N-205

COKRIGEAGE ET REGRESSION EN CORRELATION INTRINSEQUE

-=-=-=-=-=-=-=-=-=-=-=-=-=-

A. MARECHAL

FONTAINEBLEAU NOVEMBRE 1970

COKRIGEAGE ET REGRESSION EN CORRELATION INTRINSEQUE

A. Maréchal

Nous savons que dans un champ où sont données deux variables régionalisées dont les variogrammes et variogrammes croisés sont connus, il est toujours possible d'estimer par cokrigeage n'importe quelle valeur moyenne de l'une ou l'autre des deux variables, en améliorant la variance de krigeage par rapport à l'éventualité où l'on n'utiliserait les régionalisations que séparées. Le cokrigeage a été bien étudié, en particulier par J. SERRA [1]. Le but de cette note est simplement d'attirer l'attention sur les relations existant entre cokrigeage et technique de regression. Comme on pouvait s'y attendre, celle ~~n'~~étant un estimateur croisé sans biais et optimal au sens des moindres carrés, nous verrons qu'il y a équivalence entre cokrigeage et régression lorsque toutes les covariances de point à point sont nulles. Lorsque ce n'est pas le cas, la régression n'est jamais l'estimateur croisé optimal et nous le verrons sur quelques cas simples.

Pour dépasser ensuite un peu le problème de la régression, nous verrons une propriété particulière de la corrélation intrinsèque la transitivité des krigeages. Il permet de trouver la solution exacte au problème classique des mineurs qui consiste à passer d'une variable mesurée à la variable intéressante (en général par régression), puis de travailler sur ces valeurs comme si elles étaient des données d'observation.

I - NOTATIONS

Nous considérons deux V.R. $A(x)$ et $B(x)$ connues sur deux ensembles $S_A(A_i,\ i = 1,\dots n)$ et $S_B(B_i,\ i = n+1,\dots n+n')$.

Ces deux ensembles peuvent avoir des points communs, ou même être confondus. Nous travaillerons toujours sur $S = S_A + S_B$ $S(X_i,\ i = 1, \dots\ n, \dots\ n+n')$.

Nous utiliserons les notations $\sigma_{A_i B_j} = E\{(A_i - \overline{A})(B_j - \overline{B})\}$ et $\overline{A}$ et $\overline{B}$ en supposant que $\sigma_{A_i A_j}$, $\overline{A}$ et $\overline{B}$ existent.

Si A et B ne possèdent que des variogrammes, la notation $\sigma_{A_i A_j}$ représentera $- \gamma_{A_i A_j}$ et $\overline{A}$ et $\overline{B}$ les moyennes de A et B dans un grand champ quelconque.

Nous ferons parfois l'hypothèse de corrélation intrinsèque qui est la suivante

$$\sigma_A^2(x,y) = \sigma_A^2\ \sigma(x,y)$$

$$\sigma_B^2(x,y) = \sigma_B^2\ \sigma(x,y)$$

$$\sigma_{AB}(x,y) = \sigma_{AB}^2\ \sigma(x,y)$$

Dans ce cas, nous savons que le coefficient de corrélation $\rho(x,y) = \dfrac{\sigma_{AB}(x,y)}{\sqrt{\sigma_A^2(x,y)\ \sigma_B^2(x,y)}}$ a un sens intrinsèque si les V.R. sont stationnaire ou intrinsèque : en particulier, le coefficient de corrélation entre A_v(moyenne de A dans un volume v) et B_v reste égal à ρ. Dans le cas de corrélation intrinsèque, ρ est évidemment égal au coefficient de corrélation des statisticiens (c'est même le seul cas où celui-ci ait un sens en géostatistique).

Tous les calculs sont faits dans le cas stationnaire. Il est facile de retrouver les équations, dans le cas où n'existent que des variogrammes.

II - EQUATIONS GENERALES DU COKRIGEAGE SANS DERIVE.

1 - Soit $A_v = \frac{1}{v} \int_v A(x)\, dx \qquad B_{v'} = \frac{1}{v'} \int_{v'} B(x) dx$

On se propose d'estimer $\alpha A_v + \beta B_{v'}$ par une forme linéaire $(\alpha A_v + \beta B_{v'})^* = a^j A_j + b^j B_j \qquad a^j = 0 \quad j > n, \quad b^j = 0 \quad j \leq n.$

a/ Condition de non biais.

$$E[\alpha\ A_v + \beta\ B_{v'} - a^j\ A_j - b^j\ B_j] = 0$$

$\Longrightarrow \bar{A}(\sum a^j - \alpha) + \bar{B}(\sum b^j - \beta) = 0$ quelques soient les $\bar{A}$ et $\bar{B}$ inconnus

$$\Longrightarrow \begin{cases} \sum a^j = \alpha \\ \sum b^j = \beta \end{cases}$$

b/ Condition d'optimalité.

$$D^2[\alpha\ A_v + \beta\ B_{v'} - a^j\ A_j - b^j\ B_j] = E[a^j(A_j-\bar{A})+b^j(B_j-\bar{B})-\alpha(A_v-\bar{A})-\beta(B_{v'}-\bar{B})$$

$$D^2 = a^i a^j\ \sigma_{A_iA_j} + b^i b^j\ \sigma_{B_iB_j} + \alpha^2\sigma^2_{A_v} + \beta^2\sigma^2_{B_{v'}}$$

$$+ 2\ a^i b^j\ \sigma_{A_iB_j} - 2\alpha\ a^j\ \sigma_{A_jA_v} - 2\beta\ a^j\ \sigma_{A_jB_{v'}}$$

$$- 2\ \alpha\ b^j\ \sigma_{B_jA_v} - 2\ \beta\ b^j\ \sigma_{B_jB_{v'}} + 2\ \alpha\beta\ \sigma_{A_vB_{v'}}$$

On obtiendra l'optimum compte-tenu des conditions

$$\Sigma\, a^j = \alpha,\ \Sigma\, b^j = \beta,\quad a^j = 0 \quad j = n+1,\dots n'+n,\ b^j = 0 \quad j = 1,\dots$$

$$(II,1)\quad \begin{cases} a^j\, \sigma_{A_iA_j} + b^j\, \sigma_{A_iB_j} = \alpha\, \sigma_{A_iA_v} + \beta\, \sigma_{A_iB_{v'}} + \lambda_\alpha + \lambda_i \\ b^j\, \sigma_{B_iB_j} + a^j\, \sigma_{A_jB_i} = \alpha\, \sigma_{B_iA_v} + \beta\, \sigma_{B_iB_{v'}} + \mu_\beta + \mu_i \\ \Sigma\, a^j = \alpha \\ \Sigma\, b^j = \beta \\ \lambda_i = 0 \quad i = 1,\dots n \qquad a^j = 0 \quad j = n+1,\dots n+n' \\ \mu_i = 0 \quad i = n+1,\dots n+n' \quad b^j = 0 \quad j = 1,\dots n. \end{cases}$$

$$\sigma_K^2 = \sigma^2(\alpha A_v+\beta B_v) - a^i\, \sigma_{A_i(\alpha A_v+\beta B_v)} - b^i\, \sigma_{B_i(\alpha A_v+\beta B_v)}$$

$$+ \alpha\lambda_\alpha + \beta\mu_B + a^i\, \lambda^i + b^i\, \mu^i$$

2 - <u>Cokrigeage d'une erreur de mesure.</u>

On peut prendre l'exemple traité par J. SERRA [1] comme cas particulier de cokrigeage.

Soit une variable stationnaire $Y(x)$ de covariance $\sigma_Y(h)$. Au cours de la mesure, on rajoute à $Y(x)$ une erreur de mesure de moyenne nulle et de covariance $\sigma_e(h)$, qu'on supposera indépendante de la vraie variable $Y(x)$. Expérimentalement, on ne connait que les

$Z_i = Y_i + e_i$, qui permettent de calculer expérimentalement la covariance $\sigma_Z = \sigma_Y + \sigma_e$.

Soit à kriger la moyenne de Y dans v, Y_v.

. Si la présente de l'erreur de mesure est inconnue, on krige~~re~~ra à partir des Z_i et de la covariance σ_Z

$$Z_{v*} = a'^j Z_j \text{ qui donne le système (II,2)}$$

II,2,1
$$\begin{cases} a'^j(\sigma_{y_i y_j} + \sigma_{e_i e_j}) = \sigma_{Y_i Y_v} + \sigma_{e_i e_v} + \lambda' \\ \sum a'^j = 1 \end{cases}$$

$$\sigma^{2'}_K = \sigma^2_Y + \sigma^2_e - a'^j(\sigma_{Y_j Y_v} + \sigma_{e_j e_v}) + \lambda'$$

. Si, par une étude des méthodes de mesure (quartage, analyse), on peut déterminer la décomposition $\sigma_Z = \sigma_Y + \sigma_e$, on cherchera à kriger Y_v, qui est la seule variable physique.

$$Y_{v*} = a^j Z_j \text{ qui donne :}$$

II,2,2
$$\begin{cases} a^j(\sigma_{Y_i Y_j} + \sigma_{e_i e_j}) = \sigma_{Y_i Y_v} + \lambda \\ \sum a^j = 1 \qquad\qquad \sigma^2_K = \sigma^2_Y - a^j(\sigma_{Y_j Y_v} + \sigma_{e_j e_v}) + \lambda \end{cases}$$

. Si maintenant on cherche à kriger e_v (valeur moyenne de l'erreur de mesure dans v, si celui-ci était analysé de la même façon que les échantillons).

$$e_{v*} = a''^{j}\, Z_j$$

$$\text{(II,2,3)} \quad \begin{cases} a''^{j}(\sigma_{Y_iY_j} + \sigma_{e_ie_j}) = \sigma_{e_ie_v} + \lambda'' \\ \sum a''^{j} = 0 \end{cases}$$

$$\sigma^2_{K''} = \sigma^2_e - a''j\,(\sigma_{Y_jY_v} + \sigma_{e_je_v}) + \lambda''$$

en ajoutant les systèmes II,2.1 et II,2.3, on voit que :

$$a'^{j} = a^{j} + a''^{j} \qquad \text{c'est-à-dire } Z_v{*} = Y_v{*} + e{*}_v$$

$$\sigma^2_{K'} = \sigma^2_K + \sigma^2_{K''}$$

La connaissance de la décomposition $\sigma_Z = \sigma_Y + \sigma_e$ permet donc, par cokrigeage, de diminuer l'erreur d'estimation.

3 - Cas particulier : corrélation intrinsèque.

Le système II,1 s'écrit alors :

$$\text{(II,3)} \quad \begin{cases} (\sigma^2_A\, a^j + \sigma_{AB}\, b^j)\ \sigma_{ij} = \alpha\, \sigma^2_A\, \sigma_{iv} + \beta\, \sigma_{AB}\, \sigma_{iv'} + \lambda_\alpha + \lambda_i \\ (\sigma_{AB}\, a^j + \sigma^2_B\, b^j)\ \sigma_{ij} = \alpha\, \sigma_{AB}\, \sigma_{iv} + \beta\, \sigma^2_B\, \sigma_{iv'} + \mu_\beta + \mu_i \\ \sum a^j = \alpha \qquad \lambda^i = 0 \quad i = 1,\dots n \qquad a^j = 0 \quad j = n+1\dots n+n' \\ \sum b^j = \beta \qquad \mu^i = 0 \quad i = n+1,\dots n+n' \qquad b^j = 0 \quad j = 1,\dots n \end{cases}$$

Dans ce cas particulier, on s'attend à trouver des simplifications dans le cokrigeage, en particulier si le coefficient de corrélation est égal à 1.

III - COKRIGEAGE EN CORRELATION INTRINSEQUE

1 - Coefficient de corrélation égal à + 1

Réécrivons le système II,3 pour faire apparaître le coefficient de corrélation $\rho = \frac{\sigma_{AB}}{\sigma_A \sigma_B}$

$$(1-\rho^2) a^j \sigma_{ij} = \alpha(1-\rho^2)\sigma_{iv} + \frac{\lambda_\alpha + \lambda_i}{\sigma_A^2} - \frac{\rho^2}{\sigma_{AB}} (\mu_B + \mu_i)$$

$$(1-\rho^2) b^j \sigma_{ij} = \beta(1-\rho^2)\sigma_{iv'} - \frac{\rho^2}{\sigma_{AB}} (\lambda_\alpha + \lambda_i) + \frac{\mu_B + \mu_i}{\sigma_B^2}$$

La condition $\rho = \pm 1$ impose

$$\lambda_\alpha + \lambda_i = (\mu_B + \mu_i) \sqrt{\frac{\sigma_A^2}{\sigma_B}}$$

Comme $\lambda_i = 0$ si $i = 1, \dots n$ $\mu_i = 0$ si $i = n+1, \dots n+n'$

on obtient pour λ_i et μ_i les valeurs suivantes :

$$\lambda_i = 0 \qquad \mu_i = \frac{\sigma_B}{\sigma_A} \lambda_\alpha - \mu_B \qquad i = 1, \dots n$$

$$\lambda_i = \frac{\sigma_A}{\sigma_B} \mu_B - \lambda_\alpha \quad \mu_i = 0 \qquad i = n+1, \dots n+n'$$

Le système II,3 devient alors :

$$\left.\begin{aligned} (\sigma_A^2 a^j + \sigma_{AB} b^j) \sigma_{ij} &= \alpha \sigma_A^2 \sigma_{iv} + \beta \sigma_{AB} \sigma_{iv'} + \lambda_\alpha + \lambda_i \\ (\sigma_{AB} a^j + \sigma_\beta^2 b^j) \sigma_{ij} &= \alpha \sigma_{AB} \sigma_{iv} + \beta \sigma_B^2 \sigma_{iv'} + \mu_\beta + \mu_i \end{aligned}\right\}$$

Comme $\rho^2 = 1$, nous avons $\sigma_{AB} = \sigma_A . \sigma_B$

On trouve finalement le simple système :

$$\left\{\begin{array}{l} \varphi^j \, \sigma_{ij} = \dfrac{\alpha\,\sigma_A}{\alpha\,\sigma_A + \beta\,\sigma_B}\,\sigma_{iv} + \dfrac{\beta\,\sigma_B}{\alpha\,\sigma_A + \beta\,\sigma_B}\,\sigma_{iv'} + L_\alpha + L_i \\ \sum \varphi^j = 1 \\ L_i = 0 \quad i = 1,\dots n \qquad L_i = L \qquad i > n \end{array}\right.$$

en posant $\varphi^j = \dfrac{\sigma_A\, a^j + \sigma_B\, b^j}{\alpha\,\sigma_A + \beta\,\sigma_\beta}$

On voit que φ^j se met sous la forme d'une somme de termes : Appelons φ_v^j les coefficients de krigeage de v par <u>tous</u> les sondages supposés être de même nature (par exemple, tous des A_i) et $\varphi_{v'}{}^j$ ceux du krigeage de v'.

De même, soient (M_A^j, ρ_A^j) les coefficients de A_j et B_j dans le krigeage de $\bar{A}$ seul et $(M_B{}^j, \rho_B{}^j)$ dans le krigeage de $\bar{B}$ seul. Par addition des solutions, on trouve :

$$\varphi^j = \frac{\alpha\,\sigma_A\,\varphi_v{}^j + \beta\,\sigma_B\,\varphi_{v'}{}^j + \alpha\,k(\sigma_A\,M_A{}^j + \sigma_B\,\rho_A{}^j) + \beta\,k'(\sigma_A\,M_B{}^j + \sigma_B\,\rho_B{}^j)}{\alpha\,\sigma_A + \beta\,\sigma_B}$$

Soit, en identifiant les a^j et b^j, et en utilisant les conditions $\Sigma\, a^j = \alpha$, $\Sigma\, b^j = \beta$:

$$\left\{\begin{array}{l} a^j = \alpha\,\varphi_v{}^j + \beta\,\dfrac{\sigma_B}{\sigma_A}\,\varphi_{v'}{}^j + \alpha\left[1 - \Sigma - \dfrac{\beta}{\alpha}\,\dfrac{\sigma_B}{\sigma_A}\,\Sigma'\right] M_A^j + \beta\left[\Sigma' - \dfrac{\alpha}{\beta}(1-\Sigma)\dfrac{\sigma_A}{\sigma_B}\right] M_B^j \\ b^j = \alpha\,\dfrac{\sigma_A}{\sigma_B}\,\varphi_v{}^j + \beta\,\varphi_{v'}{}^j + \alpha\left[1-\Sigma - \dfrac{\beta}{a}\,\dfrac{\sigma_B}{\sigma_A}\,\Sigma'\right]\rho_A{}^j + \beta\left[\Sigma' - \dfrac{\alpha}{\beta}(1-\Sigma)\dfrac{\sigma_A}{\sigma_B}\right]\rho_B{}^j \end{array}\right.$$

où l'on a posé $\sum = \sum_1^n \varphi_v^j \qquad \sum' = \sum_1^n \varphi_{v'}^j$

Pour mieux faire apparaître l'intérêt de cette décomposition, écrivons l'estimateur $(\alpha A_v + \beta B_{v'})^*$ tout entier :

$$(\alpha A_v + \beta B_{v'})^* = \alpha[\sum_1^n \varphi_v^j A_j + \sum_{n+1}^{n+n'} \varphi_v^j [\frac{\sigma_{AB}}{\sigma_B^2} (B_j - \overline{B}^*) + \overline{A}^*]]$$
$$+ \beta[\sum_1^n \varphi_{v'}^j [\frac{\sigma_{AB}}{\sigma_A^2}(A_j - \overline{A}^*) + \overline{B}^*] + \sum_{n+1}^{n+n'} \varphi_{v'}^j B_j]$$

On voit clairement la signification de cette relation : pour estimer A_v seul, on peut kriger sur les A_j et les B_j transformés par la régression $A_j^* = \frac{\sigma_{AB}}{\sigma_B^2} (B_j - \overline{B}^*) + \overline{A}^*$, et effectuer le krigeage sur l'ensemble comme s'il était entièrement composé de A_j ; on retrouve un résultat analogue à celui de la statistique classique :

Au regard d'une estimation linéaire, quand il y a corrélation intrinsèque avec coefficient $\rho = \pm 1$, on peut considérer les variables A et B comme liées par une relation linéaire.

Il en découle un certain nombre de résultats que nous énonçons ici, et dont nous verrons plus tard qu'ils sont liés à la transitivité du cokrigeage.

- le krigeage de B_i en un point où A est connu est la régression

$$B_i = \frac{\sigma_{AB}}{\sigma_A^2} (A_i - \overline{A}^*) + \overline{B}^*.$$

- Il est équivalent, pour cokriger B_v, de cokriger d'abord A_v, puis de transformer A_v par la régression $B_v^* = \frac{\sigma_{AB}}{\sigma^2_A} (A_v^* - \overline{A}^*) + \overline{B}^*$.

- Si, sur un volume v, A_v est entièrement connu, on estimera B_v par la régression.

Nous trouvons donc la régression linéaire comme premier exemple de cokrigeage, dans le cas très particulier de corrélation intrinsèque à coefficient égal à ± 1.

Cependant, la variance d'estimation de cette régression ne sera pas 0, comme on l'aurait trouvé si $\overline{A}$ et $\overline{B}$ étaient connus a priori.

$$B_k^* = \frac{\sigma_{AB}}{\sigma^2_A} A_k + B^* - A^* \frac{\sigma_{AB}}{\sigma^2_A}$$

$$E(B_k^* - B_k)^2 = [\frac{\sigma_{AB}}{\sigma^2_A} (A_k - \overline{A}) - (B_k - \overline{B}) + (B^* - A^* \frac{\sigma_{AB}}{\sigma^2_A} - \overline{B} + \overline{A} \frac{\sigma_{AB}}{\sigma^2_A})]^2$$

On reconnaît dans le deuxième membre la variance de krigeage de $B^* - A^* \frac{\sigma_{AB}}{\sigma^2_A}$. D'après l'additivité des variances de krigeage, on a :

$$\sigma_K^2(B_k^*) = \frac{1-\rho^2}{\sigma_A^2} + \sigma_K^2 \; (B^* - A^* \frac{\sigma_{AB}}{\sigma^2_A})$$

Dans le cas où $\rho = \pm 1$, on trouve comme variance dûe à la régression ponctuelle la variance de krigeage de la moyenne $\overline{B} - \overline{A} \frac{\sigma_{AB}}{\sigma^2_A}$.

2 - Coefficient de corrélation différent de ± 1

Nous venons de voir que, quelque soit la covariance $\sigma(h)$, lorsque $\rho^2 = 1$ la régression "statistique" était le meilleur estimateur croisé. Nous allons voir maintenant sur quelques exemples particuliers que pour $\rho^2 \neq 1$, ce résultat simple est faux.

1/ Estimation de A_k quand B_k est connu.

La régression statistique conduirait au résultat :

$$A_k^* = \frac{\sigma_{AB}}{\sigma^2_B} (B_k - \overline{B}^*) + \overline{A}^*$$

soient M^j et ρ^j les coefficients de A_j et B_j dans l'estimation de $\overline{A}^* - \frac{\sigma_{AB}}{\sigma^2_B} \overline{B}^*$. Ils vérifient :

$$\begin{cases} (\sigma_A^2 M^j + \sigma_{AB} \rho^j) \sigma_{ij} = \lambda''_\alpha + \lambda''_i \\ (\sigma_{AB} M^j + \sigma_B^2 \rho^j) \sigma_{ij} = \mu''_\alpha + \mu''_i \\ \lambda''_i = 0 \quad i \leq n \qquad \sum M^j = 1 \\ \mu''_i = 0 \quad i > n \qquad \sum \rho^j = -\frac{\sigma_{AB}}{\sigma^2_B} \end{cases}$$

La régression donne comme estimateur pour A_k^*

$$A_k{}^* = M^j A_j + \frac{\sigma_{AB}}{\sigma^2_B} B_k + \rho^j B_j$$

Décomposons les coefficients du krigeage de A_k, a^j et b^j ainsi :

$$\begin{cases} a^j = M^j + \varepsilon^j & \Rightarrow \sum \varepsilon^j = 0 \\ b^j = \rho^j + \dfrac{\sigma_{AB}}{\sigma^2_B} \delta^j_k + \eta^j & \sum \eta^j = 0 \end{cases}$$

Compte tenu des équations que vérifient M^j et ρ^j, il vient :

$$\begin{cases} (\sigma^2_A M^j + \sigma_{AB} \rho^j) \sigma_{ij} = \lambda''_\alpha + \lambda''_i \\ (\sigma_{AB} M^j + \sigma^2_B \rho^j) \sigma_{ij} = \mu''_\alpha + \mu''_i \\ \sum \varepsilon^j = 0 \qquad \lambda_i, \lambda''_i = 0 \qquad i \leqslant n \\ \sum \eta^j = 0 \qquad \mu_i, \mu''_i = 0 \qquad i > n \end{cases}$$

On voit que la solution banale $\varepsilon^j = \eta^j = 0$, $\forall j = 1, \ldots n+n'$ n'existe que pour $(1-\rho^2)\sigma_{iK} = 0$. Pour $(1-\rho^2)\sigma_{iK} \neq 0$, on trouve pour ε^j et y^j des valeurs non nulles, si bien que dans ce cas là, la régression linéaire $A_k = \dfrac{\sigma_{AB}}{\sigma^2_B} (B_k - \overline{B}^*) + A^*$ <u>n'est pas</u> le cokrigeage de A_k par les B_j et A_j.

<u>2/ Estimation de A_v après régression sur les B_j.</u>

De même, supposons que pour estimer A_v, on remplace tous les B_j par des A^*_j calculés par régression

$$A^*_j = \frac{\sigma_{AB}}{\sigma^2_B} (B_j - \overline{B}^*) + \overline{A}^*$$

on a bien $E(A_j^*) = \overline{A} = E(A_j)$

Pour estimer A_v^*, on fait une pondération du type :

$$A_v^* = a'^j\ A_j + b'^j\ A_j^* \Rightarrow E(A_v^* - A_v) = 0 \Rightarrow \sum a'^j + \sum b'^j = 1.$$

$$A_v^* = a'^j\ A_j + b'^j\ [\frac{\sigma_{AB}}{\sigma^2_B}\ (B_j - B^*) + \overline{A}^*]$$

Soient M_j et ρ^j les estimateurs de $\overline{A}^* - \frac{\sigma_{AB}}{\sigma^2_B}\ \overline{B}^*$.

On forme donc l'estimateur :

$$A_v^* = a'^j\ A_j + b'^j\ [\frac{\sigma_{AB}}{\sigma^2_B}\ B_j + M^j\ A_j + \rho^j\ B_j]$$

$$A_v^* = [a'^j + M^j\ (\sum b'^j)]\ A_j + [\frac{\sigma_{AB}}{\sigma^2_B}\ b'^j + (\sum b'^j)\rho^j]\ B_j$$

Notre estimateur est donc équivalent au krigeage du type $A_v = a^j\ A_j + b^j\ B_j$ où les a^j et b^j vérifient le système II,3.

Par identification, il vient :

$$\begin{cases} a'^j + M^j \sum b'^j = a^j \\ \frac{\sigma_{AB}}{\sigma^2_B}\ b'^j + (\sum b'^j)\ \rho^j = b^j \end{cases}$$

On s'aperçoit alors que, n'ayant que la condition $\sum a'^j + \sum b'^j = 1$, il est impossible de déterminer a'^j et b'^j. On le voit encore mieux en reportant ces valeurs de a'^j et b'^j dans le système (II,3).

On trouve alors :

$$(a'^j + \rho^2 b'^j)\ \sigma_{ij} = C_i + L_\alpha + L_i$$

$$(a'^j + b'^j)\ \sigma_{ij} \quad = C_i + M_\beta + M_i$$

$$\sum a'^j + \quad b'^j = 1. \quad L_i = 0 \quad i \leqslant n \qquad M_i = 0 \qquad i \geqslant n$$

Le système n'est complet que pour $\rho^2 = 1$. Il redonne alors pour a'^j et b'^j les valeurs p^j, solution du krigeage de A_v à partir de tous les points considérés comme des A_j. Si $\rho^2 \neq 1$, le système ci-dessus est incomplet.

3/ Estimation de A_v après régression sur B_v^*.

Donnons enfin un dernier exemple d'emploi abusif de la régression qu'on rencontre souvent dans le domaine minier : Soit B_v estimé par $B_v^* = c^j\ A_j + d^j\ B_j$.

Très souvent, les mineurs estiment A_v par une régression sur B_v^*

$$A_v^* = \frac{\sigma_{AB}}{\sigma^2_B}\ (B_v^* - \overline{B^*}) + \overline{A^*}$$

Soient M^j et ρ^j les véritables estimateurs optimaux de

$$\overline{A^*} - \frac{\sigma_{AB}}{\sigma^2_B}\ \overline{B^*}$$

$$A_v^* = (M^j + \frac{\sigma_{AB}}{\sigma^2_B}\ c^j)\ A_j + (\frac{\sigma_{AB}}{\sigma^2_B}\ d^j + \rho^j)\ B^j$$

soient a^j et b^j les véritables estimateurs optimaux de A_v. Si le procédé est valable, on doit avoir :

$$a^j = M^j + \frac{\sigma_{AB}}{\sigma^2_B}\ c^j \qquad\qquad b^j = \frac{\sigma_{AB}}{\sigma^2_B}\ d^j + \rho^j$$

en reportant ces valeurs dans les équations d'optimum, il vient :

$$\begin{cases} \mu_B'' + \mu_i'' + \mu_B' + \mu_i' = \mu_B + \mu_i \\ \left(\frac{\sigma_{AB}^2}{\sigma_B^2} - \sigma_A^2\right) C_i = \lambda_\alpha - \lambda_\alpha'' + \lambda_i - \lambda_i'' + \lambda_\alpha'' + \lambda_i' \end{cases}$$

On s'aperçoit encore que le procédé n'est légitime que si $\rho^2 = 1$.

De ces différents contre-exemples, on peut tirer la conclusion que, sauf si $\rho^2 = 1$, la régression linéaire n'est pas, en général, le cokrigeage. Cependant l'étude du cokrigeage d'un point où l'une des valeurs A_j ou B_j est connue nous a montré que, dans ce cas là, on pouvait retrouver la régression comme estimateur, pourvu que les covariances $\sigma_{A_iA_j}$ et $\sigma_{A_iB_j}$ soient d'un type particulier : c'est ce que nous allons voir maintenant.

IV - LA REGRESSION COMME COKRIGEAGE AUX GRANDES MAILLES

Nous nous plaçons dans le cas général, (sans supposer de corrélation intrinsèque), mais les trois fonctions $\sigma_A^2(h)$, $\sigma_B^2(h)$ et $\sigma_{AB}(h)$ sont supposées exister et avoir des portées a, b, c inférieures à la distance entre les sondages.

Nous aurons donc, en posant $\sigma_A^2 = \sigma_A^2(0)$ etc ...

$$\sigma_{A_iA_j} = \sigma_A^2\, \delta_{ij} \qquad \sigma_{B_iB_j} = \sigma_B^2\, \delta_{ij} \qquad \sigma_{A_iB_j} = \sigma_{AB}\, \delta_{ij}$$

Le système devient :

$$(\sigma^2_A\, a^j + \sigma_{AB}\, b^j)\, \delta^{ij} = \alpha\, \sigma_{A_i A_v} + \beta\, \sigma_{A_i B_{v'}} + \lambda\alpha + \lambda_i$$

$$(\sigma_{AB}\, a^j + \sigma^2_B\, b^j)\, \delta^{ij} = \alpha\, \sigma_{B_i A_v} + \beta\, \sigma_{B_i B_{v'}} + \mu\beta + \mu_i$$

Ce qui donne :

$$\begin{cases} a^i = \alpha \dfrac{\sigma_{A_i A_v}}{\sigma^2_A} + \beta \dfrac{\sigma_{A_i B_{v'}}}{\sigma^2_A} + L\alpha \\ \sum a^i = \alpha \end{cases} \qquad i = 1, \ldots n$$

$$\begin{cases} b^i = \alpha \dfrac{\sigma_{B_i A_v}}{\sigma^2_B} + \beta \dfrac{\sigma_{A_i B_{v'}}}{\sigma^2_B} + L\beta \\ \sum b^i = \beta \end{cases} \qquad i = n+1, \ldots n+n'$$

<u>1/ Krigeage de A en un point ou B est connu.</u>

On a $\alpha = 1$, $\beta = 0$. On cherche A^*_k au point k $(n+1, \ldots n+n')$. La seule covariance du second membre est $\sigma_{A_i A_k}$ ou $\sigma_{B_i A_k}$. D'après les hypothèses que nous avons faites, $i \neq k$, $i = 1, \ldots n$ donc $\sigma_{A_i A_k} = 0 \quad \forall\ i = 1, \ldots n$

$$\sigma_{B_i A_k} = \sigma_{AB}\, \delta^{ik} \qquad i = n+1, \ldots k, \ldots n+n'$$

On trouve donc immédiatement :

$$\begin{cases} a^i = \dfrac{1}{n} \\ b^i = \dfrac{\sigma_{AB}}{\sigma^2_B} \left(\delta^{ik} - \dfrac{1}{n'}\right) \end{cases}$$

Ce qui donne comme estimateur de A_k

$$A_k^* = \frac{\sigma_{AB}}{\sigma_B^2} (B_k - \frac{\Sigma B^j}{n'}) + \frac{\Sigma A_j}{n}$$

Or, à l'aide du même système, nous trouvons comme estimateur de $\overline{A}$ et $\overline{B}$:

$$\overline{A} : \begin{cases} a^j = \frac{1}{n} \\ b^j = 0 \end{cases} \qquad \overline{B} = \begin{cases} a^j = 0 \\ b^j = \frac{1}{n'} \end{cases}$$

Dans le cas de grandes mailles, l'estimation de $\overline{A}$ et $\overline{B}$ se fait seulement à partir des A_j et des B_j respectivement. Il est d'ailleurs aisé de voir, sur les équations, que pour qu'il en soit ainsi, il suffit que $\sigma_{A_iB_j} = 0 \quad \forall$ i, j ; $\overline{A}$ est alors estimé comme si seuls les A_j étaient connus. (Cas où A et B ne sont jamais connus aux mêmes points).

A l'aide de cette remarque, nous trouvons pour A_k^*

$$A_k^* = \frac{\sigma_{AB}}{\sigma_B^2} (B_k - \overline{B}^*) + \overline{A}^*$$

c'est-à-dire la régression statistique ordinaire, puisque les σ_A^2, σ_B^2, σ_{AB} ont le sens de $\sigma_A^2(o)$, $\sigma_B^2(o)$, $\sigma_{AB}(o)$ qui sont, dans le cas des grandes mailles, pratiquement égales aux variances et covariances statistiques calculées directement sur les données.

2/Krigeage d'un grand panneau.

Nous nous plaçons dans l'hypothèse de corrélation intrinsèque. Nous supposons le grand panneau v tel que la distance de tous les sondages à sa frontière soit supérieur à la portée.

Nous savons que dans ce cas, on a :

$$\sigma_{iv} = \frac{K}{V} \quad \text{si i intérieur à V}$$

$$\sigma_{iv} = 0 \quad \text{si i extérieur à V}$$

où K est une constante donnée par : $K = \int_{R^2} C(h)dh$

Supposons que, sur les n sondages A_i, $n\nu_\alpha$ soient intérieurs à v. De même $n'\nu_B$ sondages B_i intérieurs à v.

Les équations deviennent :

$$\begin{cases} a^i = \frac{K}{V} + L\alpha & i = 1, \ldots n\,\nu_\alpha \\ a^i = L\alpha & i = n\nu_\alpha, \ldots n \\ \sum a^i = 1 & \end{cases}$$

$$\begin{cases} b^i = \frac{\sigma^2_{AB}}{\sigma^2_B}\ \frac{K}{V} + L\beta & i = n, \ \ldots n+ n'\nu_B \\ b^i = L\beta & i = n+n'\nu_B, \ldots n+n' \\ \sum b^i = 0 & \end{cases}$$

Soit :

$$\begin{cases} a^i = (1 - \nu_\alpha)\ \frac{K}{V} + \frac{1}{n} & \text{i intérieur à v} \\ a^i = - \nu_\alpha \frac{K}{V} + \frac{1}{n} & \end{cases}$$

$$\begin{cases} b^i = \frac{\sigma_{AB}}{\sigma^2_B}\ (1 - \nu_B)\ \frac{K}{V} \\ b^i = - \frac{\sigma_{AB}}{\sigma^2_B}\ \nu\ \ \frac{K}{V} \end{cases}$$

On voit immédiatement, sur cet exemple simple, que le poids accordé aux B_i est identique à celui qu'on accorderait aux $A_i^* = \frac{\sigma_{AB}}{\sigma^2_B} B_i + \overline{A}^* - \frac{\sigma_{AB}}{\sigma^2_B} \overline{B}^*$ en les considérant comme de vrais A_i.

La régression statistique n'apparaît comme un cokrigeage que dans le cas des grandes mailles ; s'il n'en est rien, elle est un estimateur non biaisé, mais pas optimal. Nous allons voir que c'est la raison pour laquelle les problèmes que nous avons essayé de faire plus haut (krigeage après régression, ou régression après krigeage) étaient insolubles : on a le droit de kriger après régression que si celle-ci est l'estimateur optimal et qu'il y a corrélation intrinsèque ; c'est une conséquence de la transitivité des cokrigeages.

V - TRANSITIVITE DU COKRIGEAGE EN CORRELATION INTRINSEQUE

Nous conservons les mêmes notations que précédemment. Supposons que dans un premier temps, on krige la valeur de A aux points où les B sont connus.

$$A_j^* = M_j^k A_k + \rho_j^k B_k \qquad j \quad n, \qquad k = 1,\dots n, n+1,\dots n+n'$$

Les M_j^k et ρ_j^k vérifient :

$$\begin{cases} M_j^k \sigma_{A_iA_k} + \rho_j^k \sigma_{A_iB_k} = \sigma_{A_iA_j} + \lambda''_{\alpha j} + \lambda''_{ij} \\ M_j^k \sigma_{A_kB_i} + \rho_j^k \sigma_{B_kB_i} = \sigma_{B_iA_j} + \mu''_{\beta j} + \mu''_{ij} \\ \sum_k M_j^k = 1 \quad \lambda''_{ij} = 0 \quad i = 1,\dots n \quad M_j^k = 0 \quad k = n+1\dots \\ \sum_k \rho_j^k = 0 \quad \mu''_{ij} = 0 \quad i = n+1,\dots n+n' \ \rho_j^k = 0 \quad k = 1,\dots n \end{cases}$$

$$\sigma^2_{kj} = \sigma^2_{A_j} - (M^k_j \, \sigma_{A_k A_j} + \rho^k_j \, \sigma_{B_k A_j}) + \lambda''_\alpha$$

On cherche alors à kriger A dans un volume v par une pondération du type :

$$A^*_v = a'^j A_j + b'^j A^*_j$$

La condition de non biais donne $\Sigma(a'^j + b'^j) = 1.$

Si on développe l'estimateur A^*_v il vient :

$$A^*_v = a'^j \, A_j + b'^j \, (M^k_j \, A_k + \rho^k_j \, B_k)$$

$$A^*_v = A_j(a'^j + b'^k \, M^j_k) + B_j \, (b'^k \, \rho^j_k)$$

Le procédé utilisé est équivalent à un krigeage direct de A_v de la forme :

$$A^*_v = a^j \, A_j + b^j \, B_j$$

On trouve donc comme solution pour a'^j et b'^j :

$$\begin{cases} a'^j + b'^k \, M^j_k = a^j \\ b'^k \, \rho^j_k = b^j \end{cases}$$

Cherchons directement, à l'aide des équations que doivent vérifier a^j et b^j, les équations que vérifient a'^j et b'^j.

$$a'^j \, \sigma_{A_i A_j} + b'^k \, (M^j_k \, \sigma_{A_i A_j} + \rho^j_k \, \sigma_{A_i B_j}) = \sigma_{A_i A_v} + \lambda_\alpha + \lambda_i$$

$$a'^j \, \sigma_{A_j B_i} + b'^k \, (M^j_k \, \sigma_{A_j B_i} + \rho^j_k \, \sigma_{B_i B_j}) = \sigma_{B_i A_v} + \mu_\beta + \mu_i$$

et à l'aide des équations de définition de M_k^j et ρ_k^j, il vient :

$$a'^j \sigma_{A_iA_j} + b'^k (\sigma_{A_iA_k} + \lambda''_{\alpha k} + \lambda''_{ik}) = \sigma_{A_iA_v} + \lambda_\alpha + \lambda_i$$

$$a'^j \sigma_{A_jB_i} + b'^k (\sigma_{B_iA_k} + \mu''_{\beta k} + \mu''_{ik}) = \sigma_{B_iA_v} + \mu_\beta + \mu_i$$

Soit finalement :

$$(a'^j + b'^j) \sigma_{A_iA_j} = \sigma_{A_iA_v} + \lambda_\alpha - b'^k \lambda''_{\alpha k} + \lambda_i - b'^k \lambda''_{ik}$$

$$(a'^j + b'^j) \sigma_{A_jB_i} = \sigma_{B_iA_v} + \mu_\beta - b'^k \mu''_{\beta k} + \mu_i - b'^k \mu''_{ik}$$

Dans le cas général, ce système n'a rien de notable. Par contre, dans le cas de corrélation intrinsèque, il vient :

$$(a'^j + b'^j) \sigma_{ij} = \sigma_{iv} + \frac{\lambda_\alpha - b'^k \lambda''_{\alpha k} + \lambda_i - b'^k \lambda''_{ik}}{\sigma^2{}_A}$$

$$(a'^j + b'^j) \sigma_{ij} = \sigma_{iv} + \frac{\mu_B - b'^k \mu''_{\beta k} + \mu_i - b'^k \mu''_{ik}}{\sigma_{AB}}$$

On doit donc avoir, pour tout i :

$$\lambda_\alpha - b'^k \lambda''_{\alpha k} + \lambda_i - b'^k \lambda''_{ik} = \frac{\sigma^2{}_A}{\sigma_{AB}} (\mu_\beta - b'^k \mu''_{\beta k} + \mu_i - b'^k \mu''_{ik})$$

Comme on a de plus $\lambda_i, \lambda''_{ik} = 0 \quad \forall\, i \leq n$

$\mu_i, \mu''_{ik} = 0 \quad \forall\, i > n$

On voit qu'on doit prendre :

$$\lambda_i = b'^k \lambda''_{ik} \qquad \mu_i = b'^k \mu''_{ik}$$

$$\sigma_{AB} (\lambda_\alpha - b'^k \lambda''_{\alpha k}) = \sigma^2_A (\mu_\beta - b'^k \mu''_{\beta k})$$

Moyennant ces conditions, les deux équations sont équivalentes, et il reste :

$$\begin{cases} (a'^j + b'^j)\ \sigma_{ij} = \sigma_{iv} + \dfrac{\lambda_\alpha - b'^k \lambda''_{\alpha k}}{\sigma^2_A} \\ \Sigma(a'^j + b'^j) = 1 \end{cases}$$

On reconnait l'équation de définition des coefficients p^j, dans le krigeage de A_v à partir de l'ensemble des points considérés comme étant des A_j.

$$\begin{cases} p^j\ \sigma_{A_i A_j} = \sigma_{A_i A_v} + \lambda \\ \Sigma\ p^j = 1 \end{cases}$$

On en tire la valeur de $\lambda_\alpha = \lambda + b'^k \lambda''_{\alpha k}$

Dans le cas de corrélation intrinsèque, il y a transitivité des résultats : on peut kriger tous les A^*_j aux points où les B^*_j sont connus et travailler ensuite sur ces A^*_j comme s'ils étaient connus directement.

On peut se demander si la relation est aussi simple pour les variances de krigeage.

Soit σ^2_k la variance du cokrigeage :

$$\sigma_k^2 = \sigma_{A_v}^2 - (a^i\ \sigma_{A_iA_v} + b^i\ \sigma_{B_iB_v}) + \lambda_\alpha$$

D'autre part, les coefficients a'^j et b'^j sont égaux à p^j tel que : $p^j\ \sigma_{ij} = \sigma_{iv} + \frac{\lambda}{\sigma^2_A}$, dont la variance de krigeage est $\sigma^2_{K'}$

$\sigma^2_{k'} = \sigma_{A_v}^2 - \sigma_A^2\ p^i\ \sigma_{iv} + \lambda.$

Le terme $(a^i\ \sigma_{A_iA_v} + b^i\ \sigma_{B_iB_v}) = (\sigma_A^2\ a^i + \sigma_{AB}\ b^i)\ \sigma_{iv}$

$$\sigma^2_k = \sigma_{A_v}^2 - (\sigma^2_A\ a^i + \sigma_{AB}\ b^i)\ (p^j\ \sigma_{ij} - \frac{\lambda}{\sigma^2_A}) + \lambda_\alpha$$

$$\sigma^2_k = \sigma_{A_v}^2 - p^j\ [(\sigma^2_A\ a^i + \sigma_{AB}\ b^i)\ \sigma_{ij}] + \lambda + \lambda_\alpha$$

Or, par définition des a^i et b^i :

$$(\sigma^2_A\ a^i + \sigma_{AB}\ b^i)\ \sigma_{ij} = \sigma^2_A\ \sigma_{jv} + \lambda\ \alpha + \lambda_j$$

$$\sigma^2_k = \sigma^2_{A_v} - \sigma^2_A\ p^j\ \sigma_{jv} - \cancel{\lambda\alpha} - p^j\ \lambda_j + \cancel{\lambda\alpha} + \lambda$$

Soit $\sigma^2_k = \sigma^2_{k'} - p^j\ \lambda_j$

Dans le cokrigeage du point j, j > n, on a

$$\sigma^2_{kj} = \sigma^2_{A_j} - (M_j^k\ \sigma_{A_kA_j} + \rho_j^k\ \sigma_{B_kA_j}) + \lambda''_{\alpha j}$$

avec $(M_j^k\ \sigma_{A_KA_j} + \rho_j^k\ \sigma_{B_kA_j}) = \sigma^2_{A_j} + \lambda''_{\alpha j} + \lambda''_{jj}$

donc $\sigma^2_{kj} = - \lambda''_{jj}.$

D'autre part, on a vu plus haut que $\lambda_i = b'^k\ \lambda''_{ik}.$

Comme pour $j > n \quad p^j = b'^j$

$$p^j \lambda_j = b'^j\, b'^k\, \lambda''_{jk} = \Sigma\, b'^j\, b'^j\, \lambda''_{jj} + \sum_{j \neq k} \lambda''_{jk}\, b'^j\, b^k$$

D'où, pour la variance du krigeage de A_v

$$\sigma^2_k = \sigma^2_{k'} + (b'^j)^2\, \sigma_{kj} - \sum_{j \neq k} b'^j\, b^k\, \lambda''_{jk}$$

Comme on pouvait s'y attendre, la variance du cokrigeage de A_v par les A_j et les A_j^* n'est pas égale à la variance du krigeage de A_v par tous les A_j et B_k considérés comme des A_k.

En conclusion, <u>dans le cas de corrélation intrinsèque</u>, on peut remplacer les données B_k par leur cokrigeage A_k^* et considérer, pour déterminer leur pondérateurs dans le krigeage de A_v, qu'ils sont équivalents à la connaissance directe de A_k ; mais la variance de ce cokrigeage n'est pas égale à la variance du krigeage de A_v par les (A_j, A_k).

V - CAS PARTICULIER : A_i et B_j sont connus aux mêmes points.

Nous nous plaçons en corrélation intrinsèque et on suppose que l'on veut estimer A et B dans le même volume v.

$$\begin{cases} (\sigma^2_A\, a^j + \sigma_{AB}\, b^j)\, \sigma_{ij} = (\alpha\, \sigma^2_A + \beta\, \sigma_{AB})\, \sigma_{iv} + \lambda \\ (\sigma_{AB}\, a^j + \sigma^2_B\, b^j)\, \sigma_{ij} = (\alpha\, \sigma_{AB} + \beta\, \sigma^2_B)\, \sigma_{iv} + \mu \\ \Sigma\, a^j = \alpha \\ \Sigma\, b^j = \beta \end{cases}$$

Soit p^j le krigeage de A_v par les A seuls.

$$p^j \sigma_{ij} = \sigma_{iv} + \frac{\nu}{\sigma^2_A}$$

Il vient alors la solution immédiate :

$$\begin{cases} a^j = \alpha\, p^j \\ b^j = \beta\, p^j \end{cases}$$

On en conclut que, pour kriger A dans un volume quelconque, on n'utilisera que les A_j. Supposons cependant, qu'ayant déjà krigé A_v à partir des A_j, on veuille trouver une relation entre B_v^* et A_v^* ; on sait que c'est impossible si on ne suppose pas de plus qu'au moins $\overline{B}^*$ est connu.

Soit M^j l'estimateur de $\overline{B}^* = M^j\, B_j$

$$\left.\begin{cases} M^j \sigma_{ij} = \lambda \\ \Sigma\, M^j = 1 \end{cases}\right\} \rightarrow M^j = \frac{\sum_i S^{ij}}{S} \qquad (S^{ij} \text{ matrice inverse de } \sigma_{ij})$$

$$S = \sum_{ij} S^{ij}$$

On sait qu'alors M^j est aussi le krigeage de $\overline{A^*} = M^j\, A_j$.

Cherchons une estimation de B_v sous la forme :

$$B_v^* = C^j\, A_j + \overline{B^*} \quad \Longrightarrow \quad \Sigma\, C^j = 0$$

$$E[B_v^* - B_v]^2 = E[C^j(A_j - \overline{A}) + M^j(B_j - \overline{B}) - (B_v - \overline{B})]^2$$

$$= C^i C^j\, \sigma_{A_i A_j} + M^i M^j\, \sigma_{B_i B_j} + \sigma^2_{B_v}$$

$$+ 2\, C^j\, M_i\, \sigma_{A_j B_i} - 2\, C^j\, \sigma_{A_j B_v} - 2\, M^i\, \sigma_{B_i B_v}$$

Soit le système :

$$\begin{cases} \sigma^2{}_A \, C^i \, \sigma_{ij} + \sigma_{AB} \, M^{ij} \, \sigma_{ij} = \sigma_{AB} \, \sigma_{jv} + \lambda \\ \Sigma \, i^i = 0 \end{cases}$$

$$\begin{cases} C^i \, \sigma_{ij} = \dfrac{\sigma_{AB}}{\sigma^2{}_A} \, \sigma_{jv} + \lambda' \\ \Sigma \, i^i = 0 \end{cases}$$

La solution est immédiate :

$$C^i = \frac{\sigma_{AB}}{\sigma^2{}_A} (p^i - M^i)$$

ce qui donne :

$$B_v^* = \frac{\sigma_{AB}}{\sigma^2{}_A} (A_v^* - \overline{A}^*) + \overline{B}^*$$

Nous trouvons donc que la meilleure estimation de B_v^*, en supposant A_v^*, $\overline{A}^*$ et $\overline{B}^*$ connus, est la régression. Mais cette estimation n'est pas la meilleure possible, qui serait $B_v'^* = p^j \, B_j$.

En effet, la variance de cette estimation se met sous la forme :

$$\sigma^2_{est} = \sigma^2_{B_v^*} + \sigma^2{}_B \, (1 - \rho^2) \, \sigma^2_u$$

où σ^2_u est une quantité positive : on retrouve la remarque que le passage de A_v^* à B_v^* par régression n'est équivalent au krigeage que si $\rho^2 = \pm 1$.

Ces résultats sont évidemment valables aussi pour l'estimation d'un point.

COKRIGEAGE PARTICULIER : TENEUR MOYENNE D'UN GISEMENT STRATIFORME

Il arrive souvent que, dans un gisement stratiforme, on ait à estimer la teneur moyenne d'un panneau, connaissant un certain nombre de sondages sur lesquels on mesure la teneur moyenne t_i sur toute la couche, ainsi que la puissance totale p_i de la couche.

Si A(x) est l'accumulation de métal au mètre carré de surface, la quantité totale de minerai dans le volume v est donnée par

$$A_V = \int_V A(x)dx \qquad \text{(intégrale de surface)}$$

et le volume du panneau V est donnée par :

$$\mathcal{V} = \int_V p(x)dx$$

La teneur moyenne τ_V du panneau V est alors définie par

$$\tau_V = \frac{\int_V A(x)dx}{\int_V p(x)dx}$$

Si on considère A(x) et p(x) comme des F.A. de l'espace à deux dimensions, on voit qu'on peut leur associer la F.A. $\tau(x) = \frac{A(x)}{p(x)}$, qui n'est autre que la teneur moyenne d'une traversée de la couche au point x. Si A(x) et p(x) sont des F.A. stationnaires, rien ne nous garantit qu'il en est de même de $\tau(x)$; c'est en général un point qu'on est obligé de vérifier expérimentalement, en construisant le variogramme de $\tau(x)$. Ce variogramme ne nous est malheureusement pas d'un grand

/

secours, puisqu'il nous permettrait d'évaluer, par krigeage, des régularisations du type :

$$t_V = \frac{1}{V} \int_V \tau(x)dx$$

alors que ce sont des quantités $\tau_V = \dfrac{\frac{1}{V}\int A(x)dx}{\frac{1}{V}\int p(x)dx}$ qui nous intéressent.

Le problème de l'estimation de la V.A. τ_V est compliqué du fait que τ_V est un quotient de deux V.A., en général non indépendantes, dont on ne peut même pas connaître l'espérance mathématique : le seul cas simple est celui où A_V et p_V sont deux lois lognormales et alors τ_V est aussi lognormale, d'espérance $\overline{\tau_V} = \dfrac{\overline{A_V}}{\overline{p_V}}$.

Nous allons cependant chercher pour τ_V une formule d'approximation linéaire.

Supposons que $A(x)$ et $p(x)$ soient deux F.A. stationnaires de moyenne A et p et soit $\tilde{A}_V$ $\tilde{p}_V$, les valeurs effectives (réalisations) de A et p dans V. Nous savons que, dès que V est grand, la variance $\sigma^2_{A_V}$ des réalisations $\tilde{A}_V$ est faible, donc que $\tilde{A}_V$ ne diffère de $\overline{A}$ que d'une faible quantité ε_V. Ecrivons donc :

$$\tilde{A}_V = \overline{A} + \varepsilon_V$$

$$\tilde{p}_V = \overline{p} + \eta_V$$

$$\tilde{\tau}_V = \frac{\overline{A} + \varepsilon_V}{\overline{p} + \eta_V}$$

Ce quotient peut se représenter, en négligeant les erreurs du second ordre, par la formule :

$$\tilde{\tau}_V \neq \frac{\bar{A}}{\bar{p}} + \frac{\varepsilon_V - \frac{\bar{A}}{\bar{p}}\eta_V}{\bar{p}}$$

L'estimateur habituel des mineurs pour τ_V est

$$\tau_V^* = \frac{\frac{1}{n}(\Sigma\ A_i)}{\frac{1}{n}(\Sigma\ p_i)}$$ s'il y a n sondages à l'intérieur de V.

en écrivant pour chaque sondage, la décomposition :

$$A_i = \bar{A} + \varepsilon_i \qquad p_i = \bar{p} + \eta_i$$

$$\tau_V^* = \frac{\bar{A} + \frac{1}{n}\Sigma\ \varepsilon_i}{\bar{p} + \frac{1}{n}\Sigma\ \eta_i}$$

$\frac{1}{n}\Sigma\ \varepsilon_i = \bar{\varepsilon}$ apparaît comme la moyenne des écarts des A_i à $\bar{A}$. Cette moyenne sera petite dans les mêmes conditions que ε V est petit, c'est-à-dire dès que V se rapproche des dimensions des différentes portées de γ_A γ_p et γ_{Ap}.

Le même développement au premier ordre pour τ_V^* donne :

$$\tau_V^* = \frac{\bar{A}}{\bar{p}} + \frac{\bar{\varepsilon} - \frac{\bar{A}}{\bar{p}}\bar{\eta}}{p}$$

On trouve évidemment $E(T_V^*) = \frac{\bar{A}}{\bar{p}} = E(\tau_V)$

La variance d'estimation sera :

$$E(\tau_V^* - \tau_V)^2 = \frac{1}{p^2} E[(\bar{\varepsilon} - \varepsilon_V) - \frac{\bar{A}}{\bar{p}}(\bar{\eta} - \eta_V)]^2$$

$$= \frac{1}{p^2}[\sigma_A^{2E} + \frac{A^2}{p^2}\sigma_p^{2E} - 2\frac{\bar{A}}{\bar{p}}\sigma_{Ap}^E]$$

$$\sigma^{2E}_{\tau_V} = \frac{A^2}{p^2}\left[\frac{\sigma^{2E}_A}{\overline{A}^2} + \frac{\sigma^{2E}_p}{\overline{p}^2} - 2\,\frac{\sigma^{E}_{Ap}}{A\ p}\right]$$

σ^{2E}_A, $\sigma^{2\,E}_{p}$, σ^E_{Ap} étant les variances d'estimation de A_V, p_V et la covariance d'estimation de A_V et p_V. Comme ces variances sont calculées d'après la position des mêmes points expérimentaux.

On a $\sigma^{2\,E}_{A} = \mathcal{E}\,(\gamma_A)$, $\sigma^{2\,E}_{p} = \mathcal{E}\,(\gamma_p)$ ...

$$\sigma^2_{\tau_V} = \frac{A^2}{p^2}\left[\frac{1}{A^2}\,\mathcal{E}\,(\gamma_A) + \frac{1}{p^2}\,\mathcal{E}\,(\gamma_p) - \frac{4}{A^p}\,\mathcal{E}\,(\gamma_{Ap})\right]$$

$$\frac{\sigma^2_{\tau_V}}{[E(\tau_V)]^2} = \mathcal{E}\left[\gamma\left(\frac{A}{\overline{A}} - \frac{p}{\overline{p}}\right]\right.$$

La variance d'estimation relative $\dfrac{\sigma^2_{\tau_V}}{{\tau_V}^2}$ de τ_V se calcule comme si on estimait τ_V par $\tau_V^* = \frac{1}{n}\,\Sigma\,\tau_i$, en utilisant le variogramme "réduit"

$$\gamma = \frac{1}{2}\,E\left\{\left(\frac{A(x+h)}{\overline{A}} - \frac{p(x+h)}{\overline{p}}\right) - \left(\frac{A(x)}{\overline{A}} - \frac{p(x)}{p}\right)^2\right\}$$

Ce variogramme réduit se calcule sur les points expérimentaux où sont connues les valeurs $\dfrac{A_i}{\overline{A}^*} - \dfrac{p_i}{\overline{p}^*}$, $\overline{A}^*$ et $\overline{p}^*$ étant elles-mêmes des estimations de $\overline{A}$ et $\overline{p}$.

Dans le cas où l'on estime τ_G, teneur moyenne de tout le gisement, cette formule est certainement plus valable, car toutes les erreurs sont faibles et les approximations faites sont alors licites.

Nous allons, dans le même ordre d'idées de simplifications, essayer de trouver la solution de l'estimation de τ_V dans un cokrigeage.

Cherchons pour τ_V un estimateur de la forme :

$\tau_V^* = \dfrac{\lambda^i A_i}{\mu^i p_i}$ Les A_i et p_i provenant des mêmes points expérimentaux.

$$\tau_V^* = \frac{\bar{A}\, \Sigma\, \lambda^i + \lambda^i\, \varepsilon_i}{\bar{p}\, \Sigma\, \mu^i + \mu^i\, \eta_i}$$

soit $\quad \Lambda = \Sigma\, \lambda^i \qquad M = \Sigma\, \mu^i$

$$\tau_V = \frac{\Lambda}{M}\, \frac{\bar{A} + \frac{\lambda^i}{\Lambda}\, \varepsilon_i}{\bar{p} + \frac{\mu^i}{M}\, \eta_i}$$

Dans les mêmes hypothèses que plus haut, ($\frac{\lambda^i}{\Lambda}\, \varepsilon_i$ et $\frac{\mu^i}{M}\, \eta_i$ petits devant $\bar{A}$ et $\bar{p}$, c'est-à-dire τ_V effectivement krigé par un grand nombre de sondages) on développe le quotient au premier ordre

$$\tau_V^* = \frac{\Lambda}{M}\, \left[\frac{\bar{A}}{\bar{p}} + \frac{\frac{\lambda^i}{\Lambda}\, \varepsilon_i - \frac{\bar{A}}{\bar{p}}\, \frac{\mu^i}{M}\, \eta_i}{\bar{p}}\right]$$

La condition de non biais, appliquée aux développements du 1er ordre, donne :

$$E(\tau_V^* - \tau_V) = \left(\frac{\Lambda}{M} - 1\right) \frac{\bar{A}}{\bar{p}}$$

La seule condition est donc $\Lambda = M$. Du fait de la forme de l'estimateur $\tau_V^* = \dfrac{\lambda^i\, A_i}{\mu^i\, p_i}$, on peut prendre pour λ^i et μ^i deux systèmes proportionnels sans changer τ_V^* : nous imposerons donc la condition $\Lambda = M = 1$.

L'estimateur s'écrit donc :

$$\begin{cases} \tau_V^* = \dfrac{\overline{A}}{\overline{p}} + \dfrac{\lambda^i \ \varepsilon_i - \dfrac{\overline{A}}{\overline{p}} \mu^i \ \eta_i}{\overline{p}} \\ \\ \Sigma \lambda^i = \Sigma \mu^i = 1 \end{cases}$$

Comme τ_V a été développé sous la forme :

$$\tau_V = \frac{\overline{A}}{\overline{p}} + \frac{\varepsilon_V - \dfrac{A}{p} \eta_V}{\overline{p}}$$

On voit qu'on est ramené au cokrigeage de ε_V et η_V par les ε_i et η_i.

On retrouve alors le système :

$$\lambda^j \ \sigma_{A_i A_j} - \frac{\overline{A}}{\overline{p}} \mu^j \ \sigma_{A_i p_j} = \sigma_{A_i A_V} - \frac{A}{p} \sigma_{A_i p_V} + \lambda_\alpha$$

$$\lambda^j \ \sigma_{A_j p_i} - \frac{\overline{A}}{\overline{p}} \mu^j \ \sigma_{p_i p_j} = \sigma_{A_V p_i} - \frac{A}{p} \sigma_{p_i p_V} + \lambda_\beta$$

$$\Sigma \lambda^i = 1$$

$$\Sigma \mu^i = 1$$

On voit qu'en posant $\quad a^j = \lambda^j \qquad \alpha = 1$

$$b^j = -\frac{A}{p} \mu^j \qquad \beta = -\frac{A}{p}$$

on tombe dans le système :

$$a^j \sigma_{A_i A_j} + b^j \sigma_{A_i p_j} = \alpha \sigma_{A_i A_v} + \beta \sigma_{A_i p_v} + \lambda\alpha$$

$$a^j \sigma_{A_j p_i} + b^j \sigma_{p_i p_j} = \alpha \sigma_{p_i A_v} + \beta \sigma_{p_i p_v} + \lambda\beta$$

$$\begin{cases} \Sigma a^j = \alpha \\ \Sigma b^j = \beta \end{cases}$$

Système de cokrigeage lorsque les paramètres sont connus aux mêmes points.

Dans la pratique, comme $A_i = p_i t_i$, il arrive souvent que les variables A_i et p_i soient en corrélation intrinsèque. On a vu alors que, si on appelle M^j les coefficients du krigeage de A_v par les A_i et p_v par les p_i que dans ce cas :

$$\begin{matrix} a^j = \alpha M^j \\ b^j = \beta M^j \end{matrix} \implies \begin{cases} \lambda^j = M^j \\ \mu^j = M^j \end{cases}$$

et l'estimateur τ_V devient :

$$\tau_V^* = \frac{M^j A_j}{M^j p_j} = \frac{A_V^*}{p_V^*}$$

On retrouve dans ce cas particulier la pratique minière habituelle qui consiste à estimer τ_V^* par quotient des estimateurs de A_V et p_V : cette pratique revient à estimer τ_V^* par une combinaison linéaire des teneurs ponctuelles tj puisque $A_j = p_j t_j$.

$$\tau_V^* = \frac{\sum_j (M^j p^j) t_j}{\sum_j M^j\ p^j} \qquad \text{soit } \tau_V^* = u^j\ t_j \qquad u^j = \frac{M^j p^j}{\Sigma\ M^j p^j}$$

Par contre, dans le cas où il n'y a pas corrélation intrinsèque, $\lambda^j \neq \mu^j$ et on a besoin de la connaissance des 3 variogrammes γ_A, γ_p et γ_{Ap} pour déterminer les coefficients λ^j et μ^j.

On peut alors chercher pour τ_V^* un autre "krigeage" (si tant est que la forme $\frac{\lambda^j\ A_j}{\mu^j\ p_j}$ soit un krigeage), plus simple du type

$$\tau_V^* = \frac{\lambda^j\ A_j}{\lambda^j\ p_j}$$

Comme plus haut, on peut choisir $\Sigma\ \lambda_i = 1$.

Le développement au 1er ordre donne :

$$\tau_V^* = \frac{\overline{A}}{\overline{p}} + \lambda^i\ \left(\frac{\varepsilon_i - \frac{A}{p}\ \eta_i}{\overline{p}}\right)$$

Posons $\quad U = \frac{\overline{A}}{\overline{p}}\ \left(\frac{A}{\overline{A}} - \frac{p}{\overline{p}}\right).$

On a $\quad \tau_V^* = \frac{\overline{A}}{\overline{p}} + \lambda^i\ U_i$

$$\tau_V = \frac{\overline{A}}{\overline{p}} + U_V$$

Le système de krigeage se réduit alors simplement à :

$$\begin{cases} \lambda^j\ \sigma_{u_i u_j} = \sigma_{u_i u_V} + C \\ \Sigma\ \lambda^i = 1. \end{cases}$$

Les covariances $\sigma_{u_i u_j}$ étant calculées à partir du variogramme

$$\gamma_U = \frac{\overline{A}^2}{p^2} \left[\frac{\gamma_A}{A^2} + \frac{\gamma_p}{p^2} - 2 \frac{\gamma_{Ap}}{A\ p}\right]$$

ou, ce qui est équivalent, à partir du variogramme "réduit"

$$\gamma_Z = \frac{\gamma_U}{\frac{\overline{A}^2}{p^2}}$$

On trouve $\begin{cases} \lambda^j \gamma_{r_i r_j} = \gamma_{r_i r_j} + C \\ \Sigma \lambda^j = 1. \end{cases}$

Ce système donne une variance de krigeage $\sigma^2_{K_r}$ et on aura

$$\sigma^2{}_K = \frac{A^2}{p^2} \ \sigma^2_{K_2}$$

S'il y a corrélation intrinsèque, on retrouve évidemment que $\gamma_p = \left[\frac{\sigma^2_A}{A^2} + \frac{\sigma^2_p}{p^2} - \frac{2\ \sigma_{Ap}}{A\ p}\right] \gamma$

et le résultat $\lambda^j = M^j$ vu plus haut.

On remarque que, dans ce cas aussi, la formule $\tau^*_V = \dfrac{\lambda^j\ A_j}{\gamma^j\ p_j}$

revient à prendre une combinaison linéaire des teneurs $\tau^*_V = u^j\ t_j$.

AUTRE PRESENTATION DU PROBLEME

Reprenons la définition de τ V

$$\tau_V = \frac{\int_V A(x)dx}{\int_V p(y)dy}$$

en interprétant A(x) comme le produit de la puissance par la teneur moyenne de la couche, τ_V peut s'écrire :

$$\tau_V = \int_V \frac{p(x)}{\int_V p(y)dy}\, t(x)dx$$

La fonction $M(x) = \frac{p(x)}{\int_V p(y)dy}$ peut être considérée, dans l'estimation d'un volume V bien défini, comme une simple fonction de pondération de somme unité, <u>déterminée mais inconnue</u>.

On peut alors, si t(x) peut être considérée comme une F.A. intrinsèque (ou stationnaire), kriger directement τ_V à partir des t_i :

$$\tau_V^* = a^i\, t_i, \text{ les } a^i \text{ vérifiant :}$$

$$\begin{cases} a^j\, \sigma_{ij} = \sigma_{pi} + \lambda \\ \Sigma \;\; a^j = 1 \end{cases}$$

avec $\sigma_{pi} = \int_V \frac{p(x)}{p(y)dy} \sigma(x-x_i)\, dx$

Les covariances σ_{pi} sont alors inconnues, et il apparaît clairement que la solution a^j dépendra fortement de la façon dont seront estimés les σ_{pi} : Montrons-le sur deux exemples.

A son tour, σ_{pi} peut être interprété en terme d'estimation transitive. Ecrivons le système sous la forme :

$$a^j\, \sigma_{ij} = \sigma_{pi} + \frac{1}{S}[1 - \sum_j \sigma_{pi}\, S^{ij}]$$

où S^{ij} est la matrice inverse de σ_{ij} et $S = \sum_{ij} S^{ij}$.

On peut écrire :

$$[\int p(y)dy]a^j \sigma_{ij} = \int_V p(x)\sigma(x-x_i)dx + \frac{1}{S}[\int p(y)dy - \sum_j S^{ij} \int_V p(x)\sigma(x-x_i)dx$$

Si nous estimons les quantités $\int p(y)dy$ et $\int p(x)\sigma(x-x_i)dx$ en transitif par les formules

$$\int_V p(y)dy = \Sigma\, p^k\, \Delta_s \qquad k \text{ intérieur à } V$$

$$\int_V p(x)\sigma(x-x_i)dx = p^k\, \sigma_{ki}\, \Delta_s$$

en posant $M^k = \frac{p^k}{\Sigma\, p^k}$, il vient simplement :

$$a^j\, \sigma_{ij} = M^k\, \sigma_{ki} + \frac{1}{S}\, [1 - \sum_j S^{ij}\, M^k\, \sigma_{ki}]$$

dont la solution est évidente et donne :

$$a^j = M^j \qquad j = 1, \dots k$$

$$a^j = 0 \qquad j > k$$

et le krigeage de τ_V est simplement :

$$\tau_V = \frac{\Sigma\, p_i\, t_i}{\Sigma\, p_i} \qquad i \text{ intérieur à } V$$

τ_V est la simple moyenne pondérée des teneurs intérieures à V par les puissances correspondantes.

Si par contre on cherche à estimer les quantités $\int_V p(y)dy$ et $\int_V p(x)\sigma(x-x_i)dx$ par un krigeage transitif du type :

$$[\int_V p(y)dy]^* = \lambda^i \; p_i \quad \forall i \quad \text{dans le gisement}$$

avec $\lambda \; g_{ij} = g_{iV}$ g covariogramme transitif des puissances

$$[\int_V p(x)\sigma(x-x_i)dx]^* = \mu_i^k \; p_k$$

$$\mu_i^k \; g_{kj} = g_j\sigma_i \qquad (g_j\sigma_i = \int_V \sigma(x-x_i)g(x-x_j)dx)$$

Le système devient alors :

$$\lambda^s \; p_s \; a^j \; \sigma_{ij} = \mu_i^{\;l} \; p_l + \frac{1}{S} [\lambda^s \; p_s - \sum_j S^{ij} \; \mu_i^{\;l} \; p_l]$$

dont la solution peut se mettre sous la forme :

$$a^j = \frac{U^{jl} \; p_l}{\lambda^s \; p_s}$$

Ce qui donne comme estimateur de τ_V

$$\tau_V^{\;*} = \frac{U^{jl} \; p_l t_j}{\lambda^s \; p_s} .$$

On voit que dans cet estimateur interviennent non seulement les $p_j t_j = A_j$ mais aussi les "accumulations croisées" $p_l t_j$. Il semblerait donc que les estimateurs $\dfrac{\lambda^i \; p_i t_i}{\mu^i \; p_i}$ que nous avons considérés auparavant ne donnent pas encore la meilleure solution. En fait, le problème de l'estimation de τ_V sort du cadre de la géostatistique dès qu'on considère $t(x)$ et $p(x)$ comme deux F.A. de l'espace à deux dimensions : dans ce cas, on peut écrire τ_V sous la forme

$$\tau_V = \check{M} * t \qquad \text{avec} \begin{cases} M(x) = \dfrac{p(x)}{\int_V p(y)dy} & \text{si } x \in V \\ M(x) = 0 & \text{si } x \notin V \end{cases}$$

$M(x)$ étant une fonction de pondération aléatoire non nécessairement stationnaire. Pour résoudre le problème, il faut probablement se donner toute la loi de $M(x)$ (par exemple en prenant pour $M(x)$ une F.A. théorique bien connue), ce qui sort légèrement des hypothèses habituelles de la géostatistique.

VI - CONCLUSION

La régression des statisticiens apparaît comme un cas particulier de cokrigeage, qu'on trouve lorsque les covariances apparaissent comme des effets de pépite : elle sera d'autant plus "optimale" que l'effet de pépite sur les 3 covariances sera fort.

Hormis ces cas particuliers, le cokrigeage n'est pas la régression mais on a vu apparaître quelques cas de simplification : corrélation intrinsèque et mieux, $\rho^2 = 1$. L'intérêt de la corrélation intrinsèque est la propriété de transitivité qu'on peut imaginer employer, par exemple pour déterminer le poids spécifique du minerai aux points de sondage : celui-ci est, en général, lié linéairement aux teneurs qui sont connues au point de sondage. Il y a donc en général corrélation intrinsèque, et on peut déterminer le poids spécifique en chacun des points du sondage par cokrigeage pour ensuite atteindre le tonnage et la quantité de métal.

-:-:-:-:-:-:-:-