

7. Krigeage d'indicatrices

On a vu précédemment que le krigeage de $Z(x)$ fournissait la meilleure estimation linéaire possible (meilleure au sens de variance d'estimation minimale). Le krigeage fournit également une variance d'estimation qui est fonction de la continuité spatiale, telle qu'exprimée par le variogramme, et de la configuration (et de la quantité) de l'information disponible.

Parfois, on veut connaître plus qu'un estimé et qu'une variance d'estimation.

Exemples:

- Dans une mine avec exploitation sélective, on veut connaître localement le tonnage et la teneur du gisement en fonctions de diverses teneurs de coupure. Il faut connaître la proportion des blocs excédant la teneur de coupure et la distribution des teneurs.
- En environnement, on veut connaître en tout point quelle est la probabilité qu'un seuil ou qu'une norme soit excédée.
- Dans une excavation dans un massif rocheux, on veut connaître la probabilité que la densité de fractures d'une certaine famille excède un seuil donné.

Dans le cas d'une distribution normale des $Z(x)$, les paramètres obtenus par krigeage correspondent à la moyenne et la variance d'une loi normale conditionnée par les valeurs prises par les observations ayant servies au krigeage. On peut donc utiliser les tables de la loi normale pour calculer les probabilités requises pour répondre aux questions précédentes.

Lorsque les $Z(x)$ ne suivent pas une distribution normale, il est de pratique courante d'effectuer une transformation qui permette d'obtenir une loi normale et d'effectuer le krigeage des variables transformées. Ces transformations peuvent être des fonctions simples (ex. $\log(Z)$) ou ne pas même avoir une forme analytique (i.e. transformation dite graphique qui associe à chaque valeur originale une valeur normale selon son rang dans la séquence ordonnée des Z). Cette approche fonctionne généralement assez bien sous deux réserves toutefois:

- i. La transformation assure, par construction, la loi normale. Elle n'assure toutefois pas que les couples, les triplets, etc...suivent une loi binormale, trinormale,...multinormale. Pour pouvoir calculer les probabilités à partir des krigeages, la distribution multinormale doit être respectée.
- ii. Même dans le cas normal, le fait de devoir fournir des énoncés de nature probabiliste pour des v.a. définies sur des supports différents de ceux des observations pose problème. Différents modèles ont été proposés pour effectuer ces changements de support. Avec le krigeage d'indicatrices, on utilise souvent la correction affine vue précédemment.

Journel (1984), a eu l'idée d'attaquer le problème précédent par une méthode la moins paramétrique possible. On a donné le nom de krigeage d'indicatrices à la méthode qu'il a développée.

L'idée de base repose sur le fait que le krigeage est un bon estimateur de l'espérance conditionnelle d'une v.a. (on rappelle que le krigeage ordinaire est presque sans biais conditionnel; dans le cas normal, le krigeage simple, une forme particulière de krigeage, est sans biais conditionnel).

Si l'on considère un seuil donné, disons "c", on peut coder, par rapport à ce seuil, la v.a. $Z(x)$ en une variable indicatrice $I(x,c)$ prenant la valeur 0 ou 1.

Supposons qu'on définisse:

$$\begin{aligned} I(x,c) &= 1 \text{ si } Z(x) \leq c \\ &= 0 \text{ si } Z(x) > c \end{aligned}$$

Considérons l'espérance de $I(x,c)$. Par définition, on a:

$$E[I(x,c)] = 1 P(Z(x) \leq c) + 0 P(Z(x) > c) = P(Z(x) \leq c) = F_Z(x,c)$$

où $F_Z(x,c)$ signifie la fonction de distribution évaluée à la localisation x pour le seuil "c".

Le problème consiste donc à estimer $I(x,c)$ en se servant de l'information disponible, les $Z(x_i)$ aux n points observations. On procède ainsi:

- i. On code les valeurs observées $Z(x_i)$ par rapport à un seuil donné "c". On obtient ainsi de nouvelles variables $I(x_i,c)$ qui sont soit des 0 soit des 1 selon les valeurs observées et le seuil considéré.
- ii. On effectue le krigeage, en un point x_0 , de $I(x_0,c)$ à partir des $I(x_i,c)$ qui sont connus. Comme le krigeage approxime l'espérance conditionnelle assez bien, la valeur krigée peut être interprétée comme une estimation de $P(Z(x_0) < c \mid Z_1, \dots, Z_n)$. Au préalable, il aura fallu calculer et modéliser le variogramme des indicatrices $I(x_i,c)$ pour pouvoir effectuer le krigeage.
- iii. On peut recommencer le processus en entier pour de nouveaux seuils c_2, c_3 et ainsi de suite. Pour chaque seuil considéré, il faut coder les valeurs originales observées, calculer et modéliser le variogramme (habituellement différent pour chaque seuil) et effectuer un nouveau krigeage.

En combinant les krigeages correspondant à divers seuils, on obtient une version discrétisée de la fonction de distribution $F_{Z(x_0, z) \mid Z_1, \dots, Z_n}^*$ que l'on appellera dorénavant $F_{KI}(x_0, z)$. On peut ensuite utiliser $F_{KI}(x_0, z)$ pour fournir divers estimés comme des probabilités, des quantiles, la médiane, une fonction de coût...enfin toute quantité imaginable qui peut se calculer à partir d'une fonction de distribution.

Notes:

- Il ne s'agit que d'une approximation puisque le krigeage ne coïncide pas exactement avec l'espérance conditionnelle sauf dans le cas normal pour le krigeage simple.
i.e., strictement:

$$E[F_{KI}(x,c)] \neq P(Z(x) \leq c \mid Z_1, \dots, Z_n)$$

De plus, en réduisant le conditionnement à la variable indicatrice on perd de l'information par rapport à la teneur. Pour retrouver la même information, il faudrait effectuer un cokrigeage en utilisant simultanément les indicatrices définies à tous les seuils (en nombre infini). Même réduit à un nombre fini de seuils, l'entreprise est rarement tentée en raison de son extrême lourdeur. Des méthodes plus simples

ont été proposées comme l'emploi de la teneur comme variable auxiliaire dans un cokrigeage ("probability kriging") ou le krigeage séparé des facteurs obtenus d'une ACP des variables indicatrices.

- Rien ne garantit que les estimés se comportent rigoureusement comme des probabilités. Il est entre autres possible d'avoir des estimés négatifs, supérieurs à 1, ou qui ne croissent pas de façon monotone avec l'augmentation du seuil. Ces problèmes sont mêmes très courants et on doit les corriger avant tout traitement subséquent. La façon de les corriger est nécessairement "ad hoc" (c'est de la cuisine). Il y a deux causes expliquant ce phénomène: les poids de krigeage peuvent être négatifs, les variogrammes modélisés aux différents seuils ne sont pas nécessairement compatibles entre eux.
- On peut fournir des estimés analogues au krigeage ordinaire, par krigeage d'indicatrices, en calculant l'espérance mathématique à partir de la loi de distribution estimée $F_{KI}(x,c)$.
- Les variogrammes des indicatrices sont souvent, dans l'ensemble, plus faciles à modéliser que celui de $Z(x)$. En effet, il n'y a pas de problème de valeurs extrêmes puisque les valeurs sont toutes 0 ou 1. La variance (de dispersion) de l'indicatrice pour un seuil correspondant à un quantile "p" est $p(1-p)$. Le plateau du variogramme devrait, lorsque le champ étudié est grand par rapport à la portée de la structure, correspondre approximativement à cette valeur (sinon il devrait être légèrement supérieur). De plus, le variogramme à l'origine ne peut être dérivable 2 fois, ce qui exclut, entre autres, le modèle gaussien. En effet, si le variogramme était dérivable 2 fois, ceci impliquerait que la variable indicatrice $I(x,c)$ serait dérivable, alors que celle-ci est en réalité discontinue (succession de 1 ou de 0).
- On ne peut choisir de seuils qui soient trop dans les extrémités de la distribution. Si on ne dispose que de 5% des valeurs qui sont des 0 ou des 1, le variogramme sera très erratique et difficile à modéliser. On n'exécède pas généralement les quantiles 10% ou 90%.
- Typiquement, il est nécessaire d'étudier entre 5 et 10 seuils différents pour estimer la fonction de distribution. Le krigeage d'indicatrices est donc conceptuellement simple, mais de réalisation assez lourde. On choisit souvent les déciles de la distribution globale comme seuils.
- Un des arguments en faveur du krigeage d'indicatrices est qu'il permet de modéliser des structures complexes, i.e. qui varient en fonction des différents seuils (par exemple une anisotropie pourrait n'être présente qu'aux seuils les plus élevés). Par opposition, les méthodes basées sur des transformations gaussiennes ne permettent qu'une seule structure. Toutefois, la variable Z est plus riche d'information que l'ensemble des variables indicatrices définies à partir de Z .
- Le problème du changement de support se pose avec autant d'acuité, et même plus, que pour les méthodes basées sur les transformations gaussiennes. La pratique courante consiste à corriger les fonctions par la correction affine.

Malgré ses défauts, le krigeage d'indicatrices a réussi à s'implanter comme une méthode d'estimation répandue. La simplicité mathématique et la facilité de sa mise en oeuvre (un programme de krigeage ordinaire suffit) justifie sa popularité auprès des utilisateurs. Sa robustesse face aux valeurs extrêmes (pour certaines tâches du moins) et sa possibilité de tenir compte de données incertaines ("soft kriging") ont aussi joué. Le KI est particulièrement intéressant lorsqu'il n'y a pas de changement de support d'impliqué.

7.1 Exemple de krigeage d'indicatrices

Soit une configuration de valeurs observées Z_1 à Z_4 , on désire effectuer le KI au point x_0 situé au centre du rectangle formé par x_1, x_2, x_3, x_4 .

$$\begin{array}{ccc} Z_1=2.2 & & Z_2=5.1 \\ & x_0 & \\ Z_3=6.4 & & Z_4=4.7 \end{array}$$

Supposons que l'on choisit les seuils aux valeurs 1,2,3,4,5,6,7.

Par symétrie, le krigeage ordinaire fournit les poids $\lambda_{i,c}=1/4$ pour tout "i" et pour tout "c". Pour d'autres configurations, ce ne serait pas le cas et il faudrait effectuer le krigeage avec le modèle de variogramme retenu pour chaque seuil.

c	$I(x_1,c)$	$I(x_2,c)$	$I(x_3,c)$	$I(x_4,c)$	$I^*(x_0,c)$
1	0	0	0	0	0
2	0	0	0	0	0
3	1	0	0	0	1/4
4	1	0	0	0	1/4
5	1	0	0	1	1/2
6	1	1	0	1	3/4
7	1	1	1	1	1

1- Quelle est la probabilité qu'au point x_0 la valeur Z soit supérieure à 3.5, à 4.3?

$$P(Z_0 > 3.5) = 1 - .25 = .75$$

$$P(Z_0 > 4.3) = 1 - [.25 + .3/1 * (.5 - .25)] = .675 \quad (\text{interpolation linéaire entre les seuils 4 et 5})$$

2- Quelle est la médiane de la distribution au point x_0 ? Quelle est la valeur correspondant au 65^e percentile de la distribution?

médiane: 5

$$65^{\text{e}} \text{ perc.: } 5 + [(.65 - .5) / (.75 - .5)] * (6 - 5) = 5.6 \quad (\text{interpolation linéaire entre les percentiles 50\% et 75\%})$$

3- Quelle est l'espérance mathématique de Z au point x_0 ?

Le résultat du KI donne les probabilités d'être dans les différentes classes définies par les seuils. Pour calculer une espérance mathématique, il faut associer une valeur représentative à chaque classe. Cette valeur peut être obtenue par l'histogramme global initial en calculant la moyenne des observations dans chaque classe. On peut aussi plus simplement prendre le milieu de chaque classe:

$$\begin{aligned} E_{KI}[Z_0] &= 0 * 0.5 + 0 * 1.5 + .25 * 2.5 + 0 * 3.5 + .25 * 4.5 + .25 * 5.5 + .25 * 6.5 \\ &= 4.75 \end{aligned}$$

Note: - Cet estimateur est sans biais pour Z_0 car $I^*(x_0, c)$ (et donc $F_{KI}(x_0, c)$) est sans biais pour $I(x_0, c)$. Ceci signifie qu'en moyenne les erreurs d'estimation sont 0 parce que, en moyenne, la fonction de distribution estimée par KI coïncide avec la fonction de distribution de $Z(x)$.

- L'estimé par krigeage ordinaire au point x_0 serait:

$$.25*(2.2+5.1+6.4+4.7) = 4.6$$

- Dans la pratique, les probabilités pour la dernière classe ne sont pas toujours zéro. Or, il est difficile de sélectionner une valeur représentative pour cette classe (cette classe est semi-ouverte). Un choix arbitraire doit être fait et il est indéniable que ce choix peut influencer le calcul de l'espérance mathématique. Un choix possible est de choisir la valeur représentative pour la dernière classe de façon à ce que la moyenne des espérances mathématiques calculées par KI coïncide avec la moyenne des valeurs krigées (par krigeage ordinaire). Un autre choix pourrait être de calculer la valeur moyenne des données originales dans cette dernière classe.

4- Supposons que l'on connaisse une fonction de coût associée aux valeurs de Z (ce pourrait être par exemple un coût de décontamination qui augmente en fonction du niveau de contamination), quelle est l'espérance du coût si $C(z) = z^2$?

On a:

$$E[C(z, x_0)] = \int_{-\infty}^{\infty} C(z) f(x_0, z) dz \approx \sum_{i=0}^{nc} z_i^2 (I^*(x_0, z_{i+1}) - I^*(x_0, z_i))$$

où nc est le nombre de seuils considérés dans le krigeage d'indicatrices, $I^*(x_0, z_i)$ est la valeur krigée pour le seuil i ($I^*=0$ pour $i=0$ et $I^*=1$ pour $i=nc+1$) et z_i est une valeur représentative de la classe i comprise entre les seuils i et $i+1$. Cette valeur peut être obtenue de l'histogramme global ou plus simplement, on peut prendre le centre de la classe. Pour les deux classes extrêmes, la détermination de z_i est assez délicate et arbitraire.

Ici, $I^*(x, z) = 1/4$ $2 < z < 3$, $4 < z < 7$, 0 ailleurs. L'espérance du coût est donc:

$$1/4*(2.5^2+4.5^2+5.5^2+6.5^2)=24.8$$

5- Quelle est la variance d'estimation si l'on estime Z_0 par l'espérance trouvée en 3?

On a:

$$\begin{aligned} \text{Var}(Z_0 - Z_0^*) &= E[(Z_0 - Z_0^*)^2] = E[Z_0^2] - Z_0^{*2} \\ &= 24.8 - 4.75^2 = 2.24 \end{aligned}$$

Note: On remarque que la variance d'estimation ici est indépendante des variogrammes choisis pour les indicatrices en raison de la symétrie et du fait qu'un krigeage ordinaire a été utilisé pour réaliser le KI. On aurait eu la même variance d'estimation que la continuité soit excellente ou non dans le champ étudié. Ce résultat étonnant, contraire aux attentes, peut être évité si l'on utilise plutôt un système de krigeage simple pour estimer chaque indicatrice.

Donc, pour estimer $I^*(x,c)$, il est recommandé d'effectuer un krigeage simple. i.e.

$$I^*(x_0, c) = \sum_{i=1}^n I_i I(x_i, c) + (1 - \sum_{i=1}^n I_i) F_Z(c)$$

où $F_Z(c)$ est la proportion, parmi l'ensemble des échantillons, inférieure à c (i.e. la fonction de distribution cumulative évaluée à " c ").

Les poids sont obtenus en solutionnant le système KS suivant:

$$\sum_{i=1}^n I_i \text{Cov}(I(x_i, c), I(x_j, c)) = \text{Cov}(I(x_0, c), I(x_j, c)) \quad \forall j$$

On doit répéter le processus pour chacun des seuils considérés. Pour chaque seuil, un variogramme d'indicatrices doit être calculé afin de pouvoir calculer les covariances dans le système précédent.

Note: Dans le cas du KI simple, si le variogramme est un effet de pépite pur, les poids seront 0 et la fonction de distribution locale sera estimée par la fonction de distribution globale. Si le variogramme montre une forte continuité spatiale alors les poids tendront vers 0.25 et les 2 krigeages d'indicatrices fourniront des résultats semblables.

7.2 Problèmes de relation d'ordre

Comme mentionné précédemment, rien ne garantit que le résultat du krigeage d'indicatrice sera une fonction obéissant aux propriétés d'une fonction de distribution:

$$0 \leq F(z) \leq 1$$

$$F(z) \leq F(z') \quad \text{si } z \leq z'$$

Il faut donc, avant tout calcul préalable, s'assurer que $F_{KI}(x_0, z)$ respecte ces propriétés, autrement des résultats farfelus comme des probabilités négatives pourraient être obtenus. Pour cela on doit effectuer des corrections "ad hoc". Il existe plusieurs recettes pour corriger les anomalies, en voici une simple:

- i. Mettre les valeurs négatives de $F_{KI}(x,c)$ égales à zéro et les valeurs supérieures à 1 égales à 1.
- ii. Considérant les seuils par ordre croissant, c_1, c_2, \dots, c_p on définit la fonction corrigée vers l'avant par:

$$F_{KI,avant}(x_0, c_1) = \max[0, F_{KI}(x_0, c_1)]$$
 Pour $i=2$ à p

$$F_{KI,avant}(x_0, c_i) = \max(F_{KI,avant}(x_0, c_{i-1}), F_{KI}(x_0, c_i)).$$
- iii. Inversant la direction, on définit la fonction corrigée vers l'arrière:

$$F_{KI,arr}(x_0, c_p) = \min[F_{KI}(x_0, c_p), 1]$$
 Pour $i=p-1$ à 1

$$F_{KI,arr}(x_0, c_i) = \min(F_{KI}(x_0, c_i), F_{KI,arr}(x_0, c_{i+1}))$$
- iv. La fonction corrigée est obtenue en faisant la moyenne des deux fonctions précédentes.

Exemple de correction

Le tableau suivant présente le détail des calculs décrits précédemment.

seuil c	$F_{KI}(x_0, c)$	$F_{KI,avant}(x_0, c)$	$F_{KI,an}(x_0, c)$	$F_{KI,corr}(X_0, c)$
1	-.01 --> 0	0	0	0
2	.13	.13	.13	.13
3	.24	.24	.234	.237
4	.238	.24	.234	.237
5	.234	.24	.234	.237
6	.237	.24	.237	.2385
7	.53	.53	.53	.53
8	.79	.79	.77	.78
9	.77	.79	.77	.78
10	1.02 -> 1.0	1.0	1.0	1.0

7.3 Modèles de changements de support

La plupart du temps, les décisions doivent être prises sur des supports différents de ceux des données. Or la fonction F_{KI} estimée précédemment est valide pour le support des observations. Ainsi dans une mine, la probabilité qu'un bloc de $125m^3$ excède une teneur de coupure donnée n'est pas la même que pour une carotte de 1m. La variance étant différente, il suit que les fonctions de distribution sont nécessairement différentes également. Il faut donc corriger la fonction F_{KI} pour tenir compte du changement de support.

Note: On pourrait croire qu'il suffit de remplacer le krigeage de points utilisés dans le KI par un krigeage de blocs. Or $I_v^*(x_0, c)$ représente une estimation de la probabilité moyenne qu'ont les points du bloc considéré d'être inférieur à "c". Il ne s'agit pas de la probabilité qu'à la valeur moyenne des points dans le bloc (i.e. la valeur du bloc) d'être inférieure à "c". Il est important de bien saisir la nuance qui signifie que la probabilité n'est pas un opérateur linéaire relativement à la moyenne, tout comme le logarithme n'est pas un opérateur linéaire par rapport à la moyenne (i.e. le logarithme d'une moyenne n'est pas la moyenne des logarithmes).

Il existe plusieurs méthodes de correction pour le changement de support. Avec le KI, la plus utilisée est la correction affine déjà vue dans le chapitre sur les ressources récupérables.

Correction affine:

L'idée de base de cette correction consiste à considérer que la distribution des blocs est identique à celle des points à une contraction près donnée par le rapport des écarts-types de dispersion des blocs et des points.

$$F_v(Z_v) = F \left((Z - m) \left\{ \frac{D^2(v/G)}{D^2(\bullet/G)} \right\}^{0.5} + m \right)$$

Où "m" désigne la moyenne locale (i.e. l'espérance obtenue de la fonction de distribution estimée par KI au niveau ponctuel).

Exemple:

Pour l'exemple précédent, on trouve une espérance conditionnelle de $m=6.33$ (m dans la formule ci-haut). Supposons que l'on a un bloc pour lequel $D^2(v/G) = 0.8 * D^2(\bullet/G)$. Quelle est la probabilité que le bloc excède la valeur 9?

Le seuil ponctuel équivalent est $(1/0.8)^{0.5}(9-6.33)+6.33 = 9.31$

Se référant au tableau précédent, on trouve que la probabilité est:

$$1 - (0.78 + 0.31(1 - 0.78)) = 0.15$$

La probabilité que le point excède le seuil 9 serait par contre:

$$1 - 0.78 = .22$$

Note: La correction affine ne peut être utilisée pour des changements de support trop importants (on demande un rapport $D^2(v/G)/D^2(\bullet/G) > 0.7$). Lorsque la correction pour la variance est trop importante, i.e. lorsque le bloc est grand par rapport à la structure présente, il n'est plus raisonnable de supposer que la forme de l'histogramme demeure inchangée. On doit en effet s'attendre alors à une diminution progressive du nombre de modes et à l'obtention d'un histogramme s'approchant d'une loi normale. On doit alors recourir à des méthodes plus sophistiquées de correction pour changement de support.

Note: Une autre méthode pour obtenir les probabilités pour des blocs consiste à effectuer des simulations ponctuelles qui présentent le même variogramme que les données et qui respectent les valeurs observées aux points échantillons (simulations conditionnelles). On peut alors regrouper les valeurs ponctuelles en blocs et déduire, après un grand nombre de simulations, leur fonction de distribution (locale). Il existe un grand nombre de techniques de simulation.

7.3 Krigeage mou (« soft kriging ») ou krigeage avec données incertaines

Le krigeage d'indicatrices permet une extension immédiate permettant d'inclure des données imprécises ou incertaines.

Ex. Considérons le problème de cartographier la profondeur du toit d'un réservoir pétrolier. Les forages pétroliers coûtent chacun une véritable fortune et peu de forages sont disponibles en général. Il arrive que certains forages doivent être abandonnés après que plusieurs centaines de mètres aient été forés mais avant que le toit du réservoir ne soit atteint. Avec le krigeage ordinaire, ce genre de données peut difficilement être inclus. Avec le KI, le traitement est immédiat :

On a :

$$Z(x_i) > 500m.$$

et $I(x,c)=1$ si $Z(x) \leq c$, 0 sinon.

Pour tous les seuils « c » inférieurs à 500m, on peut coder $I(x_i, c) = 0$. Pour les seuils supérieurs à 500m, on ne peut rien coder (donnée manquante).

ex. Toujours dans le domaine pétrolier, on peut avoir aussi des valeurs sismiques donnant la profondeur approximative du toit du réservoir. Ces profondeurs sismiques incertaines peuvent parfois être représentées par des fonctions de distribution que l'on utilise directement dans le codage des indicatrices aux différents seuils.

