall(x, y) \quad \forall(n \neq m) & \text{Cov}[\eta_n(Z(x)), \eta_m(Z(y))] = 0 \\ iii) \forall(x, y) \quad \forall(n \neq 0) & \text{Cov}[\eta_n(Z(x)), \eta_n(Z(y))] = T_n(x - y) \end{array}$$

avec des coefficients T_n dépendant de la distance $h = x - y$ sous la forme:

$$T_n(h) = \int r^n \omega_h(dr)$$

où ω_h est une loi de probabilité (dépendant de la distance) sur l'intervalle $[-1, +1]$. Cela implique que:

$$E[\eta_n(Z(x+h)) | Z(x)] = T_n(h) \eta_n(Z(x)).$$

Cette relation reste valable pour $n = 0$, avec $T_0(h) = 1$.

Exercice. On s'intéresse au cas particulier où la loi ω est de la forme:

$$\omega_h(dr) = p(h) \delta_1(dr) + (1 - p(h)) \delta_0(dr)$$

C'est donc la loi d'une variable valant 1 avec la probabilité $p(h)$, et 0 avec la probabilité complémentaire. On suppose donc que $0 \leq p(h) \leq 1$.

1. Quelle est, dans ces conditions, la forme des coefficients T_n ? (attention au cas $n = 0$). En déduire que la fonction $p(h)$ est nécessairement de type positif.

2. Si $\phi(z)$ admet le développement: $\phi(z) = \sum \phi_n \eta_n(z)$, quelle est la fonction de corrélation de ϕ :

$$\rho_\phi(h) = \frac{C_\phi(h)}{\sigma_\phi^2} = \frac{\text{Cov}[\phi(Z(x)), \phi(Z(x+h))]}{\text{var}[\phi(Z(x))]} \quad ?$$

Que vaut $E[\phi(Z(x+h)) | Z(x)]$ dans ce modèle? (attention au terme de moyenne). En déduire la forme de la loi conditionnelle $F_{z'}(dz)$ de $Z(x+h)$ à $Z(x) = z'$ fixé.

3. On cherche à estimer $\phi(Z_0)$ à partir de Z_1, \dots, Z_n . Donner le système permettant le calcul de $\eta_n^{k.d.}(Z_0)$. Montrer que le krigeage simple (à moyenne connue) de $\phi(Z_0)$ à partir des $\phi(Z_1), \dots, \phi(Z_n)$ coïncide dans ce modèle avec le krigeage disjonctif:

$$\phi^{k.d.}(Z_0) = \phi^{k.s.}(Z_0)$$

5.7 Exercice 2.

Soit à une dimension la fonction aléatoire gaussienne stationnaire $Z(x)$, de moyenne nulle, et de covariance $C(h) = e^{-a|h|}$ (avec un paramètre $a > 0$). Cette fonction est supposée connue partout, à l'exception de l'intervalle ouvert $] -1, +1[$. On note \mathcal{D} le domaine correspondant:

$$\mathcal{D} =] -\infty, -1] \cup [1, +\infty[$$

1. Montrer que dans ces conditions, l'espérance conditionnelle:

$$E[Z(0) | Z(x), x \in \mathcal{D}]$$

ne dépend que des deux valeurs $Z(-1)$ et $Z(1)$.

2. Quel est le krigeage disjonctif de $Z(0)$ à partir de l'ensemble des valeurs de Z dans \mathcal{D} ?
3. Quel est le krigeage disjonctif de $\eta_n(Z(0))$ (à partir des mêmes informations)? Montrer que lorsque $a \rightarrow \infty$ on a asymptotiquement:

$$[\eta_n(Z(0))]^{k.d.} \approx E[\eta_n(Z(0)) | Z(1)] + E[\eta_n(Z(0)) | Z(-1)]$$

4. Quelle est, dans les mêmes conditions, l'expression approchée du krigeage disjonctif de $\phi(Z(0))$?

5.8 Exercice 3. Estimation par cokrigeage simple.

On cherche à former l'estimation de $\phi(Z_0)$ par cokrigeage simple à partir des données Z_i :

$$\phi^{c.k.s.}(Z_0) = E[\phi] + \sum_i \lambda_i(\phi) (Z_i - m)$$

La dépendance par rapport à la fonction ϕ va être explicitée pour donner une représentation plus commode.

1. On suppose les régressions de Z_i à Z_0 fixé sont affines:

$$E[Z_i|Z_0] = m + r_i(Z_0 - m).$$

Expliciter les coefficients r_i en fonction des covariances des Z . Montrer que les covariances de $\phi(Z_0)$ et Z_i sont de la forme:

$$\text{Cov}(\phi(Z_0), Z_i) = r_i \text{Cov}(\phi(Z_0), Z_0) = r_i \left\{ \int z \phi(z) F(dz) - mE[\phi] \right\}$$

(F désigne la loi commune à Z_0 et aux Z_i). En déduire que les poids de cokrigeage sont de la forme:

$$\lambda_i(\phi) = \lambda_i^{k.s.} \cdot \frac{\text{Cov}(\phi, Z)}{\sigma_Z^2}$$

où les $\lambda_i^{k.s.}$ sont les poids des Z_i dans le krigeage simple de Z_0 .

2. En déduire que l'estimation de l'indicatrice est de la forme:

$$(1_{Z_0 \geq z})^{c.k.s.} = T(z) + (Z_0^{k.s.} - m) \frac{Q(z) - mT(z)}{\sigma_Z^2}$$

Critiquer l'estimateur ainsi obtenu, en considérant par exemple le cas d'une fonction aléatoire stationnaire et continue en moyenne quadratique, et où le point à estimer x_0 se rapproche d'un point informé x_1 , de telle sorte que $Z(x_0) \rightarrow Z(x_1)$.

5.9 Exercice 4. Effet d'information.

La quantité de métal qu'un exploitant minier croit récupérer sur la base d'un estimateur ultime de teneur bloc Z_V^* , est en espérance donnée par:

$$Q^{ill.}(z) = E[Z_V^* 1_{Z_V^* \geq z}]$$

tandis que, en réalité il récupère la quantité:

$$Q^{eff.}(z) = E[Z_V 1_{Z_V^* \geq z}]$$

Ces deux quantités peuvent différer sensiblement, ce qui peut être à la base de sérieuses déconvenues. On va voir qu'il est possible de modifier l'estimateur ultime de blocs Z_V^* , de façon à assurer que la quantité prévue soit effectivement récupérée.

Rappels.

Les représentations gaussiennes des teneurs d'échantillons et de blocs sont supposées connues:

$$\begin{aligned}Z_x &= \sum \phi_n \eta_n(Y_x) \\Z_V &= \sum \phi_n r^n \eta_n(Y_V)\end{aligned}$$

Et on suppose que l'estimateur ultime dépend d'une configuration d'échantillons fixe (par exemple de 4 trous de tirs intérieurs à chaque bloc): $Z_V^* = \sum \lambda_i Z_i$, avec des pondérateurs que l'on suppose tels que:

$$\begin{aligned}\forall i \quad \lambda_i &\geq 0 \\ \sum_i \lambda_i &= 1\end{aligned}$$

Ces pondérateurs sont alors interprétés comme une loi de probabilité sur l'ensemble des indices, et on note I une variable aléatoire associée. Z_I désigne la teneur d'échantillon ainsi randomisée, et $P[Z_I = Z_i] = \lambda_i$.

Questions:

1. Que vaut $Q^{eff.}$ en terme de $h(Z^*) = E[Z|Z^*]$, et de la loi $F_{V^*}(z)$ des estimateurs ultimes?
2. Donner la relation de Cartier entre Z_V^* et Z_I .
3. On prend alors pour Z_V^* la représentation gaussienne suivante:

$$Z_V^* = \sum_n \phi_n^* \eta_n(Y_V^*)$$

(les ϕ_n^* sont les coefficients de l'anamorphose gaussienne associée à la variable Z_V^* , et Y_V^* est la gaussienne correspondante). En ce qui concerne Z_I , on a le développement:

$$Z_I = \sum_n \phi_n \eta_n(Y_I)$$

Montrer que, en faisant l'hypothèse que les couples (Y_V^*, Y_I) sont gaussiens et de corrélation r^* , les coefficients ϕ_n^* se trouvent automatiquement déterminés.

4. On s'intéresse maintenant aux couples (Z_V, Z_V^*) . On fera l'hypothèse que (Y_V, Y_V^*) est un couple gaussien dont le coefficient de corrélation est noté r_{VV^*} . Donner l'expression de $h(z)$. Que valent $Q^{ill.}$ et $Q^{eff.}$, si l'estimateur Z_V^* est remplacé par $Z_V^{**} = h(Z_V^*)$?

5.10 Exercice 5. Présence de dérive.

Soit une fonction aléatoire de la forme:

$$Z(x) = m(x) \exp \left\{ \sigma Y(x) - \sigma^2/2 \right\}$$

où Y est une fonction stationnaire gaussienne centrée et normée, et m est une fonction déterministe. On suppose connue la variance logarithmique σ^2 , ainsi que la covariance de Y , notée C_Y , mais non la dérive $m(x)$. La fonction Z est mesurée en un certain nombre de points, x_1, \dots, x_N , de telle sorte que les valeurs de $Z(x_1), \dots, Z(x_N)$ sont connues, mais non les transformées $Y(x_1), \dots, Y(x_N)$.

1. On cherche à estimer $m(x)$ en un point x , et on s'oriente vers un estimateur de la forme:

$$m(x)^* = \exp \left\{ \sum_i \lambda_i \ln Z(x_i) + \lambda_0 \right\} = e^{\lambda_0} \prod_{i=1}^N Z(x_i)^{\lambda_i}$$

Montrer que, si $m(x)$ est de la forme:

$$m(x) = e^{p_k(x)}$$

où p_k est un polynôme de degré k en x , alors $m(x)^*$ est sans biais quels que soient les coefficients de p_k , si et seulement si λ_i et λ_0 vérifient:

- i) Pour tout polynôme $q_l(x)$ de degré $l \leq k$: $\sum_i \lambda_i q_l(x_i) = q_l(x)$
- ii) $\lambda_0 = -\sigma^2/2 \left\{ \sum_{i,j} \lambda_i \lambda_j C_Y(x_i - x_j) - \sum_i \lambda_i \right\}$

2. Quel est le système de krigeage permettant le calcul des pondérateurs qui minimisent:

$$\text{Var}[m(x) - m(x)^*] \quad ?$$

et que vaut cette variance?

3. On supposera les points de données en nombre suffisant pour assurer l'existence des polynômes orthogonaux, de degré inférieur ou égal à k , pour $1/N \sum_i \delta_{x_i}(dx)$, c'est à dire des polynômes $\chi_l(x)$ tels que:

$$\frac{1}{N} \sum_{i=1}^N \chi_l(x_i) \chi_m(x_i) = \delta_{lm}$$

Dans le cas où C_Y est un effet de pépite pur, que valent les pondérateurs λ_i , et la variance d'estimation en fonction des polynômes χ_l ?

Si $k = 0$, et donc que $m(x) = m_0$ est constante, que deviennent les λ_i ? Montrer que alors $\text{Var}(m_0 - m_0^*)$ tend vers 0 quand le nombre de points N augmente.

5.11 Exercice 6. Données sous forme d'intervalles.

Il arrive dans la pratique que les données se présentent sous la forme d'intervalles: on sait alors en un point informé x_i que $a < Z(x_i) < b$, par exemple, sans connaître la valeur précise de $Z(x_i)$. Ce problème envisage l'utilisation d'un krigeage disjonctif, dans le cas mixte où sont présentes des données exactes ainsi que des données de type intervalles (d'après la thèse de Valérie Langlais).

1. Soit I une variable aléatoire discrète, de loi $P[I = i] = p_i$, et soit Y une variable quelconque. Montrer que:

$$E[Y|I] = \sum_i E[Y|I = i] 1_{I=i}$$

(On se bornera à vérifier que le terme de droite est solution de l'équation de projection).

2. Soit maintenant (Y_1, Y_2) un vecteur gaussien, de lois marginales centrées normées, et de coefficient de corrélation ρ . Soit D l'intervalle $D = [a, b]$ (avec $a < b$). On notera $I_2 = 1_{Y_2 \in D}$. En utilisant le développement de l'indicatrice sur les polynômes d'Hermite, donner l'expression de:

$$E[I_2|Y_1]$$

Quelle est la loi conditionnelle de Y_2 , lorsque $I_2 = 1$? Que valent $E[H_n(Y_1)|I_2 = 1]$ et $E[H_n(Y_1)|I_2 = 0]$?

Donner finalement l'expression de:

$$E[H_n(Y_1)|I_2]$$

(Il sera commode de noter $\Delta(\phi)$ l'accroissement d'une quelconque fonction ϕ sur D , soit $\Delta(\phi) = \phi(b) - \phi(a)$).

3. Soit le vecteur gaussien centré et réduit (Y_0, Y_1, Y_2) . Nous voulons estimer Y_0 en se basant sur la connaissance de Y_1 et de l'indicatrice $I_2 = 1_{Y_2 \in D}$. On envisage pour cela le krigeage disjonctif de Y_0 :

$$Y_0^* = f_1(Y_1) + f_2(I_2) = \sum_n \lambda_n H_n(Y_1) + \mu(I_2 - E[I_2])$$

Expliciter le système linéaire permettant la détermination des coefficients λ_n et μ

4. Sous les mêmes conditions que dans la question précédente, que devient le coefficient μ lorsque l'intervalle D est symétrique ($D = [-b, b]$)? Quelle conclusion en tirer?

Annexe A

Rappels Elementaires Sur Les Projections Dans Des Espaces De Hilbert

Notations . Soit V un espace vectoriel dont les éléments seront notés u, v, w par exemple. Les lettres grecques α, β, \dots seront utilisées pour noter les scalaires. Les majuscules seront utilisées pour noter: les espaces vectoriels eux mêmes, les tableaux à une ou plusieurs dimensions, et les variables aléatoires (donc également hélas les tableaux aléatoires).

Définition 1. Un *produit scalaire* est une application notée $\langle \cdot, \cdot \rangle$, de $V \times V$ dans R (on se concentre sur le cas réel), possédant les propriétés suivantes :

i) Symétrie:

$$\forall (u, v) \in V^2 \quad \langle u, v \rangle = \langle v, u \rangle$$

ii) linéarité:

$$\forall (u, v, w) \forall (\alpha, \beta) \quad \langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$$

iii) Positivité:

$$\begin{aligned} \forall (u) \quad \langle u, u \rangle &\geq 0 \\ \text{et } \langle u, u \rangle = 0 &\iff u = 0 \end{aligned}$$

A un produit scalaire peut être associé une *norme* de la manière suivante:

$$\| u \|^2 = \langle u, u \rangle$$

Ce qui permet en particulier de dire si deux éléments u et v sont proches ou non. La *convergence* de suite a donc un sens précis:

$$\lim_{n \rightarrow \infty} u_n = u \quad \Longleftrightarrow \quad \|u_n - u\| \rightarrow 0$$

Malheureusement cette notion présente l'inconvénient que pour dire si une suite est convergente il faut d'abord en connaître la limite. On utilise alors la notion de *suite de Cauchy*. u_n est une suite de Cauchy si et seulement si:

$$\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty \\ \text{indépendamment}}} \|u_n - u_m\| = 0$$

Cette propriété, qui est plus facile à établir, n'est hélas nullement équivalente à la convergence, bien qu'elle en résulte. L'équivalence a pourtant bien lieu pour une classe particulière d'espaces vectoriels normés: les *espaces de Banach* ou espaces vectoriels complets. Notons bien que cette propriété dépend de la norme choisie: un même espace peut être complet pour une norme donnée et non pour une autre. Il en était d'ailleurs ainsi de la notion même de convergence.

Définition 2. Un *espace de Hilbert* est un espace vectoriel muni d'un produit scalaire et complet pour la norme induite par ce produit scalaire.

Deux éléments u et v non nuls seront dits *orthogonaux* et on notera $u \perp v$ si leur produit scalaire est nul. Plus généralement u sera dit orthogonal à une partie $H \subset V$ si il est orthogonal à tout élément $v \in H$.

Donnons dès à présent deux exemples, l'un fondamental pour notre compréhension intuitive, l'autre parce qu'il est au centre des applications à la géostatistique.

EXEMPLE 1. Lorsqu'un référentiel a été choisi, les vecteurs de l'espace euclidien sont définis par leur coordonnées. A tout vecteur est donc associé un tableau à trois indices, donc un élément de R^3 . Cette association est d'ailleurs compatible avec l'addition: le tableau représentatif de la somme de deux vecteurs est bien la somme des deux tableaux, en d'autres termes il y a isomorphie, ce qui justifie l'usage de la même notation pour désigner les vecteurs de la géométrie et les tableaux. (figure A.1)

Le produit scalaire est défini comme $\langle X, Y \rangle = x_1y_1 + x_2y_2 + x_3y_3$. La norme associée est bien sûr la longueur du vecteur: $\|X\| = \sqrt{x_1^2 + x_2^2 + x_3^2} = \bar{X}$, et l'orthogonalité traduit bien la propriété géométrique usuelle:

$$\begin{aligned} \{X \perp Y\} & \Longleftrightarrow \langle X, Y \rangle = \|X\| \cdot \|Y\| \cos(\theta) \\ & \Longleftrightarrow \cos(\theta) = 0 \end{aligned}$$

EXEMPLE 2. L'espace L^2 des variables aléatoires réelles admettant un moment d'ordre 2 fini:

$$L^2 = \{X : E[X^2] < \infty\}$$

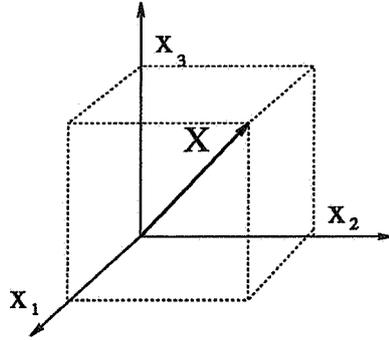


FIG. A.1 - :

Cet espace est muni du produit scalaire :

$$\langle X, Y \rangle = E[X, Y]$$

qui est la covariance non centrée entre X et Y . la norme associée $\| X \| = \sqrt{E[X^2]}$ est donc pour des variables centrées l'écart type σ_X . La convergence associée est la convergence en moyenne quadratique.

Définition : Un sous espace vectoriel de V soit $H \subset V$ est un sous ensemble de V tel que les opérations d'addition et de multiplication par des scalaires restent internes:

$$\forall (u, v) \in H^2 ; \forall (\alpha, \beta) \text{ scalaires} : \quad \alpha u + \beta v \in H$$

Lorsque V est un espace de Hilbert, les suites de Cauchy de H convergent nécessairement dans V , mais leur limite peut ne pas appartenir à H . Par contre si H est lui même complet (fermé), il contient toujours de telles limites. H est alors lui même un espace de Hilbert, et on parle de sous-espace de Hilbert.

PROPRIETE FONDAMENTALE

Soit V un espace de Hilbert et $H \subset V$ un sous espace vectoriel complet de V . Tout élément v de V a alors une **meilleure approximation** dans H , notée v^* qui est **caractérisée** par l'équation de projection suivante:

$$\forall w \in H \quad \langle v - v^*, w \rangle = 0 \tag{A.1}$$

autrement dit l'erreur d'approximation $v - v^*$ est orthogonale à H .

REMARQUE: Il est donc entendu que l'équation A.1 a toujours une solution unique. En particulier, si nous trouvons un élément de H la vérifiant, il sera toujours la meilleure approximation de v pour la norme $\| \cdot \|$.

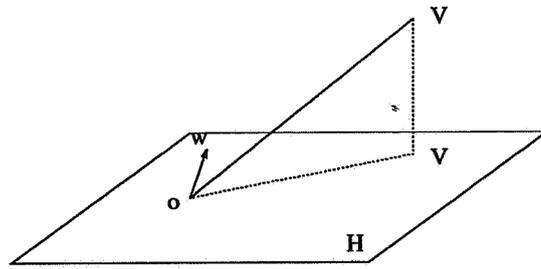


FIG. A.2 - : Projection dans un espace euclidien

REMARQUE: En ce qui concerne l'erreur d'approximation, sa norme vérifie:

$$\begin{aligned} \|v - v^*\|^2 &= \langle v - v^*, v - v^* \rangle \\ \text{mais } v^* \in H \text{ donc: } &\langle v - v^*, v^* \rangle = 0 \\ \text{il reste: } &\|v - v^*\|^2 = \langle v, v \rangle - \langle v, v^* \rangle \\ &= \|v\|^2 - \langle v, v^* \rangle \\ \text{ou encore: } &\|v - v^*\|^2 = \|v\|^2 - \|v^*\|^2 \end{aligned}$$

EXEMPLE 1. Dans R^3 un plan contenant l'origine 0 est un sous espace complet. Dire que $v - v^*$ est orthogonal à tout élément du plan est dire que v^* est le pied de la perpendiculaire menée de v au plan. On a le même résultat en remplaçant le plan par une droite passant par l'origine, qui est un autre exemple de sous espace complet. La relation concernant la variance d'erreur n'est autre que le théorème de pythagore appliqué au triangle rectangle de côtés v , v^* , et $v - v^*$. (figure A.2)

REMARQUE: Si il existe une famille génératrice de H , soit ϕ_n , le système A.1 caractérisant v^* se simplifie:

$$\forall n \quad \langle v - v^*, \phi_n \rangle = 0 \quad (\text{A.2})$$

En effet, tout élément w de H se met sous la forme d'une somme, (éventuellement infinie) $w = \sum c_n \phi_n$ où les c_n sont des scalaires. En appliquant la linéarité du produit scalaire, le résultat suit.

EXEMPLE 1. Dans R^3 , deux vecteurs linéairement indépendants du plan e_1 et e_2 forment un système générateur. La projection v^* est donc encore caractérisée par:

$$\begin{aligned} \langle v - v^*, e_1 \rangle &= 0 \\ \langle v - v^*, e_2 \rangle &= 0 \end{aligned}$$

Le gain est assez substantiel puisque l'orthogonalité n'a à être vérifiée que pour deux éléments, et non pour tous les vecteurs du plan (une infinité d'éléments).

EXEMPLE 2. Dans L^2 supposons connues N variables aléatoires Z_1, \dots, Z_N à travers leurs réalisations z_1, \dots, z_N , et soit Z_0 une variable liée d'une certaine

manière aux précédentes. On supposera ces variables centrées dans un premier temps, soit $E[Z_i] = 0$. Le problème posé est de prévoir le plus précisément possible la valeur inconnue Z_0 sur la base des Z_i ; $i \geq 1$. On se limitera dans un premier temps aux estimateurs obtenus comme combinaison linéaire des Z_i , donc:

$$Z_0^* = \sum_{i=1}^N \lambda_i Z_i$$

Ce problème est résolu en adoptant le cadre général défini précédemment: l'espace d'approximation H est justement constitué de l'ensemble des combinaisons linéaires des Z_i :

$$H = \left\{ X : X = \sum_{i \geq 1} \mu_i Z_i \right\}$$

Cet ensemble est un sous espace de Hilbert de L^2 , et les Z_i en constituent une famille génératrice. Nous sommes donc assurés de l'existence de Z_0^* , qui est solution de l'équation A.2:

$$\forall j \geq 1 \quad \langle Z_0^* - Z_0, Z_j \rangle = 0$$

et donc du système linéaire:

$$\forall j \geq 1 \quad \sum_i C_{ij} \lambda_i = C_{0j}$$

les C_{ij} étant les covariances des Z_i, Z_j . Ce système n'a de solution unique que si la matrice C_{ij} est de rang N (dans le cas contraire, toutes les solutions λ définissent le même Z_0^*).

EXEMPLE 2'. Nous nous plaçons ici dans le même contexte, mais en supposant les espérances $E[Z_i] = m_i$ connues mais non nulles. On se ramène au cas précédent en posant: $X_i = Z_i - m_i$, et en appliquant aux X_i le résultat précédent. On arrive alors à:

$$Z_0^* = m_0 + \sum_{i \geq 1} \lambda_i (Z_i - m_i)$$

Ceci constitue la base de méthodes telles que la régression linéaire multiple. Dans un cadre de fonctions aléatoires stationnaires ceci est bien sûr le krigeage à moyenne connue, ou krigeage simple.

EXEMPLE 3. On considérera dans un premier temps le problème à une seule variable informante: il s'agit d'estimer Z_0 sur la base de Z_1 seul. En revanche, nous cesserons de nous limiter aux seules fonctions affines de Z_1 (de la forme $Z_0^* = a + b Z_1$), pour considérer la classe plus générale des fonctions à variance finie:

$$H = \left\{ \psi(Z_1) : E[\psi(Z_1)^2] < \infty \right\}$$

Cet espace est encore un sous espace complet de L^2 (admis). Dans ces conditions, nous savons que la meilleure approximation en moyenne quadratique de Z_0 dans H existe et est caractérisée par:

$$\forall \psi \text{ t.q. } E[\psi(Z_1)^2] < \infty : \quad \langle Z_0 - Z_0^*, \psi(Z_1) \rangle = 0$$

Cet élément $\psi(Z_1)$ est par définition l'espérance conditionnelle de Z_0 à Z_1 fixé, et est noté $E[Z_0|Z_1]$.

REMARQUE: La définition de l'espérance conditionnelle en terme de densité conditionnelle:

$$E[X|Y] = \int x f_Y(x) dx$$

vérifie trivialement l'équation de projection A.1:

$$\begin{aligned} \langle \int x f_Y(x) dx, \psi(Y) \rangle &= E[\int x \psi(Y) f_Y(x) dx] \\ &= \int_y \int_x x \psi(y) f_y(x) f(y) dx dy \\ &= \int_y \int_x x \psi(y) f(x, y) dx dy = E[X \psi(Y)] \end{aligned}$$

(on note $f_y(x)$ la densité conditionnelle: $f(x, y)/f(y)$, et $f(y)$ la densité marginale: $\int f(x, y) dy$)

EXEMPLE 3'. Dans une généralisation au cas multivariable, l'espace H est cette fois l'ensemble des fonctions des N variables Z_1, \dots, Z_N à variance finie, et l'espérance conditionnelle $E[Z_0|Z_1, \dots, Z_N]$ est solution de:

$$\forall \psi \text{ tq } E[\psi(Z_1, \dots, Z_N)^2] < \infty : \quad \langle Z_0 - E[Z_0|Z_1, \dots, Z_N], \psi(Z_1, \dots, Z_N) \rangle = 0$$

REMARQUE: Lorsque l'espace H contient les constantes, l'estimateur obtenu par projection est toujours sans biais:

$$\langle Z_0 - Z_0^*, 1 \rangle = 0 \quad \Rightarrow \quad E[Z_0] = E[Z_0^*]$$

REMARQUE: Désignons par Π l'application qui à tout élément Z_0 associe son approximation $\Pi(Z_0) = Z_0^*$. cette application est:

i) idempotente

$$\Pi \circ \Pi = \Pi$$

ii) continue et contractante:

$$\|\Pi\| \leq 1$$

(c'est à dire: $\forall Z \quad \|\Pi(Z)\| \leq \|Z\|$)

iii) Linéaire :

$$\forall (Z, Z') \forall (\alpha, \alpha') \quad \Pi(\alpha Z + \alpha' Z') = \alpha \Pi(Z) + \alpha' \Pi(Z')$$

La première propriété est évidente. En ce qui concerne la seconde, on a vu que :

$$\| Z_0 - \Pi(Z_0) \|^2 = \| Z_0 \|^2 - \| \Pi(Z_0) \|^2 \geq 0$$

$$\text{donc : } \| \Pi(Z_0) \| \leq \| Z_0 \|$$

Enfin, il est immédiat de vérifier que $\alpha \Pi(Z) + \alpha' \Pi(Z')$ est solution de l'équation définissant $\Pi(\alpha Z + \alpha' Z')$:

$$\begin{aligned} \forall V \in H \quad \langle \alpha \Pi(Z) + \alpha' \Pi(Z'), V \rangle &= \alpha \langle \Pi(Z), V \rangle + \alpha' \langle \Pi(Z'), V \rangle \\ &= \alpha \langle Z, V \rangle + \alpha' \langle Z', V \rangle \\ &= \langle \alpha Z + \alpha' Z', V \rangle \end{aligned}$$

et donc $\Pi(\alpha Z + \alpha' Z') = \alpha \Pi(Z) + \alpha' \Pi(Z')$. De cela résultent des simplifications dont on tirera parti ultérieurement.

RECAPITULATION

- La meilleure approximation d'un élément quelconque dans un sous espace de Hilbert est caractérisée comme l'unique élément (du sous espace) tel que l'erreur d'approximation soit orthogonale au sous espace.
- Lorsque le sous espace d'approximation est constitué de l'ensemble des combinaisons linéaires de variables aléatoires connues, cette meilleure approximation est la régression linéaire multiple ou le krigeage simple, suivant le contexte.
- Si le sous espace est constitué de l'ensemble de toutes les fonctions des variables connues: $\phi(Z_1, \dots, Z_N)$ à variance finie, cette meilleure approximation est l'espérance conditionnelle.
- Nous verrons que le krigeage disjonctif entre également dans le cadre des applications de ce résultat, lorsque l'espace d'approximation est constitué de l'ensemble des sommes de fonctions de chaque variable $f_1(Z_1) + \dots + f_N(Z_N)$ (à variance finie).

Annexe B

Rappels sur les lois gaussiennes.

B.1 Loi normale à une dimension.

Une variable aléatoire Y suit loi normale centrée et réduite si sa densité est donnée par:

$$g(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\}$$

On a la formule utile suivante:

$$E[e^{aY}] = \frac{1}{\sqrt{2\pi}} \int e^{ay - \frac{1}{2}y^2} dy = e^{\frac{a^2}{2}} \int g(y - a) dy = e^{\frac{a^2}{2}}$$

En particulier la fonction caractéristique associée est:

$$\phi(u) = E[e^{iuY}] = \exp\left\{-\frac{1}{2}u^2\right\}$$

Une variable normale d'espérance m et de variance σ^2 ,

$$X = m + \sigma Y$$

a donc pour densité:

$$g_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left\{-\frac{1}{2} \left(\frac{x - m}{\sigma}\right)^2\right\}$$

et pour fonction caractéristique:

$$\phi_{m,\sigma}(u) = E[e^{iu(m+\sigma Y)}] = \exp\left\{ium - \frac{u^2\sigma^2}{2}\right\}$$

B.2 Loi normale vectorielle.

Définition. Un vecteur aléatoire V est dit gaussien si toute combinaison linéaire de ses composantes est gaussienne.

REMARQUE: Il en résulte que pour toute matrice A (ayant le nombre voulu de colonnes), le vecteur AV est encore gaussien.

Un vecteur gaussien est caractérisé par:

- Son espérance:

$$M = E[V]$$

vecteur dont les composantes sont $m_i = E[V_i]$.

- Sa matrice de variance-covariance:

$$\Sigma = E[(V - M).(V - M)^t]$$

matrice dont les termes sont les covariances:

$$\begin{aligned}\sigma_{ii} &= \text{Var}\{V_i\} \\ \sigma_{ij} &= \text{Cov}\{V_i, V_j\}\end{aligned}$$

Il est facile d'en déduire la fonction caractéristique de V :

$$\phi(U) = E[\exp\{iU^t V\}] = E[\exp\{i \sum u_i V_i\}]$$

En effet, $U^t V$ est une variable gaussienne, d'espérance $U^t M = \sum u_i m_i$, et de variance:

$$\text{Var}\{\sum u_i V_i\} = U^t \Sigma U$$

Et donc, en utilisant la fonction caractéristique de la loi de Gauss:

$$E[e^{ia(U^t V)}] = \exp\left\{ia(U^t M) - \frac{a^2 U^t \Sigma U}{2}\right\}$$

et, en faisant $a = 1$:

$$\phi(U) = \exp\left\{iU^t M - \frac{1}{2}U^t \Sigma U\right\}$$

Ce qui montre bien que M et Σ suffisent à caractériser la loi du vecteur.

De la forme de la fonction caractéristique, résulte la propriété suivante:

Propriété d'indépendance des vecteurs orthogonaux. Deux vecteurs V^1 et V^2 , dont la concaténation $V = (V^1, V^2)$ est un vecteur gaussien, sont indépendants si et seulement si $V^1 - E[V^1]$ et $V^2 - E[V^2]$ sont orthogonaux.

Cette condition signifie que les covariances entre composantes sont toutes nulles:

$$\forall (i, j) \quad \text{Cov}\{V_i^1, V_j^2\} = 0$$

En effet, la matrice de variance-covariance de V est alors bloc diagonale:

$$\Sigma_V = \begin{pmatrix} \Sigma_{V^1} & 0 \\ 0 & \Sigma_{V^2} \end{pmatrix}$$

et la fonction caractéristique de V se factorise:

$$\phi_V(U) = \exp \left\{ i(U_1^t M_1 + U_2^t M_2 - \frac{1}{2}(U_1^t \Sigma_1 U_1 + U_2^t \Sigma_2 U_2)) \right\}$$

$$\phi_V(U) = \phi_{V^1}(U_1) \phi_{V^2}(U_2)$$

D'où l'indépendance de V^1 et V^2 .

Application. Si (V_0, V_1, \dots, V_n) est un vecteur gaussien, le krigeage simple de V_0 :

$$V_0^{ks} = m_0 + \sum_{i \geq 1} \lambda_i (V_i - m_i)$$

vérifie:

$$\forall i = 1, \dots, n \quad V_0 - V_0^{ks} \perp (V_i - m_i)$$

Or $(V_0 - V_0^{ks}, V_1, \dots, V_n)$ est aussi un vecteur gaussien, ce qui implique l'indépendance du résidu $V_0 - V_0^{ks}$ et des variables informantes V_1, \dots, V_n . Alors pour toute fonction $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ assurant une variance finie à la variable aléatoire $\psi(V_1, \dots, V_n)$:

$$E[(V_0 - V_0^{ks}) \psi(V_1, \dots, V_n)] = E[V_0 - V_0^{ks}] E[\psi(V_1, \dots, V_n)] = 0$$

Ainsi V_0^{ks} vérifie l'équation de projection qui définit $E[V_0 | V_1, \dots, V_n]$.

$$V_0^{ks} = E[V_0 | V_1, \dots, V_n]$$

. Ainsi, dans le cas d'un vecteur gaussien, le krigeage simple donne l'espérance conditionnelle.

Donnons pour finir la forme de la densité de la loi gaussienne vectorielle, dans le cas régulier (matrice Σ de plein rang). Soient X_i les vecteurs propres normés de la matrice de variance-covariance:

$$\Sigma \cdot X_i = \lambda_i X_i$$

Ce qui s'écrit matriciellement:

$$\Sigma X = X \Lambda$$

En prenant $\Lambda = \text{Diag}\{\lambda_1, \dots, \lambda_n\}$, et $X = (X_1, \dots, X_n)$, matrice dont les colonnes sont constituées des X_i . Ces vecteurs peuvent être choisis orthonormés¹

$$X^t X = I$$

Comme X est régulière, cela montre que $X^{-1} \doteq X^t$, et donc aussi $XX^t = I$. Cela entraîne:

$$\Sigma = X \Lambda X^t$$

Alors, si Y est le vecteur constitué de n variables gaussiennes réduites indépendantes, le vecteur:

$$V = M + X \Lambda^{1/2} Y$$

est un vecteur gaussien (obtenu par transformation linéaire de Y), d'espérance M , et de matrice de variance-covariance:

$$E[(V - M)(V - M)^t] = X \Lambda^{1/2} E[Y \cdot Y^t] \Lambda^{1/2} X^t = X \Lambda X^t = \Sigma$$

et suit donc la loi prescrite. La loi de V s'obtient par changement de variables. Si ψ une fonction définie sur \mathbb{R}^n .

$$E[\psi(V)] = E[\psi[M + X \Lambda^{1/2} Y]] = \int \psi(M + X \Lambda^{1/2} Y) \prod_i g(y_i) dy_i$$

$$\prod_i g(y_i) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}(Y^t Y)\right\}$$

mais $Y = \Lambda^{-1/2} X^t (V - M)$, où la matrice $\Lambda^{-1/2}$ est la matrice diagonale constituée des $1/\sqrt{\lambda_i}$, quantités finies². D'où:

$$Y^t \cdot Y = (V - M)^t X \Lambda^{-1/2} \Lambda^{-1/2} X^t (V - M) = (V - M)^t \Sigma^{-1} (V - M)$$

puisque $X \Lambda^{-1} X^t = \Sigma^{-1}$. Il ne reste plus qu'à calculer le jacobien:

$$\left| \frac{DY}{DV} \right| = |\Lambda^{-1/2} X^t| = \frac{1}{\sqrt{\det \Sigma}}$$

D'où par identification:

$$g_{m, \Sigma}(V) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left\{-\frac{1}{2}(V - M)^t \Sigma^{-1} (V - M)\right\}$$

1. C'est automatiquement le cas pour des valeurs propres distinctes, puisque $\Sigma X_1 = \lambda_1 X_1$ et $\Sigma X_2 = \lambda_2 X_2$ entraîne $\langle X_1, \Sigma X_2 \rangle = \lambda_2 \langle X_1, X_2 \rangle$, mais comme on a aussi $\langle X_1, \Sigma X_2 \rangle = \langle \Sigma X_1, X_2 \rangle = \lambda_1 \langle X_1, X_2 \rangle$, $\lambda_1 \neq \lambda_2 \Rightarrow \langle X_1, X_2 \rangle = 0$. Dans le cas où il existe des valeurs propres multiples, il suffit de choisir une base orthonormée de chaque sous-espace invariant.

2. la régularité de Σ entraîne la stricte positivité des valeurs propres λ_i

Références

- [1] Lin-Ying Hu. *Mise en oeuvre du modèle gamma pour l'estimation de distributions spatiales*. Doctoral thesis, 142 p., Ecole des Mines de Paris, CG. 35 rue St. Honoré, Fontainebleau, October, 28 1988.
- [2] Christian Lajaunie. Estimation directe des paramètres de diffusion. Technical Report N-25/86/G 26 p., Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1986.
- [3] Christian Lajaunie and Christian Lantuéjoul. Setting up the general methodology for discrete isofactorial models. In M. Armstrong, editor, *Geostatistics*, volume 1, pages 323–334, Dordrecht Holland, 1989. Kluwer Academic Press.
- [4] Christian Lantuéjoul. Quelques remarques sur les facteurs de diffusion discret: une application de la positivité totale. Technical Report N-31/86/G, Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1986.
- [5] Christian Lantuéjoul. Substitution random functions. In A. Soares, editor, *Quantitative Geology and Geostatistics*, volume 1, pages 37–48, Dordrecht Holland, 1992. Kluwer Academic Press.
- [6] Georges Matheron. Leçons sur les fonctions aléatoires d'ordre deux. Ecole Nationale Supérieure des Mines de Paris, Centre de Géostatistique, 35 Rue St. Honoré . Fontainebleau, 1972. C-53.
- [7] Georges Matheron. Complément sur les modèles isofactoriels. Technical Report N-432, Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1975.
- [8] Georges Matheron. Modèles isofactoriels discrets, et modèles de walsh. Technical Report N-449, 24p, Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1975.

- [9] Georges Matheron. A simple substitute for conditional expectation: the disjunctive kriging. In M. Guarascio et al, editor, *Advanced Geostatistics in the Mining industry*, pages 221–236, Dordrecht Holland, 1976. D. Reidel.
- [10] Georges Matheron. Peut-on imposer des conditions d'universalité au krigeage disjonctif. Technical Report N-539, 24p, Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1977.
- [11] Georges Matheron. Modèles isofactoriels pour l'effet zéro. Technical Report N-659, 33p, Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1980.
- [12] Georges Matheron. La destructuration des hautes teneurs et le krigeage des indicatrices. Technical Report N-761 33 p., Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1982.
- [13] Georges Matheron. Modèles isofactoriels et changement de support. Technical Report N-821, 46p, Centre de Géostatistique de l'école des mines de Paris, 35 Rue St. Honoré . Fontainebleau, 1983.
- [14] Georges Matheron. Changement de support en modèle mosaïque. In J.J. Royer, editor, *Sciences de la terre*, volume N 20, pages 435–454. Ecole supérieure de Géologie appliquée et de prospection minière, 1984.
- [15] Georges Matheron. Isofactorial models and change of support. In G. Verly et al, editor, *Geostatistics for Natural Resources Characterisation*, volume 1, pages 449–467, Dordrecht Holland, 1984. NATO A.S.I., D. Reidel.
- [16] Georges Matheron. Une méthodologie générale pour les modèles isofactoriels discrets. In J.J. Royer, editor, *Sciences de la terre*, volume N 21, pages 1–64. Ecole supérieure de Géologie appliquée et de prospection minière, 1984.
- [17] Georges Matheron. The internal consistency of models in geostatistics. In M. Armstrong, editor, *Geostatistics*, volume 1, pages 21–38, Dordrecht Holland, 1989. Kluwer Academic Press.
- [18] Jacques Rivoirard. Models with orthogonal indicator residuals. In M. Armstrong, editor, *Geostatistics*, volume 1, pages 91–108, Dordrecht Holland, 1989. Kluwer Academic Press.