

Notes de cours de l'ISIMA, première année
<http://www.isima.fr/~leborgne>

Introduction au krigeage

Gilles LEBORGNE

19 janvier 2017

Rédaction très rapide et très incomplète.

Table des matières

1	But	1
2	Example dans \mathbb{R}	2
3	Vers la méthode de krigeage	2
3.1	Approximation affine par morceaux	2
3.2	Inverse distance weighting	2
4	Méthode de krigeage	3
4.1	Les fonctions inconnues	3
4.2	Problème du vocabulaire des probabilités	3
4.3	Mesurer l'inhomogénéité : la “matrice” de covariance	4
4.3.1	4
4.3.2	4
4.3.3	4
4.4	On préfère les fonctions croissantes : le variogramme	5
4.5	Le semi-variogramme approchée par une fonction simple	5
4.6	Retour au problème initial § 4.1	5

1 But

Calculer une fonction simple $f : P \rightarrow f(P)$ passant à peu près par des n points (P_i, z_i) . Donc les points (P_i, z_i) sont a peu près sur le graphe de f (l'entier n peut être “grand”).

Démarche.

1- En n points P_1, \dots, P_n on dispose de n valeurs z_1, \dots, z_n (mesures expérimentales entachées d'erreurs).

2- On remarque que les (P_i, z_i) semblent à peu près appartenir au graphe d'une fonction simple $f : P \rightarrow z = f(P)$: on choisit l'allure de la fonction (linéaire, quadratique, log, rationnelle...), c'est la partie “l'art de l'ingénieur”.

3- Le look présupposer de f dépend de quelques paramètres : exemple : si le graphe d'un plan semble convenir (cas où les $P_i \in \mathbb{R}^2$) alors on choisit f sous la forme $f(x, y) = ax + by + c$, et f dépend des 3 paramètres a, b, c .

4- On choisit alors (art de l'ingénieur) un critère pour calculer les paramètres (par exemple méthode des moindres carrées).

5- Et alors pour un point quelconque P on dira que sa valeur estimée est $f(P)$ (choix “raisonnable” pour estimer une mesure virtuelle de f en P).

Définition 1.1 Les étapes 2-,3-,4- sont appelées les étapes de la régression. Une régression consiste à :

- (i) se donner a priori une fonction f “simple”,
- (ii) se donner a priori un critère pour calculer f (exemple : méthode des moindres carrés = “minimiser l'énergie”), et
- (iii) calculer f à l'aide du critère choisi en (ii).

2 Example dans \mathbb{R}

1- $n = 4$ et les 4 points P_i sont $P_1 = -1$, $P_2 = 4$, $P_3 = 3$ et $P_4 = 10$, pour lesquels $z_1 = 6$, $z_2 = 0$, $z_3 = 1$, $z_4 = 12$. Dessin (matlab par exemple).

2- On peut choisir f une droite (mais visiblement c'est un mauvais choix), soit $f(x) = a_0 + a_1x$, ou on peut choisir f une parabole (ça semble meilleur), soit $f(x) = a_0 + a_1x + a_2x^2$, ou tout autre fonction. Dans le premier cas il a deux paramètres à trouver, a_0 et a_1 , et dans le second cas il y a trois paramètre à trouver.

Prenons le choix $f(x) = a_0 + a_1x$ (plus court à rédiger).

3- Choisissons la méthode des moindres carrés : il s'agit de calculer a_0 et a_1 à l'aide des quatre informations (P_i, z_i) , ce qui est brutalement impossible (une droite ne peut pas passer par quatre points non alignés). La droite choisit doit minimiser l'énergie (on a choisi la méthode des moindres carrés), donc a_0 et a_1 sont choisis t.q. qu'ils minimisent la fonction $J(a_0, a_1) = \sum_{i=1}^4 |z_i - f(x_i)|^2 = \sum_{i=1}^4 |z_i - (a_0 + a_1x_i)|^2$. Calcul matriciel usuel.

4- Une fois le minimum trouvé, i.e. une fois a_0 et a_1 trouvés, pour $x \in \mathbb{R}$, on prendra pour valeur estimée en x la valeur $f(x) = a_0 + a_1x$.

3 Vers la méthode de krigeage

3.1 Approximation affine par morceaux

Une méthode très simple pour estimer une valeur $z(P)$ à partir des n valeurs $z_i = z(P_i)$ pour $i = 1, \dots, n$ est la méthode de Lagrange (utilisée en éléments finis) : plutôt que d'essayer de trouver une fonction f simple globale, comme décrit ci-dessus, on cherche une fonction localement affine (dite fonction affine par morceaux).

La méthode de Lagrange consiste à décrire $z(P)$, pour P quelconque, à l'aide les n paramètres. Ici les n paramètres sont n "vecteurs de base", ici n fonctions de base, appelées les fonctions "chapeau" $(\varphi_i)_{i=1, \dots, n}$ (ce sont les fonctions affines par morceaux définies par $\varphi_i(P_j) = \delta_{ij}$). Et la fonction z retenue est donnée :

$$z = \sum_{i=1}^n z_i \varphi_i, \quad \text{au sens} \quad z(P) = \sum_{i=1}^n z_i \varphi_i(P). \quad (3.1)$$

Dans la méthode de krigeage, les inconnues correspondent aux fonctions φ_i , et elles sont souvent notées w_i pour "weighting functions" : ce ne sont pas des fonctions affines par morceaux : la méthode de krigeage est une méthode très particulière pour les calculer. Et d'ailleurs la méthode de krigeage ne les calcule pas vraiment (les w_i) : elle les calcule uniquement en quelques points.

N.B. : ici il n'y a pas d'étape de régression : on donne a priori les fonctions φ_i , donc on se donne a priori z .

Exercice 3.1 Donner et dessiner les $n = 4$ fonctions chapeau de base de l'exemple du paragraphe 2. ■

3.2 Inverse distance weighting

L'idée est : "Things which are closer to each other are more related than distant ones".

Ici au lieu considérer une approximation affine par morceaux, cf. (3.1), on définit z par, si $P = P_i$ alors $z(P_i) = z_i$ et si $P \neq P_i$ alors :

$$z(P) = \sum_{i=1}^n \lambda_i(P) z_i, \quad (3.2)$$

où les fonctions de pondérations λ_i (les "weighting functions") sont données par :

$$\lambda_i(P) = \frac{1}{M_i} \frac{1}{\|P - P_i\|^2} \quad \text{où} \quad M_i = \frac{1}{\sum_{j \neq i} \frac{1}{\|P - P_j\|^2}}, \quad (3.3)$$

où donc $\sum_{i=1}^n \lambda_i(P) = 1$: les $\lambda_i(P)$ sont des pourcentages, la somme valant $\frac{100}{100} = 100\% = 1$.

Ainsi, pour chaque P , pour calculer $z(P)$ on pondère les valeurs z_i par des poids inversement proportionnelle à la distance de P à P_i .

N.B. : ici il n'y a pas d'étape de régression : on donne a priori les $\lambda_i(P)$, donc z .

4 Méthode de krigeage

4.1 Les fonctions inconnues

On cherche toujours à estimer $z(P)$ en fonction des z_i . La forme de $z : P \rightarrow z(P)$ cherchée est la même que dans (3.1) et (3.2) :

$$z = \sum_{i=1}^n w_i z_i, \quad \text{au sens} \quad z(P) = \sum_{i=1}^n w_i(P) z_i, \quad (4.1)$$

et les inconnues sont les fonctions $w_i : P \rightarrow w_i(P)$.

Suivant les références, z est noté v , ou ..., les points P sont notés x ou u ou...

Dans le cas d'un "milieu assez homogène", les fonctions w_i trouvées ne seront "pas très différentes" des fonctions λ_i données en (3.3) : la méthode de krigeage n'a aucun intérêt dans ce cas. La méthode de krigeage est intéressante dans le cas d'un milieu "suffisamment non homogène".

4.2 Problème du vocabulaire des probabilités

1- La valeur moyenne d'une fonction f sur un ensemble Ω est formellement :

$$\text{valeur moyenne de } f = m(f) = \frac{1}{\int_{\Omega} dm} \int_{x \in \Omega} f(x) dm(x), \quad (4.2)$$

où dm est la "densité" et $\int_{\Omega} dm$ est la "masse de Ω ".

1'- Quand on ramène la masse totale à $1 = \frac{100}{100} = 100\%$, la démarche utilisée est souvent celle des probabilités, avec son vocabulaire particulier... On pose alors :

$$dp = \frac{1}{\int_{\Omega} dm} dm, \quad \text{où donc} \quad \int_{x \in \Omega} dp(x) = 1 = \frac{100}{100} = 100\%, \quad (4.3)$$

et la valeur moyenne est appelée l'espérance :

$$\text{espérance de } f = E(f) = \int_{x \in \Omega} f(x) dp(x) \stackrel{\text{noté}}{=} \bar{f}, \quad (4.4)$$

2- Le moment d'ordre 2 de f autour de sa valeur moyenne est formellement :

$$M_2(f) = \frac{1}{\int_{\Omega} dm} \int_{x \in \Omega} (f(x) - m(f))^2 dm(x), \quad (4.5)$$

2'- En probabilité c'est appelé la variance :

$$V(f) = \int_{x \in \Omega} (f(x) - \bar{f})^2 dp(x). \quad (4.6)$$

3- Le moment $M_2(f)$ défini trivialement une semi-norme associée au semi-produit scalaire :

$$(f, g)_2 = \frac{1}{\int_{\Omega} dm} \int_{x \in \Omega} (f(x) - m(f))(g(x) - m(g)) dm(x). \quad (4.7)$$

On dispose alors de la notion de "presque-orthogonal" quand $(f, g)_2 = 0$.

Comme $(\cdot, \cdot)_2$ est un produit scalaire sur le sous-ensemble des fonctions de valeur moyenne nulle, on a donc, dans ce sous-ensemble, f est orthogonal à g ssi $(f, g)_2 = 0$.

3'- En probabilité ce semi-produit scalaire est appelé la covariance :

$$\text{cov}(f, g) = (f, g)_{\text{cov}} = \int_{x \in \Omega} (f(x) - \bar{f})(g(x) - \bar{g}) dp(x). \quad (4.8)$$

On dispose donc de la notion de "presque-orthogonalité".

Et on a la notion d'orthogonalité sur l'ensemble des fonctions d'espérance nulle : deux fonctions orthogonales (de moyenne nulle) sont dites indépendantes.

Donc $(f, g)_{\text{cov}}$ mesure "la dépendance ou l'indépendance relative de f et de g ", pour les fonctions de moyenne (espérance) nulle.

4.3 Mesurer l'inhomogénéité : la “matrice” de covariance

Soit $z : P \rightarrow z(P)$ la fonction (encore inconnue cf. (4.1)) qui au point P associe sa valeur $z(P)$.

Pour illustrer, on peut penser à $z(P)$ “température à P ”, ou à $z(P)$ “porosité à P ”.

On ne connaît (grâce aux mesures) que des valeurs approchées des $z(P_i)$ pour $i = 1, \dots, n$.

4.3.1

Soit un P_i fixé. Soit $h > 0$. Soit l'ensemble des points à la distance h de P_i :

$$E(i, h) = \{\text{les } P_j \text{ t.q. } \|P_i - P_j\| = h \text{ où } j > i\}, \quad \text{et} \quad N(i, h) = \text{cardinal de } E(i, h). \quad (4.9)$$

Pour $N(i, h) \neq 0$, la valeur moyenne des z_j sur $E(i, h)$ et la variance sont :

$$\begin{aligned} m(i, h) &= \frac{1}{N(i, h)} \sum_{P_j \in E(i, h)} z(P_j), \\ V(i, h) &= \frac{1}{N(i, h)} \sum_{P_j \in E(i, h)} (z(P_j) - m(i, h))^2. \end{aligned} \quad (4.10)$$

Remarque 4.1 Dans la pratique on considère les P_j à une distance $h \pm \varepsilon$ de P_i où $\varepsilon \ll h$ où ε est à définir par “l'art de l'ingénieur”. Donc dans la suite on pourra remplacer $\|P_i - P_j\| = h$ par $\|P_i - P_j\| \in [h - \varepsilon, h + \varepsilon]$. ■■

4.3.2

On considère tous les couples (P_i, P_j) t.q. $\|P_i - P_j\| = h$.

Pour un problème de notation dans la suite, redéfinissons $E(i, h)$ par (i et h sont fixés) :

$$E(i, h) = \{\text{les couples } (P_i, P_j) \text{ t.q. } \|P_i - P_j\| = h \text{ où } j > i\}. \quad (4.11)$$

On peut maintenant considérer $\bigcup_{i=1}^n E(i, h) =$ l'ensemble de tous les couples (P_i, P_j) t.q. $\|P_i - P_j\| = h$.

Mais pour ne pas compter les couples deux fois, on réduit l'union à, pour $h > 0$:

$$E(h) = \text{l'ensemble des couples } (P_i, P_j) \text{ t.q. } \|P_i - P_j\| = h \text{ où } j > i, \quad (4.12)$$

et on note :

$$N_h = \text{cardinal}(E(h)), \quad (4.13)$$

le nombre des couples (P_i, P_j) t.q. $\|P_i - P_j\| = h$.

En pratique : pour $h > 0$ on forme la matrice triangulaire supérieure stricte avec P_i en ligne et $\|P_j - P_i\|$ à la colonne j , et regarde les termes de la matrice qui valent h .

Et on pose $E(0) =$ les couples (P_i, P_i) . Et on note $N_0 = n$.

On note m_h la valeur moyenne dans E_h (moyenne des valeurs moyennes) :

$$\begin{aligned} m(h) &= \frac{1}{n} \sum_{i=1}^n m(i, h) = \frac{1}{N_h} \sum_{(P_i, P_j) \in E(h)} z(P_j) \stackrel{\text{noté}}{=} m_h, \quad \text{et donc,} \\ m(0) &= \frac{1}{n} \sum_{i=1}^n m_i \stackrel{\text{noté}}{=} m_0. \end{aligned} \quad (4.14)$$

4.3.3

N.B. : les couples $(P_i, P_j) \in E_h$ sont des couples ordonnés : la première composante P_i représente le point P_i un point et le point P_j un des points à la distance h de P_i .

On note X_0 et X_h les fonctions (appelées variables aléatoires par les probabilistes) définie sur E_h par : si $(P_i, P_j) \in E_h$ alors :

$$X_0(P_i, P_j) = z(P_i), \quad X_h(P_i, P_j) = z(P_j). \quad (4.15)$$

On dispose donc de deux fonctions (deux variables aléatoires). On peut donc considérer leur covariance :

$$\text{cov}(X_0, X_h) = \sum_{(P_i, P_j) \in E(h)} (X_0(P_i, P_j) - m_0)(X_h(P_i, P_j) - m_h) \stackrel{\text{noté}}{=} C(h). \quad (4.16)$$

Cette quantité est notée de manière abusive (usuelle) :

$$C(h) = \frac{1}{N_h} \sum_{k=1}^{N_h} (z_k - m_0)(z_{k+h} - m_h). \quad (4.17)$$

N.B. : la fonction $C : h \rightarrow C(h)$ n'est pas une matrice de co-variance (une co-qqc se réfère à deux éléments...). Elle est néanmoins appelée une fonction de covariance...

4.4 On préfère les fonctions croissantes : le variogramme

Il se trouve que la fonction $C : h \rightarrow C(h)$ définie en (4.16) est décroissante.

Et qu'on préfère utiliser les fonctions croissantes...

On pose alors :

$$\Gamma(h) = C(0) - C(h). \quad (4.18)$$

Cette fonction $\Gamma : h \rightarrow \Gamma(h)$ est appelée le "variogramme".

Et que l'usage fait qu'on utilise le "semi-variogramme" :

$$\gamma(h) = \frac{1}{2}\Gamma(h) = \frac{C(0) - C(h)}{2}. \quad (4.19)$$

4.5 Le semi-variogramme approchée par une fonction simple

On regarde sur un graphe à quoi ressemble le semi-variogramme, on choisit une fonction simple qui a l'air d'approcher : c'est la partie "régression" : choix de la fonction simple et son calcul, voir définition 1.1.

On obtient :

$$\hat{\gamma}(h) = \text{approximation simple de } \gamma(h). \quad (4.20)$$

4.6 Retour au problème initial § 4.1

Tout ce qu'on vient de présenter : c'est pour calculer les fonctions w_i de (4.1) (on suppose que les méthodes des § 3.1 et 3.2 ne sont pas satisfaisantes).

En fait ce n'est pas pour calculer la fonction w_i , mais uniquement en un point $P = \stackrel{\text{noté}}{=} P_0$. Donc dans (4.1) on n'obtiendra qu'une seule valeur estimée :

$$z(P_0) = \sum_{i=1}^n w_i(P_0) z_i. \quad (4.21)$$

Le calcul est simple : on forme la matrice $G = [G_{ij}]$ où :

$$G_{ij} = \hat{\gamma}(\|P_i - P_j\|) \stackrel{\text{noté}}{=} \hat{\gamma}(h_{ij}). \quad (4.22)$$

On forme le vecteur $\vec{v} = \begin{pmatrix} v_1 = \hat{\gamma}(\|P_1 - P_0\|) = \stackrel{\text{noté}}{=} \hat{\gamma}(h_{10}) \\ \vdots \\ v_n = \hat{\gamma}(\|P_n - P_0\|) = \stackrel{\text{noté}}{=} \hat{\gamma}(h_{n0}) \end{pmatrix}$. On résout le système :

$$G \cdot \vec{w} = \vec{v}, \quad (4.23)$$

où l'inconnue est $\vec{w} = \begin{pmatrix} w_1(P_0) \\ \vdots \\ w_n(P_0) \end{pmatrix}$.