

FONTAINEBLEAU/CG

N-55/88/G

UN MODELE D'INTERVALLES DE CONFIANCE
POUR LA MOYENNE GLOBALE

C. LAJAUNIE

J. RIVOIRARD

Décembre 1988

UN MODELE D'INTERVALLES DE CONFIANCE

POUR LA MOYENNE GLOBALE

C. Lajaunie et J. Rivoirard

1 - Introduction

Le problème des intervalles de confiance n'est pas nouveau pour l'industrie minière. Les compagnies Sud-Africaines, devant décider de l'ouverture d'une mine profonde sur la base de quelques sondages seulement, en sont conscientes depuis longtemps. Et Sichel, depuis une quarantaine d'années, a développé des méthodes bien connues qui permettent d'obtenir des intervalles de confiance. Celles-ci sont basées sur une hypothèse d'indépendance et de lognormalité des teneurs mesurées.

Ces dernières dizaines d'années ont également vu l'éclosion de la géostatistique, théorie des variables régionalisées, apte à prendre en compte la structure spatiale, et donc en particulier la dépendance entre données voisines. Pour des données structurées il est bien sûr souhaitable de prendre en compte cette structure lors du calcul d'intervalles de confiance, et donc de proposer des méthodes géostatistiques d'estimation de tels intervalles.

La nécessité de disposer d'intervalles de confiance nous semble apparaître dans des contextes différents qu'il convient de préciser :

- Au niveau de la reconnaissance, lors d'un calcul de grandeur moyenne sur une partie importante du gisement lui même. Il est assez fréquent que en raison du caractère très fragmentaire de l'échantillonnage, ou d'un fort coefficient de variation σ/m dans la distribution statistique, les moyennes et variances de dispersion (et donc le variogramme) soient assez mal connues. Cette imprécision doit être prise en compte lorsque l'on cherche à quantifier l'erreur sur l'estimation globale.

- Lors de l'estimation locale, les paramètres de distribution et le variogramme, qui sont basés sur des données plus abondantes, peuvent être considérés comme connus avec une précision suffisante. Dans ces conditions, l'estimation locale consiste à tenir compte des données de voisinage pour corriger la distribution globale. L'erreur d'estimation due au caractère imparfait de ce conditionnement étant prépondérante devant l'erreur dans le calcul des paramètres globaux, un intervalle de confiance est donc obtenu en considérant un intervalle de probabilité pour la distribution conditionnelle.

Différentes méthodes ont été proposées pour obtenir des intervalles de confiance dans un contexte de variables régionalisées. Les plus élémentaires reposent sur une hypothèse

distributionnelle (le plus souvent gaussienne) concernant l'erreur d'estimation, d'autres proposent des distributions conditionnelles obtenues par anamorphose de lois gaussiennes (Dowd 1989), ou des approximations de ces distributions conditionnelles par des méthodes non linéaires classiques (krigeage disjonctif dans (Kleingeld 1987) et dans (Dowd 1989) et krigeage d'indicatrices dans (Switzer et Xiao, 1988)). Ces dernières méthodes se concentrent sur le problème de l'estimation locale, tandis que dans le premier cas, en ignorant l'incertitude sur la variance de la variable régionalisée lors de l'estimation globale, il semble qu'une composante essentielle du problème soit ignorée.

Ici nous commencerons par considérer deux points de vue possibles, qui conduisent à des procédures différentes. Puis nous développerons le premier point de vue. L'accent sera mis sur l'estimation globale, et donc sur la prise en compte de l'incertitude concernant les deux premiers moments de la distribution régionale. Ce point de vue qui a déjà été esquissé dans (Lajaunie et Rivoirard 1987) est ici exposé de façon plus détaillée et plus générale.

1.1 - INTERVALLES POUR UNE GRANDEUR DETERMINISTE

La démarche sera illustrée par un problème type. On notera V un volume du gisement, fixé à l'avance, et dont il s'agit d'estimer la teneur $z(V)$. Ce volume est reconnu par un certain nombre de prélèvements à maille régulière, dont les teneurs respectives seront notées:

$$z(x_0 + ap)$$

x_0 désigne l'origine de la maille, p est un vecteur à valeurs dans $\{0, \dots, N-1\} \times \{0, \dots, M-1\}$ par exemple lorsque V est un rectangle, et a désigne la matrice dont les lignes sont constituées des vecteurs pas d'échantillonnage dans les différentes directions de la maille (par exemple (a_1, a_2) à deux dimensions). La figure 1 illustre cette situation.

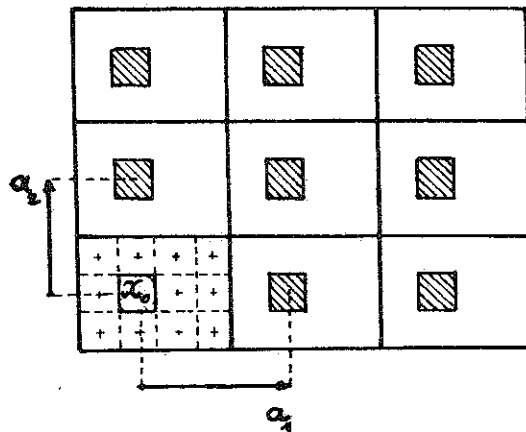


figure 1. Configuration d'échantillonnage

Enfin on notera w l'ensemble des positions de la maille de reconnaissance: $w = \{x : x = x_0 + ap\}$. La moyenne des valeurs de la maille, notée $z(w, x_0)$ fournit un estimateur de la

teneur $z(V)$:

$$z^*(V) = z(w, x_0) = \frac{1}{NM} \sum_p z(x_0 + ap)$$

En faisant varier l'origine de la maille dans un bloc de taille $a_1 \times a_2$, l'échantillonnage sera supposé décrire exhaustivement le volume V . En langage probabiliste, on dira que si l'origine est choisie selon une loi uniforme dans son bloc, la variable $z(w, X_0)$ vérifie:

$$E[z(w, X_0)] = z(V)$$

De même, l'indice p peut être choisi selon une loi uniforme sur $\{0, \dots, N-1\} \times \{0, \dots, M-1\}$ (soit alors P la variable aléatoire associée) nous avons:

$$E[z(X_0 + aP) \mid z(w, X_0)] = z(w, X_0)$$

Expérimentalement, seule la distribution à $X_0 = x_0$ fixé est observable, et ainsi qu'il a été dit, ses moment d'ordre 1 et 2 peuvent être sensiblement différents de ceux de la distribution exhaustive dans V . A titre d'exemple donnons les valeurs observées sur une simulation bidimensionnelle échantillonnée le long de lignes parallèles équidistantes, suivant le schéma de la figure 2. Cette simulation a fait l'objet d'une description dans (Lajaunie et Lantuéjoul, 1988). La distance entre les lignes étant égale à 10 fois leur largeur, chaque échantillonnage couvre donc le 1/10 ième du volume V total, et nous avons 10 distributions possibles, en faisant varier l'origine de la maille. La moyenne et variance exhaustives sont de 1,47 et 14,1 respectivement, ce qui donne un coefficient de variation σ/m de 2,6.

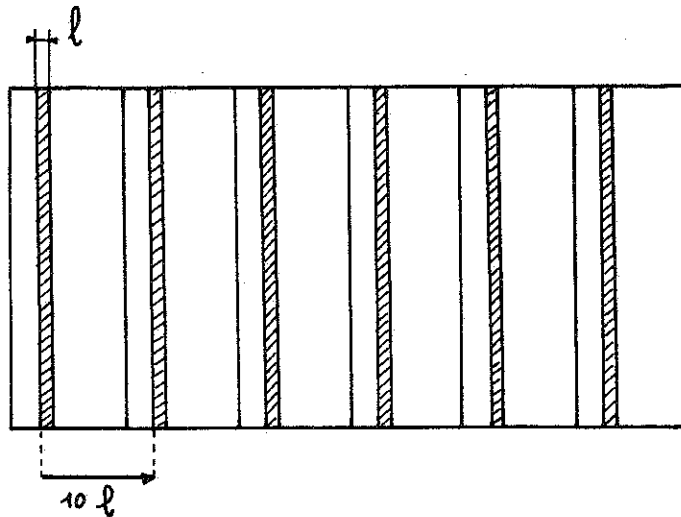


figure 2. Echantillonnage de la simulation

Les 10 valeurs échantillonnées sont données par le tableau suivant:

numéro d'échantillon	m	σ^2	σ/m
1	1,54	16,2	2,61
2	1,57	22,2	3,00
3	1,43	15,7	2,77
4	1,31	9,1	2,31
5	1,31	10,0	2,41
6	1,38	12,5	2,56
7	1,51	13,1	2,40
8	1,61	14,6	2,37
9	1,58	15,3	2,48
10	1,46	11,7	2,34

De même que ces statistiques, le variogramme expérimental calculé le long des lignes fluctue fortement d'un échantillonnage à l'autre, ce que la figure 3 illustre.

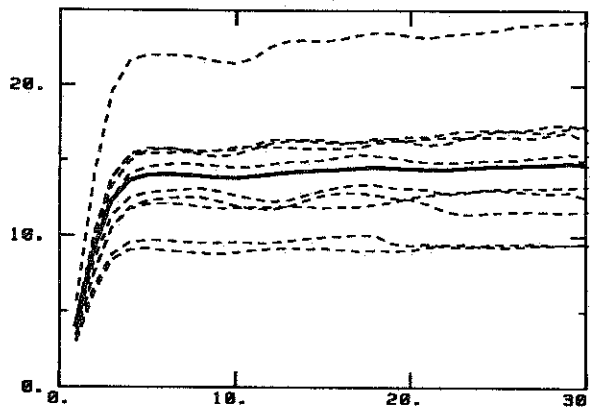


Figure 3. Les 10 variogrammes expérimentaux.

Sur la base de l'échantillonnage disponible, pour une valeur donnée x_0 , nous définirons à l'aide d'une certaine procédure un intervalle autour de $z(w, x_0)$, soit $I(x_0)$. La même procédure appliquée aux autres échantillons (et donc aux autres origines x_0 possibles) nous donnerait une famille d'intervalles. Si une proportion supérieure à $1 - \alpha$ parmi ces intervalles contiennent la valeur inconnue $z(V)$, nous dirons que l'intervalle aléatoire $I(X_0)$ constitue un intervalle de confiance à $1 - \alpha$ pour $z(V)$. En termes probabilistes:

$$P[I(X_0) \ni z(V)] \geq (1 - \alpha)$$

On observera que la longueur de ces intervalles est en général variable d'une position x_0 à l'autre.

1.2 - POUR UNE VARIABLE ALEATOIRE

Le second point de vue impose une interprétation probabiliste de $z(V)$ ainsi que des autres variables $z(w)$ et $z(x_i)$ ($z(x_i)$ est utilisée comme abréviation de $z(x_0 + ap)$). Pour cette raison ces variables seront désignées par des majuscules $Z(V), Z(w), \dots$. On supposera connue la distribution à priori de $Z(V)$. Celle ci peut être donnée "subjectivement" par l'utilisateur (point de vue Bayésien), en tenant compte des analogies pouvant exister entre le gisement à étudier et d'autres gisements déjà connus, ou bien résulter d'une reconnaissance du gisement sur des zones plus importantes (on rejoint alors le point de vue de l'estimation locale). Dans ce cas, l'intérêt se porte sur des modèles réalistes permettant d'obtenir l'expression de la loi conditionnelle de la teneur $Z(V)$ lorsque les données d'échantillons sont connues:

$$f_{z_1, z_2, \dots, z_k}(z_V)$$

Les intervalles de confiance pour $Z(V)$ en découlent. Ceci nécessite en principe la manipulation de distributions multivariées, ce que l'on ne sait faire que dans le cas multigaussien anamorphosé (Dowd 1989). Aussi peut-on avoir recours en pratique à deux démarches d'approximation.

La première consiste à approcher la fonction de répartition conditionnelle à l'aide d'une méthode adaptée à l'estimation de variables indicatrices :

$$F^*(y) = 1_{Y_0 < y}^*$$

La méthode pourra être un krigeage disjonctif, un krigeage d'indicatrice, ou un cokrigeage indicatrice-rang (p.k.). En ce qui concerne le k.d., voir (Kleingeld 1987) pour une utilisation du modèle binomial négatif à cette fin, ou (Dowd 1989) pour une comparaison avec des résultats de simulation, dans un contexte de modèle gaussien. Une étude du krigeage d'indicatrices est présentée dans (Switzer et Xiao, 1988), enfin, on trouvera dans (Lajaunie, 1989) une comparaison des trois méthodes, dans un esprit assez proche.

La seconde démarche d'approximation consiste à résumer l'information de $Z_1, Z_2, \dots, Z_k$ par celle apportée par l'estimateur $Z^*(V) = \sum \lambda_i Z_i$ de la teneur sur V , de façon à ne manipuler que la loi bivariable du couple $(Z(V), Z^*(V))$: c'est une optique "conditionnement uniforme".

2. Principe de construction pour un paramètre

On se concentre ici sur le premier point de vue, de telle sorte que le seul élément aléatoire est l'origine de la maille X_0 . Considérons l'ensemble des valeurs à priori possibles pour $z(V)$. La distribution de $z(w, X_0)$ dépend de cette valeur inconnue, ce que l'on illustre par sa fonction de répartition:

$$F_{z(V)}(z) = P_{z(V)}[z(w, X_0) \leq z]$$

(il ne s'agit pas de distribution conditionnelle puisque $z(V)$ est un paramètre). Choisissons

alors Z_1 et Z_2 (aléatoires) tels que:

$$F_{Z_1}[z(w, X_0)] = 1 - \frac{\alpha}{2}$$

$$F_{Z_2}[z(w, X_0)] = \frac{\alpha}{2}$$

Chose possible lorsque $F_z(u)$ est continue en u , et dépend de manière monotone (par exemple croissante) et continue de z . La figure 4 illustre cette situation (f_z est alors la densité de F_z).

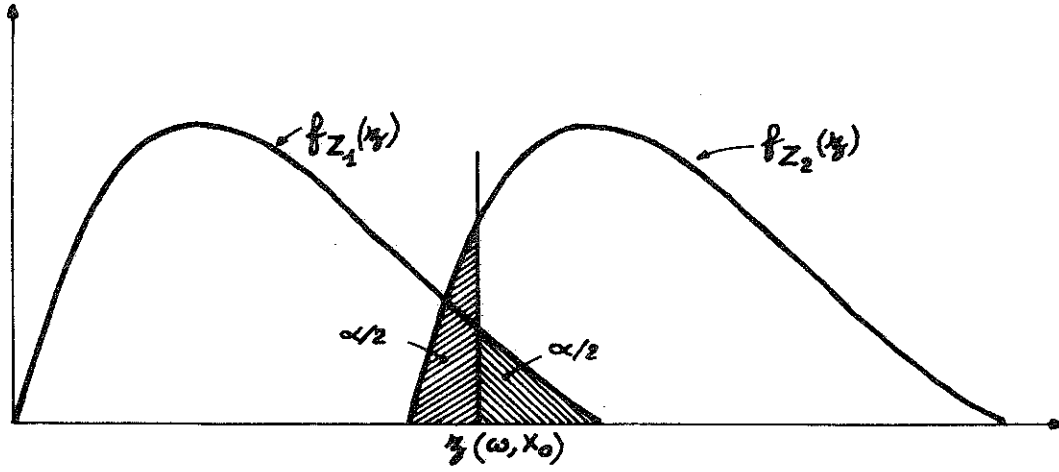


figure 4.

Dans ces conditions l'intervalle $[Z_1, Z_2]$ est un intervalle de confiance à $1 - \alpha$ pour $z(V)$: En effet, pour tout $z(V)$, la variable $U = F_{z(V)}[z(w, X_0)]$ est uniforme sur $[0, 1]$, de telle sorte que:

$$P \left[\frac{\alpha}{2} \leq F_{z(V)}[z(w, X_0)] \leq 1 - \frac{\alpha}{2} \right] = 1 - \alpha$$

mais on a l'équivalence (monotonie de F_z par rapport à z):

$$\left\{ \frac{\alpha}{2} = F_{Z_1}[z(w, X_0)] \leq F_{z(V)}[z(w, X_0)] \right\} \iff \{ Z_1 \leq z(V) \}$$

de telle sorte que:

$$P[Z_1 \leq z(V) \leq Z_2] = 1 - \alpha$$

Plus généralement, la construction pratique d'intervalles de confiance est possible si l'on a pu mettre en évidence une variable pivot, c'est à dire lorsqu'une fonction $\psi(x, y)$ à été trouvée telle que:

- i) $T = \psi(z(w, X_0); z(V))$ ait une distribution indépendante de $z(V)$
- ii) $\psi(x, y)$ soit une fonction monotone inversible de $y \in \mathbb{R}^+$,
pour toutes les valeurs possibles de $x > 0$

Ce rôle était joué dans l'exemple précédent par $F_y(x)$. Dans ces conditions, pour tout jeu de valeurs t_i et $t_s \in]0, 1[$ il est possible de déterminer les seuils Z_i et Z_s tels que:

$$\{t_i \leq T \leq t_s\} \iff \{z(V) \in [Z_i, Z_s]\}$$

Les seuils Z_i et Z_s sont fonctions de X_0 et sont déterminés par les équations:

$$\begin{aligned}\psi(z(w, X_0); S_1) &= t_i \\ \psi(z(w, X_0); S_2) &= t_s\end{aligned}$$

Avec : $Z_i = \text{Min}(S_1, S_2)$ et $Z_s = \text{Max}(S_1, S_2)$. Alors, si $P[T \in [t_i, t_s]] \geq (1 - \alpha)$ on a bien : $P[[Z_i, Z_s] \ni z(V)] \geq (1 - \alpha)$. Dans cette inégalité, l'élément variable est l'intervalle, qui dépend de l'échantillonnage choisi, et non la teneur $z(V)$, considérée comme fixe. Le choix des seuils t_i et t_s pour une valeur de $1 - \alpha$ laisse une certaine liberté, que l'on pourra exploiter pour minimiser la longueur de l'intervalle $[Z_i, Z_s]$.

3. Un modèle possible: le fractionnement itéré

Désignons ici par V un bloc quelconque dans le gisement (de taille également quelconque), et $Z(V)$ sa teneur. Cette teneur pourra selon le contexte être considérée comme variable aléatoire, si V est par exemple un bloc de position aléatoire dans le gisement, ou paramètre déterministe s'il s'agit d'une partie fixe du gisement. Le bloc peut être considéré comme l'union d'un certain nombre de sous blocs de même taille, par exemple v_i ; $i = 1, \dots, k$. En choisissant de manière uniforme un bloc parmi ces k , on a une teneur que nous noterons $Z(v)$. Dans le modèle proposé, les variables $Z(V)$ et $Z(v)$ se déduisent l'une de l'autre par une relation du type:

$$Z(v) = T_{v,V} Z(V)$$

relation dans laquelle les variables $T_{v,V}$ et $Z(V)$ sont indépendantes. La relation de Cartier impose alors la condition de cohérence sur la variable $T_{v,V}$: $E[T_{v,V}] = 1$. Cette relation doit être valable pour toute partition $v \subset V$. En particulier, si v' est une sous-partition de v : $v'_{ij} \subset v_i$, alors v' est encore une partition de V , et donc:

$$T_{v',V} = T_{v',v} T_{v,V}$$

avec encore indépendance entre $T_{v',v}$ et $T_{v,V}$. Les variables T intervenant dans le calcul auront leur distributions modélisées par une famille paramétrique. Afin que la factorisation soit possible pour toutes les partitions envisageables, les variables $T_{v',v}$ et $T_{v,V}$ seront modélisées comme des produits d'un certain nombre de variables indépendantes et de même loi, représentant le fractionnement élémentaire. Soit T_e une telle variable. Il sera commode de dissocier les valeurs positives et les valeurs nulles de T en adoptant une représentation :

$$T_e = J Y \quad \text{en notant} \quad J = 1_{T_e > 0}$$

avec une variable positive Y et indépendante de l'indicatrice J . Pour un produit de telles variables, noté T^n , on aura:

$$T^n = \prod_{k=1}^n J_k \prod_{l=1}^n Y_l$$

ou, avec une notation évidente: $T^n = J^n Y^n$. Ce qui suggère de prendre pour loi de Y^n une loi log-indéfiniment divisible. Dans le cas log-normal, la famille de distributions correspondante comportera donc deux paramètres $p = P[T = 0]$ et σ :

$$T = J \frac{1}{1-p} \exp\{\sigma U - \frac{\sigma^2}{2}\} \quad (1)$$

avec U variable normale indépendante de J . En ce qui concerne l'échantillonnage du domaine V , nous avons alors :

$$\begin{aligned} z(w, X_0) &= T_{w,V} z(V) \\ z(X_0 + aP) &= T_{x,w} z(w, X_0) \end{aligned}$$

avec $T_{x,w}$ et $T_{w,V}$ suivant des lois de type (1). Les valeurs observées de $z(x_0 + ap) / z(w, x_0)$ constituent un échantillon (non indépendant: elles sont liées algébriquement) pour la distribution de $T_{x,w}$. Le modèle ne sera donc utilisable que dans le cas où les teneurs positives suivent approximativement une loi log-normale. L'estimation des paramètres p et σ de $T_{x,w}$ est alors possible.

Par contre, nous ne connaissons pas a priori les paramètres de la distribution de $T_{w,V}$, et nous devons faire des hypothèses raisonnables pour y parvenir. La première hypothèse concerne la distribution des valeurs nulles. Admettant que si l'échantillonnage a été raisonnablement réalisé, il est très peu vraisemblable d'observer une teneur moyenne $z(V)$ nulle, on fera la simplification $p = 0$. Il reste donc à déterminer la variance de $T_{w,V}$.

3.1 CALCUL DE LA VARIANCE DE $T_{w,V}$

La variance de la teneur d'un point aléatoire uniforme dans V admet la décomposition suivante:

$$Var[z(X_0 + aP)] = E[Var[z(X_0 + aP)|X_0]] + Var\{E[z(X_0 + aP)|X_0]\} \quad (2)$$

La variance à X_0 fixé s'obtient exactement en considérant le variogramme sur la maille de reconnaissance:

$$\gamma_{X_0}(aq) = \frac{1}{2N(aq)} \sum_p [z(X_0 + ap) - z(X_0 + ap + aq)]^2$$

($N(aq)$ est le nombre de couples dans la sommation). La relation classique (cf annexe A) donne alors:

$$Var\{z(X_0 + aP)|X_0\} = \sum_{p,q} \gamma_{X_0}(a(q-p)) = \bar{\gamma}_{X_0}(w, w)$$

en prenant l'espérance en X_0 :

$$E[Var[z(X_0 + aP)|X_0]] = E[\bar{\gamma}_{X_0}(w, w)] = \bar{\gamma}_v(w, w)$$

puisque $E[\gamma_{X_0}(pa)] = \gamma_v(pa)$, variogramme régional dans V au pas pa . Enfin, de la même manière on a exactement:

$$Var[z(X_0 + aP)] = \bar{\gamma}_v(V, V)$$

de telle sorte que (2) se réduit à:

$$Var[z(w, X_0)] = \bar{\gamma}_v(V, V) - \bar{\gamma}_v(w, w)$$

Le variogramme régional γ_v n'étant pas accessible expérimentalement, on sera amené à faire une hypothèse de proportionnalité sur les variogrammes limités à w :

$$\gamma_{X_0}(pa) = \gamma_v(pa) \cdot K(X_0)$$

hypothèse qui semble tout à fait réaliste sur l'exemple de simulation présenté. Enfin, on supposera que cette relation reste approximativement vérifiée par l'ajustement de variogramme, que l'on désignera encore par $\gamma_{X_0}(h)$:

$$\gamma_{X_0}(h) \approx \gamma_v(h) \cdot K(X_0)$$

Il en résulte:

$$Var[z(w, X_0)] = \left\{ \bar{\gamma}_{X_0}(V, V) - \bar{\gamma}_{X_0}(w, w) \right\} \frac{E[Var\{z(X_0 + aP)|X_0\}]}{Var[z(X_0 + aP)|X_0]}$$

Le modèle de fractionnement itéré permet de donner une expression à ces variances:

$$\begin{aligned} z(X_0 + aP) &= T_{xw} z(w, X_0) \\ Var\{z(X_0 + aP)|X_0\} &= Var(T_{xw}) z(w, X_0)^2 \\ E[Var\{z(X_0 + aP)|X_0\}] &= Var(T_{xw}) (Var(z(w, X_0) + z(V)^2) \\ Var(z(w, X_0)) &= Var(T_{wV}) z(V)^2 \end{aligned}$$

La variance de T_{wV} en résulte:

$$Var(T_{wV}) = \frac{D_{x_0}^2[w|V]}{z(w)^2 - D_{x_0}^2[w|V]} \quad (3)$$

en ayant noté:

$$D_{x_0}^2[w|V] = \bar{\gamma}_{x_0}(V, V) - \bar{\gamma}_{x_0}(w, w)$$

3.2 LES INTERVALLES DE CONFIANCE DANS CE MODELE

La loi de $T_{w,V}$ est alors entièrement déterminée, et il est facile de trouver des seuils t_i et t_s tels que:

$$P[T_{w,V} \in [t_i, t_s]] \geq (1 - \alpha)$$

Les intervalles générés sont alors du type:

$$\left[\frac{z(w, X_0)}{t_s}, \frac{z(w, X_0)}{t_i} \right]$$

Leur taille est donc proportionnelle à $Z(w)$.

4. CONCLUSION

Le modèle proposé demande à être testé sur des exemples pratiques où la réalité est connue presque exhaustivement. De tels exemples pourront être fournis par des données de trous de tir si nous cherchons un exemple dans le domaine minier, par des images numériques (intéressantes pour leur exhaustivité), ou bien par des simulations comme dans notre exemple introductif. L'examen de ces situations pourra d'ailleurs suggérer d'autres modèles. L'hypothèse distributionnelle fondamentale réside dans l'extrapolation de la loi de $T_{x,w}$, qui est observée à celle de $T_{w,v}$, laquelle constitue notre loi pivot. Dans cette étape, le modèle log-normal avec effet zéro ajouté n'est bien sur qu'une suggestion parmi d'autres possibles.

References

- DOWD, P. (1989) *Some observations on confidence intervals and kriging errors*. Proc. of third int. Geost. congress. Avignon, Septembre, 1988.
- KLEINGELD, W.J. (1987) *Geostatistics and the discrete particles*, Doctorate Thesis, Ecole des Mines de Paris.
- LAJAUNIE, C. (1989) *Some comparisons of non linear methods to estimate conditional distribution and quantiles*. Rapport interne Centre de Géostatistique.
- LAJAUNIE, C. et RIVOIRARD, J. (1987) *Confidence Intervals on the Global Mean Grade* Section 7 p. 27 in RIVOIRARD, J. *Advanced Geostatistics*, Cours Afrique du Sud. Centre de Géostatistique C131.
- SWITZER, P. et XIAO, H. (1988) *Approximate prediction intervals for spatial interpolation using information: an experiment*. Stanford university, Dept of statistics . Technical report N 118

APPENDICE A -

Variance de dispersion en terme de variogramme régional

Le variogramme régional dans V est défini pour les distances h telles que $|V \cap V_{-h}| > 0$ comme:

$$\gamma_v(h) = \frac{1}{2|V \cap V_{-h}|} \int_{V \cap V_{-h}} [z(x) - z(x+h)]^2 dx$$

en utilisant les notations de la théorie transitive:

$$\begin{aligned}
k(x) &= 1_{x \in V} \\
K(h) &= \int k(x) k(x+h) dx = |V \cap V_{-h}|
\end{aligned}$$

on a l'écriture équivalente:

$$\gamma_v(h) = \frac{1}{2K(h)} \int [z(x) - z(x+h)]^2 k(x) k(x+h) dx$$

La valeur moyenne dans V du variogramme régional est donc:

$$\begin{aligned}
\bar{\gamma}_v(V, V) &= \frac{1}{|V|^2} \int_V \int_V \gamma_v(x-y) dx dy \\
&= \frac{1}{|V|^2} \int \int \gamma_v(x-y) k(x) k(y) dx dy
\end{aligned}$$

(une valeur quelconque finie peut être assignée à $\gamma(h)$ lorsque $|V \cap V_{-h}| = 0$) .Avec le changement de variables $h = x - y$:

$$\begin{aligned}
\bar{\gamma}_v(V, V) &= \frac{1}{|V|^2} \int \int k(y) k(y+h) \gamma_v(h) dy dh \\
&= \frac{1}{|V|^2} \int \int k(y) k(y+h) \int \frac{1}{2K(h)} k(u) k(u+h) [z(u) - z(u+h)]^2 du dy dh \\
&= \frac{1}{2|V|^2} \int \int k(u) k(u+h) [z(u) - z(u+h)]^2 du dh
\end{aligned}$$

puis en posant $v = u + h$:

$$\begin{aligned}
\bar{\gamma}_v(V, V) &= \frac{1}{2|V|^2} \int_V \int_V [z(u) - z(v)]^2 du dv \\
&= \frac{1}{|V|} \int_V z(u)^2 du - \left[\frac{1}{|V|} \int_V z(u) du \right]^2
\end{aligned}$$

Qui est bien la variance régionale dans V . Ce résultat reste valable lorsque le domaine V est constitué d'un ensemble de points isolés comme la maille de reconnaissance w .