C-99

REMARQUES PRATIQUES A PROPOS DES VARIANCES ET DU VARIOGRAMME.

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

J.RIVOIRARD JANVIER 1983

REMARQUES PRATIQUES A PROPOS DES VARIANCES ET DU VARIOGRAMME

Cette note vise à donner des indications pratiques et précises pour le calcul et l'usage des variances et du variogramme.

1 - Le variogramme

Une fois précisé par la critique des données, le champ S sur lequel portera l'analyse structurale, la structure de la variable régionalisée z(x) est représentée par le variogramme régional :

$$\gamma_r(h) = \frac{1}{2} \frac{1}{K(h)} \int_{S \cap S-h} \left(z(x+h) - z(x)\right)^2 dx$$

où S-h est le translaté par (-h) du champ S et K(h) la mesure de $S \cap S-h$.

Cette intégrale d'espace est naturellement inconnue. En pratique on l'estime par le variogramme expérimental :

$$\gamma_e(h) = \frac{1}{2} \frac{1}{N(h)} \sum_{1}^{N(h)} \left[z(x_i+h) - z(x_i)\right]^2$$

qui, calculé sur les points de mesure, est en réalité une discrétisation de l'intégrale d'espace correspondante.

C'est sur ce variogramme expérimental, supposé bien représenter le variogramme régional, que sera ajusté le modèle de variogramme $\gamma(h)$

2 - Le modèle probabiliste

La variable régionalisée z(x) sera alors considérée comme une réalisation limitée au champ S, de la Fonction Aléatoire Intrinsèque Z(x) définie par ses accroissements :

$$\left.\begin{aligned} E(Z(x+h) - Z(x)) &= 0 \\ Var(Z(x+h) - Z(x)) &= 2\gamma(h) \end{aligned}\right\} \text{ indépendants de } x$$

A l'intérieur de ce modèle probabiliste de Fonction Aléatoire - dont la spécification ne nécessite que le modèle de variogramme - on peut calculer la variance de toute combinaison linéaire $\sum \lambda^{\alpha} Z(x_{\alpha})$ par la formule de base :

$$\mathrm{Var}\left[\sum \lambda^{\alpha} Z(x_{\alpha})\right] = - \sum_{\alpha} \sum_{\beta} \lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta}$$

à condition que : $\sum \lambda^{\alpha}$ soit nul

($\gamma_{\alpha\beta}$ représente $\gamma\,(x_\alpha - x_\beta)$)

Une variance devant être nécessairement positive, n'importe quelle fonction $\gamma(h)$ ne peut pas servir de modèle de variogramme : c'est pourquoi l'ajustement d'un variogramme expérimental se fait à l'aide de certaines fonctions connues pour pouvoir servir de modèles.

3 - Variance expérimentale et variogramme expérimental

La variance expérimentale :

$$\sigma^2 = \frac{1}{n} \sum_{1}^{n} \left[z(x_i) - \left(\frac{1}{n} \sum_{1}^{n} z(x_j) \right) \right]^2$$

peut aussi s'écrire :

$$\sigma^2 = \frac{1}{2n^2} \sum_i \sum_j \left(z(x_i) - z(x_j) \right)^2$$

C'est donc la demi-moyenne des carrés des différences de tous les couples qu'on peut former avec les données. En classant les $(z(x_i) - z(x_j))$ selon la distance $|x_i - x_j|$, on reconnait les couples qui interviennent pour le calcul, dans les différentes classes de distance, du variogramme expérimental.

Par conséquent, si tous les couples $(z(x_i) - z(x_j))$ sont utilisés pour le calcul de variogramme, on s'aperçoit que la variance n'est rien d'autre que la moyenne de toutes les valeurs prises par le variogramme expérimental. Chaque point du variogramme expérimental, correspondant par exemple à une classe de distance, doit être naturellement pondéré par le nombre de couples de la classe. Noter qu'interviennent dans la moyenne les n couples, parmi les n^2 couples possibles, correspondant à i = j, c'est-à-dire à :

$$z(x_i) - z(x_j) = 0,$$

et qui représentent la valeur du variogramme pour la distance zéro qui, on le sait, est nulle par définition : $\gamma(0) = 0$

Ainsi, avec une structure isotrope de portée grande par rapport au champ, il y aura beaucoup de points du variogramme situés au dessous du palier et la moyenne de tous les points - c'est-à-dire la variance expérimentale - sera inférieure au palier.

Si la portée est petite, palier et variance se confondront (A condition que la variogramme se stabilise bien autour du palier et ne présente pas de dérive aux grandes distances).

Par contre, si un variogramme n'est pas calculé avec tous les couples, il n'y a aucune raison pour que la moyenne de ses points représente la variance. Ainsi dans un gisement à 3 dimensions, le variogramme vertical moyen peut très bien présenter un palier beaucoup plus bas que la variance. Il y a donc une anisotropie zonale et on doit s'attendre à ce que les variogrammes dans d'autres directions soient plus hauts que la variance.

4 - Variance de dispersion

Le modèle de Fonction Aléatoire Intrinsèque - donc le modèle de variogramme - permet de calculer la variance de dispersion $D^2(v/V)$ de la variable définie sur le support v à travers le domaine V.

0, v, V désignant des supports de plus en plus grands (0 sera le support ponctuel) la relation d'addivité permet de sommer les variances :

$$D^2(0/V) = D^2(0/v) + D^2(v/V)$$

De plus : $D^2(0/v) = \overline{\gamma}(v,v)$ noté aussi $\overline{\gamma}_{vv}$, valeur de variogramme $\gamma(x-y)$ lorsque les points x et y décrivent indépendamment le support v :

De même : $D^2(0/V) = \overline{\gamma}(V,V)$

On remarque que si le variogramme possède un palier, et si la portée est petite visà vis du domaine V, on a :

$$D^2(0/V) = \overline{\gamma}(V,V) \# \text{ Palier du variogramme } = \gamma(\infty)$$

Si par contre la portée est grande, et si $\gamma(h)$ est croissant jusqu'à son palier :

$$D^2(0/V) = \overline{\gamma}(V,V) < \text{ Palier } \gamma(\infty)$$

La formule d'additivité se met sous la forme :

$$D^2(v/V) = D^2(0/V) - D^2(0/v)$$
$$= \overline{\gamma}(V,V) - \overline{\gamma}(v,v)$$

ce qui permet de passer, dans le modèle, de la variance de dispersion d'un point dans un champ V à celle d'un support v dans le même champ.

En pratique minière cette formule est constamment utilisée pour effectuer le changement de support entre les échantillons et les unités de sélection.

Le second membre de la formule se compose de 2 termes:

- $D^2(0/V)$, qui représente la variance du point dans V, peut effectivement se calculer par la formule théorique $\overline{\gamma}(V,V)$, si on fait confiance à la représentativité du modèle de variogramme à grandes distances. Mais on peut lui préférer aussi la variance des points connue expérimentalement.

- $D^2(0/v) = \overline{\gamma}(v,v)$ se calcule toujours à l'aide du modèle du variogramme (et ne nécessite d'ailleurs que son comportement à faibles distances). Il représente la perte de dispersion quand on passe du point au bloc, et est indépendant de la valeur affectée à $D^2(0/V)$.

Une méthode simple pour effectuer un changement de support sera donc de :

- calculer $\bar{\gamma}(v,v)$ d'après le modèle de variogramme.
- soustraire cette quantité de la variance expérimentale ponctuelle.

Cette méthode présente l'avantage d'être utilisable même dans le cas où variance et palier ne coïncident pas.

5 - Variance d'estimation

Si on estime la variable Z_V définie sur un support V par la variable connue Z_v définie sur un support v, le modèle de variogramme permet d'apprécier la qualité de l'estimation grace au calcul de la variance d'estimation, c'est-à-dire de la variance de l'erreur d'estimation :

$$\sigma^2 = E(Z_v - Z_V)^2 = 2\,\bar{\gamma}_{vV} - \bar{\gamma}_{vv} - \bar{\gamma}_{VV}$$

$\bar{\gamma}_{vV}$ désigne la valeur moyenne de $\gamma(x-y)$ lorsque x parcourt v et y parcourt V.

Si l'estimateur est formé par une combinaison linéaire $\sum\lambda^{\alpha}Z(x_{\alpha})$, il est nécessaire, pour pouvoir en calculer la variance, que l'erreur ($\sum\lambda^{\alpha}Z(x_{\alpha}) - Z_V$) satisfasse à la condition : $\sum\lambda^{\alpha}-1=0$ (autrement dit : somme des poids λ^{α} égale à 100%). On a alors :

$$\sigma_E^2 = 2\sum_{\alpha}\lambda^{\alpha}\,\gamma_{\alpha V} - \sum_{\alpha\beta}\sum\lambda^{\alpha}\lambda^{\beta}\gamma_{\alpha\beta} - \bar{\gamma}_{VV}$$

6 - Régularisation du variogramme

Le variogramme régularisé sur support v : γ_v se calcule à partir du variogramme ponctuel γ par la formule :

$$\gamma_v(h) = \bar{\gamma}_{vv_h} - \bar{\gamma}_{vv}$$

$\bar{\gamma}_{vv_h}$: valeur moyenne de $\gamma(x-y)$ lorsque se décrit v et y décrit le translaté de v par le vecteur h.

7 - Fonctions auxiliaires

Les fonctions auxiliaires représentées sous forme d'abaques pour les modèles courants de variogramme, permettent le calcul "manuel" de variance.

Exemple:

$\gamma(h)$ est un sphérique isotrope de portée a = 50 mètres, de palier C = 7

L'abaque $F\left(\frac{1}{a}\ \frac{1}{a}\right)$ permet de calculer la quantité :

$$\bar{\gamma}(v,v) = 7 \times F\left(\frac{10}{50}, \frac{10}{50}\right) = 7 \times 0.155 = 1.085$$

lorsque v est un carré de 10x10m

Remarque: toutes les formules de variances sont linéaires en γ
Si γ se présente comme une somme de plusieurs schémas $\gamma = C_0+\gamma_1+\gamma_2$ on pourra donc effectuer le calcul de variance schéma après schéma et en faire la somme. En particulier on traitera l'effet de pépite à part.

8 - Notion de support élémentaire non ponctuel

La théorie géostatistique permet de calculer les différentes variances à l'aide du modèle de variogramme ponctuel :

- variance de dispersion de v dans V :

$$D^2(v/V) = \bar{\gamma}(V,V) - \bar{\gamma}(v,v)$$

- variances d'estimation de V par v :

$$\sigma_E^2 = 2\ \bar{\gamma}_{vV} - \bar{\gamma}_{VV} - \bar{\gamma}_{vv}$$

On peut également régulariser le modèle de variogramme ponctuel sur un support v :

$$\gamma_v(h) = \bar{\gamma}_{vv_h} - \bar{\gamma}_{vv}$$

En réalité, en géostatistique minière, le variogramme ponctuel n'est pas accessible expérimentalement. Même lorsqu'on a accès à des données quasi-ponctuelles (par exemple enregistrement radiométrique digitalisé à un pas de 10 ou 20 cm), on pourra faire des études structurales détaillées, mais on modèlisera plutôt des variogrammes expérimentaux régularisés. On utilisera d'ailleurs comme modèles des schémas sphériques, exponentiels ou autres, et non des schémas régularisés de sphériques...

Il est alors commode de travailler sur un support élémentaire non ponctuel v_o. Ainsi dans un gisement à 3 dimensions, reconnu par des sondages verticaux, si l'estimation se fait par blocs de 3 mètres de haut, il est commode de travailler sur un support élémentaire non ponctuel qui sera ici une carotte de 3 mètres de haut.

A condition de n'utiliser alors que des supports qui soient multiples du support élémentaire (on pourra utiliser des supports de hauteur 3, 6 ou 9 mètres mais non 1.5 ou 5 mètres), il est facile de voir que les formules rappelées plus haut demeurent valables si on utilise le variogramme correspondant au support élémentaire v_o (soit : γ_{vo}).

En effet, comme $\gamma_{vo}(h) = \bar{\gamma}_{vo\ voh} - \bar{\gamma}_{vo\ vo}$

on a par exemple :

$$D^2(v/V) = \bar{\gamma}(V,V) - \bar{\gamma}(v,v) = \bar{\gamma}_{v_o}(V,V) - \bar{\gamma}_{v_o}(v,v)$$

9 - La Pratique Informatique

Dans une telle optique, une carotte de 3 mètres de haut sera représentée par son point central, un bloc de 10x10x3m ou un niveau de 3 mètres par son plan horizontal médian, et enfin un gisement par l'ensemble des plans représentant chacun des niveaux.

Ceci correspond bien à la pratique informatique puisque les logiciels de krigeage, de calcul de variance, ou de régularisation, fonctionnent le plus souvent d'après ce principe.

Par exemple le calcul de $\bar{\gamma}(v,v)$, v étant une carotte de 6 mètres de haut se calculera en discrétisant cette carotte. Si on utilise un variogramme ponctuel γ, il faudra beaucoup discrétiser. Si par contre on utilise le variogramme γ_{vo} défini sur le support élémentaire v_o (carotte de 3m de haut), on considèrera que la carotte de 6 mètres est constitué de 2 carottes de 3 mètres et donc sera discrétisée par 2 points. $\bar{\gamma}_{v_o}(v,v)$ sera alors la moyenne de

$\gamma_{v_0}(x-y)$ lorsque x décrit les 2 points et y également, d'où :

$$\bar{\gamma}_{v_0}(v,v) = \frac{1}{4}\left[\gamma_{v_0}(0) + \gamma_{v_0}(0) + \gamma_{v_0}(3) + \gamma_{v_0}(3)\right] = \frac{1}{2}\gamma_{v_0}(3)$$

On voit donc la différence qui existe entre le calcul informatique nécessitant une discrétisation, et le calcul manuel par fonctions auxiliaires qui utilise un variogramme défini ponctuellement.

10 - Les précautions à prendre lorsqu'on utilise un support élémentaire non ponctuel.

L'usage d'un support élémentaire ponctuel n'est rigoureux que dans le cas où tous les supports utilisés sont des multiples du support élémentaire.

Si par exemple on dispose de sondages verticaux et également de sondages horizontaux (percés à partir de galeries), le seul support élémentaire théoriquement valide sera la point (on néglige ici le diamètre des sondages). En pratique, on sera obligé de travailler sur un support régularisé (par exemple 1m) mais il serait dangereux de représenter par un même point un tronçon de carotte de 10 mètres qui pourrait être en réalité vertical aussi bien qu'horizontal.

Dans le cas où le choix d'un support élémentaire non ponctuel se justifie il faudra faire attention aux anisotropies. Un phénomène isotrope ponctuellement pourra se montrer anisotrope lorsqu'on passe au régularisé : ainsi, avec une régularisation sur grande carotte verticale, le variogramme vertical régularisé sera plus continu que le variogramme horizontal.

A propos de l'ajustement, on peut remarquer que le modèle de variogramme ajusté ne sera utilisé, dans la direction verticale, que pour les distances multiples de la longueur de la carotte (par ex. 3,6,9,12...m).

11 - L'effet de pépite

Dans les formules géostatistiques, il est extrêmement simple de prendre en compte l'effet de pépite lorsque l'on considère :

- que le modèle de variogramme est défini sur un support élémentaire non ponctuel v_o.
- que les supports utilisés sont des multiples finis de support élémentaire.

Ainsi, le support élémentaire étant la carotte de 3 m., une carotte de 6 mètres sera constituée de 2 tels supports élémentaires. Mais également un bloc de 10x10x3m sera considéré comme un nombre fini n de tels supports élémentaires : dans ce cas après avoir développé la formule utilisée, on considèrera dans le résultat final que n est très grand, donc, pratiquement, infini.

Exemple: Que devient la composante pépitique lorsqu'on régularise sur un support v ?

Un effet de pépite connu C_o est observé sur le support élémentaire considéré comme nonponctuel v_o, et v sera a priori constitué de n supports v_o.

Quand on régularise sur v, on observe une perte de dispersion dont la composante pépitique vaut :

$$\bar{\gamma}_{v_o}(v,v)$$

avec γ_{v_o} défini sur support v_o, valant donc :

- Entre un support donné v_o et lui-même: $\gamma_{v_o}(0) = 0$ par définition

- Entre un support donné v_o placé en x_i et un autre support v_o placé en $x_j \# x_i$:

$$\gamma_{v_o}(x_i - x_j) = C_o$$

$\bar{\gamma}_{vo}(v,v)$ est la moyenne des $\gamma_{vo}(x_i - x_j)$ pour tous les couples (x_i, x_j) de supports élémentaires constituant v.

Sur les n^2 couples, n correspondent à $i = j$: d'où 0

n^2-n à $i \# j$: d'où C_o

En moyenne $\bar{\gamma}_{vo}(v,v) = \frac{n^2-n}{n^2} C_o = \frac{n-1}{n} C_o$

La composante pépitique, initialement égale $= C_o$, a diminué de $\frac{n-1}{n} C_o$

Il en reste : $\frac{C_o}{n} = C_o \frac{v_o}{v}$

D'où la loi : l'effet de pépite varie en raison inverse du volume.

Si v est très grand par rapport à v_o, la composante pépitique $\frac{C_o}{n} = C_o \frac{v_o}{v}$ devient nulle.

<u>Autre exemple</u> : Composante pépitique dans l'estimation d'un volume V par un échantillon v_o.

$$\sigma_o^2 = 2\,\bar{\gamma}_{vo}(v_o,V) - \bar{\gamma}_{v_o}(v_o,v_o) - \bar{\gamma}_{v_o}(V,V)$$

$$\bar{\gamma}_{v_o}(v_o,v_o) = \gamma_{v_o}(o) = 0$$

$$\bar{\gamma}_{v_o}(V,V) = \frac{n-1}{n} C_o = \frac{V-v_o}{V} C_o = (1-\frac{v_o}{V}) C_o$$

Si v_o est intérieur à V :

$$\gamma_{v_o}(v_o,V) = \frac{n-1}{n} C_o = \frac{V-v_o}{V} C_o = (1-\frac{v_o}{V})C_o$$

D'où : $\sigma_o^2 = C_o\,(1 - \frac{v_o}{V})$

Si v_o est extérieur à V :

$\bar{\gamma}_{v_o}(v_o,V) = C_o$

D'où $\sigma_o^2 = C_o(1 + \frac{v_o}{V})$

<u>REMARQUE</u> : Traiter l'effet de pépite avec des réunions finies de supports élémentaires non ponctuels revient à travailler avec des variables indépendantes comme en stat.classiques : c'est ce qui explique la simplicité des formules.