

**Les Cahiers du Centre de Morphologie Mathématique
de Fontainebleau**

Fascicule 1

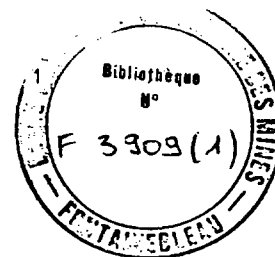
LE KRIGEAGE UNIVERSEL

par

G. MATHERON

1969

Edité par l'Ecole Nationale Supérieure des Mines de Paris



LE KRIGEAGE UNIVERSEL

(RECHERCHE D'ESTIMATEURS OPTIMAUX EN PRESENCE D'UNE DERIVE)

O - INTRODUCTION

O-1 CRITIQUE DES METHODES DE MOINDRES CARRÉS

Nous nous proposons, dans ce qui suit, de formuler en termes de fonctions aléatoires non stationnaires le problème de l'estimation des dérivées ¹ (ou tendances), et de montrer que ce problème admet une solution optimale, bien différente de celles auxquelles conduisent les méthodes dites de "trend surface analysis" : ces dernières, qui consistent en général à ajuster un polynôme par une méthode de moindres carrés, donnent un sentiment de malaise; on a l'impression, en effet, de faire violence à la nature, en lui imposant de force une expression polynomiale, qui n'a aucune raison a priori de présenter le moindre rapport avec la structure réelle du phénomène que l'on veut représenter. D'une manière plus précise, nous adresserons trois reproches principaux à ces méthodes de moindres carrés :

Tout d'abord, ces méthodes entraînent souvent une confusion du mode opératoire et du concept. Peu d'auteurs se donnent la peine de définir la signification de ce "trend" qu'ils estiment par une méthode de moindres carrés. On a souvent l'impression que ce fameux trend n'est rien de plus que le résultat numérique auquel conduit un mode opératoire - c'est-à-dire, peut-être, un pur et simple artefact - A l'analyse, il semble que le terme peu clair de "trend" se réfère tour à tour, et parfois même simultanément dans un même contexte, à trois notions bien distinctes au moins : si $Z(x)$ est la variable régionalisée à laquelle on s'intéresse, et $P(x)$ le polynôme ajusté par moindres carrés à partir des valeurs expérimentalement connues aux points $x_1, x_2, \dots$, on peut attribuer à la valeur en x de $P(x)$ l'une ou l'autre des significations incompatibles suivantes :

1 - Pour éviter tout anthropomorphisme, nous utiliserons toujours le terme "dérive" au lieu de "tendance", ou "trend".

a/ $P(x)$ est (une estimation de) l'espérance à priori $E(Z(x))$: c'est toujours ce sens a/ que nous attribuerons au terme "dérive".

b/ $P(x)$ est une estimation de la vraie valeur (inconnue) $Z(x)$ prise par la variable régionalisée au point x : c'est à ce sens b/ que se rattache notre terme "krigeage ponctuel". Plus précisément, le krigeage ponctuel sera le meilleur estimateur linéaire de $Z(x)$ construit à partir des données disponibles $Z(x_1), Z(x_2), \dots$.

c/ On attribue parfois, enfin, à $P(x)$ le sens d'une "moyenne mobile". $P(x_0)$ serait alors (une estimation de) la valeur moyenne de $Z(x)$ dans une zone plus ou moins grande (à préciser) entourant un point donné x_0 . Dans notre terminologie, le krigeage désignera le meilleur estimateur linéaire de cette moyenne mobile.

Ces distinctions ont une grande importance théorique et pratique. Il est essentiel de définir avec précision l'objectif que l'on vise (a/, b/, ou c/). En cartographie sous-marine [1], ce sont les vraies valeurs $Z(x)$ que l'on doit estimer et cartographier (cas b/). En exploitation minière [3], [4], [5], on s'intéresse à la teneur moyenne d'un panneau de taille donnée (cas c/). Dans certaines études à caractère plus fondamental, enfin, où l'on cherche à reconstituer les mécanismes généraux qui ont donné naissance au phénomène que l'on étudie, c'est plutôt dans la dérive elle-même (cas a/) que l'on peut espérer trouver un reflet de la structure de ces mécanismes.

Nous en arrivons ainsi au deuxième grief que l'on peut formuler à l'encontre des méthodes de moindres carrés. Il n'est pas possible que le même polynôme $P(x)$ résolve à la fois les trois problèmes a/, b/ et c/. Nous verrons que $P(x)$ constitue en fait une solution du problème a/. Mais, en général, ce n'est pas la meilleure solution possible : ces méthodes passe-partout qui utilisent toujours les mêmes polynômes, quelles que soient les caractéristiques structurales du phénomène que l'on étudie, n'ont aucune chance en général, de conduire à un optimum.

En troisième lieu, enfin, les méthodes de moindre carré ne permettent en aucune façon d'évaluer l'erreur que l'on commet en estimant la dérive à l'aide du

polynôme $P(x)$. La variance des résidus, contrairement à ce que l'on croit parfois, n'est pas une variance d'estimation : la variance des écarts $Z(x_i) - P(x_i)$ aux points x_i où les données sont connues est, par construction, systématiquement plus faible (et même beaucoup plus faible) que celle de l'écart $Z(x) - P(x)$ en un point x différent des x_i . Ainsi, la variance des résidus n'est pas la variance d'estimation de $Z(x)$ (problème b/). Elle n'est pas davantage la variance d'estimation de la dérive (problème a/), car il n'y a cette fois plus aucun lien conceptuel entre un écart $Z(x_i) - P(x_i)$ et la qualité de $P(x_i)$ considéré comme un estimateur de l'espérance à priori $E[Z(x_i)]$.

Cependant, si les méthodes de moindres carrés se révèlent peu satisfaisantes, l'objectif qu'elles visaient, et qu'elles n'atteignent pas ou atteignent mal, correspond, lui, à un problème très réel et très important. Il existe réellement des phénomènes que l'on ne peut absolument pas assimiler à des (réalisations de) fonctions aléatoires stationnaires. Pour reprendre l'exemple de la cartographie des fonds sous-marins, il est certain que la profondeur va en augmentant lorsqu'on s'éloigne des côtes. Nous allons essayer de formuler et de résoudre dans le cadre de la théorie des fonctions aléatoires non stationnaires les problèmes importants que posent ces dérivées.

O-2 POSITION DU PROBLEME ET HYPOTHESES GENERALES.

La variable régionalisée $z(x)$ que l'on étudie sera interprétée comme une réalisation d'une fonction aléatoire $Z(x)$ non stationnaire en général. Dans un premier groupe d'hypothèses, nous supposerons que $Z(x)$ admet des moments d'ordre 1 et 2 :

$$(I) \quad \begin{cases} E[Z(x)] = m(x) \\ E[Z(x) Z(y)] = m(x) m(y) + C(x,y) \end{cases}$$

Nous dirons que la fonction $m(x)$ est la dérivée de la F.A. $Z(x)$. Il nous semble, en effet, que la seule définition conceptuellement claire de la notion de dérive est bien celle-ci : moment d'ordre 1 d'une F.A. non stationnaire. La fonction $C(x,y)$ - qui dépend séparément des deux points x et y est la covariance (non stationnaire) habituelle.

Dans bien des cas (comme le montre la pratique de la Géostatistique([3], [5])) ces hypothèses (I) seront encore trop restrictives, et devront être remplacées par les hypothèses (I') suivantes qui expriment que les accroissements de la fonction aléatoire $Z(x)$ (et non $Z(x)$ elle-même) admettent des moments d'ordre 1 et 2 :

$$(I') \quad \begin{cases} E[Z(x) - Z(y)] = m(x) - m(y) \\ \frac{1}{2} D^2[Z(x) - Z(y)] = \gamma(x, y) \end{cases}$$

La dérive $m(x)$ n'est plus ici déterminée qu'à une constante près, car, en général, l'espérance $E[Z(x)]$ n'existe plus : le point de vue des hypothèses (I') revient donc à étudier la fonction aléatoire $Z(x)$ à une constante additive près. La fonction $\gamma(x, y)$ est le demi-variogramme habituel. (Dans l'espace à une seule dimension, des processus aussi usuels que le mouvement brownien ou les processus de Poisson donnent des exemples simples de F.A. sans espérance, vérifiant les hypothèses (I') mais non les hypothèses (I)).

Nous avons à résoudre les deux problèmes (liés, mais distincts) suivants : connaissant les valeurs numériques prises sur un ensemble S (l'ensemble des points où l'on dispose de données expérimentales) par une réalisation $z(x)$ de la F.A. $Z(x)$ nous devons :

1°/ estimer la fonction $m(x)$ (sur S et à l'extérieur de S) : c'est le problème a/ ci-dessus.

2°/ estimer $z(x)$ en un point $x \notin S$ (problème b/), ou, plus généralement estimer une "moyenne mobile" $\int \mu(dx) z(x)$ où μ est une mesure donnée dont le support est disjoint de S (problème c/).

De plus, nous devons être capables de représenter par des variances d'estimation les erreurs commises dans ces opérations, et nous efforcer de choisir nos estimateurs de manière à minimiser (si faire se peut) cette variance d'estimation.

En ce qui concerne ce dernier point, indiquons que nous nous limiterons en fait à rechercher le meilleur estimateur linéaire de $m(x)$ ou de $\int \mu(dx) z(x)$ que

l'on puisse former à partir des valeurs numériques des $z(y)$, $y \in S$: les estimateurs non linéaires sont beaucoup trop compliqués pour qu'il soit possible de les mettre en oeuvre dans les applications, et d'autre part leurs propriétés ne sont plus liées aux seuls moments d'ordre 1 et 2, qui figurent dans les hypothèses (I) ou (I'), mais font intervenir la totalité de la loi spatiale de la F.A. $Z(x)$ - (dans le cas où cette loi spatiale est gaussienne, on sait d'ailleurs que le meilleur estimateur possible coïncide avec le meilleur estimateur linéaire).

Ainsi formulé en toute généralité (les fonctions $m(x)$ et $C(y,x)$ étant complètement inconnues) notre problème est manifestement insoluble, et d'ailleurs probablement dépourvu de sens. Mais, dans les problèmes concrets, la notion de dérive ne peut présenter une signification réelle que si la fonction $m(x)$ correspondante varie d'une manière continue et régulière relativement à l'échelle à laquelle on travaille (et aux données expérimentalement disponibles) : si la fonction $m(x)$ était irrégulière et chaotique à cette échelle, on devrait la considérer elle-même comme une réalisation d'une nouvelle fonction aléatoire. Du point de vue de l'interprétation physique (et non mathématique) la notion de dérive est ainsi manifestement liée à celle d'échelle. Si l'on étudie des cotes topographiques, à l'échelle de la dizaine de mètres, la notion de montagne se traduit par une dérive ; à l'échelle de la dizaine de kilomètres, il ne lui correspond plus qu'une fonction aléatoire, et la dérive, à cette échelle, exprimera plutôt la notion de chaîne de montagnes. (cf. aussi la notion de structures gigogne, [7]).

Mais cette condition de régularité imposée à priori à la fonction $m(x)$ pour que la notion de dérive ait un contenu physique réel, entraîne aussi qu'une estimation de $m(x)$ doit toujours être plus ou moins possible localement. Cette condition exprime en effet que - sur un certain voisinage V d'un point x_0 donné - la fonction $m(x)$ peut être approchée avec une excellente précision par une fonction de la forme :

$$(II) \quad m(x) \sim f(x) = \sum_{\ell=0}^k a_{\ell} f^{\ell}(x)$$

où les $f^{\ell}(x)$ sont des fonctions connues, choisies une fois pour toutes (par exemple

des polynômes), et les a_k des coefficients inconnus : quant au voisinage V de x_0 sur lequel l'approximation (II) est acceptable, sans être très grand, il doit - si le problème a un sens - contenir un nombre suffisant de points expérimentaux pour qu'il soit possible d'estimer les $k+1$ coefficients inconnus $a_0, a_1, \dots, a_k$ -

Il reste le problème de la fonction $C(x,y)$, ou $\gamma(x,y)$ sur laquelle on ne sait rien, non plus, à priori. Mais, ici encore, et pour les mêmes raisons, on peut supposer que C ou γ sont localement assimilables à des fonctions de type connu, et ne se déforment qu'assez lentement dans l'espace à l'échelle à laquelle on travaille.

Le cas le plus favorable sera celui d'une fonction γ (ou C) de la forme $\gamma(x,y) = \omega \gamma_0(x-y)$ où γ_0 est une fonction connue, et ω un facteur lentement variable, que l'on pourra regarder comme constant sur le voisinage V du point x_0 : les équations qui déterminent les estimateurs optimaux étant linéaires et homogènes, ces estimateurs ne dépendront que de γ_0 et non du facteur ω ; ce dernier ne se répercutera donc que sur les variances d'estimation, et non sur les estimateurs eux-mêmes. Dans certains cas, on pourra prendre l'expression très simple :

$$\gamma(x,y) = \omega r \quad (r = |x-y|)$$

(variogramme linéaire). Il suffit, pour cela, en effet, que le vrai variogramme ou la vraie covariance ait un comportement linéaire au voisinage de $x = y$ et jusqu'à des distances comparables aux dimensions du voisinage V ci-dessus : cette circonstance se rencontre plus souvent qu'on ne croit. (cf. [1]).

En général, pourtant, en plus du facteur ω , il conviendra d'introduire un ou plusieurs paramètres supplémentaires. Par exemple, on prendra

$$\gamma(x,y) = \omega r^{\alpha-1} \quad (0 < \alpha < 2)$$

ou encore

$$C(x,y) = \omega e^{-\alpha r}$$

Dans le premier cas, le paramètre α est en relation avec le degré de continuité de la variable régionalisée. Dans le deuxième (covariance exponentielle), le paramètre a , ou plutôt son inverse, donne la mesure de la portée du phénomène (distance au delà de laquelle les corrélations s'éteignent).

Naturellement, le contrôle expérimental d'une hypothèse de ce genre, et l'estimation des paramètres correspondants (α et surtout α ou a) posera des problèmes assez délicats de statistique mathématique : en raison de leur extrême importance pour les applications, nous consacrerons à ce problème la section III de cette étude.

En résumé le problème que nous devons traiter se trouve schématisé comme suit : on a une F.A. $Z(x)$ vérifiant les hypothèses (I) ou (I'). Sur un voisinage V d'un point x_0 , la dérivée $m(x)$ est de la forme (II) avec des coefficients a_j inconnus. On connaît à priori (éventuellement à un facteur près) la covariance $C(x,y)$ ou le variogramme $\gamma(x,y)$. Enfin, on connaît les valeurs numériques de (la réalisation de) $Z(x)$ pour les points x appartenant à un ensemble $S \subset V$. On veut former les meilleurs estimateurs linéaires :

- 1°/ des coefficients a_j inconnus de la dérivée.
- 2°/ de la valeur numérique de (la réalisation de) $Z(x)$ en $x_0 \notin S$ (mais $x_0 \in V$) ou de $\int \mu(dx) Z(x)$ pour une mesure μ dont le support est disjoint de S .
- 3°/ nous voulons de plus contrôler la validité de l'hypothèse que nous avons faite en choisissant pour la covariance ou le variogramme une expression mathématique particulière, et estimer les paramètres dont cette expression dépend.

Pour faciliter la compréhension (et quitte à énoncer deux fois les mêmes résultats), nous donnerons dans la section I un exposé simplifié, en supposant que l'ensemble S des données expérimentales est fini. Dans la section II nous ferons le même exposé, avec davantage de rigueur, en supposant S compact (fini ou infini). La section III sera consacrée aux problèmes de statistique mathématique dont nous avons parlé. Certains points particuliers, et notamment un examen de la méthode des moindres carrés, seront traités en annexes.

SECTION I

CAS OU L'ENSEMBLE S DES DONNEES EST FINI

Nous traiterons d'abord l'estimation de la dérive (1-1) et le krigeage (1-2 et 1-3) dans le cas où il existe une covariance (connue) $C(x,y)$, c'est-à-dire dans le cadre des hypothèses (I). Dans la sous-section 1-4, nous transposerons ces résultats au cas où il n'existe qu'un variogramme $\gamma(x,y)$. Enfin nous traiterons l'exemple du variogramme linéaire $\gamma(x,y) = \omega |x-y|$ (1-5).

1-1 ESTIMATION OPTIMALE DE LA DERIVE (Covariance connue)

Soit $Z(x)$ une F.A. vérifiant les hypothèses (I) avec :

$$m(x) = \sum_{\ell=0}^k a_{\ell} f^{\ell}(x)$$

Les $f^{\ell}(x)$ sont des fonctions connues, mais les coefficients a_{ℓ} sont inconnus, et le problème consiste à les estimer. L'ensemble S où l'on connaît les valeurs prises par la réalisation de $Z(x)$ est constitué de n points x_{α} , $\alpha = 1, 2, \dots, n$.

Soit x_0 un point (appartenant, ou non, à S), et $b^{\ell} = f^{\ell}(x_0)$. Pour estimer la dérive en x_0 , soit :

$$m(x_0) = \sum_{\ell} a_{\ell} b^{\ell}$$

nous allons former une combinaison linéaire :

$$(1-1) \quad M = \sum_{\alpha=1}^n \lambda^{\alpha} Z_{\alpha}$$

et choisir les coefficients λ^{α} de manière à remplir les deux conditions suivantes, quels que soient les coefficients a_{ℓ} ($\ell = 0, 1, \dots, k$) :

- ~ L'espérance $E(M)$ doit être égale à $m(x_0) = \sum a_{\ell} b^{\ell}$
- ~ La variance $D^2(M)$ doit être minimale.

Pour abréger les notations, nous poserons :

$$Z(x_{\alpha}) = Z_{\alpha}, \quad f^{\ell}(x_{\alpha}) = f_{\alpha}^{\ell}, \quad C(x_{\alpha}, x_{\beta}) = \sigma_{\alpha\beta}$$

La première condition exprime que M est un estimateur sans biais. Elle s'écrit :

$$\sum_{\alpha} \sum_{\ell} \lambda^{\alpha} a_{\ell} f_{\alpha}^{\ell} = \sum_{s} a_s b^s$$

Cette condition est vérifiée quels que soient les a_{ℓ} si et seulement si on a :

$$(1-2) \quad \sum_{\alpha} \lambda^{\alpha} f_{\alpha}^{\ell} = b^{\ell} \quad (\ell = 0, 1, \dots, k)$$

Nous dirons qu'un estimateur (1-1) de $m(x_0)$ est universel si les coefficients λ^{α} vérifient (1-2), c'est-à-dire si cet estimateur est sans biais quels que soient les a_{ℓ} . L'estimateur des moindres carrés, par exemple, est un estimateur universel (Annexe 3). Mais ce n'est pas le meilleur possible.

Moyennant la condition (1-2), on a $E(M) = m(x_0)$, et par suite $E[(M - m(x_0))^2] = D^2(M)$: la variance de M , qui ne dépend que des $\sigma_{\alpha\beta}$ et non des a_{ℓ} , donne une mesure de l'erreur commise en estimant $m(x_0)$ à l'aide de M . Cette variance a pour expression :

$$(1-3) \quad D^2(M) = \sum_{\alpha, \beta} \lambda^{\alpha} \lambda^{\beta} \sigma_{\alpha\beta}$$

L'estimateur universel optimal correspond donc aux coefficients λ^{α} qui minimisent la forme quadratique (1-3) compte tenu des conditions (1-2). Le formalisme classique de Lagrange conduit au système :

$$(1-4) \quad \begin{aligned} \sum_{\beta} \lambda^{\beta} \sigma_{\alpha\beta} &= \sum_{\ell} \mu_{\ell} f_{\alpha}^{\ell} & (\alpha = 1, 2, \dots, n) \\ \sum_{\alpha} \lambda^{\alpha} f_{\alpha}^{\ell} &= b^{\ell} & (\ell = 0, 1, \dots, k) \end{aligned}$$

La matrice des $\sigma_{\alpha\beta} = C(x_{\alpha}, x_{\beta})$ est toujours strictement définie positive. Nous verrons dans la section 2 que le système (1-4) de $1 + k + n$ équations à $1 + k + n$ inconnues (les n coefficients λ^{α} et les $k+1$ paramètres de Lagrange) est régulier si et seulement si les fonctions (connues) $f^{\ell}(x)$ sont linéairement indépendantes sur S , autrement dit si :

$$\sum_{\ell} c_{\ell} f_{\alpha}^{\ell} = 0 \quad (\alpha = 1, 2, \dots, n) \Rightarrow c_{\ell} = 0 \quad (\ell = 0, 1, \dots, k)$$

Nous supposons toujours cette condition remplie. La solution unique de (1-4) donne alors l'estimateur universel optimal cherché. La variance de cet estimateur optimal est donnée par :

$$(1-5) \quad D^2(M) = \sum_{\ell} \mu_{\ell} b^{\ell}$$

On le voit sans peine en multipliant par λ^{α} la première relation (1-4) et en sommant en α . Comme cette variance ne dépend que des paramètres de Lagrange μ_{ℓ} , on l'obtient comme un sous-produit immédiat de la résolution du système (1-4).

D'autre part, la condition (1-2) est linéaire en b^{ℓ} . Par suite, les coefficients λ^{α} et les paramètres de Lagrange sont eux-mêmes linéaires en b^{ℓ} , soit :

$$\begin{cases} \lambda^{\alpha} = \sum_{\ell} \lambda_{\ell}^{\alpha} b^{\ell} \\ \mu_{\ell} = \sum_{s} \mu_{\ell s} b^s \end{cases}$$

Les matrices λ_{ℓ}^{α} et $\mu_{\ell s}$ constituent la solution unique du système suivant :

$$(1-D) \quad \begin{cases} \sum_{\beta} \lambda_{\ell}^{\beta} \sigma_{\alpha\beta} = \sum_s \mu_{\ell s} r_{\alpha}^s & (\alpha = 1, \dots, n, \ell = 0, 1, \dots, k) \\ \sum_{\alpha} \lambda_{\ell}^{\alpha} r_{\alpha}^s = \delta_{\ell}^s \end{cases}$$

(avec $\delta_{\ell}^s = 1$ pour $\ell = s$ et 0 pour $\ell \neq s$). Le vecteur :

$$(1-6) \quad A_{\ell} = \sum_{\alpha} \lambda_{\ell}^{\alpha} Z_{\alpha}$$

constitue alors l'estimateur universel optimal du vecteur dérive a_{ℓ} :

On a $E(A_{\ell}) = a_{\ell}$ quelles que soient les vraies valeurs inconnues des a_{ℓ} , et la variance $D^2(\sum_{\ell} A_{\ell} b^{\ell})$ est minimale pour tout vecteur b^{ℓ} . De plus, la matrice des paramètres de Lagrange $\mu_{\ell s}$ coïncide avec la matrice des covariances des A_{ℓ} :

$$(1-7) \quad \mu_{\ell s} = \text{Cov}(A_{\ell}, A_s) = E(A_{\ell} A_s) - a_{\ell} a_s$$

On l'obtient à titre de sous-produit de la résolution du système fondamental (1-D). Notons enfin que ce système de $n(k+1) + (k+1)^2 = (k+1)(n+k+1)$ équations à $n(k+1) + (k+1)^2$ inconnues (les λ^α_ℓ et les paramètres de Lagrange μ_{ℓ_S}) est régulier si, et seulement si, les $f^\ell(x)$ sont linéairement indépendantes sur S. On notera aussi, d'après (1-7), que l'on a toujours $\mu_{\ell_S} = \mu_S \ell$.

Remarque : Dans le cas gaussien, l'estimateur optimal universel de la dérive, solution du système fondamental (1-D) coïncide avec l'estimateur du maximum de vraisemblance (voir Annexe 2).

1-2 LE KRIGEAGE UNIVERSEL (à covariance connue).

Proposons-nous maintenant de procéder à l'estimation d'une "moyenne pondérée" de $Z(x)$, c'est-à-dire d'une expression de la forme :

$$(1-8) \quad Z = \int Z(x) p(dx)$$

où $p(dx)$ est une mesure donnée dont le support est disjoint de S, et où $Z(x)$ est la réalisation de notre fonction aléatoire. Nous allons pour cela former une combinaison linéaire des données disponibles (les $Z_\alpha = Z(x_\alpha)$), soit :

$$(1-9) \quad Z^* = \sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$$

et déterminer les coefficients λ^{α} grâce à des conditions convenables. Si l'on considère $Z(x)$ non plus comme une réalisation, mais comme la F.A. elle-même, Z , Z^* et les Z_{α} sont des variables aléatoires. Nous allons imposer aux λ^{α} les deux conditions suivantes :

- ~ quels que soient les a_{ℓ} , $E(Z^* - Z) = 0$
- ~ quels que soient les a_{ℓ} , $D^2(Z^* - Z)$ est minimale

Au lieu de $E(Z - Z^*) = 0$, on aurait pu s'imposer la condition beaucoup plus forte : espérance conditionnelle de Z relativement aux $Z_{\alpha} = Z^*$. Mais cette condition plus forte nous ferait sortir du cadre des estimateurs linéaires, et nécessiterait l'intervention de la loi spatiale de $Z(x)$ au lieu de sa seule covariance.

La première condition ($E(Z^* - Z) = 0$) se traduit par :

$$\sum_{\ell} a_{\ell} \int p(dx) f^{\ell}(x) = \sum_{\ell} \sum_{\alpha} a_{\ell} \lambda^{\alpha} f^{\ell}_{\alpha}$$

Elle est réalisée quels que soient les a_{ℓ} si et seulement si :

$$(1-10) \quad \sum_{\alpha} \lambda^{\alpha} f^{\ell}_{\alpha} = \int p(dx) f^{\ell}(x)$$

Cette condition d'universalité (1-10) étant remplie, la variance d'estimation $D^2(Z^* - Z)$ ne dépend plus des a_{ℓ} et vaut :

$$(1-11) \quad D^2(Z - Z^*) = \sigma_Z^2 - 2 \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha, Z} + \sum_{\alpha, \beta} \lambda^{\alpha} \lambda^{\beta} \sigma_{\alpha\beta}$$

avec

$$\begin{cases} \sigma_Z^2 = \int C(x, y) p(dx) p(dy) \\ \sigma_{\alpha, Z} = \int C(x_{\alpha}, y) p(dy) \\ \sigma_{\alpha\beta} = C(x_{\alpha}, x_{\beta}) \end{cases}$$

Elle est minimale, compte tenu des conditions (1-10) pour des λ^{α} vérifiant le système :

$$(1-K) \quad \begin{cases} \sum_{\beta} \lambda^{\beta} \sigma_{\alpha\beta} = \sigma_{\alpha, Z} + \sum_{\ell} \mu_{\ell} f^{\ell}_{\alpha} \\ \sum_{\alpha} \lambda^{\alpha} f^{\ell}_{\alpha} = \int p(dx) f^{\ell}(x) \end{cases}$$

Telles sont les équations du krigeage universel. Le système (1-K) comporte $n+1+k$ équations à $n+1+k$ inconnues (les λ^{α} et les paramètres de Lagrange μ_{ℓ}). La covariance $C(x, y)$ étant de type strictement défini positif, ce système est régulier pourvu seulement que les $f^{\ell}(x)$ soient linéairement indépendantes sur S . La variance σ_U^2 de l'estimateur Z^* du krigeage universel est donnée par :

$$(1-12) \quad \begin{cases} \sigma_U^2 = \sigma_Z^2 - \sum_{\alpha} \lambda^{\alpha} \sigma_{\alpha, Z} + \sum_{\ell} \mu_{\ell} b^{\ell} \\ (b^{\ell} = \int p(dx) f^{\ell}(x)) \end{cases}$$

Krigeage Ponctuel - Dans le cas où la variable Z à estimer est la valeur $Z(x)$ en un point $x \notin S$, le système (1-K) se réduit à :

$$(1-13) \quad \begin{cases} \sum_{\beta} \lambda^{\beta} \sigma_{\alpha\beta} = C(x_{\alpha}, x) + \sum_{\ell} \mu_{\ell} f_{\alpha}^{\ell} \\ \sum_{\alpha} \lambda^{\alpha} f_{\alpha}^{\ell} = f^{\ell}(x) \end{cases}$$

et l'estimateur $Z^*(x)$ du krigeage universel de $Z(x)$:

$$Z^*(x) = \sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$$

admet la variance

$$\sigma_U^2 = C(x, x) - \sum_{\alpha} \lambda^{\alpha} C(x_{\alpha}, x) + \sum_{\ell} \mu_{\ell} f^{\ell}(x)$$

Remarque : En un point x_{α} où $Z_{\alpha} = Z(x_{\alpha})$ est connu, le krigeage ponctuel donne $Z^*(x_{\alpha}) = Z_{\alpha}$, comme on le vérifie sans peine : le krigeage ponctuel constitue donc une méthode d'interpolation (et même la meilleure possible) entre des points connus.

1-3 LE THEOREME D'ADDITIVITE

Examinons les rapports existants entre les estimateurs optimaux de la dérive - système (1-D)-et du krigeage universel - système (1-K) - En l'absence de dérive ($a_{\ell} = 0$); l'estimateur optimal de

$$Z = \int p(dx) Z(x)$$

serait de la forme :

$$(1-14) \quad Z_K = \sum_{\alpha} \lambda_K^{\alpha} Z_{\alpha}$$

avec des coefficients λ_K^{α} minimisant la variance

$$(1-15) \quad D^2(Z_K - Z) = \sigma_Z^2 - 2 \sum_{\alpha} \lambda_K^{\alpha} \sigma_{\alpha, Z} + \sum_{\alpha, \beta} \lambda_K^{\alpha} \lambda_K^{\beta} \sigma_{\alpha\beta}$$

c'est-à-dire vérifiant le système :

$$(1-16) \quad \sum_{\beta} \lambda_K^{\beta} \sigma_{\alpha\beta} = \sigma_{\alpha Z}$$

et la variance correspondante serait :

$$\sigma_K^2 = \sigma_Z^2 - \sum_{\alpha} \lambda_K^{\alpha} \sigma_{\alpha, Z}$$

Comparons les systèmes (1-K) et (1-16). Posons :

$$Z^* = Z_K + Z_D$$

Z_D représente la correction de dérive. Z_K , solution du krigage simple (en l'absence de dérive) étant donnée par (1-14) avec les coefficients λ_K^{α} solution de (1-16), on voit que Z_D est de la forme

$$Z_D = \sum_{\alpha} \lambda_D^{\alpha} z_{\alpha}$$

avec des coefficients λ_D^{α} vérifiant :

$$(1-17) \quad \begin{cases} \sum_{\beta} \lambda_D^{\beta} \sigma_{\alpha\beta} = \sum_{\ell} \mu_{\ell} r_{\alpha}^{\ell} \\ \sum_{\alpha} \lambda_D^{\alpha} r_{\alpha}^{\ell} = \int p(dx) r^{\ell}(x) - \sum_{\alpha} \lambda_K^{\alpha} r_{\alpha}^{\ell} \end{cases}$$

Ce système est identique à (1-4), à condition de prendre

$$b^{\ell} = \int p(dx) r^{\ell}(x) - \sum_{\alpha} \lambda_K^{\alpha} r_{\alpha}^{\ell}$$

Si donc nous désignons par A_{ℓ} :

$$A_{\ell} = \sum_{\alpha} \lambda_{\ell}^{\alpha} z_{\alpha}$$

L'estimateur optimal de la dérive a_{ℓ} , avec λ_{ℓ}^{α} vérifiant (1-D), nous voyons que les coefficients λ_D^{α} du terme correctif sont donnés par :

$$\lambda_D^{\alpha} = \sum_{\ell} \lambda_{\ell}^{\alpha} \left[\int p(dx) r^{\ell}(x) - \sum_{\beta} \lambda_K^{\beta} r_{\beta}^{\ell} \right]$$

L'estimateur $Z^* = Z_K + Z_D$ se met donc sous la forme

$$\begin{aligned} Z^* &= \sum_{\alpha} \lambda_K^{\alpha} z_{\alpha} + \sum_{\alpha} \lambda_D^{\alpha} z_{\alpha} = \\ &= \sum_{\alpha} \lambda_K^{\alpha} \left(z_{\alpha} - \sum_{\ell, \beta} r_{\alpha}^{\ell} \lambda_{\ell}^{\beta} z_{\beta} \right) + \sum_{\alpha, \ell} \lambda_{\ell}^{\alpha} z_{\alpha} \int p(dx) r^{\ell}(x) \end{aligned}$$

Compte tenu de l'expression (1,6) de l'estimateur A_ℓ de la dérive, ceci s'écrit encore :

$$(1-18) \quad Z^* = \sum_{\ell} A_{\ell} \int p(dx) f^{\ell}(x) + \sum_{\alpha} \lambda_K^{\alpha} [Z_{\alpha} - \sum_{\ell} A_{\ell} f^{\ell}_{\alpha}]$$

Le sens de cette relation est le suivant : on obtient l'estimateur Z^* du krigeage universel en appliquant le krigeage ordinaire (coefficient λ_K^{α}) aux résidus $Z(x) - \sum_{\ell} A_{\ell} f^{\ell}(x)$, c'est-à-dire aux valeurs de la F.A. $Z(x)$ corrigées de la dérive estimée $\sum_{\ell} A_{\ell} f^{\ell}(x)$. Si les valeurs vraies des a_{ℓ} étaient connues, l'estimateur optimal serait évidemment :

$$(1-19) \quad Z^* = \sum_{\ell} a_{\ell} \int p(dx) f^{\ell}(x) + \sum_{\alpha} \lambda_K^{\alpha} (Z_{\alpha} - \sum_{\ell} a_{\ell} f^{\ell}_{\alpha})$$

La relation (1-18) est identique à (1-19), à ceci près que les a_{ℓ} sont remplacées par leurs estimations (optimales) A_{ℓ} . Ce théorème d'additivité n'est évidemment valable que si les A_{ℓ} sont effectivement l'estimateur optimal de la dérive, solution de (1-K). Il ne s'étendrait pas à des estimateurs quelconques des a_{ℓ} (par exemple aux estimateurs de moindre carré).

Dans les applications, si l'objectif est d'estimer une "moyenne mobile" $\int p(dx) Z(x)$ il suffit de résoudre directement le système (1-K) ($n+1+k$ équations) sans qu'il soit nécessaire d'expliciter l'estimation de la dérive (ce qui nécessiterait la résolution du système (1-D) à $(k+1)(n+k+1)$ équations). Mais le théorème d'additivité (1-18) montre que le système plus simple (1-K) tient compte, implicitement, de cette dérive.

Dans le cas particulier du krigeage ponctuel (estimation de $Z(x)$ en $x \notin S$), l'estimateur $Z^*(x)$ obtenu en résolvant (1-13) se met sous la forme :

$$Z^*(x) = \sum_{\ell} A_{\ell} f^{\ell}(x) + \sum_{\alpha} \lambda_K^{\alpha} (Z_{\alpha} - \sum_{\ell} A_{\ell} f^{\ell}_{\alpha})$$

Cette relation exprime que l'on a le droit de prolonger à l'extérieur de S l'estimation optimale $\sum_{\ell} A_{\ell} f^{\ell}(x)$ de la dérive, et de kriger les résidus comme s'il n'y avait pas de dérive.

Ce théorème d'additivité s'étend aux variances d'estimation.

En effet, avec $Z^* = Z_K + Z_D$, on trouve :

$$D^2[Z-Z^*] = D^2(Z-Z_K) + D^2(Z_D) - 2 \text{Cov}(Z-Z_K, Z_D)$$

Mais les équations (1-16) du krigeage ordinaire expriment que $Z-Z_K$ a une covariance nulle avec chacun des Z_α , donc aussi avec Z_D (qui est une combinaison linéaire des Z_α). Il reste donc :

$$D^2[Z-Z^*] = D^2(Z-Z_K) + D^2(Z_D)$$

et on voit que la variance du krigeage universel Z^* est la somme :

$$(1-20) \quad \sigma_U^2 = \sigma_K^2 + \sigma_D^2$$

de la variance σ_K^2 du krigeage simple :

$$\sigma_K^2 = \sigma_Z^2 - \sum_{\alpha} \lambda_K^{\alpha} \sigma_{\alpha Z}$$

et de la variance de la correction Z_D de dérive, soit :

$$\sigma_D^2 = \sum_{\ell, s} \mu_{\ell s} b_{\ell}^{\ell} b_s^s$$

avec $\mu_{\ell s} = \text{Cov}(A_{\ell}, A_s)$ et $b^{\ell} = \int p(dx) f^{\ell}(x) - \sum_{\alpha} \lambda_K^{\alpha} f_{\alpha}^{\ell}$

1-4 CAS OU IL N'EXISTE QU'UN VARIOGRAMME $\gamma(x, y)$

Plaçons-nous maintenant dans le cadre des hypothèses (I') de l'introduction : il existe un demi-variogramme $\gamma(x, y)$, mais non, en général, de covariance $C(x, y)$, ni même d'espérance $E[Z(x)] = m(x)$. Pour des coefficients λ^{α} quelconques, la combinaison $\sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$ n'a donc pas, en général, de variance finie, ni même d'espérance. Moyennant la condition :

$$(1-21) \quad \sum_{\alpha} \lambda^{\alpha} = 0$$

on voit cependant que $\sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$ est une combinaison linéaire d'accroissements de la F.A. $Z(x)$: dans ce cas, $\sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$ admet une espérance et une variance finies.

La dérive

$$m(x) = \sum_{\ell=0}^k a_{\ell} f^{\ell}(x)$$

n'est ici déterminée qu'à une constante près. Si l'on prend $f^0(x) = 1$ (conformément à un usage naturel), on voit que le coefficient a_0 est indéterminé, les a_{ℓ} n'étant réellement définis que pour $\ell = 1, 2, \dots, k$.

Estimation de la dérive.— Pour obtenir les estimateurs optimaux A_{ℓ} des coefficients a_{ℓ} ($\ell = 1, 2, \dots, k$) on raisonne comme dans la section 1-1, à cette réserve près que toutes les combinaisons linéaires des Z_{α} doivent vérifier la condition supplémentaire (1-21). L'indice ℓ et les indices latins ne prenant plus cette fois que les valeurs $1, 2, \dots, k$, on voit facilement que ces estimateurs sont de la forme

$$(1-22) \quad A_{\ell} = \sum_{\alpha} \lambda_{\ell}^{\alpha} Z_{\alpha}$$

avec une matrice λ_{ℓ}^{α} ($\alpha = 1, 2, \dots, n$; $\ell = 1, 2, \dots, k$) vérifiant le système :

$$(1-D') \quad \begin{cases} \sum_{\beta} \lambda_{\ell}^{\beta} \gamma_{\alpha\beta} = - \sum_{\beta} \mu_{\beta\ell} f_{\alpha}^{\beta} - C_{\ell} \\ \sum_{\alpha} \lambda_{\ell}^{\alpha} f_{\alpha}^{\beta} = \delta_{\ell}^{\beta} \\ \sum_{\alpha} \lambda_{\ell}^{\alpha} = 0 \end{cases} \quad \begin{pmatrix} \alpha, \beta = 1, 2, \dots, n \\ \ell, s = 1, 2, \dots, k \\ \gamma_{\alpha\beta} = \gamma(x_{\alpha}, x_{\beta}) \end{pmatrix}$$

Compte tenu du fait que ℓ ne prend plus la valeur 0, ce système ne diffère pas réellement de (1-D) (avec $\sigma_{\alpha\beta} = -\gamma_{\alpha\beta}$ et $C_{\ell} = \mu_{0\ell}$), à condition d'exclure de celui-ci les équations correspondant à $\ell = 0$.

Les paramètres de Lagrange $\mu_{\ell s}$ coïncident encore avec la matrice des covariances :

$$\mu_{\ell s} = \text{Cov}(A_{\ell}, A_s) = E(A_{\ell} A_s) - a_{\ell} a_s$$

et l'on a, bien entendu, $E(A_{\ell}) = a_{\ell}$ quels que soient les a_{ℓ} .

Remarque : On peut même compléter le système (1-D') en réintroduisant les indices $\ell = 0$.

On obtient (1-D), avec $\sigma_{\alpha\beta} = -\gamma_{\alpha\beta}$. Mais l'estimation

$$A_0 = \sum_{\alpha} \lambda_0^{\alpha} Z_{\alpha}$$

de la constante a_0 (qui est, en réalité, indéterminée) ne vérifie pas la condition (1-21), puisque l'on a $\sum_{\alpha} \lambda_0^{\alpha} = 1$: elle conduit donc, en général, à une variance infinie, ce qui est bien conforme à l'hypothèse faite (indétermination de a_0 , ou inexistence de $E(Z(x))$).

Krigeage Universel - De la même manière, pour estimer une moyenne pondérée $\int p(dx) Z(x)$, avec $\int p(dx) = 1$, on formera une expression de la forme

$$(1-23) \quad Z^* = \sum_{\alpha} \lambda^{\alpha} Z_{\alpha}$$

avec cette fois $\sum_{\alpha} \lambda^{\alpha} = 1$, de manière à ce que l'espérance $E(Z-Z^*)$ existe toujours. En imposant à cette espérance d'être nulle quels que soient les a_{ℓ} , et en écrivant que la variance de $Z-Z^*$ est minimale, on obtient le système

$$(1-K') \quad \begin{cases} \sum_{\beta} \lambda^{\beta} \gamma_{\alpha\beta} = \int \gamma(x_{\alpha}, y) p(dy) - \sum_{\ell=1}^k \mu_{\ell} f_{\alpha}^{\ell} - C \\ \sum_{\alpha} \lambda^{\alpha} f_{\alpha}^{\ell} = \int p(dx) f^{\ell}(x) \\ \sum_{\alpha} \lambda^{\alpha} = 1 \end{cases}$$

La variance de $Z^* - Z$ est alors :

$$(1-24) \quad \sigma_U^2 = - \iint p(dx) p(dy) \gamma(x, y) + \sum_{\alpha} \lambda^{\alpha} \int \gamma(x_{\alpha}, y) p(dy) + \sum_{\ell=1}^k \mu_{\ell} f_{\alpha}^{\ell} + C$$

Dans le cas du krigeage ponctuel, l'estimateur optimal $Z^*(x)$ de $Z(x)$, $x \notin S$ est de la forme (1-23), avec des λ^{α} vérifiant :

$$(1-25) \quad \begin{cases} \sum_{\beta} \lambda^{\beta} \gamma_{\alpha\beta} = \gamma(x_{\alpha}, x) - \sum_{\ell=1}^k \mu_{\ell} f_{\alpha}^{\ell} - C \\ \sum_{\alpha} \lambda^{\alpha} f_{\alpha}^{\ell} = f^{\ell}(x) \\ \sum_{\alpha} \lambda^{\alpha} = 1 \end{cases}$$

et la variance correspondante est (compte tenu de $\gamma(x, x) = 0$)

$$(1-26) \quad \sigma_U^2 = \sum_{\alpha} \lambda_{\alpha}^2 \gamma(x_{\alpha}, x) + \sum_{\ell=1}^k \mu_{\ell}^2 \ell_{\alpha}^2 + c$$

Le théorème d'additivité subsiste sans changement : si l'on désigne par

$$z_K = \sum_{\alpha} \lambda_K^{\alpha} z_{\alpha}$$

la solution du krigeage simple (en l'absence de dérive), les λ_K^{α} vérifient le système :

$$\begin{aligned} \sum_{\beta} \lambda_K^{\beta} \gamma_{\alpha\beta} &= \int \gamma(x_{\alpha}, x) p(dx) - c_K \\ \sum_{\alpha} \lambda_K^{\alpha} &= \int p(dx) \end{aligned}$$

On met Z^* sous la forme $Z^* = z_K + z_D$, et la correction de dérive z_D est encore de la forme

$$z_D = \sum_{\ell=1}^k A_{\ell} \int f_{\ell}(x) p(dx) - \sum_{\alpha} \lambda_K^{\alpha} \sum_{\ell=1}^k A_{\ell} f_{\ell}^{\alpha}$$

et on a encore :

$$D^2(z - Z^*) = D^2(z - z_K) + D^2(z_D)$$

On trouvera dans la section 2-10 un exemple de calcul explicite (dérive et krigeage universel) dans le cas d'un variogramme linéaire $\gamma(x; y) = \omega|x-y|$ et de l'espace à une seule dimension.

SECTION 2

THEORIE GENERALE2-1 INTRODUCTION

Nous avons en vue le problème suivant : soit $Z_\omega(x)$ ($x \in \mathbb{R}^n$) une fonction aléatoire sur un espace $(\Omega, \mathcal{A}, P)$, dont nous savons qu'elle est de la forme :

$$(2-1) \quad Z_\omega(x) = Y_\omega(x) + a_\ell f^\ell(x)$$

(avec $a_\ell f^\ell$ au lieu de $\sum_\ell a_\ell f^\ell$, selon la convention de sommation). Les $f^\ell(x)$ ($\ell = 0, 1, \dots, k$) sont des fonctions connues de $x \in \mathbb{R}^n$, les a_ℓ des coefficients constants dont nous ne connaissons pas la valeur numérique; $Y_\omega(x)$, enfin, est une F.A. admettant une espérance nulle et une covariance $C(x, y)$:

$$(2-2) \quad E(Y_\omega(x)) = 0, \quad E(Y_\omega(x) Y_\omega(y)) = C(x, y)$$

On suppose connue une réalisation de $Z_\omega(x)$ pour $x \in S$ ($S \subset \mathbb{R}^n$, ensemble fini ou non de points : en général, S sera compact). On cherche :

- 1 - à former un estimateur A optimal et sans biais du vecteur dérive $a = (a_\ell)$ à l'aide d'opérations linéaires effectuées sur la réalisation $Z_\omega(x)$, $x \in S$.
- 2 - à former, dans les mêmes conditions, un estimateur optimal et sans biais de $Z_\omega(z)$, $z \notin S$, ou, plus généralement, de $Z = \int \mu(dx) Y(x)$, μ étant une mesure dont le support est disjoint de S .
- 3 - à examiner les rapports existant entre ces deux estimateurs (en particulier, a-t-on le droit de travailler sur les résidus $Z(x) - A_\ell f^\ell(x)$, et quelles sortes de biais résultent du fait que la vraie dérive a_ℓ a été remplacée par son estimation A_ℓ).

Ces estimateurs méritent le qualificatif d'universel, puisqu'ils doivent être utilisables quelle que soit la vraie dérive a_ℓ . Il s'agit, cependant, d'une "universalité" limitée puisque nous supposons connues les fonctions $f^\ell(x)$ et la covariance $C(x, y)$ (celle-ci, éventuellement, à un facteur multiplicatif près, comme nous le verrons).

2-2 LES ESPACES $\bar{H}$ ET $\bar{H}_f$

Pour abréger les notations, nous désignerons par :

$\sim$ Λ l'ensemble des mesures $\lambda(dx)$ à support fini inclus dans S

$\sim$ M l'ensemble des mesures $\mu(dx)$ à support inclus dans S

Ainsi une expression du type $\int \lambda(dx) Z(x)$, $\lambda \in \Lambda$ représente une combinaison linéaire finie $\sum_1 \lambda_1 Z(x_1)$ avec $x_1 \in S$. Nous allons considérer une fonction aléatoire $Y_\omega(x)$ vérifiant les hypothèses (2-2) (espérance nulle et covariance $C(x,y)$). On supposera, toujours, que la covariance $C(x,y)$ est de type strictement positif, en ce sens que pour toute mesure μ on a :

$$(2-3) \quad \iint \mu(dx) C(x,y) \mu(dy) = 0 \quad \Rightarrow \quad \mu = 0$$

Le plus souvent, $C(x,y)$ sera continue - et par suite $Y_\omega(x)$ continue en moyenne quadratique.

Soit alors $f(x)$ une fonction donnée (par exemple $f(x) = \sum a_\ell f^\ell(x)$), et la F.A. $Z_\omega(x)$ définie par :

$$(2-4) \quad Z_\omega(x) = Y_\omega(x) + f(x)$$

A $x \in S$ fixé (2-4) définit une application d'une partie de $L^2(\Omega, \mathcal{A}, P)$ dans $L^2(\Omega, \mathcal{A}, P)$ dans $L^2(\Omega, \mathcal{A}, P)$. Nous allons prolonger cette application à certains sous-espaces de L^2 . Pour cela nous nous introduirons :

$\sim$ l'espace H des combinaisons linéaires finies du type : $\int \lambda(dx) Z(x)$, $\lambda \in \Lambda$

$\sim$ l'espace H_f des V.A. de la forme $\int \lambda(dx) [Y(x) + f(x)]$, $\lambda \in \Lambda$, $f(x)$ désignant la fonction donnée ci-dessus.

Si S est fini, H et H_f sont des sous-espaces fermés de $L^2(\Omega, \mathcal{A}, P)$, donc des espaces de Hilbert, et les discussions se simplifient beaucoup. Si S est infini, H et H_f ne seront pas fermés en général, et nous devons considérer leurs fermetures dans $L^2(\Omega, \mathcal{A}, P)$:

$\sim$ les espaces $\bar{H}$ et $\bar{H}_f$, fermetures dans $L^2(\Omega, \mathcal{A}, P)$ des précédents : ce sont des espaces de Hilbert.

Remarque : Tout $Y \in \bar{H}$ est limite dans L^2 d'une suite $Y_n \in H$. Chaque Y_n est combinaison linéaire finie des $Y(x)$; $x \in S$, soit $Y_n = \int \lambda_n(dx) Y(x)$, $\lambda_n \in \Lambda$. Mais la convergence $Y_n \rightarrow Y$ dans L^2 n'entraîne pas que les mesures λ_n convergent vaguement vers une mesure $\mu \in M$. En général, donc, un $Y \in \bar{H}$ quelconque n'admettra pas de représentation de la forme

$$(2-5) \quad Y = \int \mu(dx) Y(x) \quad , \quad \mu \in M$$

L'ensemble H_M des $Y \in \bar{H}$ admettant la représentation (1) pour une mesure $\mu \in M$ est un sous-espace de $\bar{H}$ non fermé dans $\bar{H}$ en général.

Il sera capital, dans les applications, de savoir si un $Y \in \bar{H}$ est ou non dans H_M . En effet, si Y est un estimateur, on ne pourra calculer sa valeur numérique à partir d'une réalisation que si $Y \in H_M$. On trouvera toujours une suite $\lambda_n \in \Lambda$ de mesures à support fini telles que

$$\|Y - \int \lambda_n(dx) Y(x)\| \rightarrow 0$$

mais cela n'impliquera la convergence numérique

$$Y_n = \int \lambda_n(dx) Y_\omega(x) \rightarrow Y_\omega$$

pour presque tout $\omega \in \Omega$ que si $Y \in H_M$ (on peut, théoriquement, extraire de la suite Y_n une suite partielle convergeant presque sûrement vers Y , mais, dans les applications, on ne saura pas construire effectivement cette suite à partir d'une réalisation). Si $Y \notin H_M$, la suite numérique des Y_n construites à partir d'une réalisation sera divergente, en général, bien que les variances des estimateurs Y_n successifs tendent en décroissant vers la variance de l'estimateur optimal Y . Ces remarques s'appliquent aussi à $\bar{H}_F$.

L'Application Φ_F - Considérons l'application de $L^2(\Omega, \mathcal{A}, P)$ dans lui-même définie par :

$$(2-6) \quad \Phi(Z) = Z - E(Z)$$

Elle est continue ($\|Z - E(Z)\|^2 = \|Z\|^2 - (E(Z))^2 \leq \|Z\|^2$). Soit Φ_F la restriction de Φ à H_F . Φ_F applique H_F sur H , puisque tout $Y \in H$ est de la forme $Y = \int \lambda(dx) Y(x)$, $\lambda \in \Lambda$, donc est l'image de :

$$(2-7) \quad \Phi_F^{-1}(Y) = Y + \int \lambda(dx) f(x) \in H_F$$

Elle est injective : $Z - E(Z) = Z' - E(Z')$ pour Z et $Z' \in H_f$ entraîne $E(Z) - E(Z') = Z - Z' \in H_f$. Si $Z \neq Z'$, on a donc $E(Z) - E(Z') \neq 0$ et $1 \in H_f$. Il existe donc une mesure $\lambda \in \Lambda$ avec $1 = \int \lambda(dx) Y(x) + \int \lambda(dx) f(x)$, d'où $0 = \int \lambda(dx) Y(x)$ avec $\lambda \neq 0$, ce qui contredit le fait que C est strictement de type positif.

Ainsi Φ_f est un isomorphisme continu de H_f sur H . En tant qu'application continue, Φ_f se prolonge en une application continue de $\bar{H}_f$ dans $\bar{H}$ que nous désignerons encore par Φ_f . On a $\Phi_f(Z) = Z - E(Z)$, et les inclusions

$$H = \Phi_f(H) \subset \Phi_f(\bar{H}_f) \subset \bar{H}$$

montrent que l'espace image $\Phi_f(\bar{H}_f)$ est dense dans $\bar{H}$: mais, en général, il n'est pas fermé et ne coïncide pas avec $\bar{H}$: si une suite $\lambda_n \in \Lambda$ vérifie le critère de Cauchy dans $L^2(\Omega, \sigma, P)$ (existence de $\lim_{n,m} \iint \lambda_n(dx) \lambda_m(dy) C(x,y)$), il n'en résulte pas nécessairement que la suite $\int \lambda_n(dx) f(x)$ soit convergente, de sorte que $Y_n \rightarrow Y$ n'entraîne pas que $\Phi_f^{-1}(Y_n)$ converge dans $\bar{H}_f$.

La proposition suivante, dont on trouvera la démonstration en Annexe 2, donne les conditions moyennant lesquelles on a $\Phi_f(\bar{H}_f) = \bar{H}$: dans tout ce qui suit, nous supposons que ces conditions sont satisfaites - . Elles expriment que les espaces de Hilbert $\bar{H}$ et $\bar{H}_f$ sont homéomorphes.

Proposition 2-2 - Les 4 conditions suivantes sont équivalentes :

- a/ $1 \notin \bar{H}_f$
- b/ la forme linéaire $\lambda \rightarrow \int \lambda(dx) f(x)$ est continue sur Λ munie de la norme $\|\lambda\|^2 = \iint \lambda(dx) \cdot \lambda(dy) C(x,y)$
- c/ il existe un élément $Y_f \in \bar{H}$ tel que l'on ait :

$$(2-8) \quad \langle Y_f, Y(x) \rangle = f(x) \quad , \quad \forall x \in S$$
- d/ L'application $\Phi_f : Z \rightarrow Z - E(Z)$ est un homéomorphisme de $\bar{H}_f$ sur $\bar{H}$.

Remarques - 1/ Lorsque ces conditions sont vérifiées, l'application inverse Φ_f^{-1} de $\bar{H}$ sur $\bar{H}_f$ est donnée par :

$$(2-9) \quad \Phi_f^{-1}(Y) = Y + \langle Y_f, Y \rangle$$

D'après (2-8), cette formule se réduit à (2-7) pour $Y \in H$.

2/ Si la covariance $C(x,y)$ est continue, les propriétés de la proposition ne peuvent être vérifiées que si f est une fonction continue sur S . En effet, les $Y(x)$ sont continues m.q., et $x_n \rightarrow x$ entraîne $Y(x_n) \rightarrow Y(x)$ dans L^2 , et a fortiori $\langle Y(x_n), Y_f \rangle$ tend vers $\langle Y(x), Y_f \rangle$, c'est-à-dire $f(x_n) \rightarrow f(x)$.

D'ailleurs, si $f(x)$ n'est pas continue en $x_0 \in S$, soit $x_n \rightarrow x_0$ et

$$\lim_{x_n \rightarrow x_0} [f(x_n) - f(x_0)] = C \neq 0$$

Comme $Y(x_n) \rightarrow Y(x_0)$ dans $\bar{H}$, on a :

$$\lim_{x_n \rightarrow x_0} [Y(x_n) - Y(x_0) + f(x) - f(x_0)] = C \in \bar{H}_f, \text{ et } 1 \in \bar{H}_f$$

3/ Toujours en supposant $C(x,y)$ continue et de plus S compact, la convergence vague $\mu_n \rightarrow \mu$ d'une suite de mesures à support dans S entraîne la convergence dans $\bar{H}_f$ de $\int \mu_n(dx) [Y(x) + f(x)]$ vers $\int \mu(dx) [Y(x) + f(x)]$, comme on le vérifie facilement à partir du critère de Cauchy.

La réciproque n'est pas vraie : la convergence dans H_M n'entraîne pas la convergence vague des mesures μ_n , comme nous l'avons déjà indiqué - et les inclusions $H \subset H_M \subset \bar{H}$ sont en général strictes.

En particulier, le vecteur Y_f , qui existe dès que $1 \notin \bar{H}_f$, n'est pas forcément dans H_M .

4/ La condition $Y_f \in H_M$, c'est-à-dire l'existence d'une mesure $\mu_f \in M$ avec :

$$Y_f = \int \mu_f(dx) Y(x)$$

entraîne pour tout $y \in S$:

$$\langle Y_f, Y(y) \rangle = \int \mu_f(dx) \langle Y(x), Y(y) \rangle$$

c'est-à-dire :

$$f(y) = \int \mu_f(dx) C(x,y)$$

Inversement, si $f(y)$ est de cette forme, le vecteur $Y_f = \int \mu_f(dy) Y(x) \in H_M$ vérifie

bien la condition $\langle Y_f, Y(y) \rangle = f(y)$, $\forall y \in S$.

Dans les applications, on s'efforcera de réaliser cette condition $Y_f \in H_M$, (moyennant laquelle l'estimation numérique de la dérive à partir d'une réalisation sera possible). Il suffit, pour cela, que la fonction $f(x)$ soit de la forme

$$\int \mu(dy) C(x,y) , \quad \mu \in M$$

5/ Prenons $f(x) = a_\ell f^\ell(x)$, et posons $\bar{H}_a = \bar{H}_f$. A chaque vecteur $a = (a_\ell)$ est associé un espace $\bar{H}_a$ homéomorphe à $\bar{H}$ par l'homéomorphisme :

$$(2-10) \quad \begin{cases} \Phi_a(Z) = Z - E(Z) & (Z \in \bar{H}_a) \\ \Phi_a^{-1}(Y) = Y + a_\ell \langle Y^\ell, Y \rangle & (Y \in \bar{H}) \end{cases}$$

où Y^ℓ est le vecteur vérifiant :

$$(2-11) \quad \langle Y^\ell, Y(x) \rangle = f^\ell(x) , \quad \forall x \in S$$

Soit $Y \in H_M$, une V.A. admettant la représentation $Y = \int \mu(dx) Y(x)$, $\mu \in M$, .

Compte tenu de (2-11), l'image $\Phi_a^{-1}(Y)$ de Y dans $\bar{H}_a$ est :

$$\Phi_a^{-1}(Y) = \int \mu(dx) [Y(x) + a_\ell f^\ell(x)] = \int \mu(dx) Z(x)$$

On l'obtient en effectuant sur la F.A. $Z(x) = Y(x) + a_\ell f^\ell(x)$ la même opération linéaire $\mu \in M$. Si la mesure μ est connue, et si on connaît une réalisation de la F.A. $Z(x)$ admettant une dérive (inconnue) on est donc capable d'estimer

$$\Phi_a^{-1}(Y) = \int \mu(dx) Z(x)$$

bien que le vecteur a soit inconnu.

2-3 L'ESTIMATEUR OPTIMAL DE LA DERIVE

Soit $Z(x) = Y(x) + a_\ell f^\ell(x)$ une F.A. déduite de la F.A. $Y(x)$ précédente par addition d'une dérive $a_\ell f^\ell(x)$, où les $f^\ell(x)$ ($\ell = 0, 1, \dots, k$) sont $k+1$ fonctions connues (linéairement indépendantes pour $x \in S$, et telles que $1 \notin \bar{H}_f^\ell$) et les a_ℓ des coefficients constants, mais inconnus.

Nous nous proposons de former l'estimateur optimal de ces coefficients a_ℓ , c'est-à-dire de trouver $k+1$ variables aléatoires $A_\ell \in \bar{H}$ (appartenant à l'espace de

Hilbert engendré par les combinaisons linéaires des $Z(x)$, $x \in S$, qui constitue la seule réalité accessible expérimentalement) et vérifiant les conditions:

$$\begin{cases} E(A_\ell) = a_\ell \\ D^2(b^\ell_{A_\ell}) \text{ minimum quels que soient les } k+1 \text{ nombres } b^\ell \end{cases}$$

ces conditions devant être de plus réalisées dans tous les $\bar{H}_a$ à la fois (quels que soient les a_ℓ) : c'est un tel estimateur que nous appelons universel.

D'après la proposition 2-2, il existe des V.A. $Y^\ell \in \bar{H}$ ($\ell = 0, 1, \dots, k$) telles que

$$\langle Y(x), Y^\ell \rangle = f^\ell(x) \quad \forall x \in S$$

Ces vecteurs de $\bar{H}$ sont linéairement indépendants (puisque les $f^\ell(x)$ le sont) et le déterminant des $\langle Y^i, Y^j \rangle$ est différent de 0. La matrice des $\langle Y^i, Y^j \rangle$ admet donc une inverse, soit g_{ij} , et les

$$Y_i = g_{ij} Y^j$$

forment, pour l'espace vectoriel engendré par les Y^ℓ , la base duale de (Y^ℓ) :

$$\langle Y_i, Y^j \rangle = \delta_i^j$$

a/ Considérons, dans $\bar{H}$, la variété linéaire fermée $\bar{V}(b)$ définie par les équations

$$\langle Y, Y^\ell \rangle = b^\ell \quad (\ell = 0, 1, \dots, k)$$

A tout $X \in \bar{V}(b)$ est associé dans chaque $\bar{H}_a$ l'élément :

$$Z_a = X + a_\ell \langle X, Y^\ell \rangle = X + a_\ell b^\ell \in \bar{H}_a$$

et cet élément vérifie par construction :

$$E(Z_a) = a_\ell b^\ell$$

Inversement, si une famille $(Z_a)_{a \in \mathbb{R}^{k+1}}$ d'éléments, vérifiant $Z_a \in \bar{H}_a$ et $Z_a - E(Z_a) = Z_0$ pour tout $a \in \mathbb{R}^{k+1}$, remplit de plus la condition $E(Z_a) = a_\ell b^\ell$ pour tout $a \in \mathbb{R}^{k+1}$, on a

$$E(Z_a) = a_\ell \langle Z_0, Y^\ell \rangle = a_\ell b^\ell, \quad \forall a \in \mathbb{R}^{k+1}$$

donc

$$\langle Z_0, Y^\ell \rangle = b^\ell, \text{ et } Z_0 \in \bar{V}(b)$$

Ainsi, la variété $V(b)$ est constituée des éléments $X \in \bar{H}$ dont les images inverses par chaque ϕ_a sont dans $\bar{H}_a$ des estimateurs sans biais de $a_\ell b^\ell$. Cette variété engendre donc les estimateurs universels que nous cherchons.

Si $X \in \bar{V}(b)$ est de plus dans H_M , soit :

$$X = \int \mu(dx) Y(x) \quad , \quad \mu \in M$$

sa représentation. L'élément $Z_a = X + \langle X, Y^\ell \rangle a_\ell = X + a_\ell b^\ell$ est alors de la forme

$$Z_a = \int \mu(dx) [Y(x) + a_\ell f^\ell(x)] = \int \mu(dx) Z(x)$$

et l'on voit que cet estimateur est accessible numériquement à partir de la réalisation de $Z(x)$, c'est-à-dire à partir des seules données effectivement disponibles.

b/ Parmi les estimateurs sans biais Z_a que nous venons de construire, il reste à choisir le meilleur, c'est-à-dire celui qui minimise la variance $D^2(Z_a) = \|Z_a - E(Z_a)\|^2$. Comme $Z_a - E(Z_a) = X$ ne dépend pas de a , il suffit de chercher dans $\bar{V}(b)$ l'élément $X = X(b)$ de norme minimale : son image inverse dans chaque $\bar{H}_a$ constituera l'estimateur optimal cherché.

Projetons donc 0 dans la variété fermée $\bar{V}(b)$: d'après les propriétés des espaces de Hilbert (cf. par exemple [6] ou [8]), $X(b)$ est l'unique élément de $\bar{V}(b)$ vérifiant :

$$\langle Y, X(b) \rangle = 0 \quad \text{pour tout } Y \in \bar{V}(0) \quad (\langle Y, Y^\ell \rangle = 0)$$

Si Y est un élément quelconque de $\bar{H}$, posons

$$\begin{cases} \langle Y, Y^\ell \rangle = C^\ell \\ C_i = \varepsilon_{ij} C^j \quad (C^\ell = C_i \langle Y^i, Y^\ell \rangle) \end{cases}$$

Le vecteur $Y - C_\ell Y^\ell$ est alors orthogonal aux Y^i , donc appartient à $\bar{V}(0)$, et $X(b)$ vérifie

$$\langle Y, X(b) \rangle = C_\ell \langle Y^\ell, X(b) \rangle \quad , \quad \forall Y \in \bar{H}$$

Il suffit d'ailleurs de considérer les éléments $Y(x)$, $x \in S$, qui engendrent $\bar{H}$. Pour

$Y(x)$, on a par définition de Y^ℓ :

$$C^\ell = \langle Y, Y^\ell \rangle = f^\ell(x)$$

Ainsi $X(b)$ est l'unique élément de $\bar{V}(b)$ vérifiant :

$$\langle Y(x), X(b) \rangle = \langle X, Y^i \rangle g_{i\ell} f^\ell(x), \quad \forall x \in S$$

c'est-à-dire une relation de la forme :

$$\langle Y(x), X(b) \rangle = \mu_\ell f^\ell(x), \quad \forall x \in S$$

Inversement, si $X \in \bar{V}(b)$ vérifie cette relation, pour des constantes μ_ℓ données, on a :

$$X = \mu_\ell Y^\ell$$

et par suite $\langle X, Y^i \rangle = \mu_\ell \langle Y^\ell, Y^i \rangle$, d'où résulte bien $\mu_\ell = \langle X, Y^i \rangle g_{i\ell}$ et $X = X(b)$.

En résumé, l'estimateur universel optimal cherché de $a_\ell b^\ell$ est l'image dans $\bar{H}_a$ de l'unique élément X de $\bar{H}$ vérifiant le système :

$$(2-12) \quad \begin{cases} \langle Y(x), X(b) \rangle = \mu_\ell f^\ell(x) & \forall x \in S \\ \langle X(b), Y^\ell \rangle = b^\ell & (\ell = 0, 1, \dots, k) \end{cases}$$

Les μ_ℓ sont les paramètres de Lagrange habituels, et les secondes relations (2-12) permettent de les déterminer.

c/ Comme les Y^ℓ vérifient par définition $\langle Y(x), Y^\ell \rangle = f^\ell(x)$ l'unique élément $X(b)$ de $\bar{H}$ vérifiant (2-12) est donné par :

$$\begin{cases} X(b) = \mu_\ell Y^\ell \\ \mu_\ell = g_{\ell i} b^i \end{cases}$$

et $X(b) = g_{\ell i} b^i Y^\ell$ dépend linéairement des constantes arbitraires b^i . Introduisons donc le vecteur aléatoire Y_i défini par

$$Y_i = g_{\ell i} Y^\ell$$

Ce vecteur est l'unique solution du système :

$$(2-D) \quad \begin{cases} \langle Y(x), Y_i \rangle = \mu_{i\ell} f^\ell(x) & , \quad x \in S \\ \langle Y_i, Y^\ell \rangle = \delta_i^\ell \end{cases}$$

et, quelles que soient les constantes b^ℓ , on a $X(b) = b^i Y_i$. On note aussi que les paramètres de Lagrange :

$$\mu_{ij} = \langle Y_i, Y_j \rangle = \varepsilon_{ij}$$

constituent la matrice des covariances des Y_i - L'image inverse de Y_i dans chaque $\tilde{H}_a$ constitue l'estimateur optimal universel A_i du vecteur dérive a_i : on a donc, d'après (2-10)

$$A_i = Y_i + a_\ell \langle Y_i, Y^\ell \rangle$$

c'est-à-dire, compte tenu de la seconde relation (2-D)

$$A_i = Y_i + a_i$$

Dans chaque $\tilde{H}_a$, le vecteur A vérifie donc $E(A) = a$ et $D^2(A_i b^i)$ minimale $\forall b \in \mathbb{R}^{k+1}$.

Cas particuliers - Supposons que les Y^i ou, ce qui revient au même, les Y_i soient dans H_M , autrement dit qu'il existe des mesures $\lambda_i \in M$ avec $Y_i = \int \lambda_i(dx) Y(x)$. Ces mesures λ_i constituent l'unique solution du système d'équations intégrales :

$$(2-DM) \quad \begin{cases} \int \lambda_i(dy) C(x,y) = \mu_{i\ell} f^\ell(x) & , \quad x \in S \\ \int \lambda_i(dx) f^\ell(x) = \delta_i^\ell \end{cases}$$

et les constantes de Lagrange $\mu_{i\ell}$ sont données par :

$$\mu_{ij} = \iint \lambda_i(dx) C(x,y) \lambda_j(dy)$$

Dans chaque $\tilde{H}_a$, l'estimateur universel A_i admet alors la représentation :

$$A_i = \int \lambda_i(dx) Z(x)$$

et se calcule numériquement à partir de la réalisation $Z(x) = Y(x) + a_\ell f^\ell(x)$ effectivement disponible. De plus, la matrice des covariances des A_i est identique à la matrice μ_{ij} des paramètres de Lagrange du système (2-DM)

$$E((A_i - a_i)(A_j - a_j)) = \mu_{ij}$$

Dans le cas général où les Y_i ne sont pas dans H_M , A_i est cependant limite dans L^2 d'une suite :

$$A_i^n = \int \lambda_i^n(dx) Z(x) \quad , \quad \lambda_i^n \in \Lambda$$

de combinaisons linéaires des $Z(x)$. Les A_i^n constituent une suite d'estimateurs sans biais dont les variances, décroissantes, tendent vers la variance minimale; mais cette suite n'est pas nécessairement convergente numériquement pour une réalisation donnée de $Z(x)$.

2-4 EXEMPLE : ESTIMATION DE L'ESPERANCE D'UNE F.A. STATIONNAIRE

A titre d'exemple, considérons le cas simple où $Z(x)$ est une F.A. stationnaire d'ordre 2, d'espérance (inconnue) $E[Z(x)] = a$ et de covariance (connue) $C(x,y) = C(x-y)$. Prenons $Y(x) = Z(x) - a$. Pour estimer $a = E[Z(x)]$, formons l'estimateur universel A . L'image $A - E(A) = A - a$ de A dans $\tilde{H}$ est l'élément X vérifiant

$$\begin{cases} \langle Y(x), X \rangle = \mu & , \quad \forall x \in S \\ \langle X, Y_1 \rangle = 1 \end{cases}$$

(Ici, $f(x) = 1$, et Y_1 est l'élément vérifiant $\langle Y(x), Y_1 \rangle = 1$ pour tout $x \in S$). Il en résulte que l'on a simplement :

$$X = \frac{Y_1}{\|Y_1\|^2} \quad \text{et} \quad \mu = \frac{1}{\|Y_1\|^2} = \|X\|^2 = D^2(X)$$

Si $Y_1 \in H_M$, soit $Y_1 = \int \lambda_1(dx) Y(x)$, $\lambda_1 \in M$, on a alors :

$$A = \int \lambda(dx) Z(x)$$

dans chaque $\tilde{H}_a$, avec une mesure λ vérifiant :

$$\begin{cases} \int \lambda(dy) C(x-y) = \mu & , \quad \forall x \in S \\ \int \lambda(dx) = 1 \end{cases}$$

La condition pour qu'il existe une telle mesure λ , à savoir l'existence de $\lambda_1 \in M$ vérifiant $\int \lambda_1(dy) C(x-y) = 1$ pour $x \in S$ exprime que le noyau $C(x-y)$ admet sur S un

potentiel d'équilibre. La capacité de S est alors $\int \lambda_1(dx) = \frac{1}{\mu} = \frac{1}{D^2(A)}$

2-5 LE KRIGEAGE UNIVERSEL

Enoncé en toute généralité, le problème que nous avons à résoudre est le suivant : soit $Z_0 \in L^2(\Omega, \mathcal{A}, P)$ une V.A. d'espérance nulle n'appartenant pas à notre espace $\bar{H}$, et soient b^ℓ ($\ell = 0, 1, \dots, k$) un vecteur (donné) de $\mathbb{R}^{k+1}$. Pour chaque vecteur $a \in \mathbb{R}^{k+1}$, nous faisons correspondre à Z_0 la V.A. :

$$Z_a = Z_0 + a_\ell b^\ell$$

On a donc $E(Z_a) = a_\ell b^\ell$. Le problème du krigage en présence d'une dérive $a_\ell f^\ell(x)$ (avec des coefficients a_ℓ inconnus) consiste à trouver un élément $Y^* \in \bar{H}$ vérifiant les deux conditions suivantes :

1/ Pour chaque $a \in \mathbb{R}^{k+1}$, l'image inverse $\Phi_a^{-1}(Y^*)$ de Y^* dans H_a est l'élément $Z_a^* = Y^* + a_\ell b^\ell$.

2/ De plus, pour chaque $a \in \mathbb{R}^{k+1}$, la variance $D^2(Z_a - Z_a^*) = D^2(Y^* - Z_0)$ doit être minimale. (Lorsque la condition 1/ est réalisée, cette deuxième condition ne fait plus intervenir le vecteur a).

La première condition ($Y^* + a_\ell b^\ell \in \bar{H}_a$, $\forall a \in \mathbb{R}^{k+1}$) exprime, comme nous l'avons vu en 2-2, que Y^* appartient à la variété fermée $V(b)$ définie par :

$$Y \in V(b) \Leftrightarrow Y \in \bar{H}, \quad \langle Y, Y^\ell \rangle = b^\ell, \quad \ell = 0, 1, \dots, k$$

La seconde condition ($\|Y^* - Z_0\|$ minimale) montre ensuite que Y^* est la projection de Z_0 dans cette variété fermée $V(b)$. Autrement dit, Y^* est l'unique élément de $V(b)$ vérifiant :

$$\langle Z_0 - Y^*, Y \rangle = 0 \quad \forall Y \in V(0)$$

En procédant exactement comme nous l'avons fait pour établir le système (2-D), on en déduit que Y^* est l'unique solution du système :

$$(2-K) \quad \begin{cases} \langle Y(x), Y^* \rangle = \langle Y(x), Z_0 \rangle + \mu_\ell f^\ell(x), & \forall x \in S \\ \langle Y^*, Y^\ell \rangle = b^\ell \end{cases}$$

(les Y^ℓ sont, comme toujours, les éléments de $\bar{H}$ vérifiant $\langle Y(x), Y^\ell \rangle = f^\ell(x)$, $\forall x \in S$).

Lorsque $Y^* \in H_M$, on a $Y^* = \int \lambda(dx) Y(x)$ avec une mesure $\lambda \in M$ vérifiant le système d'équations intégrales :

$$(2-K,M) \quad \begin{cases} \int \lambda(dy) C(x,y) = \langle Y(x), Z_0 \rangle + \mu_\ell f^\ell(x), & \forall x \in S \\ \int \lambda(dx) f^\ell(x) = b^\ell \end{cases}$$

L'image inverse Z_a^* de Y^* dans chaque $\bar{H}_a$ (c'est-à-dire l'estimateur optimal) s'écrit alors, quel que soit a :

$$Z_a^* = \int \lambda(dx) Z(x)$$

D'après la seconde relation (2-K,M), on a aussi :

$$Z_a^* = Y^* + a_\ell b^\ell$$

Cas particulier : krigeage ponctuel - Si la variable Z_a à estimer est $Z(z)$, valeur en $z \notin S$ de la F.A. $Z(x)$, on a : $Z(z) = Y(z) + a_\ell f^\ell(z)$, et :

$$\begin{cases} \langle Y(x), Z_0 \rangle = C(x,z) \\ b^\ell = f^\ell(z) \end{cases}$$

Le système (2-K,M), par exemple, est alors remplacé par :

$$(2-13) \quad \begin{cases} \int \lambda(dy) C(x,y) = C(x,z) + \mu_\ell f^\ell(x), & \forall x \in S \\ \int \lambda(dx) f^\ell(x) = f^\ell(z) \end{cases}$$

2-6 RELATION ENTRE L'ESTIMATION DE LA DERIVE ET LE KRIGEAGE UNIVERSEL

En l'absence de dérive, le meilleur estimateur de Z_0 est la projection de Z_0 dans $\bar{H}$. C'est l'élément unique $Y_K \in \bar{H}$ vérifiant les équations du krigeage en l'absence de dérive :

$$(2-14) \quad \langle Y(x), Y_K \rangle = \langle Y(x), Z_0 \rangle, \quad \forall x \in S$$

Posons alors

$$Y^* = Y_K + Y_D$$

et substituons dans (2-K) : l'élément Y_D représente la correction que l'on doit effectuer en présence d'une dérive (dont les coefficients a_ℓ sont inconnus) pour obtenir un estimateur universel. Il doit vérifier :

$$(2-15) \quad \begin{cases} \langle Y(x), Y_D \rangle = \mu_\ell f^\ell(x) & , \quad \forall x \in S \\ \langle Y_D, Y^\ell \rangle = b^\ell - \langle Y_K, Y^\ell \rangle = b^\ell - \langle Z_0, Y^\ell \rangle \end{cases}$$

Or ce système (2-15) est identique à (2-12), à ceci près que l'on a remplacé b^ℓ par $b^\ell - \langle Y_K, Y^\ell \rangle$. Par conséquent, en désignant par Y_i le vecteur aléatoire solution du système (2-D), la correction de dérive Y_D est :

$$Y_D = Y_i [b^i - \langle Y_K, Y^i \rangle]$$

L'image inverse par Φ_a de Y_i est l'estimateur optimal $A_i \in \bar{H}_a$ de la dérive ; celle de Z_D est donc $A_i b^i - A_i \langle Y_K, Y^i \rangle$. Soient $Z_a^* = \Phi_a^{-1}(Y^*)$ l'estimateur cherché et $Z_K = \Phi_a^{-1}(Y_K)$ l'image inverse dans $\bar{H}_a$ de Y_K : la relation $Y^* = Y_K + Y_D$ donne dans $\bar{H}_a$:

$$(2-16) \quad Z_a^* = Z_K - A_i \langle Y_K, Y^i \rangle + A_i b^i$$

Or, si l'on connaissait les vraies valeurs des a_i , le meilleur estimateur de Z_a serait :

$$Y_K + a_\ell b^\ell = Z_K - a_i \langle Y_K, Y^i \rangle + a_\ell b^\ell$$

On voit la signification de la relation (2-16) : elle exprime que l'on a le droit de calculer Z_a^* comme si $A_\ell f^\ell(x)$ représentait la véritable dérive $a_\ell f^\ell(x)$, et non une simple estimation de celle-ci. Autrement dit, le krigeage universel coïncide avec le krigeage ordinaire effectué sur les résidus $Y(x) - A_\ell f^\ell(x)$.

Exemple : Krigeage Ponctuel - Soit à estimer $Z(z)$ en $z \notin S$. Le système (2-K) donne :

$$(2-17) \quad \begin{cases} \langle Y(x), Y^* \rangle = C(x, z) + \mu_\ell f^\ell(x) & , \quad \forall x \in S \\ \langle Y^*, Y^\ell \rangle = f^\ell(z) \end{cases}$$

En l'absence de dérive, l'estimateur du krigeage Y_K est la solution de :

$$\langle Y(x), Y_K \rangle = C(x, z) \quad , \quad \forall x \in S$$

et l'estimateur universel (2-16) s'écrit :

$$Z^*(z) = Z_K - \langle Y_K, Y^1 \rangle A_1 + A_2 f^\ell(z)$$

Plaçons-nous dans le cas où A_1 et $Y_K \in H_M$. On a :

$$\begin{cases} A_1 = \int \lambda_1(dx) [Y(x) + a_\ell f^\ell(x)] = \int \lambda_1(dx) Y(x) + a_\ell \\ \int \lambda_1(dx) f^\ell(x) = \delta_1^\ell \end{cases}$$

et

$$Y_K = \int \lambda_K(dx) Y(x) \quad , \quad \lambda_K \in M$$

avec une mesure λ_K vérifiant :

$$\int \lambda_K(dy) C(x,y) = C(x,z) \quad , \quad \forall x \in S$$

On voit que la condition $Y_K \in H_M$ exprime que le problème du balayage sur S est soluble pour le noyau potentiel $C(x,y)$. L'estimateur cherché se met sous la forme :

$$(2-18) \quad Z^*(z) = A_2 f^\ell(z) + \int \lambda_K(dx) [Z(x) - A_1 f^1(x)]$$

Cette relation exprime que l'on peut prolonger l'estimation $A_2 f^\ell(x)$ de la dérive à l'extérieur de S , et kriger les résidus (à l'aide de la covariance $C(x,y)$ comme s'il n'y avait pas de dérive.

2-7 VARIANCE DU KRIGEAGE UNIVERSEL

Désignons par σ_U^2 la variance $D^2(Z_a - Z_a^*) = D^2(Z_0 - Y^*)$ associée à notre estimateur optimal, solution de (2-K). On a :

$$\sigma_U^2 = \|Y^* - Z_0\|^2 = \|Y^*\|^2 + \|Z_0\|^2 - 2 \langle Y^*, Z_0 \rangle$$

Mais, d'après (2-K), Y^* vérifie

$$\langle Y^*, Y \rangle = \langle Z_0, Y \rangle + \mu_\ell \langle Y^\ell, Y \rangle \quad , \quad \forall Y \in \bar{H}$$

En particulier, en prenant $Y = Y^*$, on a :

$$\|Y^*\|^2 = \langle Z_0, Y^* \rangle + \mu_\ell \langle Y^\ell, Y^* \rangle = \langle Z_0, Y^* \rangle + \mu_\ell b^\ell$$

Par suite, la variance σ_U^2 est :

$$(2-19) \quad \sigma_U^2 = \|z_0\|^2 + \mu_{\ell} b^{\ell} - \langle Y^*, z_0 \rangle$$

Par exemple, dans le cas du krigeage ponctuel et pour $Y^* = \int \lambda(dx) Y(x) \in H_M$, on trouve :

$$\sigma_U^2 = C(z, z) + \mu_{\ell} r^{\ell}(z) - \int \lambda(dx) C(x, y)$$

D'autre part, on a vu, en 2-6, que $Y^* = Y_K + Y_D$ est la somme de l'estimateur Y_K du krigeage sans dérive et de la correction de dérive $Y_D = A_i b^i - A_i \langle Y_K, Y^i \rangle$. Par définition de Y_K , $Z_0 - Y_K$ est orthogonal à tous les éléments de H , et, en particulier, à Y_D . On a donc additivité des variances

$$(2-20) \quad \sigma_U^2 = \sigma_K^2 + \sigma_D^2$$

$$(\sigma_K^2 = \|z_0 - Y_K\|^2, \quad \sigma_D^2 = \|Y_D\|^2)$$

2-8 CAS OU IL N'EXISTE QU'UN VARIOGRAMME $\gamma(x, y)$

Supposons, maintenant, que la F.A. $Y(x)$ vérifie seulement les hypothèses :

$$\begin{cases} E[Y(y) - Y(x)] = 0 \\ \frac{1}{2} D^2[Y(y) - Y(x)] = \gamma(x, y) \end{cases}$$

et considérons, comme plus haut, les F.A. $Z(x) = Y(x) + a_{\ell} r^{\ell}(x)$, $a = (a_{\ell}) \in \mathbb{R}^k$ ($\ell = 1, 2, \dots, k$: l'indice $\ell = 0$ est supprimé, car le coefficient a_0 associé à la fonction constante $r^0(x) = 1$ est superflu, puisque l'on n'étudie plus ici $Z(x)$ qu'à une constante près). On peut reprendre les raisonnements précédents à condition de remplacer $\bar{H}$ et $\bar{H}_f$ par les espaces de Hilbert engendrés par les combinaisons linéaires finies $\int \lambda(dx) Y(x)$, $\int \lambda(x) Z(x)$ avec $\lambda \in \Lambda_0$, ensemble des mesures à support fini inclus dans S vérifiant la condition supplémentaire $\int \lambda(dx) = 0$ (moyennant laquelle espérances et variances existent (cf. [3])). On peut aussi se ramener aux cas déjà traités en remplaçant $Y(x)$ par $Y(x) - Y(x_0)$, pour un $x_0 \in S$ choisi une fois pour toutes. Cette nouvelle F.A. a une espérance nulle, et admet la covariance :

$$C(x, y) = -\gamma(x, y) + \gamma(x, x_0) + \gamma(y, x_0)$$

Le système (2-D) est alors remplacé par le suivant :

$$(2-D') \quad \begin{cases} \langle Y(x) - Y(x_0), Y_i \rangle = \mu_{i\ell} [f^\ell(x) - f^\ell(x_0)] \\ \langle Y_i, Y^\ell \rangle = \delta_i^\ell \end{cases}$$

les Y^ℓ désignant les vecteurs vérifiant :

$$\langle Y(x) - Y(x_0), Y^\ell \rangle = f^\ell(x) - f^\ell(x_0), \quad \forall x \in S$$

L'estimateur universel optimal $A_i = Y_i + a_\ell \langle Y_i, Y^\ell \rangle = Y_i + a_i$ vérifie alors $E(a_i) = a_i$ et $D^2(b^1 A_i)$ minimal $\forall b \in \mathbb{R}^k$. Les paramètres de Lagrange coïncident encore avec la matrice des covariances des A_i .

Dans le cas $Y_i \in H_M$, on a :

$$A_i = \int \lambda_i(dx) Z(x)$$

avec des mesures $\lambda_i \in M$ vérifiant pour $i, \ell = 1, 2, \dots, k$:

$$(2-D', M) \quad \begin{cases} - \int \lambda_i(dy) \gamma(x, y) = \mu_{i\ell} f^\ell(x) + c_i, \quad \forall x \in S \\ \int \lambda_i(dx) f^\ell(x) = \delta_i^\ell \\ \int \lambda_i(dx) = 0 \end{cases}$$

On voit que ce système ne diffère de (2-D, M) que par la suppression des équations relatives à λ_0 [on pourrait rétablir ces équations, et former $A_0 = \int \lambda_0(dx) Z(x)$: mais la variance de A_0 serait, en général, infinie]. On a de plus :

$$(2-21) \quad \begin{cases} E(A_i) = a_i \quad (\forall a \in \mathbb{R}^k) \\ E(A_i A_j) - a_i a_j = \mu_{ij} = - \int \lambda_i(dx) \lambda_j(dy) \gamma(x, y) \end{cases}$$

Les résultats relatifs au krigeage se transposent de même sans difficulté. En particulier, on a encore le droit de prolonger l'estimation $A_\ell f^\ell(x)$ de la dérive à l'extérieur de S et de kriger les résidus $Z(x) - A_\ell f^\ell(x)$ comme s'il n'y avait pas de dérive. Explicitons seulement les systèmes d'équations du krigeage ponctuel dans le cas $Y^* \in H_M$.

Pour estimer $Z_a(z), z \notin S$, on forme un estimateur de la forme

$$Z^*(z) = \int \lambda(dx) Z(x)$$

avec la mesure $\lambda \in M$ vérifiant le système d'équations intégrales :

$$(2-K') \quad \begin{cases} \int \gamma(x,y) \lambda(dy) = \gamma(x,z) - \mu_{\ell} f^{\ell}(x) - C, & \forall x \in S \\ \int \lambda(dy) f^{\ell}(y) = f^{\ell}(z) \\ \int \lambda(dy) = 1 \end{cases}$$

La variance de cette estimation est :

$$\sigma_U^2 = \int \lambda(dy) \gamma(y,z) + \mu_{\ell} f^{\ell}(z) + C$$

Elle vérifie encore la relation d'additivité (2-20), avec, comme en (2-18), $Z^*(z) = Z_K^*(z) + Z_D^*(z)$:

$$\begin{cases} Z_K(z) = \int \lambda_K(dx) Z(x) \\ Z_D(z) = A_{\ell} f^{\ell}(z) - \int \lambda_K(dx) A_1 f^1(x) \end{cases}$$

2-9 LE VARIOGRAMME DES RESIDUS

Dans ce qui suit, nous désignons par $k(x)$ l'indicatrice de S ($k(x) = 1$ pour $x \in S$, $= 0$ pour $x \notin S$), et nous désignons par $k(h)$ la mesure de l'intersection de S et de son translaté par le vecteur h :

$$K(h) = \int k(x) k(x+h) dx$$

Nous allons établir un résultat qui aura une grande importance dans les problèmes de contrôle des hypothèses abordés dans la section 3. Plaçons-nous dans les hypothèses de la section 2-8 ($\gamma(x,y)$ connu). Ayant formé l'estimateur optimal A_{ℓ} de la dérive, nous connaissons les résidus $Z(x) - A_{\ell} f^{\ell}(x)$, et nous pouvons construire le variogramme des résidus :

$$(2-22) \quad 2 \gamma^*(h) = \frac{1}{K(h)} \int k(x) k(x+h) [Z(x+h) - A_{\ell} f^{\ell}(x+h) - Z(x) + A_{\ell} f^{\ell}(x)]^2 dx$$

Nous poserons aussi

$$\gamma(h) = \frac{1}{K(h)} \int k(x) k(x+h) \gamma(x, x+h) dx$$

(lorsque $\gamma(x,y) = \gamma(x+h, y+h)$ est invariant par translation, on a simplement $\gamma(x,y) = \gamma(x-y)$). Si, au lieu des A_ℓ , figuraient dans (2-22) les vraies valeurs des coefficients a_ℓ , on aurait $E(\gamma^*(h)) = \gamma(h)$. Mais le fait que ces vraies valeurs inconnues ont été remplacées par leurs estimations A_ℓ va entraîner un biais irréductible. Calculons, en effet, l'espérance du second membre de (2-22). Comme les A_ℓ vérifient $E(A_\ell) = a_\ell$, on peut supposer $a_\ell = 0$ (ce qui revient à remplacer $Z(x)$ par $Y(x) = Z(x) - a_\ell f^\ell(x)$). Avec cette simplification, cherchons l'espérance de l'expression :

$$\begin{aligned} [Z(x+h) - Z(x) - A_\ell(f^\ell(x+h) - f^\ell(x))]^2 &= [Z(x+h) - Z(x)]^2 \\ &- 2 A_\ell[f^\ell(x+h) - f^\ell(x)] [Z(x+h) - Z(x)] \\ &+ A_\ell A_s[f^\ell(x+h) - f^\ell(x)] [f^s(x+h) - f^s(x)] \end{aligned}$$

Le premier terme a pour espérance $2 \gamma(x, x+h)$, le troisième :

$$\mu_{\ell s} [f^\ell(x+h) - f^\ell(x)] [f^s(x+h) - f^s(x)]$$

les $\mu_{\ell s} = \text{Cov}(A_\ell, A_s)$ désignant les covariances des A_ℓ . Calculons l'espérance du second terme, et pour cela tout d'abord $E[A_\ell(Z(x+h) - Z(x))]$. Or, d'après (2-D'), on a ici $A_\ell = \mu_{\ell s} Y^s$ avec un vecteur Y^s vérifiant :

$$\langle Y, Y^s \rangle = \int \lambda(dx) f^s(x)$$

pour tout $Y = \int \lambda(dx) Z(x)$ avec $\lambda \in \Lambda$ et $\int \lambda(dx) = 0$. En prenant $Y = Z(x+h) - Z(x)$, on trouve donc :

$$E[A_\ell(Z(x+h) - Z(x))] = \mu_{\ell s} [f^s(x+h) - f^s(x)]$$

Il en résulte qu'au signe près l'espérance du second terme est le double de celle du troisième. En portant dans (2-22) et en intégrant en x , il vient donc

$$(2-23) \quad \begin{cases} 2 E(\gamma^*(h)) = 2 \gamma(h) - \mu_{\ell s} T^{\ell s}(h) \\ T^{\ell s}(h) = \frac{1}{K(h)} \int k(x) k(x+h) [f^\ell(x+h) - f^\ell(x)] [f^s(x+h) - f^s(x)] dx \end{cases}$$

Comme la matrice $\mu_{\ell s}$ est de type positif, le terme $\mu_{\ell s} T^{\ell s}(h)$ qui représente le biais est toujours positif, et on a toujours

$$E(\gamma^*(h)) < \gamma(h)$$

On verra, dans l'application ci-dessous, que ce biais peut être considérable. Ce biais peut être si important qu'il peut inciter (à tort) l'expérimentateur à conclure à l'indépendance des résidus. Cet effet est évidemment général, et se manifeste encore lorsque l'on estime la dérive par un procédé quelconque (non optimal), par exemple par une méthode de moindres carrés : il donne à l'expérimentateur l'impression (erronée) que le phénomène qu'il étudie se réduit à la simple somme d'une dérive et d'un bruit de fond.

2-10 EXEMPLE DU VARIOGRAMME LINEAIRE

A titre d'exercice, nous allons traiter explicitement un exemple simple, d'ailleurs très utile dans les applications : prenons le variogramme $\gamma(x,y) = |x-y|$ dans l'espace à une dimension, une dérive polynomiale $a_1 x + \dots + a_k x^k$, et pour S le segment $[-R, +R]$ de longueur $2R$.

Désignons par δ_R et δ_{-R} les mesures de Dirac (masse unité) placées en R et en $-R$. On a pour tout $y \in [-R, +R]$:

$$(a) \quad \begin{cases} \frac{1}{2R} \int [\delta_R(dx) + \delta_{-R}(dx)] |x-y| = 1 \\ \frac{1}{2} \int [\delta_R(dx) - \delta_{-R}(dx)] |x-y| = -y \end{cases}$$

De même, la mesure de densité x^p sur $[-R, +R]$ (et 0 à l'extérieur) vérifie la relation :

$$(b) \quad \int_{-R}^R x^p |x-y| dx = \frac{2 y^{p+2}}{(p+1)(p+2)} - \frac{y R^{p+1}}{p+1} (1 + (-1)^{p+1}) + \frac{R^{p+2}}{p+2} (1 + (-1)^{p+2})$$

En particulier, pour $p = 0$:

$$(b') \quad \int_{-R}^R |x-y| dx = y^2 + R^2$$

Lorsque la dérive est un polynôme d'ordre k , $\sum_{n=1}^k a_n x^n$, l'estimateur optimal A_n du coefficient a_n est de la forme

$$(c) \quad A_n = \int_{-R}^R \lambda_n(dx) Z(x)$$

avec des mesures λ_n vérifiant le système $(2-D', M)$:

$$(d) \quad \begin{cases} -\int \lambda_n(dx) |x-y| = \sum_l \mu_n l y^l + c_n & (-R \leq y \leq R) \\ \int \lambda_n(dx) x^l = \delta_n^l \\ \int \lambda_n(dx) = 0 \end{cases}$$

Il suffit de comparer la première relation (d) à (a) et (b) pour voir que ces mesures λ_n sont nécessairement des combinaisons linéaires des mesures :

$$\frac{\delta_R + \delta_{-R}}{2R}, \quad \frac{\delta_R - \delta_{-R}}{2}, \quad x^p dx \quad (p = 0, 1, \dots, k-2)$$

Les deux dernières relations (d) permettront de déterminer les coefficients de ces combinaisons linéaires, de sorte que l'on saura toujours former la solution explicite du système (d) pour une dérive polynomiale de degré k donné.

Par exemple, pour $k=1$ (dérive $a_1 x$ purement linéaire), on doit avoir $A_1 = B \int \frac{\delta_R(dx) - \delta_{-R}(dx)}{2} \quad Z(x) = B \frac{Z(R) - Z(-R)}{2}$. La condition $B \int \lambda_1(dx) x = BR = 1$ donne $B = \frac{1}{R}$, et notre estimateur (très simple) est donc

$$A_1 = \frac{Z(R) - Z(-R)}{2R}$$

La première relation (d) permet, par identification, de déterminer μ_{11} qui est la variance de A_1 : d'après (a), en effet, on trouve

$$-\frac{1}{2R} \int [\delta_R(dx) - \delta_{-R}(dx)] |x-y| = \frac{y}{R}$$

D'où :

$$\mu_{11} = D^2(A_1) = \frac{1}{R}$$

Traisons aussi le cas d'une dérive quadratique $a_1 x + a_2 x^2$ ($k = 2$). D'après (a) et (b'), on doit chercher des mesures de la forme :

$$\begin{cases} \lambda_1 = B \frac{\delta_R - \delta_{-R}}{2} \\ \lambda_2 = E \frac{\delta_R + \delta_{-R}}{2R} + F dx \end{cases}$$

Ces mesures vérifient la première relation (d). Pour λ_1 , on a déjà $\int \lambda_1(dx) = \int x^2 \lambda_1(dx) = 0$. La condition $\int x \lambda_1(dx) = 1$ donne $BR = 1$, soit :

$$B = \frac{1}{R}$$

Pour λ_2 , on a déjà $\int x \lambda_2(dx) = 0$. Les relations $\int \lambda_2(dx) = 0$ et $\int x^2 \lambda_2(dx) = 1$ donnent de plus :

$$\begin{cases} \frac{E}{R} + F \cdot 2R = 0 \\ ER + \frac{2}{3} F R^3 = 1 \end{cases}$$

D'où $F = -\frac{3}{4R^3}$ et $E = \frac{3}{2R}$. Posant $\bar{Z} = \frac{1}{2R} \int_{-R}^R Z(x) dx$, on voit que les estimateurs optimaux sont :

$$(e) \quad \begin{cases} A_1 = \frac{1}{2R} (Z(R) - Z(-R)) \\ A_2 = \frac{3}{4R^2} (Z(R) + Z(-R)) - \frac{3}{2R^2} \bar{Z} \end{cases}$$

Cherchons la matrice des covariances μ_{ij} : il suffit pour cela d'identifier les coefficients de y et y^2 au second membre de la première relation (d). D'après (a) et (b'), on a :

$$\begin{aligned} -\frac{1}{R} \int \frac{\delta_R(dx) - \delta_{-R}(dx)}{2} |x-y| &= \frac{y}{R} \\ -\frac{3}{2R} \int \frac{\delta_R(dx) + \delta_{-R}(dx)}{2R} |x-y| + \frac{3}{4R^3} \int_{-R}^R |x-y| dx &= \frac{3}{2R} + \frac{3}{4R} + \frac{3}{4} \frac{y^2}{R^3} \end{aligned}$$

On a donc :

$$(f) \quad \begin{cases} \mu_{11} = D^2(A_1) = \frac{1}{R} \\ \mu_{12} = \mu_{21} = \text{Cov}(A_1, A_2) = 0 \\ \mu_{22} = D^2(A_2) = \frac{3}{4R^3} \end{cases}$$

Calculons maintenant l'espérance du variogramme des résidus $Z(x) - A_1 x - A_2 x^2$.

D'après (2-23), nous devons calculer les termes :

$$\begin{cases} T^{11}(h) = \frac{1}{2R-h} \int_{-R}^{R-h} h^2 dx = h^2 \\ T^{22}(h) = \frac{1}{2R-h} \int_{-R}^{R-h} (h^2 + 2hx)^2 dx = \frac{h^2}{3} (2R-h)^2 \\ T^{12}(h) = T^{21}(h) = 0 \end{cases}$$

(avec $h \geq 0$), et on a :

$$E[2 \gamma^*(h)] = 2h - \frac{h^2}{R} - \frac{h^2}{4R^3} (2R-h)^2$$

soit

$$(g) \quad E[\gamma^*(h)] = h - \frac{h^2}{R} + \frac{1}{2} \frac{h^3}{R^2} - \frac{1}{8} \frac{h^4}{R^3}$$

La pente à l'origine n'est pas altérée, mais le graphe de $E[\gamma^*(h)]$ est une sorte de parabole aplatie, symétrique autour de $h = 2R$ ($E[\gamma^*(R)] = \frac{3}{8}R$, $E[\gamma^*(2R)] = 0$)

Passons maintenant au krigeage universel, et limitons-nous pour abrégé au krigage ponctuel. Pour estimer $Z(z)$ en $z > R$, en l'absence de dérive, on doit prendre

$$Z_K(z) = Z(R)$$

et la variance de ce krigage serait $\sigma_K^2 = 2(z-R)$ (on peut vérifier directement que cet estimateur satisfait au système d'équations du krigage ordinaire; il est plus simple de remarquer que, dans le cas gaussien, le processus de Wiener-Lévy dont le variogramme est $\gamma(h) = h$ est à accroissements indépendants et stationnaires). La correction de dérive est ici

$$Z_D(z) = A_1 z + A_2 z^2 - A_1 R - A_2 R^2$$

Posant $h = z - R > 0$, on a donc

$$(h) \quad Z^*(z) = Z(R) + A_1 h + A_2 h(2R + h)$$

D'après le théorème d'additivité, la variance $\sigma_U^2 = D^2(Z(z) - Z^*(z))$ du krigage est

$$\sigma_U^2 = \sigma_K^2 + \sigma_D^2 = 2h + \mu_{11} h^2 + \mu_{22} h^2 (2R + h)^2$$

soit, d'après (f) :

$$(i) \quad \sigma_U^2 = 2h + \frac{4h^2}{R} + \frac{3h^3}{R^2} + \frac{3h^4}{4R^3}$$

On voit qu'elle augmente rapidement avec h . En présence d'une dérive, il n'est jamais possible d'extrapoler loin des données expérimentales : à supposer (ce qui

n'est pas toujours le cas) que la dérive reste égale à $\sum a_\ell f^\ell(x)$ avec les mêmes coefficients a_ℓ qu'au voisinage de S, l'estimation $\sum A_\ell f^\ell(x)$ devient vite assez mauvaise lorsqu'on s'éloigne de S.

Cas discret - Supposons que l'on connaisse la réalisation de $Z(x)$ aux $n+1$ points $x_i = ih$ ($i = 0, 1, \dots, n$) (nous ne plaçons plus l'origine des coordonnées au centre des données expérimentales, pour ne pas être obligé de distinguer les cas où n est pair ou impair). On posera $L = nh$. Les mesures v_1 et v_2 définies par :

$$\begin{cases} \int v_1(dx) f(x) = f(nh) - f(0) \\ \int v_2(dx) f(x) = \sum_{i=0}^n f(ih) - \frac{n+1}{2} [f(nh) - f(0)] \end{cases}$$

vérifient (pour $y = jh$) les relations :

$$\begin{cases} \int v_1(dx) |x - jh| = (n - 2j)h \\ \int v_2(dx) |x - jh| = -j(n-j)h \end{cases}$$

Pour estimer a_1 et a_2 , on prendra donc

$$A_1 = \int \lambda_1(dx) Z(x), \quad A_2 = \int \lambda_2(dx) Z(x)$$

avec des mesures λ_1 et λ_2 de la forme :

$$\begin{cases} \lambda_1 = A v_1 + B v_2 \\ \lambda_2 = C v_1 + D v_2 \end{cases}$$

La condition $\int x \lambda_2(dx) = 0$ donne $C = 0$, et $\int x^2 \lambda_2(dx) = 1$ donne

$$D = \frac{6}{n(1 - n^2) h^2}$$

Pour λ_1 , on trouve le système :

$$\begin{aligned} \int x \lambda_1(dx) &= A n h = 1 \\ \int x^2 \lambda_1(dx) &= h^2 [A n^2 + B \frac{n(1 - n^2)}{6}] = 0 \end{aligned}$$

D'où :
$$A = \frac{1}{n h} \quad , \quad B = \frac{6}{h(n^2-1)}$$

On en déduit les estimateurs cherchés, soit (avec $Z_1 = Z(ih)$) :

$$\begin{cases} A_1 = \frac{Z_n - Z_0}{n h} + \frac{6}{(n-1) h} \left(\bar{Z} - \frac{1}{2} (Z_n + Z_0) \right) \\ A_2 = - \frac{6}{h^2 n(n-1)} \left[\bar{Z} - \frac{1}{2} (Z_n + Z_0) \right] \end{cases}$$

(pour $n \rightarrow \infty$ à $nh = L$ fixé, ces estimateurs s'identifient aux estimateurs (e), compte tenu du fait que l'origine n'est plus placée au centre du segment L).

Pour obtenir la matrice des covariances μ_{ij} , on calcule les expressions :

$$\begin{aligned} - \int \lambda_1(dx) |x - jh| &= -1 + \frac{8 n^2 - 1}{n(n^2 - 1)} j - \frac{6 j^2}{n^2 - 1} \\ - \int \lambda_2(dx) |x - jh| &= - \frac{6}{n^2 - 1} \frac{j}{h} + \frac{6}{n(n^2 - 1)} \frac{j^2}{h} \end{aligned}$$

et on identifie les coefficients des termes en j et j^2 , d'où :

$$\begin{cases} \mu_{11} = \frac{8 n^2 - 1}{n(n^2 - 1)} \frac{1}{h} \\ \mu_{12} = \mu_{21} = - \frac{6}{n^2 - 1} \frac{1}{h^2} \\ \mu_{22} = \frac{6}{n(n^2 - 1)} \frac{1}{h^3} \end{cases}$$

Krigeage dans le cas discret - Dans les mêmes hypothèses que ci-dessus, proposons-nous d'estimer la valeur $Z((i+\epsilon)h)$ ($0 \leq \epsilon < 1$) de la V.R. au point $x = ih + \epsilon h$ compris entre x_i et x_{i+1} - En l'absence de dérive, l'estimateur optimal serait :

$$Z_K(x) = \epsilon Z_{i+1} + (1-\epsilon) Z_i$$

[comme ci-dessus, on peut le vérifier directement, ou déduire cette expression de la propriété markovienne du mouvement brownien]. Effectuons la correction de dérive, en utilisant les estimateurs A_1 et A_2 . On trouve :

$$Z^*(x) = \varepsilon [Z_{i+1} - A_1(i+1)h - A_2(i+1)^2 h^2] + \\ + (1-\varepsilon) [Z_i - A_1 i h - A_2 i^2 h^2] + A_1 (i+\varepsilon)h + A_2 (i+\varepsilon)^2 h^2$$

En développant, on constate que le terme en A_1 s'évanouit, et l'estimateur du kri-geage universel (en présence d'une dérive quadratique) se met sous la forme :

$$Z^*(x) = \varepsilon Z_{i+1} + (1-\varepsilon) Z_i - A_2 h^2 \varepsilon(1-\varepsilon)$$

Avec $\gamma(h) = |h|$, la variance σ_K^2 du krigage Z_K sans dérive est :

$$\sigma_K^2 = 2 \varepsilon(1-\varepsilon) h$$

La variance du terme correctif en présence de dérive est :

$$\sigma_D^2 = h^4 \varepsilon^2(1-\varepsilon)^2 D^2(A_2) = \frac{6 \varepsilon^2(1-\varepsilon)^2}{n(n^2-1)} h$$

Le théorème d'additivité donne ensuite la variance du krigage universel :

$$\sigma_U^2 = 2 \varepsilon(1-\varepsilon)h + \frac{6 \varepsilon^2(1-\varepsilon)^2}{n(n^2-1)} h$$

Si le variogramme est ωh , cette variance est évidemment multipliée par ω .

SECTION 3

ESTIMATION DES PARAMETRES ET CONTROLE DES HYPOTHESES3-1 DIFFICULTES DE CES PROBLEMES

Nous avons déjà souligné l'extrême importance pour les applications des problèmes que pose la vérification de l'hypothèse que l'on introduit en choisissant une expression mathématique particulière pour la covariance ou le variogramme, ainsi que la détermination des paramètres essentiels dont ces expressions dépendent. Nous allons nous heurter à deux sortes de difficultés, les unes techniques, les autres (peut-être beaucoup plus graves) méthodologiques et conceptuelles.

En premier lieu, l'inférence statistique est plus difficile pour les F.A. que pour les variables aléatoires indépendantes que manipule la statistique usuelle. On voit surgir de tous côtés des biais énormes dont notre intuition, formée à l'école de la statistique usuelle, ne nous faisaient même pas prévoir l'existence : on en a déjà rencontré un exemple dans la section 2-9, on en trouvera d'autres ci-après et en Annexe 4. Toutefois, il ne s'agit là que de difficultés techniques, que l'on peut toujours résoudre, en principe, si l'on dispose de la patience et des moyens de calculs nécessaires. C'est à certaines de ces difficultés que nous consacrons cette section 3.

Plus graves, peut-être, sont les difficultés méthodologiques, liées à une certaine ambiguïté de la notion de dérive. Il ne s'agit pas ici de difficultés mathématiques : dans le cadre des hypothèses faites dans les sections 1 et 2, les problèmes sont théoriquement bien posés et nous en avons construit la solution. Mais le contenu physique de ces hypothèses est parfois difficile à préciser.

Rappelons ces hypothèses : tout point x_0 admet un voisinage V sur lequel la dérive est bien représentée par une expression de la forme

$$m(x) = a_0 + \sum_{\ell=1}^k a_{\ell} f^{\ell}(x)$$

Sur V , la F.A. $Z(x) - m(x)$, d'espérance nulle (ou dont les accroissements ont une espérance nulle), admet une covariance $C(x,y)$ (ou un variogramme $\gamma(x,y)$) que l'on peut supposer connu (moyennant la résolution des problèmes techniques posés par l'inférence statistique). Enfin, l'ensemble $S \subset V$ des points expérimentaux contenus dans V est assez riche pour permettre le krigeage et l'estimation des a_ℓ .

Nous avons déjà souligné le caractère relatif de ces hypothèses, et combien elles sont liées au choix d'une échelle de travail. Selon que le voisinage V est pris plus ou moins grand, certains traits structuraux du phénomène seront pris en charge par la dérive ou par la covariance. En pratique, pourtant, le choix de V est le plus souvent à peu près imposé : on prend pour V le domaine le plus grand possible (pour qu'il contienne beaucoup de points expérimentaux) qui conduise à un variogramme expérimental des résidus conforme au $E[\gamma^*(h)]$ théorique du schéma que l'on a adopté. Si le schéma théorique comporte une covariance, et si la portée de cette covariance (distance à partir de laquelle les corrélations deviennent négligeables) se trouve être inférieure aux dimensions de V , l'estimation des a_ℓ , y compris celle de la constante a_0 , sera possible dans des conditions acceptables. Mais si cette portée dépasse les dimensions de V - ou bien si elle est infinie, c'est-à-dire s'il n'existe qu'un variogramme $\gamma(h) -$ l'estimateur A_0 de la constante a_0 prend une variance très grande, ou même infinie. Dans ces conditions, l'expression $A_0 + \sum_{\ell} A_\ell f^\ell(x)$ n'a plus qu'un rapport lointain, ou même pas de rapport du tout, avec la "vraie" dérive $m(x)$ (si tant est que, dans ce cas, il soit possible de parler de vraie dérive). Les variations locales (partie $\sum_{\ell=1}^k A_\ell f^\ell(x)$) de cette expression ont des chances d'être conformes aux variations locales de $m(x)$, mais le "niveau général" (A_0) représente alors plutôt une "moyenne mobile" des données disponibles dans V que le "vrai" niveau général a_0 de $m(x)$. A la limite où l'on a un variogramme, mais pas de covariance, ce niveau a_0 est non seulement indéterminable, mais même réellement indéterminé (puisque $m(x)$, comme la F.A. $Z(x)$ à accroissements d'ordre 2, n'est définie qu'à une constante près). Si l'on attribue à la dérive le sens simplement pragmatique d'une sorte de moyenne mobile (le niveau général est plus élevé en tel point qu'en tel autre) l'expression $A_0 + \sum_{\ell} A_\ell f^\ell(x)$ est peut-être utilisable (encore qu'il soit sans doute préférable de lui substituer une vraie moyenne mobile estimée par le krigeage). Mais si l'on espère déduire de la dérive estimée des informations sur la structure des mécanismes qui ont engendré le phénomène - on risque de ne recueillir que des artefacts. L'exemple du mouvement brownien (sans dérive) le montre clairement [4] :

ce processus ne représente rien de plus que l'effet cumulatif de petits mouvements aléatoires indépendants. Il possède un variogramme $\gamma(h) = |h|$, mais pas d'espérance a priori. Il ne comporte aucun mécanisme régulateur qui tendrait à faire revenir la particule au voisinage de sa position initiale, de sorte que les réalisations de ce processus montrent toujours de magnifiques exemples de pseudo-dérives parfaitement aléatoires. Les techniques d'estimation de la dérive (y compris nos estimateurs optimaux) lorsqu'on les applique à une telle réalisation réussissent, malheureusement, beaucoup trop bien : elles donnent de très belles dérivées, mais il ne s'agit que d'artefacts (puisque la "vraie" dérive est nulle par construction, et que les mécanismes aléatoires qui provoquent le mouvement brownien n'ont ni structure ni orientation).

On notera que ces difficultés ne concernent que la dérive elle-même. L'estimation d'une valeur ponctuelle, ou d'une moyenne mobile - c'est-à-dire le krigeage - est toujours possible dans des conditions à peu près acceptables. (Cela provient du fait que le krigeage universel intègre l'effet de la partie mobile $\sum_{\ell=1}^{\infty} a_{\ell} f'_{\ell}(x)$ de la dérive, mais ne dépend pas de la constante a_0).

Nous allons passer, maintenant, à la partie technique de cette section. Nous traiterons d'abord le cas le plus simple où le seul paramètre à estimer est un facteur ω (section 3-2), puis le problème plus difficile de l'estimation d'un paramètre essentiel (section 3-3). Nous terminerons par deux exemples : variogramme en $r^{\alpha-1}$ et covariance en e^{-br} .

3-2 ESTIMATION DU FACTEUR ω

L'hypothèse la plus simple que l'on puisse avoir à tester est celle d'un variogramme de la forme

$$(3-1) \quad \gamma(x,y) = \omega \gamma_0(r) \quad (r = |x-y|)$$

avec une fonction $\gamma_0(r)$ connue (par exemple, $\gamma_0(r) = r$) et un facteur ω à ajuster. On notera que (3-1) implique une hypothèse d'isotropie (du moins si le facteur ω n'est pas considéré comme une fonction du vecteur $x - y$) : plus exactement, (3-1) implique que les anisotropies éventuelles seront prises en charge par la dérive, et ne se répercutent pas sur le variogramme. Il est facile de tester cette hypothèse. On a vu

que les estimateurs optimaux de la dérive (système 2-D), dans une telle hypothèse, ne dépendent que de $\gamma_0(r)$ et non du facteur ω . On peut donc les former, construire le variogramme expérimental des résidus et le comparer à l'expression théorique $\omega E[\gamma_0^*(h)]$, formule (2-23). Cette méthode permet à la fois de tester l'hypothèse et d'estimer le facteur ω .

En pratique, on effectuera non pas une seule, mais plusieurs déterminations du variogramme expérimental dans plusieurs domaines $V_1, V_2 \dots$ égaux et disjoints, et c'est la moyenne de ces variogrammes expérimentaux que l'on comparera à $E[\gamma_0^*(h)]$: on en déduira une confirmation de l'hypothèse (3-1), et une estimation de la valeur moyenne de ω .

Si les points expérimentaux sont alignés sur des courbes à peu près rectilignes, on prendra sur ces courbes un certain nombre de segments de même longueur $L = 2R$ et de même direction θ . Pour chacun de ces segments, on effectuera la correction de dérive et on construira le variogramme des résidus. On prendra la moyenne $\gamma^*(h)$ de ces variogrammes expérimentaux et on la comparera à $\omega E[\gamma_0^*(h)]$ (par exemple, pour $\gamma_0^*(h) = r$, à $\omega[h - \frac{2h^2}{L} + \frac{2h^3}{L^2} - \frac{h^4}{L^4}]$, d'après 2-10). Il faudra vérifier que le coefficient ω ainsi ajusté ne varie pas trop avec la direction θ des lignes. Enfin, on exécutera cette opération pour diverses valeurs croissantes de L : la valeur L_0 au delà de laquelle l'accord cesse d'être bon fixera les dimensions du voisinage V sur lequel il convient de travailler.

Cette méthode ne permet d'estimer qu'un facteur ω moyen. Si l'on désire estimer la valeur locale $\omega(x_0)$ au voisinage $V(x_0)$ d'un point x_0 , on pourra utiliser la variance expérimentale des résidus : on estimera la dérive A_ℓ à partir des données disponibles (ensemble $S \subset V(x_0)$) et on calculera :

$$s^2 = \frac{1}{S} \int_S [Z(x) - A_\ell f^\ell(x) - m]^2 dx$$

$$(S, \text{ mesure de l'ensemble } S, m = \frac{1}{S} \int_S [Z(x) - A_\ell f^\ell(x)] dx$$

Par un calcul analogue à celui de la section 2-9, on montre que cette variance expérimentale a pour espérance :

$$E(s^2) = \frac{1}{S^2} \int K(h) E[\gamma^*(h)] dh = \sigma^2 - \frac{1}{2} \mu_s \frac{1}{S^2} \int K(h) T^{\ell_s}(h)$$

$(K(h) = \int k(x) k(x+h) dx, \text{ si } k(x) \text{ est l'indicatrice de } S; \sigma^2 \text{ est la variance des vrais résidus } Z(x) - a_{\ell} f^{\ell}(x) : \text{ on voit que } s^2 \text{ est un estimateur biaisé de } \sigma^2).$
On en déduit l'estimateur :

$$\omega^* = \frac{s^2}{\frac{1}{S^2} \int K(h) E[\gamma^*(h)] dh}$$

On peut aussi utiliser des méthodes plus élaborées, (par exemple, maximum de vraisemblance), mais nous n'insisterons pas sur ce point.

3-3 ESTIMATION DE PARAMETRES ESSENTIELS

Traitions maintenant le problème de l'estimation d'un paramètre essentiel (par exemple, paramètre α d'un variogramme $\propto r^{\alpha-1}$, paramètre a d'une covariance $\propto e^{-ah}$). Les opérations doivent se dérouler en deux étapes.

1/ On choisit un estimateur sans biais (qui ne sera pas évidemment pas optimal en général) de la dérive, en prenant par exemple :

$$(3-2) \quad A_{\ell} = \int \lambda_{\ell}(dx) Z(x) \quad (\ell = 1, 2, \dots, k)$$

avec des mesures λ_{ℓ} à support dans S vérifiant :

$$(3-3) \quad \begin{cases} \int \lambda_{\ell}(dx) = 0 \\ \int \lambda_{\ell}(dx) f^S(x) = \delta_{\ell}^S \end{cases}$$

On note qu'il n'est pas nécessaire de former un estimateur A_0 de la constante a_0 (puisque l'on travaillera sur des variogrammes), de sorte que la méthode est à l'abri des difficultés signalées en 3-1. Avec ces estimateurs A_{ℓ} de la dérive, on construit la variogramme des résidus $\gamma^*(h)$ dans plusieurs zones $V_1, V_2, \dots$, on en déduit un variogramme moyen et on le compare à son espérance $\propto E[\gamma_{\alpha}^*(h)]$ calculée dans l'hypothèse où le vrai variogramme est $\propto \gamma_{\alpha}(h)$ (α est le paramètre à estimer). Pour ramener l'effet du facteur inconnu $\propto$ à une translation verticale, on construira les courbes $E[\gamma_{\alpha}^*(h)]$ en portant en ordonnée une échelle logarithmique : une comparaison directe de la courbe expérimentale avec le réseau des $E[\gamma_{\alpha}^*(h)]$ construites pour diverses

valeurs de α permettra de choisir α et ω .

2/ Ayant choisi α , on recommence l'opération en prenant cette fois pour A_ℓ l'estimateur optimal de la dérive (optimal pour $\gamma_\alpha(h)$), et on doit vérifier que le nouveau variogramme expérimental des résidus est conforme à son espérance (qui est cette fois donnée par la formule (2-23)).

La mise en oeuvre de la première partie de cette méthode suppose que l'on soit capable de calculer, pour un variogramme $\gamma(h)$ donné, l'espérance du variogramme expérimental des résidus obtenus à partir de l'estimation (3-2) non optimale de la dérive. C'est ce problème théorique que nous allons traiter.

Désignons par $k(x)$ l'indicatrice de S , par $K(h) = \int k(x) k(x+h) dx$ la mesure de S et de son translaté par h . Le variogramme expérimental des résidus est :

$$2 \gamma^*(h) = \frac{1}{K(h)} \int k(x) k(x+h) [Z(x+h) - Z(x) - A_\ell f^\ell(x+h) + A_\ell f^\ell(x)]^2 dx$$

Les A_ℓ constituent l'estimateur sans biais (bien que non optimal) défini en (3-2) - (3-3). Nous pouvons donc calculer l'espérance de $2 \gamma^*(h)$ en supposant $a_\ell = E(A_\ell) = 0$. Nous voyons apparaître trois termes :

$$E[2 \gamma^*(h)] = T_0 - 2 T_1 + T_2$$

Le premier est manifestement

$$T_0 = 2 \gamma(h)$$

Les deux autres représentent le biais. Ils se mettent sous la forme :

$$(3-4) \quad \begin{cases} T_1 = \frac{1}{K(h)} \int k(x) k(x+h) T_1(x;h) dx \\ T_1(x;h) = [f^\ell(x+h) - f^\ell(x)] E[A_\ell (Z(x+h) - Z(x))] \end{cases}$$

$$(3-5) \quad \begin{cases} T_2 = \mu_{\ell s} T_2^{\ell s}(h) \\ T_2^{\ell s}(h) = \frac{1}{K(h)} \int k(x) k(x+h) [f^\ell(x+h) - f^\ell(x)][f^s(x+h) - f^s(x)] dx \\ \mu_{\ell s} = \text{Cov}(A_\ell, A_s) \end{cases}$$

Le terme T_2 est identique à celui qui figure en (2-23), à ceci près que les $\mu_{\ell s}$ représentent les covariances des estimateurs non optimaux (3-2). On a évidemment :

$$(3-6) \quad \mu_{\ell s} = - \iint \gamma(x-y) \lambda_{\ell}(dx) \lambda_s(dy)$$

et le calcul de T_2 ne pose pas de difficultés particulières. Comme A_{ℓ} n'est pas optimal, la simplification $T_1 = T_2$ ne se produit pas, et il faut calculer à part le terme T_1 . De :

$$A_{\ell} [Z(x+h) - Z(x)] = (Z(x+h) - Z(x)) \int \lambda_{\ell}(dy) Z(y)$$

on déduit d'abord, en passant aux espérances :

$$(3-7) \quad E[A_{\ell}(Z(x+h) - Z(x))] = \int \lambda_{\ell}(dy) [\gamma(x-y) - \gamma(x+h-y)]$$

En portant cette expression dans (3-4), on en déduit ensuite $T_1(x;h)$ puis T_1 .

Cas où S est un segment de droite $(-R, +R)$ de longueur $2R$ - Le cas où les données expérimentales s'alignent sur des courbes à peu près rectilignes est particulièrement favorable à l'application de la méthode ci-dessus. En procédant comme indiqué en 3-2, on peut même mettre en évidence une anisotropie éventuelle du facteur ω . Pour obtenir un estimateur sans biais (3-2) de la dérive à partir des données disponibles sur un segment $(-R, +R)$, le plus simple est d'utiliser l'estimateur obtenu en 2-10 (qui est de plus optimal si par hasard le variogramme est linéaire).

Pour une dérive purement linéaire ($a_1 x$), on prendra donc :

$$(3-8) \quad A_1 = \frac{1}{2R} [Z(R) - Z(-R)]$$

Pour une dérive quadratique ($a_1 x + a_2 x^2$), on prendra :

$$(3-9) \quad \begin{cases} A_1 = \frac{1}{2R} [Z(R) - Z(-R)] \\ A_2 = \frac{3}{4R^2} [Z(R) + Z(-R)] - \frac{3}{2R^2} \bar{Z} \end{cases}$$

c'est-à-dire :

$$(3-10) \quad \begin{cases} \lambda_1(dx) = \frac{\delta_R - \delta_{-R}}{2R} \\ \lambda_2(dx) = \frac{3}{4R^2} [\delta_R - \delta_{-R}] - \frac{3}{4R^3} dx \quad (-R \leq x \leq R) \end{cases}$$

Explicitons le calcul dans ces hypothèses :

Calcul de T_2 . De (3-6), on tire sans peine :

$$(3-11) \quad \begin{cases} \mu_{11} = D^2(A_1) = \frac{1}{2R^2} \gamma(2R) \\ \mu_{12} = \mu_{21} = \text{Cov}(A_1, A_2) = 0 \\ \mu_{22} = D^2(A_2) = \frac{9}{8R^4} [4 \chi(2R) - 2 F(2R) - \gamma(2R)] \end{cases}$$

avec les fonctions auxiliaires χ et F définies comme suit :

$$\begin{cases} \chi(h) = \frac{1}{h} \int_0^h \gamma(x) dx \\ F(h) = \frac{2}{h^2} \int_0^h x\chi(x) dx = \frac{2}{h^2} \int_0^h (h-x) \gamma(x) dx \end{cases}$$

Pour une dérive purement linéaire, on trouve $T_2 = \frac{h^2}{2R^2} \gamma(2R)$. Pour une dérive quadratique, avec les $T^{\text{ls}}(h)$ calculés en 2-10, on trouve :

$$(3-12) \quad T_2 = \frac{h^2}{2R^2} \gamma(2R) + \frac{3}{8R^4} h^2(2R-h)^2 [4 \chi(2R) - 2 F(2R) - \gamma(2R)]$$

Passons au calcul de T_1 - De (3-7), on tire :

$$\begin{cases} E_1(x, h) = E[A_1(Z(x+h) - Z(x))] = \frac{1}{2R} [\gamma(R-x) - \gamma(R+x) - \gamma(R-x-h) + \gamma(R+x+h)] \\ E_2(x, h) = E[A_2(Z(x+h) - Z(x))] = \frac{3}{4R^2} [\gamma(R-x) + \gamma(R+x) - \gamma(R-x-h) - \gamma(R+x+h)] \\ \quad - \frac{3}{4R^2} \int_{-R}^R [\gamma(x-y) - \gamma(x+h-y)] dy \end{cases}$$

D'après (3-4), on a ensuite ($h > 0$) :

$$T_1 = \frac{1}{2R-h} \int_{-R}^{R-h} [h E_1(x, h) + h(h+2x) E_2(x, h)] dx$$

Après quelques manipulations un peu longues mais sans difficultés particulières on trouve (dans le cas de la dérive quadratique) :

$$(3-13) \quad \left\{ \begin{aligned} T_1 &= \frac{1}{2R-h} \int_h^{2R} \left[4 \frac{h}{R} - \frac{3}{2} \frac{h^2}{R^2} + \frac{3}{2} \frac{h^2}{R^3} z - \frac{3h}{2R^3} z^2 \right] \gamma(z) dz \\ &- \frac{1}{2R-h} \int_0^{2R-h} \left[4 \frac{h}{R} - \frac{3}{2} \frac{h^2}{R^2} - \frac{3}{2} \frac{h^2}{R^3} z - \frac{3h}{2R^3} z^2 \right] \gamma(z) dz \end{aligned} \right.$$

Dans le cas d'une dérive simplement linéaire, on aurait :

$$T_1 = \frac{1}{2R-h} \frac{h}{R} \left[\int_h^{2R} \gamma(z) dz - \int_0^{2R-h} \gamma(z) dz \right]$$

Connaissant T_1 - formule (3-13), et T_2 - formule (3-12), on en déduit, dans le cas de la dérive quadratique :

$$(3-14) \quad 2 E[\gamma^*(h)] = 2 \gamma(h) - 2 T_1 + T_2$$

Nous allons appliquer ces résultats à deux cas particuliers.

3-4 EXEMPLE D'UN VARIOGRAMME EN $r^{\alpha-1}$

Nous supposons les données $Z(x)$ disponibles sur l'intervalle $(-R, +R)$, et nous allons traiter le cas d'un variogramme $\gamma(h) = \omega|h|^{\alpha-1}$ avec $0 < \alpha < 2$. Le cas $\alpha = 2$ correspond au variogramme linéaire étudié en 2-10. Pour $\alpha = 1$, un passage à la limite sur $\frac{h^{\alpha-1}-1}{\alpha-1}$ conduit au variogramme de Wijsien $\gamma(h) = \log h$. Pour $\alpha = 0$, enfin, un passage à la limite sur $\frac{r^{\alpha-1}}{\Gamma(\frac{\alpha}{2})}$ donne la mesure de Dirac (bruit de fond, ou effet de pépite pur) : dans ce dernier cas, on doit retrouver les estimateurs des moindres carrés (cf. Annexe 3). Pour $\alpha < 1$, le facteur ω doit être pris négatif. On a alors une covariance du type $C(h) = -\gamma(h) = |h|^{\alpha-1}$ qui est une mesure et non plus une fonction : le système représente alors une mesure aléatoire, et non plus une F.A., ce qui n'a d'ailleurs aucune espèce d'inconvénient (il suffit de régulariser cette mesure aléatoire par une fonction continue de support très petit pour retomber sur une F.A.).

Formule de Landkoff - Pour $|y| < R$, l'intégrale

$$(3-15) \quad \int_{-R}^R \frac{|x-y|^{\alpha-1}}{(R^2-x^2)^{\frac{\alpha}{2}}} dx = \frac{\pi}{\sin \frac{\pi \alpha}{2}} \quad (0 < \alpha < 2)$$

est constante et indépendante de y . On démontre cette formule [2], en effectuant une inversion de centre y . Cette formule de Landkoff montre que la mesure de densité $(R^2 - x^2)^{-\frac{\alpha}{2}}$ sur $(-R, +R)$ réalise le potentiel d'équilibre pour le noyau $|x-y|^{\alpha-1}$ et l'intervalle $(-R, +R)$. Par des calculs analogues, on montre que les intégrales :

$$I_{2n}(y) = \int_{-R}^R \frac{|x-y|^{\alpha-1}}{(R^2-x^2)^{\frac{\alpha}{2}-n}} dx$$

$$I_{2n-1}(y) = \int_{-R}^R \frac{x|x-y|^{\alpha-1}}{(R^2-x^2)^{1+\frac{\alpha}{2}-n}} dx$$

se réduisent respectivement, sur $(-R, +R)$ à un polynôme pair de degré $2n$ en y et à un polynôme impair de degré $2n-1$. Comme nous n'envisageons de traiter qu'une dérive quadratique, nous n'explicitons que $I_2(y)$ et $I_1(y)$: pour $-R \leq y \leq R$, on a :

$$(3-16) \quad \begin{cases} \int_{-R}^R x(R^2-x^2)^{-\frac{\alpha}{2}} |x-y|^{\alpha-1} dx = -\frac{(\alpha-1)\pi}{\sin \frac{\alpha\pi}{2}} y \\ \int_{-R}^R (R^2-x^2)^{1-\frac{\alpha}{2}} |x-y|^{\alpha-1} dx = \frac{\pi(1-\frac{\alpha}{2})}{\sin \frac{\alpha\pi}{2}} (R^2 + (\alpha-1)y^2) \end{cases}$$

Ces formules (3-15) et (3-16) vont nous permettre de former l'estimateur optimal de la dérive. Nous pourrions même (mais cela sera inutile ici) résoudre les équations du krigeage universel à partir d'une autre formule de Landkoff :

$$\int_{-R}^R \frac{(z^2-R^2)^{\frac{\alpha}{2}}}{(R^2-y^2)^{\frac{\alpha}{2}} |z-y|} |x-y|^{\alpha-1} dy = \frac{\pi}{\sin \frac{\pi\alpha}{2}} |x-z|^{\alpha-1} \quad (z > R)$$

qui montre que la mesure de densité $(z^2-R^2)^{\alpha/2} (R^2-y^2)^{-\frac{\alpha}{2}} |y-z|^{-1}$ résoud le problème du balayage pour l'intervalle $(-R, R)$ et le noyau $|x-y|^{\alpha-1}$.

Estimation optimale d'une dérive quadratique

Les mesures λ_1, λ_2 vérifiant la première équation (2-D) :

$$(3-17) \quad - \int_{-R}^R \lambda_i(dy) |x-y|^{\alpha-1} dy = \sum_{\ell=1}^2 \mu_{i\ell} x^\ell + C_i \quad (i = 1, 2)$$

sont, d'après (3-15) et (3-16), des combinaisons linéaires des mesures de densités :

$$(R^2 - x^2)^{-\frac{\alpha}{2}}, \quad x(R^2 - x^2)^{-\frac{\alpha}{2}}, \quad (R^2 - x^2)^{1-\frac{\alpha}{2}}$$

Les conditions habituelles :

$$\begin{cases} \int \lambda_i(dx) = 0 \\ \int x^l \lambda_i(dx) = \delta_i^l \end{cases}$$

permettent de déterminer les coefficients de ces combinaisons linéaires, et l'on obtient ainsi les estimations optimales :

$$(3-18) \quad \begin{cases} A_1 = B \int_{-R}^R x(R^2 - x^2)^{-\frac{\alpha}{2}} Z(x) dx \\ A_2 = D \int_{-R}^R (R^2 - x^2)^{-\frac{\alpha}{2}} Z(x) dx + F \int_{-R}^R (R^2 - x^2)^{1-\frac{\alpha}{2}} Z(x) dx \end{cases}$$

avec des coefficients B, D, F donnés par :

$$\begin{cases} B = \frac{2}{\sqrt{\pi}} \frac{\Gamma(1 + \frac{3-\alpha}{2})}{\Gamma(1 - \frac{\alpha}{2})} R^{\alpha-3} \\ D = \frac{2}{\sqrt{\pi}} \frac{\Gamma(2 + \frac{3-\alpha}{2})}{\Gamma(1 - \frac{\alpha}{2})} R^{\alpha-3} \\ F = -\frac{3-\alpha}{\sqrt{\pi}} \frac{\Gamma(2 + \frac{3-\alpha}{2})}{\Gamma(2 - \frac{\alpha}{2})} R^{\alpha-5} \end{cases}$$

($\Gamma(x)$ désigne la fonction eulérienne habituelle).

Les covariances $\mu_{ij} = \text{Cov}(A_i, A_j)$ s'obtiennent facilement en portant les expressions de ces mesures λ_1 et λ_2 dans le premier membre de (3-17) et en identifiant les coefficients de x et de x^2 . Tous calculs faits, on trouve :

$$(3-19) \quad \begin{cases} \mu_{11} = \frac{2(\alpha-1)}{\sqrt{\pi}} \Gamma(\frac{\alpha}{2}) \Gamma(1 + \frac{3-\alpha}{2}) R^{\alpha-3} \\ \mu_{12} = \mu_{21} = 0 \\ \mu_{22} = \frac{(\alpha-1)(3-\alpha)}{\sqrt{\pi}} \Gamma(\frac{\alpha}{2}) \Gamma(2 + \frac{3-\alpha}{2}) R^{\alpha-5} \end{cases}$$

La Table I donne les valeurs numériques de B , D , F , μ_{11} et μ_{22} en fonction de α . Pour $\alpha < 1$, ces formules donnent des valeurs négatives à μ_{11} et μ_{22} , mais le facteur ω du variogramme $\omega r^{\alpha-1}$ est négatif pour $\alpha < 1$, de sorte que les variances $D^2(A_1)$ et $D^2(A_2)$ sont positives.

Pour $\alpha = 2$, on peut vérifier que (3-18) et (3-19) redonnent les résultats déjà obtenus en 2-10 (variogramme linéaire). Pour $A = 1$ (passage à la limite $\frac{r^{\alpha-1}-1}{\alpha-1} \rightarrow \log r$) on trouve les formules du cas de Wijsien $\gamma(h) = \log h$. On peut faire directement $\alpha = 1$ dans (3-18). Pour les covariances (3-19), il faut au préalable effectuer la division par $\alpha-1$. Il vient :

$$\begin{cases} A_1 = \frac{2}{\pi R^2} \int_{-R}^R \frac{x Z(x)}{\sqrt{R^2-x^2}} dx \\ A_2 = \frac{4}{\pi R^4} \int_{-R}^R \frac{(2x^2-R^2)}{\sqrt{R^2-x^2}} Z(x) dx \end{cases}$$

$$\mu_{11} = \frac{2}{R^2}, \quad \mu_{12} = \mu_{21} = 0, \quad \mu_{22} = \frac{4}{R^4}$$

Cas $\alpha = 0$: effet de pépite, ou bruit de fond - Pour $\alpha \rightarrow 0$, $\frac{r^{\alpha-1}}{\Gamma(\frac{\alpha}{2})}$ tend vers $-\delta$

(mesure de Dirac). Les estimations (3-18) coïncident dans ce cas avec celles de la méthode des moindres carrés (cf. Annexe 3). Pour les covariances, il convient de diviser (3-19) par $\Gamma(\frac{\alpha}{2})$ avant de faire $\alpha = 0$. Nous examinerons leur signification dans l'annexe 3. On trouve ainsi :

$$\begin{cases} A_1 = \frac{3}{2R^3} \int_{-R}^R x Z(x) dx \\ A_2 = -\frac{15}{8R^3} \int_{-R}^R Z(x) dx + \frac{45}{8R^5} \int_{-R}^R x^2 Z(x) dx \end{cases}$$

$$\mu_{11} = \frac{3}{2R^3}, \quad \mu_{12} = \mu_{21} = 0, \quad \mu_{22} = \frac{45}{8R^5}$$

Le variogramme des résidus (pour les estimations optimales ci-dessus) admet une espérance $E[\gamma^*(h)]$ que l'on déduit facilement de (2-23) et des résultats ci-dessus. On trouve :

$$(3-20) \quad E[\gamma^*(h)] = h^{\alpha-1} - R^{\alpha-1} \frac{(\alpha-1)}{2\sqrt{\pi}} \Gamma\left(\frac{\alpha}{2}\right) \Gamma\left(1+\frac{3-\alpha}{2}\right) \left[2 \frac{h^2}{R^2} + \left(\frac{1-\alpha}{3}\right) \left(1 + \frac{3-\alpha}{2}\right) \frac{h^2}{R^2} \left(2 - \frac{h}{R}\right)^2 \right]$$

Cette formule a été tabulée pour diverses valeurs de α (Abaque I. N.B./ce n'est pas cet abaque que l'on doit utiliser pour estimer α dans la phase 1 de la section 3-3; on peut l'utiliser dans la phase 2, lorsqu' α est déjà déterminé, pour tester le variogramme des résidus optimaux)

Estimation du paramètre α - Lorsque le vrai variogramme est $r^{\alpha-1}$ et que l'on utilise les estimateurs (sans biais, mais non optimaux) écrits en (3-9), le demi-variogramme des résidus expérimentaux a pour espérance :

$$(3-21) \quad E[\gamma^*(h)] = h^{\alpha-1} - T_1 + \frac{1}{2} T_2$$

On calcule T_1 et T_2 par les formules (3-13) et (3-12). Cela donne :

$$(3-22) \quad \left| \begin{aligned} T_1 &= \frac{1}{\alpha} \left(4 \frac{h}{R} - \frac{3}{2} \frac{h^2}{R^2} \right) \frac{(2R)^{\alpha-h^{\alpha}}}{2R-h} + \frac{3}{2} \frac{h^2}{R^3} \frac{(2R)^{1+\alpha-h^{1+\alpha}}}{(2R-h)(1+\alpha)} - \frac{3h}{2R^3} \frac{(2R)^{2+\alpha-h^{2+\alpha}}}{(2+\alpha)(2R-h)} \\ &- \frac{1}{\alpha} \left(4 \frac{h}{R} - \frac{3}{2} \frac{h^2}{R^2} \right) (2R-h)^{\alpha-1} + \frac{3}{2} \frac{h^2}{R^3} \frac{(2R-h)^{1+\alpha}}{1+\alpha} + \frac{3h}{R^3} \frac{(2R-h)^{\alpha+2}}{\alpha+2} \\ T_2 &= (2R)^{\alpha-1} \left[\frac{h^2}{2R^2} + \frac{3(3-\alpha)}{8(1+\alpha)} \frac{h^2(2R-h)^2}{R^4} \right] \end{aligned} \right|$$

La tabulation de ces formules (3-21) et (3-22) permet, dans les applications, d'estimer les paramètres α et ω [voir abaque II]

3-5 EXEMPLE D'UNE COVARIANCE EN e^{-ah}

Dans les mêmes conditions que ci-dessus (données disponibles sur l'intervalle $(-R, +R)$), traitons le cas d'une dérive quadratique et d'une covariance $\omega e^{-a|h|}$ - c'est-à-dire d'un variogramme

$$\gamma(h) = \omega [1 - e^{-a|h|}]$$

Pour a très petit (et ω a fini), on retombe sur le cas du variogramme linéaire de la section 2-10. Pour a très grand, au contraire, on retombera sur l'effet de pépité déjà rencontré en 3-4. Nous ferons les calculs avec $\omega = 1$.

Les relations suivantes, valables pour $-R \leq y \leq R$:

$$\begin{cases} \frac{1}{2} \int [\delta_R(dx) + \delta_{-R}(dx)] e^{-a|x-y|} = e^{-aR} \text{chay} \\ \int_{-R}^R e^{-a|x-y|} dx = \frac{2}{a} - \frac{2}{a} e^{-aR} \text{chay} \end{cases}$$

montrent que la mesure v_0 à support dans $[-R, +R]$ définie comme suit :

$$(3-23) \quad v_0(dx) = \frac{a}{2} dx + \frac{1}{2} (\delta_R + \delta_{-R})$$

vérifie l'équation intégrale :

$$(3-24) \quad \int v_0(dx) e^{-a|x-y|} = 1 \quad (-R \leq y \leq R)$$

Cette mesure v_0 réalise donc le potentiel d'équilibre sur $(-R, +R)$ pour le noyau exponentiel $e^{-a|x-y|}$.

Estimation de l'espérance d'une F.A. stationnaire dont la covariance est $C(h) = e^{-|h|}$:

comme $\int v_0(dx) = 1 + aR$, la formule (3-23) montre que l'estimateur optimal $m^* = A_0$ de l'espérance $m = a_0$ de cette F.A. (connaissant la réalisation sur $(-R, +R)$) est :

$$\begin{cases} m^* = \frac{aR}{1+aR} \bar{Z} + \frac{1}{1+aR} \frac{Z(R) + Z(-R)}{2} \\ D^2(m^*) = \frac{1}{1+aR} \end{cases}$$

($\bar{Z} = \frac{1}{2R} \int_{-R}^R Z(x) dx$ est la "moyenne empirique"). Pour cette même F.A. stationnaire (d'espérance inconnue) le meilleur estimateur de $Z(z)$, $z > R$ est la solution des équations du krigeage universel dans le cas où il n'y a qu'une fonction de dérive, soit $f^0(x) = 1$. En l'absence de dérive (à $m = 0$ connue), on aurait $Z_K(z) = e^{-a(z-R)} Z(R)$, à cause de la propriété markovienne de la covariance exponentielle, et $\sigma_K^2 = 1 - e^{-2a(z-R)}$. En effectuant la correction de dérive, on voit que l'estima-

teur du krigeage universel est :

$$Z^*(z) = e^{-a(z-R)} Z(R) + [1 - e^{-a(z-R)}] m^*$$

et le théorème d'additivité permet de calculer la variance :

$$D^2[Z^*(z) - Z(z)] = 1 - e^{-2a(z-R)} + [1 - e^{-a(z-R)}] D^2(m^*)$$

Estimation optimale d'une dérive quadratique.

On peut résoudre les équations (2-D) dans le cas où la dérive est un polynôme de degré n quelconque. Nous nous limitons ici au cas d'une dérive parabolique $a_0 + a_1 x + a_2 x^2$. Les mesures v_1 et v_2 concentrées sur $[-R, +R]$ et définies par :

$$(3-25) \quad \begin{cases} v_1(dx) = \frac{a}{2} x dx + \frac{1+aR}{2a} (\delta_R - \delta_{-R}) \\ v_2(dx) = \frac{a}{2} x^2 dx + \frac{1}{a^2} (1 + aR + \frac{1}{2} a^2 R^2) (\delta_R + \delta_{-R}) \end{cases}$$

vérifient, pour $-R \leq y \leq R$, les relations :

$$(3-26) \quad \begin{cases} \int_{-R}^R v_1(dx) e^{-a|x-y|} = y \\ \int_{-R}^R v_2(dx) e^{-a|x-y|} = y^2 + \frac{2}{a^2} \end{cases}$$

Les estimateurs $A_i = \int_{-R}^R \lambda_i(dx) Z(x)$ sont optimaux si les mesures λ_i ($i = 0, 1, 2$) vérifient le système habituel :

$$(3-27) \quad \begin{cases} \int_{-R}^R \lambda_i(dx) e^{-a|x-y|} = \sum_{\ell=0}^2 \mu_{i\ell} y^\ell & (-R \leq y \leq R) \\ \int_{-R}^R \lambda_i(dx) x^\ell = \delta_i^\ell & (i, \ell = 0, 1, 2) \end{cases}$$

(3-24) et (3-26) montrent que les solutions de ce système sont des combinaisons linéaires des trois mesures v_0, v_1, v_2 - Les secondes relations (3-27) permettent de déterminer les coefficients de ces combinaisons linéaires. Enfin, le calcul explicite

du premier membre de la première relation (3-27) permet ensuite de déterminer les covariances μ_{ij} en identifiant les termes de même degré en y . Ces calculs sont relativement longs, mais sans difficulté aucune. Nous n'en reproduisons que le résultat.

Les estimateurs optimaux de la dérive sont de la forme :

$$(3-28) \quad \begin{cases} A_0 = C \bar{Z} + D \frac{1}{2R} \int_{-R}^R x^2 Z(x) dx + D' \frac{Z(R) + Z(-R)}{2} \\ A_1 = B \frac{1}{2R} \int_{-R}^R x Z(x) dx + B' \frac{Z(R) - Z(-R)}{2} \\ A_2 = E \bar{Z} + F \frac{1}{2R} \int_{-R}^R x^2 Z(x) dx + F' \frac{Z(R) + Z(-R)}{2} \end{cases}$$

avec des coefficients $C, D, \dots, F'$ donnés par :

$$\begin{aligned} C &= \frac{3}{2} \frac{1 + aR + \frac{1}{2} a^2 R^2 + \frac{1}{10} a^3 R^3}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\ D &= -\frac{3}{4} a^2 \frac{1 + \frac{1}{3} aR}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\ D' &= -\frac{1}{2} \frac{1 + aR + \frac{1}{5} a^2 R^2}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\ B &= \frac{a^2}{1 + aR + \frac{1}{3} a^2 R^2} \\ B' &= \frac{1 + aR}{R(1 + aR + \frac{1}{3} a^2 R^2)} \\ E &= -\frac{3}{2R^2} \frac{1 + aR + \frac{1}{2} a^2 R^2 + \frac{1}{6} a^3 R^3}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\ F &= \frac{3a^2}{4R^2} \frac{1 + aR}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\ F' &= \frac{3}{2R^2} \frac{1 + aR + \frac{1}{3} a^2 R^2}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \end{aligned}$$

Les covariances $\mu_{ij} = \text{Cov}(A_i, A_j)$, de leur côté, sont :

$$\begin{aligned}
 (3-29) \quad \mu_{00} &= \frac{1 + \frac{3}{4} aR + \frac{3}{20} a^2 R^2}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\
 \mu_{01} &= \mu_{10} = \mu_{12} = \mu_{21} = 0 \\
 \mu_{20} &= \mu_{02} = -\frac{3a}{4R} \frac{1 + \frac{aR}{3}}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3} \\
 \mu_{11} &= \frac{a}{R(1 + aR + \frac{a^2 R^2}{3})} \\
 \mu_{22} &= \frac{3a}{4R^3} \frac{1 + aR}{1 + aR + \frac{2}{5} a^2 R^2 + \frac{1}{15} a^3 R^3}
 \end{aligned}$$

Pour a très petit, (3-28) redonne les estimateurs de la section 2-10, et, après division par a , (3-29) donne les covariances correspondantes (sauf μ_{00} qui devient infini, conformément à la règle générale que nous avons déjà rencontrée. Mais l'estimateur A_0 , bien que sa variance soit infinie, reste parfaitement défini : on trouve

$$A_0 = \frac{3}{2} \bar{Z} - \frac{1}{4} [Z(R) + Z(-R)]$$

Pour a très grand, au contraire, on retombe sur les estimateurs des moindres carrés déjà rencontrés en 3-4. (voir aussi Annexe 3)

Le variogramme des résidus (optimaux) - Avec les valeurs μ_{11} et μ_{22} calculées ci-dessus, la formule (2-23) montre que le variogramme des résidus ($Z(x)$ corrigé de l'estimation optimale de la dérive) a pour espérance :

$$(3-30) \quad E[\gamma^*(h)] = 1 - e^{-ah} - \frac{1}{2} [\mu_{11} h^2 + \mu_{22} \frac{h^2}{3} (2R-h)^2]$$

(formule représentée par l'abaque [1]).

Estimation du paramètre a - En vue de l'estimation de a , donnons l'espérance du variogramme des résidus non optimaux, calculés à partir de l'estimation non optimale (3-9). On trouve :

$$(3-31) \quad E[\gamma^*(h)] = 1 - e^{-ah} - T_1 + \frac{1}{2} T_2$$

avec des termes T_1 et T_2 qui se déduisent de (3-13) et (3-12) :

$$T_1 = -\frac{1}{2R-h} \left\{ \left(4 \frac{h}{R} - \frac{3}{2} \frac{h^2}{R^2} \right) \frac{e^{-ah} - e^{-2aR}}{a} + \frac{3}{2} \frac{h^2}{a^2 R^3} [(1+ah) e^{-ah} - (1+2aR) e^{-2aR}] \right. \\ \left. - \frac{3h}{a^3 R^3} \left[\left(1+ah + \frac{a^2 h^2}{2} \right) e^{-ah} - (1+2aR + 2a^2 R^2) e^{-2aR} \right] \right\} \\ + \frac{1}{2R-h} \left\{ \left(4 \frac{h}{R} - \frac{3}{2} \frac{h^2}{R^2} \right) \frac{1 - e^{-a(2R-h)}}{a} - \frac{3}{2} \frac{h^2}{a^2 R^3} [1 - (1+a(2R-h)) e^{-a(2R-h)}] \right. \\ \left. - \frac{3h}{a^3 R^3} \left[1 - (1+a(2R-h) + \frac{a^2 (2R-h)^2}{2}) e^{-a(2R-h)} \right] \right\}$$

et :

$$T_2 = \frac{h^2}{2R^2} \gamma(2R) + \frac{3}{8} \frac{h^2 (2R-h)^2}{R^4} [4 \chi(2R) - 2F(2R) - \gamma(2R)]$$

$$\gamma(h) = 1 - e^{-ah}$$

$$\chi(h) = \frac{e^{-ah} - 1+ah}{ah}$$

$$F(h) = \frac{2}{a^2 h^2} \left[1 - ah + \frac{a^2 h^2}{2} - e^{-ah} \right]$$

La formule (3-31) est représentée par l'abaque V

ANNEXE 1

COMPARAISON AVEC L'ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE

Dans le cas gaussien, l'ensemble S où les données sont disponibles étant fini, l'estimateur du maximum de vraisemblance coïncide avec l'estimateur optimal de la dérive défini en 1-1. En effet, soient Z_α ($\alpha = 1, 2, \dots, n$) n variables gaussiennes, $\sigma_{\alpha\beta}$ leurs covariances, $B^{\alpha\beta}$ la matrice inverse des $\sigma_{\alpha\beta}$, et soient enfin :

$$(A-1,1) \quad m_\alpha = E(Z_\alpha) = \sum_i a_i f_\alpha^i \quad (i = 0, 1, \dots, k)$$

leurs espérances, avec une matrice f_α^i connue et des coefficients a_i inconnus (à estimer).

Le logarithme de la densité Ψ de probabilité des n variables Z_α a pour expression:

$$\log \Psi = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log \text{Det}(\sigma_{\alpha\beta}) - \frac{1}{2} \sum_{\alpha, \beta} B^{\alpha\beta} (z_\alpha - m_\alpha)(z_\beta - m_\beta)$$

(avec $m_\alpha = \sum_i a_i f_\alpha^i$). Annulons les dérivées partielles en a_i : on voit que l'estimateur A_i du maximum de vraisemblance est solution du système :

$$(A-1,2) \quad \sum_{j, \alpha, \beta} A_j f_\beta^j f_\alpha^i B^{\alpha\beta} = \sum_{\alpha, \beta} f_\alpha^i B^{\alpha\beta} z_\beta$$

à $k+1$ équations et $K+1$ inconnues (A_i , $i = 0, 1, \dots, k$). La matrice $B^{\alpha\beta}$ est strictement définie positive. Si les vecteurs $f^i = (f_\alpha^i, \alpha = 1, 2, \dots, n)$ sont linéairement indépendants, la matrice $f B f$ est elle-même strictement définie positive, et le système (A-1,2) admet une solution unique.

D'autre part, la première relation (1-D) équivaut à :

$$(A-1,3) \quad \lambda_\ell^\beta = \sum_{\alpha, s} B^{\beta\alpha} f_\alpha^s \mu_s \ell$$

et la deuxième relation (1-D), qui se met sous la forme :

$$(A-1,4) \quad \sum_{\alpha, \beta, s} f_\beta^i B^{\alpha\beta} f_\alpha^s \mu_s \ell = \delta_\ell^i$$

exprime que la matrice $\mu_s \ell$ est l'inverse de la matrice $f B f$ du système (A-1-2).

Ainsi, la solution de (A,1,2) est :

$$A_i = \sum_{\ell, \alpha, \beta} \mu_{i\ell} f_{\alpha}^{\ell} B^{\alpha\beta} z_{\beta}$$

soit, compte tenu de (A,1,3) :

$$A_i = \sum_{\beta} \lambda_i^{\beta} z_{\beta}$$

c'est-à-dire, précisément, l'estimateur optimal de la dérive.

ANNEXE 2

DEMONSTRATION DE LA PROPOSITION 2-2

Posons d'abord quelques lemmes :

Lemme 1 : On a $Z \in \bar{H} \cap \bar{H}_f$ si et seulement si $Z \in \bar{H}_f$ et $E(Z) = 0$. En effet, $E(Z) = 0$ et $Z \in \bar{H}_f$ entraînent $\Phi_f(Z) = Z \in \bar{H}$. Inversement, soit $Z \in \bar{H} \cap \bar{H}_f$: $Z \in \bar{H}$ est limite dans $\bar{H}$ d'une suite $Z_n \in H$ avec $E(Z_n) = 0$, donc on a $E(Z) = 0$.

Lemme 2 : $\bar{H} \cap \bar{H}_f = \overline{H \cap H_f}$

Il suffit de montrer $\bar{H} \cap \bar{H}_f \subset \overline{H \cap H_f}$, l'inclusion inverse étant toujours vraie. Soit $Z \in \bar{H} \cap \bar{H}_f$, et une suite $Z_n = \int \lambda_n(dx) [Y(x) + f(x)] \in H_f$ convergeant vers Z dans $\bar{H}_f$. On a $E(Z_n) \rightarrow E(Z) = 0$, c'est-à-dire $\int \lambda_n(dx) f(x) \rightarrow 0$. Choisissons une mesure $\mu \in \Lambda$ avec :

$$\int f(x) \mu(dx) = 1, \quad \iint \mu(dx) \mu(dy) C(x,y) = a < \infty$$

Les variables $Y_n = \int (\lambda_n - E(Z_n)\mu) (Y(x) + f(x))$ vérifient $Y_n \in H_f$, $Y_n \in H$ (car $\int (\lambda_n - E(Z_n)\mu) f(x) = 0$) et $Y_n \rightarrow Z$ dans L^2 : par suite $Z \in \overline{H \cap H_f}$.

Lemme 3 : Tout $Z \in \bar{H}_f$ est limite d'une suite $Z_n \in H_f$, avec $E(Z_n) = E(Z)$. Si $E(Z) = 0$, cela résulte des deux lemmes ci-dessus. Soit $E(Z) \neq 0$, et $Z = \lim \int \lambda_n(dx) [Y(x) + f(x)]$ avec $\lambda_n \in \Lambda$. On a $\int \lambda_n(dx) f(x) \rightarrow E(Z)$, et la suite

$$Z_n = \frac{E(Z)}{\int \lambda_n(dx) f(x)} \int \lambda_n(dx) [Y(x) + f(x)]$$

converge vers Z dans $\bar{H}_f$ et vérifie $E(Z_n) = E(Z)$.

Passons maintenant à la proposition principale :

Proposition : Les 4 propriétés suivantes sont équivalentes :

a/ $1 \notin H_f$

b/ la forme linéaire $\lambda \rightarrow \int \lambda(dx) f(x)$ est continue sur Λ munie de la norme $\|\lambda\|^2 = \iint \lambda(dx) C(x,y) \lambda(dy)$

c/ il existe un élément $Y_f \in \bar{H}$ tel que $\langle Y_f, Y(x) \rangle = f(x)$ pour tout $x \in S$.

d/ l'application $\Phi_f : Z \rightarrow Z - E(Z)$ est un homéomorphisme de $\bar{H}_f$ sur $\bar{H}$.

Montrons $a \Leftrightarrow b$: $1 \in \bar{H}_f$ signifie que l'on peut trouver une suite $\lambda_n \in \Lambda$ avec $Z_n = \int \lambda_n(dx) (Y(x) + f(x)) \rightarrow 1$, et, d'après le lemme 3, on peut même supposer $\int \lambda_n(dx) f(x) = 1$. De plus, on a $\int \lambda_n(dx) Y(x) \rightarrow 0$ dans $\bar{H}$. Par suite, $1 \notin \bar{H}_f$ équivaut à :

$$\exists \varepsilon > 0, \forall \lambda \in \Lambda, \int \lambda(dx) f(x) = 1 \Rightarrow \int \lambda(dx) \lambda(dy) C(x,y) \geq \varepsilon,$$

donc aussi à :

$$\varepsilon \leq \left| \int \lambda(dx) f(x) \right|^2 \leq \int \lambda(dx) \lambda(dy) C(x,y)$$

c'est-à-dire à b/

$b \Leftrightarrow c$: D'autre part, Λ , muni de la norme $\|\lambda\|$ ci-dessus, est homéomorphe à H . Si la forme linéaire : $\lambda \rightarrow \int \lambda f$ est continue, elle se prolonge sur $\bar{H}$, et par suite il existe $Y_f \in \bar{H}$ tel que l'on ait $\langle Y, Y_f \rangle = \int \lambda(dx) f(x)$ pour tout $Y = \int \lambda(dx) Y(x) \in H$. En particulier, on a

$$\langle Y(x), Y_f \rangle = f(x)$$

pour tout $x \in S$. Inversement, si $f(x) = \langle Y(x), Y_f \rangle$ pour un $Y_f \in \bar{H}$ et tout $x \in S$, on a $\int \lambda(dx) f(x) = \langle Y, Y_f \rangle$ pour $Y = \int \lambda(dx) Y(x)$, et cette forme linéaire est continue sur Λ pour la norme $\|\cdot\|$.

L'application $\Phi_f : Z \rightarrow Z - E(Z)$ de $\bar{H}_f$ dans $\bar{H}$ est continue et linéaire. D'après les propriétés des espaces de Hilbert, [6], [8], elle constitue un homéomorphisme si, et seulement si, elle est bijective.

Montrons que a/ ou c/ entraîne d/ - Soient Z et $Z' \in \bar{H}_f$ avec $Z - E(Z') \in \bar{H}$. On en déduit $E(Z) - E(Z') = Z - Z' \in \bar{H}_f$, et $Z \neq Z'$ si et seulement si $E(Z) - E(Z') \neq 0$, ce qui entraîne $1 \in \bar{H}_f$. Donc $1 \notin \bar{H}_f$ entraîne que Φ_f est injective.

Montrons maintenant que c/ entraîne Φ_f surjective. Soit $Y \in \bar{H}$, avec $Y = \lim Y_n$, $Y_n = \int \lambda_n(dx) Y(x) \in H$ ($\lambda_n \in \Lambda$). D'après la continuité du produit scalaire, on a :

$$\langle Y, Y_f \rangle = \lim \langle Y_n, Y_f \rangle = \lim \int \lambda_n(dx) f(x)$$

Par suite, $Y + \langle Y, Y_f \rangle$ est dans $\bar{H}_f$, et cet élément a pour image dans H l'élément Y

lui-même. Donc Φ_f est surjective.

Si a/, ou, ce qui revient au même, c/ est vraie, l'application linéaire continue Φ_f est donc bijective. D'après les propriétés des espaces de Hilbert, Φ_f est alors un homéomorphisme.

Montrons enfin d = c : Si Φ_f est un homéomorphisme, Φ_f^{-1} est continue, et la forme linéaire $Y \rightarrow E[\Phi_f^{-1}(Y)]$ est continue sur $\bar{H}$. D'après les propriétés des espaces de Hilbert, il existe alors $Y_f \in \bar{H}$ tel que :

$$E[\Phi_f^{-1}(Y)] = \langle Y, Y_f \rangle, \quad \forall Y \in \bar{H}$$

En particulier, pour $Y = Y(x)$, $x \in S$, on a $\Phi_f^{-1}(Y(x)) = Y(x) + f(x)$, et par suite $f(x) = \langle Y(x), Y_f \rangle$.

Caractérisation de Y_f - Soit Z la projection de 1 dans $\bar{H}_f$. On a :

$$\langle Z - 1, Y(x) + f(x) \rangle = 0, \quad \forall x \in S$$

c'est-à-dire $\langle Z, Y(x) \rangle = (1 - E(Z)) f(x)$. Posons $X = Z - E(Z)$ ($X \in \bar{H}$), on a aussi :

$$\langle X, Y(x) \rangle = (1 - E(Z)) f(x)$$

On voit que la condition $E(Z) = 1$ équivaut à $1 \in \bar{H}_f$. En effet, $E(Z) = 1$ entraîne $\langle X, Y(x) \rangle = 0$ pour tout $x \in S$, donc $X = 0$ puisque $X \in \bar{H}$, c'est-à-dire $Z = E(Z) = 1 \in \bar{H}_f$. Inversement, si $1 \in \bar{H}_f$, on a $Z = 1$ et a fortiori $E(Z) = 1$.

Si $E(Z) \neq 1$, c'est-à-dire si $1 \notin \bar{H}_f$, on a alors $Y_f = \frac{X}{1 - E(Z)}$, c'est-à-dire, en désignant par Π_f le projecteur de $\bar{H}_f$:

$$Y_f = \frac{\Pi_f 1 - E(\Pi_f 1)}{1 - E(\Pi_f 1)}$$

ANNEXE 3

COMPARAISON AVEC LA METHODE DES MOINDRES CARRES

Pour estimer les coefficients a_ℓ d'une dérive $a_\ell f^\ell(x)$, où les fonctions connues $f^\ell(x)$ sont, (par exemple) des polynômes, on a parfois recours à la méthode des moindres carrés. Nous exposerons cette méthode (A-3,1) et nous examinerons l'efficacité des estimateurs (non optimaux en général) auxquels elle conduit (A-3,2), en explicitant les calculs dans un cas simple (A-3,3). Enfin, nous examinerons dans quels cas la méthode des moindres carrés conduit au même estimateur optimal que le krigeage universel.

A-3-1 THEORIE DES MOINDRES CARRES

Soit $Z(x)$ une F.A. admettant une dérive :

$$m(x) = a_0 + \sum_{\ell=1}^k a_\ell f^\ell(x)$$

avec des coefficients a_ℓ inconnus. On connaît la réalisation de $Z(x)$ sur un ensemble S dont l'indicatrice est $k(x)$ ($= 1$ si $x \in S$, $= 0$ si $x \notin S$). A partir de cette réalisation, on détermine les $k+1$ nombres $B_0, B_1, \dots, B_k$ réalisant le minimum de l'intégrale :

$$I = \int k(x) [Z(x) - B_\ell f^\ell(x)]^2 dx$$

Ces nombres vérifient le système obtenu en annulant des dérivées partielles $\frac{\partial I}{\partial B_\ell}$, soit

$$B_j \int k(x) f^j(x) f^\ell(x) dx = \int k(x) Z(x) f^\ell(x) dx$$

Posons

$$T^{\ell j} = \frac{1}{S} \int k(x) f^\ell(x) f^j(x) dx$$

$S = \int k(x) dx$ désignant la mesure de S . Le système précédent s'écrit (avec sommation de $j = 0$ à $j = k$) :

$$(A-3,1) \quad B_j T^{j\ell} = \frac{1}{S} \int k(x) Z(x) f^\ell(x) dx$$

Dans les applications, on choisit toujours les fonctions $f^\ell(x)$ de manière à ce que la matrice des $T^{j\ell}$ admette une inverse $S_{j\ell}$. Les B_j sont donc donnés (avec sommation de

$\ell = 0$ à $\ell = k$) par :

$$(A-3,2) \quad B_j = \frac{1}{S} S_{j\ell} \int_{k(x)} Z(x) f^\ell(x) dx$$

On voit que ceci peut s'écrire :

$$B_j = \int \lambda_j(dx) Z(x)$$

avec des mesures λ_j dont le support est S et qui admettent les densités :

$$\frac{\lambda_j(dx)}{dx} = \frac{1}{S} k(x) S_{j\ell} f^\ell(x)$$

Ces mesures vérifient les secondes relations (2-D) (celles qui expriment que l'on a un estimateur sans biais de la dérive). En effet, on trouve :

$$\int \lambda_j(dx) f^s(x) = \frac{1}{S} S_{j\ell} \int_{k(x)} f^\ell(x) f^s(x) dx = S_{j\ell} T^{ls}$$

Comme les matrices S et T sont inverses l'une de l'autre, on a bien :

$$(A-3,3) \quad \int \lambda_j(dx) f^s(x) = \delta_j^s \quad (j, s = 0, 1, \dots, k)$$

A-3-2 ESPERANCES ET COVARIANCES DES B_j

Plaçons-nous d'abord dans le cas le plus simple, où $Z(x)$ admet une fonction de covariance $C(x,y)$. Les B_j , définies par (A-3,2), qui représentaient des nombres lorsque $Z(x)$ désignait la réalisation de la F.A. deviennent des variables aléatoires lorsque $Z(x)$ est considérée comme la F.A. elle-même (les intégrales stochastiques qui figurent au second membre de (A-3,2) ont un sens, puisqu'il existe une covariance). La relation (A-3,3) exprime, comme nous l'avons déjà remarqué, que les B_j sont des estimateurs sans biais des vrais coefficients inconnus a_j : en ce sens, on peut dire que les estimateurs des moindres carrés sont des estimateurs universels (bien qu'ils ne soient pas optimaux en général). On a ainsi :

$$E(B_j) = a_j \quad (j = 0, 1, \dots, k)$$

Calculons maintenant la matrice des covariances

$$\beta_{ij} = E[B_i B_j] - a_i a_j$$

Comme les B_j sont des estimateurs sans biais, on peut faire le calcul comme si on avait $a_j = 0$ (ce qui revient à remplacer $Z(x)$ par $Z(x) - a_0 f^\ell(x)$). On déduit de (A-3,2) :

$$B_i B_j = \frac{1}{S^2} S_{i\ell} S_{js} \iint k(x) k(y) Z(x) Z(y) f^\ell(x) f^s(y) dx dy$$

d'où, en passant aux espérances :

$$(A-3,4) \quad \beta_{ij} = \frac{1}{S^2} S_{i\ell} S_{js} \iint k(x) k(y) f^\ell(x) f^s(y) C(x,y) dx dy \quad (i,j = 0,1,\dots,k)$$

Examinons maintenant le cas où il n'existe qu'un demi-variogramme $\gamma(x,y)$, $Z(x)$ elle-même n'ayant plus de moments d'ordre 2, ni même, en général; d'espérance. Pour $\ell = 0$, l'intégrale stochastique qui figure au second membre de (A-3,2) n'existe plus (au sens de la convergence en moyenne quadratique), car, d'après (A-3,3), on a $\int \lambda_0(dx) = 1$. Cela équivaut à dire que la variance $D^2(B_0)$ est infinie, et que l'espérance $E(B_0)$ n'existe pas non plus en général : on retrouve ici une circonstance que nous avons déjà rencontrée en plusieurs occasions. Par contre, pour $j \neq 0$, la relation (A-3,3), écrite avec $s = 0$, donne $\int \lambda_j(dx) = 0$ ($j = 1,2,\dots,k$). Cette condition nous garantit que l'intégrale stochastique (A-3,2) existe (au sens de la convergence en moyenne quadratique) et qu'elle admet une variance finie. La matrice des covariances β_{ij} , restreinte aux indices $i, j \neq 0$, est alors donnée par la relation suivante, analogue à (A-3,4) :

$$(A-3,5) \quad \beta_{ij} = -\frac{1}{S^2} S_{i\ell} S_{js} \iint k(x) k(y) f^\ell(x) f^s(y) \gamma(x,y) dx dy \quad (i,j = 1,2,\dots,k)$$

A-3-3 EXEMPLE DU VARIOGRAMME LINEAIRE

Prenons $\gamma(x,y) = |x-y|$, plaçons-nous sur la droite réelle, et cherchons à estimer une dérive quadratique $a_0 + a_1 x + a_2 x^2$ à partir d'une réalisation de la F.A. $Z(x)$ donnée sur l'intervalle $(-R, +R)$.

Il n'y a aucune difficulté à calculer la matrice des T^{ij} et son inverse S_{ij} ($i,j = 0,1,2$). On trouve :

$$T = \begin{bmatrix} 1 & 0 & \frac{R^2}{3} \\ 0 & \frac{R^2}{3} & 0 \\ \frac{R^2}{3} & 0 & \frac{R^4}{5} \end{bmatrix} \quad S = \begin{bmatrix} \frac{9}{4} & 0 & -\frac{15}{4R^2} \\ 0 & \frac{3}{R^2} & 0 \\ -\frac{15}{4R^2} & 0 & \frac{45}{4R^4} \end{bmatrix}$$

Les estimateurs des moindres carrés sont donc :

$$\begin{cases} B_0 = \frac{9}{4} \frac{1}{2R} \int_{-R}^R Z(x) dx - \frac{15}{4R^2} \frac{1}{2R} \int_{-R}^R x^2 Z(x) dx \\ B_1 = \frac{3}{R^2} \frac{1}{2R} \int_{-R}^R x Z(x) dx \\ B_2 = -\frac{15}{4R^2} \frac{1}{2R} \int_{-R}^R Z(x) dx + \frac{45}{4R^4} \frac{1}{2R} \int_{-R}^R x^2 Z(x) dx \end{cases}$$

En ce qui concerne la matrice des covariances restreintes aux indices $i, j = 1, 2$ (β_{00} serait ici infini), on trouve

$$\beta_{11} = \frac{6}{5R}, \quad \beta_{12} = \beta_{21} = 0, \quad \beta_{22} = \frac{15}{14R^3}$$

On comparera avec les relations (f) de la section 2-10 ($\frac{6}{5R}$ contre $\frac{1}{R}$ pour l'estimateur optimal, et $\frac{15}{14R^3}$ contre $\frac{3}{4R^3}$) : les estimateurs des moindres carrés sont nettement plus variants que les estimateurs optimaux.

A-3-4 CONDITION POUR QUE L'ESTIMATEUR DES MOINDRES CARRÉS SOIT OPTIMAL

Parmi les estimateurs de la forme $B_j = \int \lambda_j(dx) Z(x)$ (support de λ_j inclus dans S), les estimateurs des moindres carrés sont caractérisés par les deux conditions suivantes :

- ~ ils sont sans biais ($E(B_j) = a_j, \forall a_j$)
- ~ les mesures λ_j ont des densités qui sont des combinaisons linéaires des fonctions $k(x)$ et $f^L(x)$.

On a déjà vu que l'estimateur des moindres carrés vérifie ces conditions. Inver-

fient (A-3,3). Mais, si les λ_j ont des densités de la forme $\frac{k(x)}{S} S'_{j\ell} f^\ell(x)$, ces conditions (A-3,3) expriment que la matrice des $S'_{j\ell}$ coïncide avec l'inverse $S_{j\ell}$ de la matrice des T^{ij} , de sorte que les estimateurs B_j sont bien donnés par (A-3,2).

Un estimateur vérifiant ces deux propriétés vérifie par définition les secondes relations du système (2-D). Pour qu'il soit optimal, il faut et il suffit qu'il satisfasse également aux premières relations (2-D), c'est-à-dire que l'on puisse trouver des constantes μ_{js} telles que :

$$(A-3,5) \quad \frac{1}{S} S_{j\ell} \int k(x) f^\ell(x) C(x,y) dx = \mu_{js} f^s(y), \quad \forall y \in S$$

Il en résulte que l'estimateur des moindres carrés coïncide avec l'estimateur optimal de la dérive si et seulement si les $k+1$ fonctions $f^\ell(x)$ sont des combinaisons linéaires de $k+1$ fonctions propres distinctes du noyau $k(x) C(x,y) k(y)$, considéré comme un opérateur agissant sur les fonctions à support dans S .

En particulier, si les $f^\ell(x)$ sont elles-mêmes des fonctions propres $(\int k(x) f^\ell(x) C(x,y) dx = \mu^\ell f^\ell(y) \text{ pour } y \in S)$, la relation (A-3,5) se réduit à :

$$(A-3,6) \quad \mu_{js} = \frac{1}{S} S_{js} \mu^s \quad (\text{sans sommation en } s)$$

Dans le cas où il n'existe pas de covariance $C(x,y)$, mais seulement un variogramme $\gamma(x,y)$, la condition est un tout petit peu moins simple. Si l'on désigne par $\bar{Z} = \frac{1}{S} \int k(x) Z(x) dx$ la moyenne (spatiale) de $Z(x)$ sur S , $Z(x) - \bar{Z}$ est une F.A. admettant des moments d'ordre 1 et 2. Sa covariance $\bar{C}(x,y)$ est donnée par :

$$\begin{aligned} \bar{C}(x,y) = & -\gamma(x,y) + \frac{1}{S} \int k(y) \gamma(x,y) dx + \frac{1}{S} \int k(x) \gamma(x,y) dy \\ & - \frac{1}{S^2} \iint k(x) k(y) \gamma(x,y) dx dy \end{aligned}$$

On peut alors montrer que les B_j des moindres carrés ($j = 1, 2, \dots, k$) coïncident avec l'estimateur optimal si et seulement si les k fonctions $f^\ell(x)$ ($\ell = 1, 2, \dots, k$) sont des combinaisons linéaires de k fonctions propres du noyau $\bar{C}(x,y)$ correspondant à des valeurs propres non nulles.

Cette condition est de nature assez complexe, puisqu'elle fait intervenir à la fois les fonctions $f^\ell(x)$, le noyau $C(x,y)$ et la géométrie de l'ensemble S où les

données sont disponibles. Mais il y a un cas où cette condition est automatiquement vérifiée. C'est le cas où la partie aléatoire de $Z(x)$ se réduit à un simple bruit de fond, (ou, en terminologie géostatistique, à un effet de pépité pur). Dans ce cas, la covariance $C(x,y)$ se réduit à $\omega\delta(x-y)$, δ désignant la mesure de Dirac et ω un facteur constant : toute fonction $f(x)$ continue sur S est alors fonction propre (pour la valeur propre ω) et cela quel que soit l'ensemble S . Dans ce cas, l'estimateur des moindres carrés est toujours optimal, et les covariances μ_{js} , d'après (A-3,6) sont données par :

$$(A-3,7) \quad \mu_{js} = \frac{\omega}{S} S_{js}$$

En fait, une covariance $\omega\delta(x-y)$ caractérise une mesure, et non plus une fonction aléatoire. Ce n'est d'ailleurs pas gênant si l'ensemble S est continu, puisqu'en fait on ne travaille alors que sur des régularisées de cette mesure aléatoire (qui sont à nouveau des F.A.). En pratique, il suffit pour que l'estimateur des moindres carrés soit (presque) optimal que la covariance soit de la forme $C(x-y)$ avec une portée très faible (par exemple, $C(h) = e^{-ah}$ avec $\frac{1}{a}$ petit relativement aux dimensions de S). En particulier, si l'ensemble S est fini, il suffira que cette portée soit inférieure à la plus petite des distances séparant les points de S .

Dans ce cas (covariance $C(h)$ de portée très petite), l'estimateur des moindres carrés est (presque) optimal, et les covariances μ_{ij} sont encore données par la relation (A-3,7), à condition de prendre comme facteur ω l'intégrale de $C(h)$ étendue à l'espace entier, soit :

$$(A-3,8) \quad \mu_{ij} = \frac{S_{js}}{S} \int C(h) dh$$

S_{js} désignant toujours l'inverse de

$$T^{js} = \frac{1}{S} \int k(x) f^j(x) f^s(x) dx$$

A-3-5 CAS OU L'ENSEMBLE S EST FINI

Nous avons donné ci-dessus les formules des moindres carrés dans le cas d'un ensemble S continu. Si l'ensemble S est fini, ces formules se modifient de manière évidente. Supposons S constitué de n points x_α ($\alpha = 1, 2, \dots, n$). Posons

$$(A-3,9) \quad T^{ij} = \frac{1}{n} \sum_{\alpha} f^i(x_{\alpha}) f^j(x_{\alpha})$$

et désignons par S_{ij} l'inverse de la matrice des T^{ij} . Les estimateurs des moindres carrés sont alors :

$$(A-3,10) \quad B_j = \frac{1}{n} \sum_{\ell} S_{j\ell} \sum_{\alpha} f^{\ell}(x_{\alpha}) Z(x_{\alpha})$$

Dans le cas où la covariance $C(x-y)$ a une portée plus petite que la distance séparant deux points expérimentaux ($C(x_{\alpha} - x_{\beta}) = 0$ pour $\alpha \neq \beta$, les estimateurs (A-3-9) sont optimaux. Posant $\sigma^2 = C(0)$, on voit que la matrice des covariances $\mu_{ij} = \text{Cov}(B_i B_j)$ est alors :

$$\mu_{ij} = \frac{\sigma^2}{n^2} \sum_{\ell} \sum_s \sum_{\alpha} S_{i\ell} S_{js} f^{\ell}(x_{\alpha}) f^s(x_{\alpha})$$

Compte tenu de l'expression (A-3,9) des T^{ij} et du fait que la matrice S est inverse de la matrice T , ceci donne :

$$\mu_{ij} = \frac{\sigma^2}{n} \sum_{\ell s} S_{i\ell} S_{js} T^{\ell s}$$

c'est-à-dire

$$(A-3,11) \quad \mu_{ij} = \frac{\sigma^2}{n} S_{ij}$$

Avant d'utiliser cette formule (A-3,11) dans une application, il faut :

1/ S'assurer que la portée de la covariance $C(h)$ est réellement inférieure à la distance entre points expérimentaux.

2/ Estimer la variance σ^2 .

Examinons rapidement ces deux points.

1/ Conformément à la méthode générale, on construira le variogramme des résidus. Posant :

$$2 \gamma_{\alpha\beta}^* = (Z_{\alpha} - B_j f_{\alpha}^j - Z_{\beta} + B_j f_{\beta}^j)^2$$

($Z_{\alpha} = Z(x_{\alpha})$ et $f_{\alpha}^j = f^j(x_{\alpha})$), on trouve, compte tenu de (A-3,11) et de la condition $\text{Cov}(Z_{\alpha} Z_{\beta}) = 0$ ($\alpha \neq \beta$) :

$$2 E[\gamma_{\alpha\beta}^*] = 2 \sigma^2 - \frac{\sigma^2}{n} S_{ij} (f_{\alpha}^j - f_{\beta}^j) (f_{\alpha}^i - f_{\beta}^i)$$

Expérimentalement, on devra comparer à cette expression la moyenne des $\gamma_{\alpha\beta}^*$ prises sur les couples (x_{α}, x_{β}) pour lesquels $x_{\alpha} - x_{\beta}$ est (à peu près) le même vecteur h . On notera que cette espérance $2 E(\gamma_{\alpha\beta}^*)$ ne se réduit pas à la constante $2 \sigma^2$.

2/ Pour estimer σ^2 , le plus simple est de former (comme dans la section 3-2) la variance expérimentale s^2 des résidus :

$$s^2 = \frac{1}{n} \sum_{\alpha} (Z_{\alpha} - B_j f_{\alpha}^j)^2$$

Développons, et passons à l'espérance : on trouve :

$$E(s^2) = \sigma^2 - \frac{2\sigma^2}{n^3} \sum_{\alpha} f_{\alpha}^j S_{j\ell} f_{\alpha}^{\ell} + \frac{\sigma^2}{n^3} \sum_{\alpha} S_{ij} f_{\alpha}^i f_{\alpha}^j$$

soit, puisque S est l'inverse de T :

$$E(s^2) = \frac{n-1}{n} \sigma^2$$

L'estimateur sans biais que l'on doit utiliser est donc $\frac{n}{n-1} s^2$ exactement comme en statistique classique.

ANNEXE 4

PROBLEMES D'INFERENCE STATISTIQUE POUR LES F. A.

Nous avons déjà rencontré plusieurs exemples de biais inattendus et graves qui surgissent lorsque l'on procède à une inférence statistique sur une réalisation d'une F.A. En particulier, dans la section 2-9, nous avons signalé que ces biais avaient systématiquement pour effet de donner une justification expérimentale apparente à l'hypothèse simple selon laquelle la F.A. se décomposerait en une dérive et un bruit de fond - et cela, même lorsque cette hypothèse est tout-à-fait fausse. Un biais analogue vient renforcer, de la même manière illusoire, la croyance en l'existence d'une covariance $C(h)$. On a souvent, en effet, l'habitude d'admettre, sans examen approfondi que les V.R. que l'on étudie peuvent être considérées comme des réalisations de F.A. stationnaires d'ordre 2 (alors que bien souvent une étude plus fine montre que seuls les accroissements de cette F.A. ont des moments d'ordre 2 - autrement dit, bien souvent, il n'existe pas de covariance $C(h)$, mais seulement un variogramme $\gamma(h)$). Or il se trouve que les procédés d'inférence statistique que l'on utilise ont systématiquement pour effet, par suite des biais qu'ils entraînent, de rendre cette hypothèse plausible, même lorsqu'elle est tout-à-fait fausse.

Nous nous contenterons de montrer ce phénomène sur un exemple : celui du mouvement brownien, de demi-variogramme $\gamma(h) = |h|$, dont on connaît une réalisation sur l'intervalle $(0, L)$. On suppose donc :

$$\begin{cases} E[Y(x+h) - Y(x)] = 0 \\ \frac{1}{2} D^2[Y(x+h) - Y(x)] = |h| \end{cases}$$

Dans la pratique usuelle, pour estimer l'hypothétique covariance $C(h)$, qui en fait n'existe pas ici, on calcule successivement les valeurs (expérimentales) des expressions :

$$\bar{Y} = \frac{1}{L} \int_0^L Y(x) dx$$

$$(A-4, 1) \quad C^*(x, y) = (Y(x) - \bar{Y})(Y(y) - \bar{Y})$$

et on en déduit la "covariance expérimentale"

$$(A-4,2) \quad C^*(h) = \frac{1}{L-h} \int_0^{L-h} C^*(x+h, x) dx$$

Or, avec le demi-variogramme $\gamma(h) = |h|$, l'espérance de $C^*(x, y)$ est :

$$E[C^*(x, y)] = \frac{2}{3} L + \frac{x^2 + y^2}{L} - 2 \text{ Sup}(x, y)$$

Pour $h \geq 0$, on a donc

$$E[C^*(x+h, x)] = \frac{2}{3} L + \frac{x^2 + (x+h)^2}{L} - 2x - 2h$$

et, en intégrant en x , on voit que la "covariance expérimentale" admet l'espérance :

$$(A-4,3) \quad E[C^*(h)] = \frac{1}{3} L - \frac{4}{3} h + \frac{2}{3} \frac{h^2}{L}$$

C'est une parabole, de pente $-\frac{4}{3}$ à l'origine. On trouvera donc une variance apparente $C(0) = \frac{1}{3} L$, liée uniquement à la longueur L de l'intervalle sur lequel on travaille, et qui constitue un pur artefact (puisque la variance réelle est infinie). Bien qu'il n'existe pas de covariance stationnaire, mais seulement un variogramme linéaire, les biais qu'introduit ce procédé d'estimation ont pour effet de donner une confirmation apparente de l'existence de cette covariance. On note que la structure du phénomène est profondément déformée : non seulement la droite est remplacée par une parabole, mais même la pente à l'origine se trouve affectée ($\frac{4}{3}$ au lieu de 1). Le $C^*(h)$ est un pur artefact, et ne conserve à peu près plus rien de la structure réelle du processus.

On notera que le variogramme expérimental :

$$2 \gamma^*(h) = \frac{1}{L-h} \int_0^{L-h} [Y(x+h) - Y(x)]^2 dx$$

a comme espérance

$$E[\gamma^*(h)] = \gamma(h) = |h|$$

il n'est pas affecté par le biais précédent, et constitue par conséquent pour l'expérimentateur un outil plus sûr que la covariance.

Plus généralement, soit $Y(x)$ une F.A. sans dérive et admettant un variogramme $\gamma(x-y)$. L'expression (A-4,1) admet l'espérance :

$$E[C^*(x, y)] = -\gamma(x-y) + \frac{1}{L} \int_0^L \gamma(x-y) dy + \frac{1}{L} \int_0^L \gamma(x-y) dx - \frac{1}{L^2} \int_0^L \int_0^L \gamma(x-y) dx dy$$

En posant :

$$\chi(h) = \frac{1}{h} \int_0^h \gamma(x) dx, \quad F(h) = \frac{2}{h^2} \int_0^h x \chi(x) dx$$

on voit facilement que la "covariance expérimentale" $C^*(h)$ définie par (A-4,2) admet l'espérance :

$$(A-4,5) \quad E[C^*(h)] = -\gamma(h) - F(L) + \frac{1}{L(L-h)} \left[\int_0^L dz \int_0^{L-h} \gamma(x-z) dx + \int_0^L dz \int_h^L \gamma(x-z) dx \right]$$

En particulier, la variance, si elle existe, vaut $\gamma(\infty)$: mais le procédé d'inférence statistique conduit à une "variance expérimentale" dont l'espérance est :

$$E[C^*(0)] = F(L)$$

C'est seulement dans le cas où $\gamma(h)$ est borné et admet une portée petite devant L que l'on a $F(L) \neq \gamma(\infty)$. Dans tous les autres cas, $F(L)$ est inférieur à la variance réelle $\gamma(\infty)$ (finie ou non).

De même, supposons que $\gamma(h)$ admette une pente à l'origine (dérivée à droite) $\gamma'_+(0)$. En dérivant (A-4,5), on trouve pour $E[C^*(h)]$ la pente :

$$\left. \frac{d^+}{dh} E[C^*(h)] \right|_{h=0} = -\gamma'_+(0) + \frac{2}{L} F(L) - \frac{2}{L} \chi(L)$$

Cette pente est biaisée (sauf toujours si $\gamma(h)$ admet une portée petite devant L).

Ainsi, pour une F.A. dérivable en moyenne quadratique $\gamma'_+(0) = 0$, et $\gamma(h)$ deux fois dérivable à l'origine, on trouve un $E[C^*(h)]$ qui présente une pente oblique à l'origine et fait conclure à la non-dérivabilité m.q. : la structure du phénomène est profondément défigurée.

BIBLIOGRAPHIE

- [1] - JOURNEL, A. (1969) - Rapport d'Etude sur l'estimation d'une variable régionalisée : application à la cartographie automatique - (Service Hydrographique de la Marine, Paris).
- [2] - LANDKOFF, N.S. (1966) - Osnovy sovremennoi teorii potentziala, Moscou, Izd. Nauka, 516 pages.
- [3] - MATHERON, G. (1965) - Les variables régionalisées et leur estimation, Paris, Masson, 306 pages.
- [4] - MATHERON, G. (1967) - Kriging, or polynomial interpolation procedures?. Can. Min. and Met. Bull., LXX, 240-244.
- [5] - MATHERON, G. (1968) - Osnovy prikladnoi Geostatistiki, Moscou, Editions Mir, 408 pages.
- [6] - PLESNER, A.I. (1965) - Spektralnaia teoria lineinikh operatorov. Moscou. Izd. Nauka; 624 pages.
- [7] - SERRA, J. (1968) - Les structures gigogne : morphologie mathématique et interprétation métallogénique - Mineralium Deposita, 3, 135-154.
- [8] - YOSIDA, K. (1968) - Functional Analysis, Springer, Berlin, 466 pages.

LE KRIGEAGE UNIVERSEL

(RECHERCHE D'ESTIMATEURS OPTIMAUX EN PRESENCE D'UNE DERIVE)

Table des Matières

<u>O - INTRODUCTION</u>	1
O-1 Critique des Méthodes de Moindre Carré	1
O-2 Position du Problème et Hypothèses Générales	3
 <u>SECTION 1 - CAS OU L'ENSEMBLE S DES DONNEES EST FINI</u>	 8
1-1 Estimation Optimale de la Dérive (covariance connue)	8
Système 1-D	10
1-2 Le Krigeage Universel (à covariance connue)	11
Système 1-K	12
Krigeage Ponctuel	13
1-3 Le Théorème d'additivité	13
1-4 Cas où il n'existe qu'un variogramme $\gamma(x,y)$	16
Estimation de la Dérive (système 1-D')	17
Krigeage Universel (système (1-K'))	18
 <u>SECTION 2 - THEORIE GENERALE</u>	 20
2-1 Introduction	20
2-2 Les Espaces $\bar{H}$ et $\bar{H}_f$	21
Proposition 2-2	23
2-3 L'Estimateur Optimal de la Dérive	25
Système (2-D)	29
Système (2-DM)	29
2-4 Exemple: Estimation de l'Espérance d'une P.A. stationnaire	30
2-5 Le Krigeage Universel	31
Systèmes (2-K) et (2-KM)	32
Krigeage ponctuel	32
2-6 Relation entre l'Estimation de la Dérive et le Krigeage Universel	32
Exemple: Krigeage Ponctuel	33
2-7 Variance du Krigeage Universel	34

2-8	Cas où il n'existe qu'un variogramme $\gamma(x,y)$	35
	Systèmes (2-D') et (2-D'M)	36
	Système (2-K')	37
2-9	Le Variogramme des Résidus	37
2-10	Exemple du Variogramme Linéaire	39
	Estimation d'une Dérive Quadratique	40
	Krigeage Universel	42
	Cas discret	43
<u>SECTION 3 - ESTIMATION DES PARAMETRES, ET CONTROLE DES HYPOTHESES</u>		46
3-1	Difficultés de ces Problèmes	46
3-2	Estimation du Facteur ω	49
3-3	Estimation de Paramètres Essentiels	51
	Cas où S est un segment de droite	53
3-4	Exemple d'un Variogramme en $r^{\alpha-1}$	55
	Formule de Landkoff	55
	Estimation optimale d'une dérive quadratique	56
	Estimation du paramètre α	59
3-5	Exemple d'une covariance en $e^{-a h }$	59
	Estimation de l'espérance d'une F.A. stationnaire	60
	Estimation optimale d'une dérive quadratique	61
	Estimation du paramètre a	63
<u>ANNEXE 1 - COMPARAISON AVEC L'ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE</u>		65
<u>ANNEXE 2 - DEMONSTRATION DE LA PROPOSITION 2-2</u>		67
<u>ANNEXE 3 - COMPARAISON AVEC LES METHODES DE MOINDRES CARRES</u>		70
A-3-1	Théorie des Moindres Carrés	70
A-3-2	Espérances et Covariances des B_j	71
A-3-3	Exemple du Variogramme Linéaire	72
A-3-4	Condition pour que l'estimateur des Moindres Carrés soit optimal	73
A-3-5	Cas où l'Ensemble S est fini	75
<u>ANNEXE 4 - PROBLEMES D'INFERENCE STATISTIQUE POUR LES F.A.</u>		78
<u>BIBLIOGRAPHIE</u>		81