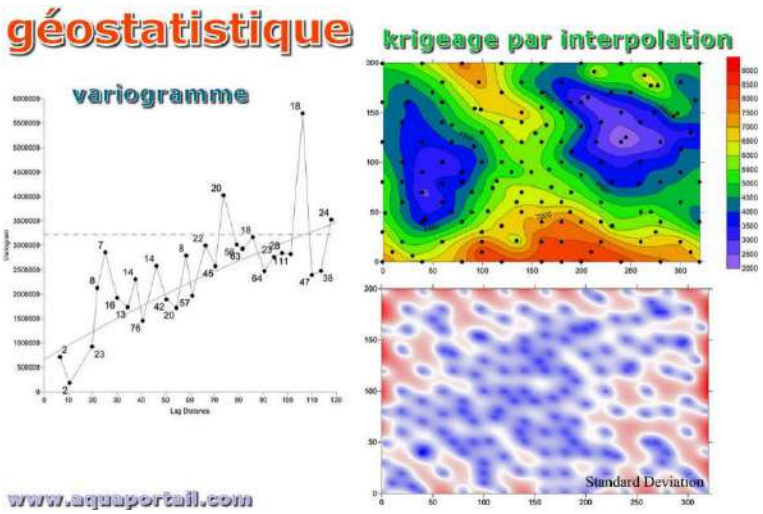
Matière : Géostatistique
Coeff. 2 Crédit. 4
Cours : 01h30 TD : 1h30

Cours de Géostatistique



L3/S5

Responsable du cours Dr. Cheriet Manel

Chapitre 1/ Introduction à la Géostatistique

1-Introduction à la géostatistique

La géostatistique consiste à étudier les phénomènes corrélés dans l'espace, au moyen
D'un outil probabiliste : ***" la théorie de variables régionalisées ".***
La géostatistique est une application de la théorie des fonctions aléatoires à des données localisées dans un espace géographique
Il existe deux définitions de la notion « géostatistique » :

1- La Géo-statistique = Statistique appliquée aux sciences géologiques et sciences de la terre

2-La géostatistique (Matheron 1971) : La géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et à l'estimation des phénomènes naturels repartis dans l'espace (phénomènes régionalisés) et/ou dans le temps (Minéralisation, pollution, propriété physique de roches,...).

1.1. Méthodes géostatistiques

Les méthodes géostatistiques, ont été initialement proposées en exploration minière et pétrolière telles que le krigeage.
La géostatistique est classiquement subdivisée en géostatistique linéaire et multivariable, géostatistique non-linéaire, simulations géostatistiques.

1.2. Modèles et types de la géostatique

La géostatistique étudie des phénomènes naturelles répartie dans l'espace (phénomènes régionalisés) et/ou dans le temps (Minéralisation, pollution, propriété physique de roches, pluviométrie.....) La géostatistique étudie des phénomènes naturelles répartie dans l'espace (phénomènes régionalisés) et/ou dans le temps (Minéralisation, pollution, propriété physique de roches, pluviométrie.....)
1 - Si au point x_i, la variable régionalisée $Z(x_i)$ est considérée comme valeur unique (valeur vraie), dans ce cas, la géostatistique étudiera la corrélation spatiale de la V.R. Z(x) et la structure de cette variable dans l'espace (fig.1)

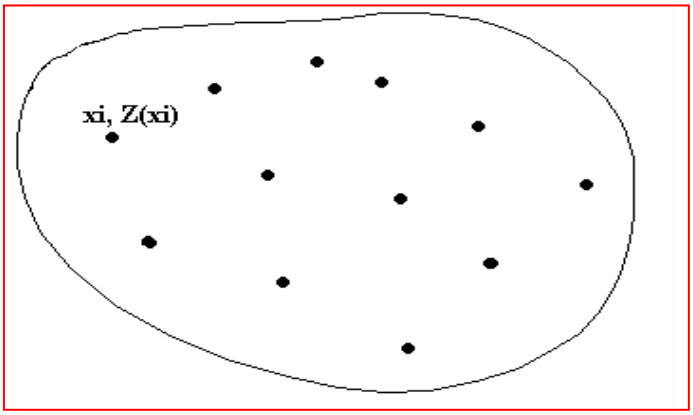


Fig. 1 – Localisation des points de mesures X**i (**Répartition des données)

1.2. Rappelles sur les statistiques descriptives

1.2.1. Typologie des variables

On appelle variable le caractère sur lequel porte une étude d'un ensemble d'individus et qui change de l'un a l'autre. Si le changement de ce caractère est imprévisible, la variable est dite *variable aléatoire*. Si cette variable aléatoire est répartie dans l'espace, dite *variable régionalisée*

1- **Variable qualitative** : La variable est dite qualitative quand les modalités sont des catégories.
2- **Variable qualitative nominale** : La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées.
3- **Variable qualitative ordinale** : La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées. Le fait de pouvoir ou non ordonner les modalités est parfois discutable. Par exemple : dans les catégories socioprofessionnelles, on admet d'ordonner les modalités : 'ouvriers', 'employés', 'cadres'. Si on ajoute les modalités 'sans profession', 'enseignant', 'artisan', l'ordre devient beaucoup plus discutable.
4- **Variable quantitative** : Une variable est dite quantitative si toutes ses valeurs possibles sont numériques.
5- **Variable quantitative discrète** : Une variable est dite discrète, si l'ensemble des valeurs possibles est dénombrable.
6- **Variable quantitative continue** : Une variable est dite continue, si l'ensemble des valeurs possibles est continu.

Exemple : le nombre de pièce d'une maison est une variable discontinue, alors que la teneur d'un élément chimique dans la croute terrestre est une variable continue

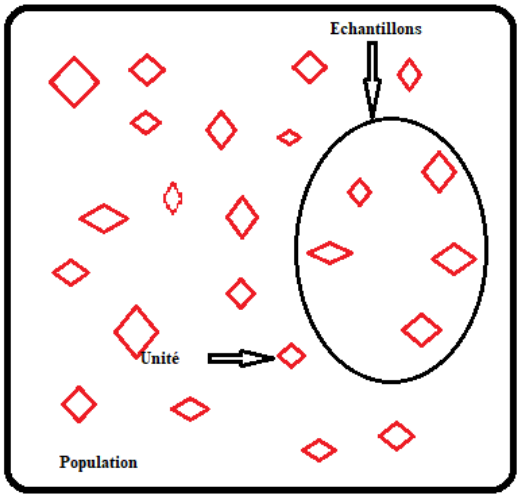


Figure 1. Schéma montrant la population, l'échantillon et les individus statistiques

1.2.2. Classes et intervalles de classes

Les unités d'une population statistique peuvent être représentées individuellement ou regroupés par classes (intervalles) de valeurs ou de critères qualitatifs. Pour les critères quantitatifs, chaque classe possède **un centre (valeur centrale)** qui est la moyenne des deux bornes de la classe. L'**amplitude** d'une classe est la différence des bornes de la classe.

1.2.3. Effectifs et fréquences

* L'effectif ou la fréquence absolue *ni* est le nombre de fois qu'une valeur ou un critère se répète dans une série statistique. La somme des effectifs absolue est l'effectif total *N*.

1.2.3.1. Effectif cumulé croissant associé à la classe *i* : c'est le nombre d'individus ayant un caractère inférieur ou égal à ceux de la classe *i :*

$$N_i^{c\uparrow} = \sum_{j=1}^{i} n_j$$

1.2.3.2. Effectif cumulé décroissant associé à la classe *i* : c'est le nombre d'individus ayant un caractère supérieur ou égal à ceux de la classe *i* :

$$N_i^{c\downarrow} = \sum_{j=i}^{p} n_j$$

* La fréquence relative est égale à la fréquence absolue divisée par la fréquence totale :

$$fi = ni/N$$

1.2.3.3. La fréquence cumulée croissante associée à la classe *i* : est la somme des effectifs des modalités qui lui sont inférieures ou égales à ceux de la classe *i :*

$$F_i^{c\uparrow} = \sum_{j=1}^{i} f_j$$

1.2.3.4. La fréquence cumulée décroissante associée à la classe *i* : est la somme des effectifs des modalités qui lui sont supérieures ou égales à ceux de la classe *i.*

$$F_i^{c\downarrow} = \sum_{j=i}^{p} f_j$$

1.3. Rappel sur les statistiques linéaires (mono et bivariable)

I.3.1. Paramètres de position centrale

Un indicateur de position est un nombre réel permettant de situer les valeurs d'une série statistique d'une variable quantitative. Les principaux paramètres de position centrale sont :

*** le mode** : le mode d'une série statistique est une valeur du caractère correspondant au plus grand effectif (ou à la plus grande fréquence) par rapport aux autres caractères qui les entourent.

Une série statistique peut avoir plusieurs modes. Si la variable est continue, ses modalités sont des classes de valeurs**.** Le mode de distribution ne pourra pas être une modalité représentant une valeur précise de cette variable mais sera une classe de valeurs. On appelle alors **classe modale** la classe constituant **le mode de la distribution**.

Exemple :

Soit la série statistique suivante : le mode correspond à la valeur qui a l'effectif le plus élevé, qui donc la valeur 27.

	Xi	*ni (effectif)*
	12	2
	16	13
	22	15
mode	27	◄ 17
	32	14
	35	10
	45	9

La médiane La médiane est la valeur de la variable qui permet de partager la population étudiée en deux telle que la moitié des individus de la population prenne une valeur qui lui soit inférieure, l'autre moitié des individus de la population prenant par conséquent une valeur qui lui soit supérieure.
On note généralement la médiane : **Mé**.
La valeur de la médiane est déterminée par la formule $N/2$ (N effectif total). Si la variable est continue la médiane est donc la classe correspond à l'effectif $N/2$ (ou $N/2$ fait partie).

* **La moyenne** d'une série statistique est calculée par la formule :

$$\bar{X} = \sum_{i=1}^{p} ni * xi/N$$

Donc on peut écrire la moyenne en fonction de fréquence relative :

$$\bar{X} = \sum_{i=1}^{p} ni * xi/N$$

I.3. Paramètres de dispersion

*** Étendue :** C'est la différence des valeurs extrêmes de la série (en valeur absolue).
Exemple : soit la série S = {4 ; 2; 3; 0; 2; 1; 3; -1; 3}. L'étendue vaut 4 - (-1) = 5.

*** l'écart-type et la variance :** La variance d'une série est la quantité notée var(X) ou S2 calculé par la formule *:* $\boldsymbol{var}(\boldsymbol{X})=\mathbf{1}\ \boldsymbol{N}\ \Sigma(\boldsymbol{xi}-\bar{\boldsymbol{X}})\mathbf{2}$

L'écart-type est la racine au carré de la variance : $\boldsymbol{S(X)} = \sqrt{(\boldsymbol{X})}$

**** Quartiles***
Soit *X* une série, on définit les quartiles *Q1, Q2 et Q3* de la manière suivante :
 Q1 est une valeur du caractère telle que 25% de la population a un caractère inférieur à *Q1*.

 *Q*2 est une valeur du caractère telle que 50% de la population a un caractère inférieur à *Q2:*

 *Q*3 est une valeur du caractère telle que 75% de la population a un caractère inférieur à *Q*3.

Remarque :

 On note que Q2 = Me

 L'intervalle [Q1-Q3] s'appelle l'intervalle interquartile. Il contient 50% de la population.

*** Fonctions de répartition**

Soit *X* une série. On appelle fonction de répartition F, la fonction qui a une valeur du caractère x cumulée croissante jusqu'à *xi*

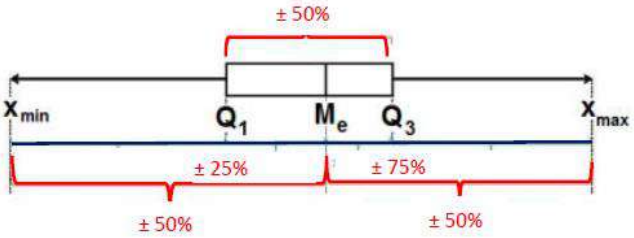
$F : \{caractères\} \rightarrow [0;1]$ donc : $X \mapsto fXc\uparrow = fxi<X$

Avec cette définition, les quartiles sont simplement définis par :

$Q1 = F\text{-}1(0;25)$; $Q2 = F\text{-}1(0;5)$ *et* $Q3 = F\text{-}1(0;75)$

*** La boîte à moustaches**

Dans les représentations graphiques de données statistiques, la boîte à moustaches (aussi appelée diagramme en boîte, boîte de Tukey ou box plot) est un moyen rapide de figurer le profil essentiel d'une série statistique quantitative. Une boîte à moustaches nous indique de façon simple et visuelle quelques traits marquants de la série observée. Ce graphe permet de comparer plusieurs séries d'un seul coup d'oeil.



La boîte à moustaches

1.4. L'analyse bivariée

L'analyse bivariée est une méthode statistique qui examine la relation entre deux variables. Elle permet de comprendre comment une variable est associée à une autre et de déterminer la nature et la force de cette association (Pieretti and Weiland 1996).

1.4.1. Méthodes d'analyse bivariée

1.4.1.1.Diagramme de dispersion (Scatter Plot) : Un graphique montrant les points de données pour deux variables quantitatives, permettant d'observer visuellement la relation.

Exemple :

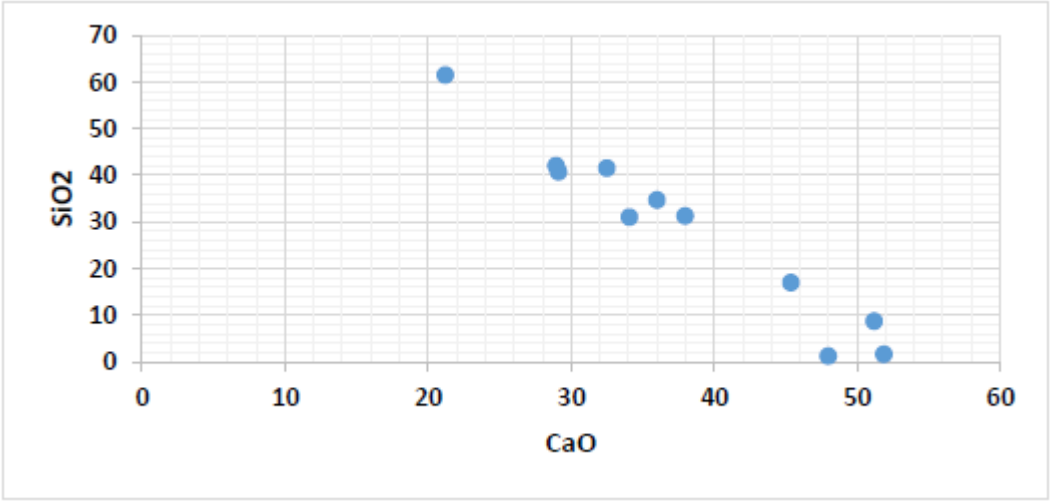
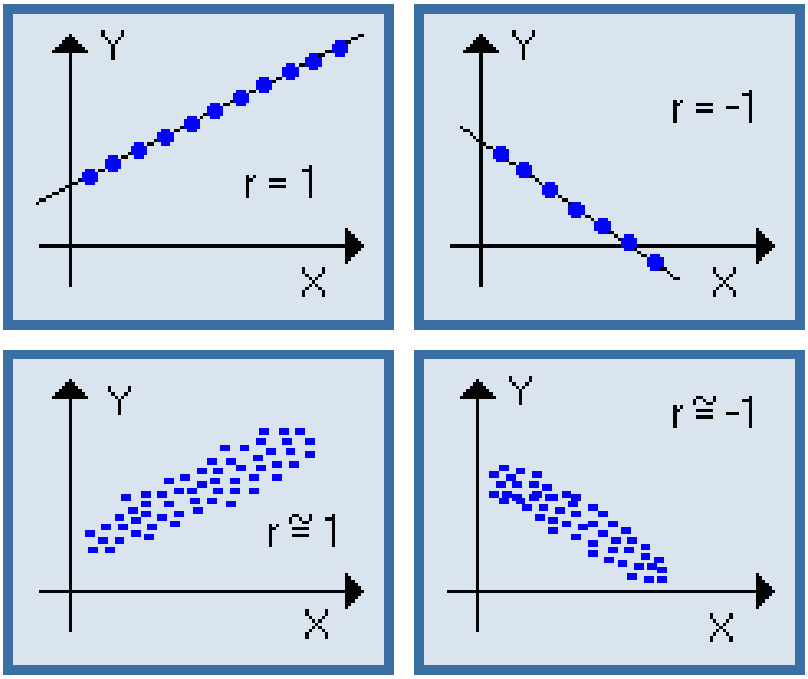


Figure .2 Digramme de dispersion de deux variables : SiO2 et CaO dans une roche

A. **Coefficient de Corrélation de Pearson :** Mesure de la force et de la direction de la relation linéaire entre deux variables quantitatives.

$$r(x,y) = \frac{\mathrm{cov}(x.y)}{\sigma x.\, \sigma y}$$

- r varie de -1 à 1.
- $r > 0$ indique une relation positive.
- $r < 0$ indique une relation négative.
- $r = 0$ n'indique aucune relation linéaire.



Remarque : Pour pouvoir parler de forte liaison entre x et y il faut que la valeur absolue de r atteigne au moins 0.87

Régression Linéaire : Modélisation de la relation entre une variable indépendante *X* et une variable dépendante *Y* pour prédire *Y* en fonction de *X*.

$$yi=axi+b$$

- *a* est l'ordonnée à l'origine (intercept).
- *b* est la pente (slope).

$$a = \frac{\mathrm{cov}(x, y)}{\mathrm{V}(x)}$$

$$b=\overline{y}-a\overline{x}$$

Pour la régression linéaire simple, les coefficients *a* (l'ordonnée à l'origine) et *b* (la pente) sont calculés en utilisant la méthode des moindres carrés. Cette méthode minimise la somme des carrés des différences entre les valeurs observées et les valeurs prédites par le modèle.
La méthode des moindres carrés minimise la somme des carrés des erreurs (SCE : résidus) :

$$\mathrm{SCE} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

où *Yi* est la valeur observée et *Yi* est la valeur prédite par le modèle.

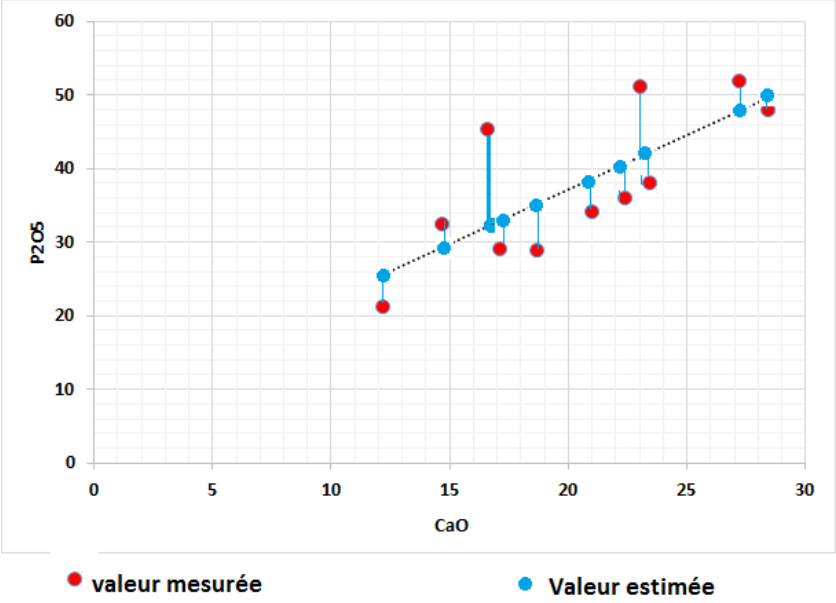


Figure 3. Principe de la méthode des moindres carrés

Les formules pour les coefficients sont :

$$b = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}$$

$$a = \bar{Y} - b\bar{X}$$

Chapitre 2: Les Probabilités

Probabilité

Une probabilité correspond à une fonction permettant de « mesurer » la chance de réalisation d'un évènement de P(Ω) (ou plus généralement d'une tribu A).

Définition : Soit (Ω, A) un espace probabilisable. Une probabilité sur (Ω,A) est une application $P \longrightarrow [0,1]$ satisfaisant les 3 conditions suivantes (Mountassir 2014):

$\mathbf{0 \leq (A) \leq 1}$
$\mathbf{P(\Omega) = 1}$
$\mathbf{P(\cup Ai) = \Sigma P(Ai)}$

Dès lors que P est définie, (Ω, A, P) s'appelle un espace de probabilité.

Opérations sur les probabilités

*** $(\phi)=0$**
*** $(\bar{A})=1-(A)$**
*** $0 \leq (A) \leq 1$**

II.4. Espérance mathématique

L'espérance mathématique ou **moyenne théorique**, noté *E(x),* est égale à la somme des produits des probabilités successives par leur valeur.

$$(x) = \Sigma xi \times (X = xi)\ ki = 1$$

Si $P(X = xi)$ est remplacé par ni/N (qui est *fi*), l'espérance mathématique peut alors s'écrire :

$$(x) = \Sigma xi \times fiki = 1$$

On obtient ainsi une identité entre les notions de moyenne arithmétique et l'espérance mathématique.

*** Propriété de l'espérance mathématique :**
soit a et b des constantes :

**** E(ax) = a x E(x)***
**** E(ax + b) = a x E(x) + b***
**** E(x + y) = E(x) + E(y)***
E[XY] = E[X].E[Y] + Cov(X; Y)

Cov(X; Y)*: c'est la covariance des deux variables X et Y.*

On définit la *covariance* par la quantité :

Cov($X; Y$) = E[(X - E[X])(Y - E[Y])]

 Moment d'ordre n (Saporta 2006)**:**

- Définition : Si X est une variable aléatoire, on appelle moment d'ordre k, s'il existe, le nombre $E(x_k)$. Donc l'espérance mathématique est le moment d'ordre 1.
Si X est une variable aléatoire discrète, son moment d'ordre k se calcule par la formule :

$$m_k = \sum_{i=1}^{q} x_i^k P(X = x_i).$$

Si X est une variable aléatoire continue, alors ce même moment se calcule de la façon suivante :

$$m_k = \int_{\mathbb{R}} x^k f(x) dx.$$

Il existe encore différents types de moments :

 le moment centré d'ordre k :

$$\mu_k = E\left[(X - E(X))^k\right].$$

La variance d'une variable aléatoire est donc le moment centré d'ordre 2.
On peut donc écrire :
$\sigma^2 = \Sigma[(x_i - E(x))^2 . P(X = x_i)]_{ki=1}$
En remplaçant $P(X = x_i)$ par f_i *et* $E(x)$ *par* $\bar{X}$ on obtient donc *:* $S^2 = \sigma^2 = \Sigma[(x_i - \bar{X})^2 . f_i]_{ki=1}$

II.5. Loi d'une variable aléatoire

Une variable aléatoire est totalement définie par sa loi de probabilité. Cette dernière est caractérisée par (Mountassir 2016):

1. l'ensemble des valeurs qu'elle peut prendre (son domaine de définition D_X);
2. les probabilités attribuées à chacune de ses valeurs :

$\boldsymbol{P(X = x_i) : 0 \leq (X = x_i) \leq 1}$

2. Éléments du calcul des probabilités

2.1.Vocabulaire probabiliste

2.1.1.Expérience aléatoire

Une expérience est dite aléatoire si :

a- On ne peut prédire avec certitude son résultat

b- On peut décrire l'ensemble de tous les résultats possibles.

***Exemple* : jet d'un dé ; lancer d'une pièce de monnaie, comportement d'achat d'une personne.**

2.1.2. Ensemble fondamental

(Appelé également univers des possibles, espace échantillonnal ou référentiel) représente l'ensemble des résultats possibles d'une expérience aléatoire ; il est noté**Ω.**

***Exemple* : Si on lance un dé une seule fois, l'ensemble des résultats possibles sont**

Ω = {1, 2, 3, 4, 5, 6}.

2.2.3.Evènement

C'est un élément ou sous ensemble deΩ. On distingue l'événement élémentaire : obtenir 2 de l'événement composé, obtenir un nombre impair.

2.2.Définition classique d'une probabilité

Soit Ω un ensemble fondamental et A un événement quelconque de Ω :

$$P(A) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} = \frac{\text{Card A}}{\text{Card } \Omega}$$

- **2.2.1. Définition fréquentielle**

Soit Ω un ensemble fondamental et A un événement quelconque de Ω.

$$P(A) = \lim_{n \to \infty} f_n(A)$$

avec

n : nombre de fois que l'expérience se répète

et

$f_n(A) = \frac{n(A)}{n}$: fréquence de la réalisation de l'événement A au cours des n répétitions.

Exemple :

Un professeur de statistique a enseigné à 12848 personnes, parmi celles-ci 542 ont échoué

La probabilité d'échouer est 542/12848=0.0422

2.2.2. Les règles de calcul des probabilités

- La probabilité de réalisation d'un événement impossible est égale à 0.
- La probabilité de réalisation d'un événement certain est égale à 1.
- Si A et B sont deux événements incompatibles, alors la probabilité de la réalisation simultanée des deux événements est la somme des probabilités : P (A ∪B) = P(A) + P(B).
- La probabilité de l'événement contraire de A est 1-P(A)

Remarque :

Si A et b ne sont pas deux événements compatibles, alors :

P(A ∪B) = P(A) +P(B)-P (A ∩ B)

Exemple :

On jette un dé une seule fois, soient les deux événements suivants :

A : obtenir un chiffre pair

B : obtenir un chiffre inférieur à 3

Calculer p (A /B) ?

Solution :

$P(A) = 3/6$
$P(B) = 3/6$
$P(A \cap B) = 1/6$
$P(A/B) = (1/6) / (3/6) = 1/3$

Si A est dépendant de B, cela signifie que si B s'est produit, la probabilité que A se produise n'est pas la même que si B ne l'est pas.

En retenant les données de l'exemple précédent, on peut dire que A et B sont deux événements dépendants car : $p(A) \neq p(A/B)$

2.3.Notion de variable aléatoire

Une variable aléatoire est une grandeur numérique attaché au résultat d'une expérience aléatoire. Chacune de ses valeurs est associé à une probabilité d'apparition.

Exemple 1 : On jette une pièce de monnaie deux fois et on s'intéresse au nombre de fois que pile apparaît au cours des deux jets.

On à quatre résultats possibles : PP, PF, FP, FF

Le nombre de fois que Pile peut apparaître est 0, 1 ou 2.

La variable aléatoire retenue peut donc prendre ces trois valeurs, son ensemble de définition est donc : {0, 1, 2}

Une VA peut être discrète ou continue :

- Une VA est dite discrète si l'ensemble des valeurs qu'elle est susceptible de prendre est fini ou infini dénombrable.
- Une VA est dite continue si elle peut prendre toute valeur à l'intérieur d'un intervalle donné.

2.3.1.Les caractéristiques d'une variable aléatoires discrètes

a-Loi de probabilité *:*

On appelle loi de probabilité de X l'ensemble des couples (xi, pi).

b-Fonction de répartition :

On, appelle fonction de répartition, la fonction F définie par :

F: IR -->[0,1]

c-Espérance mathématique:

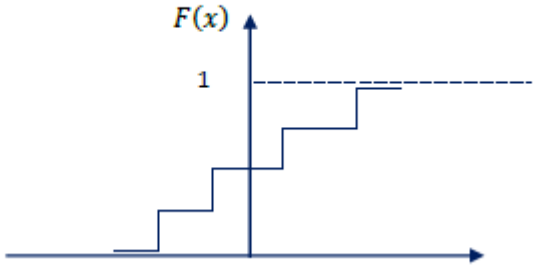
On appelle espérance mathématique de X et on note E(X) la moyenne des valeurs possibles pondérées par leurs probabilités :

E(X) = Σ *xi.pi.*

Pour une variable discrète : $(x) = (X \leq xi) = \Sigma(X = i)_{ni=1}$

(x) est une fonction en escalier, continue à droite.

$F(x)$ est une fonction en escalier, continue à droite.



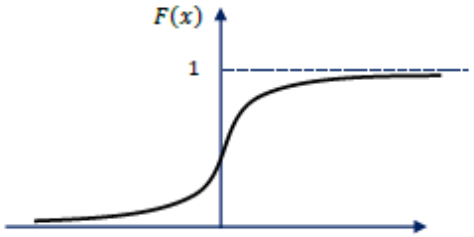
Graphe de fonction de répartition d'une v.a. discrète

*** Pour une variable continue :**

La probabilité ponctuelle $P(X = x) = f(x)$ est appelée la fonction de densité.

La fonction de répartition est : $(x) = (X \leq xi) = \int (x)dx_{xi-\infty}$

(x) est une fonction continue



Graphe de fonction de répartition d'une v.a. continue

2.4. Variance et écart type

On appelle variance de la VA X le nombre réel défini par :

$V(X) = E[X - E(X)]^2 = E(X^2) - E(X)^2$

On appelle écart type, la racine carrée de la variance

2.4.1.Les caractéristiques d'une variable aléatoires continue

Fonction de densité de probabilité *: On* appelle fonction de densité de probabilité toute fonction satisfaisant aux 2 conditions suivantes :

$$\forall x \in IR, f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

La densité de probabilité d'une variable aléatoire continue est la dérivée première par rapport à x de la fonction de répartition. Cette dérivée prend le nom de fonction de densité.

La loi de d'une variable aléatoire X est définie par sa fonction de densité $f(x)$ de R dans R. sa fonction de densité $f(x)$ de R dans R. Cette fonction est caractérisé par :

$(x)\geq 0$

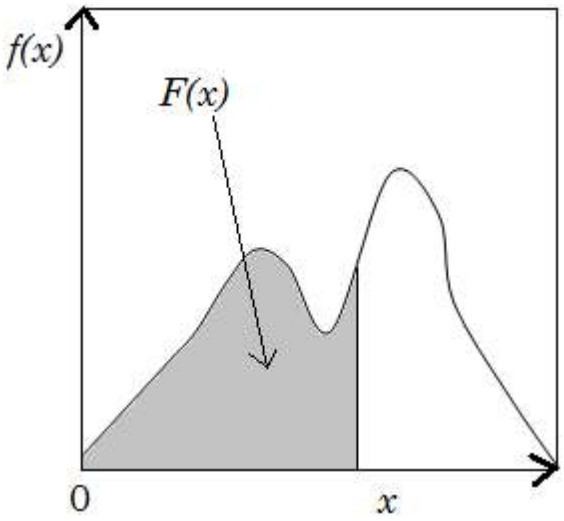
$\int(x)=1+\propto$

Fonction de répartition *: Soit* X une VA continue et f sa densité de probabilité. La fonction de répartition de X est la fonction F telle que :

$$E(x)=\int_{-\infty}^{+\infty} xf(x)dx$$

$$V(x)=\int_{-\infty}^{+\infty}(x-E(x))^2 f(x)dx=\int_{-\infty}^{+\infty} x^2 f(x)dx-(\int_{-\infty}^{+\infty} xf(x)dx)^2$$

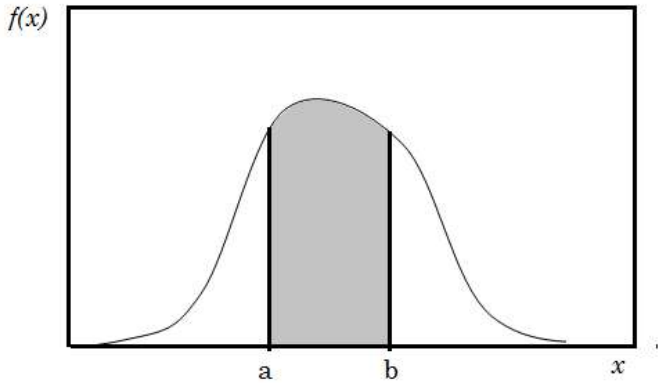
Rappelant que la fonction de répartition est une intégrale de la fonction de densité. Il est également possible d'utiliser cette dernière pour représenter graphiquement la fonction de répartition qui correspond donc à une surface dans ce cas (Figure)



La fonction de densité

2.5. Probabilité d'un intervalle

Graphiquement, elle correspond à la surface comprise entre a et b sur le graphe de la fonction de densité (Hurlin and Mignon 2022).



Probabilité d'un intervalle [a,b]

Analytiquement, il s'agit de :

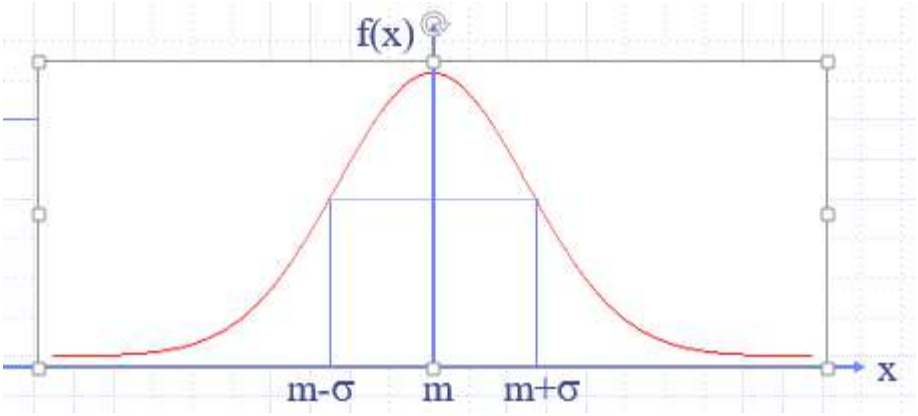
$$P(a < X < b) = \int f(x)dx_{ba}$$

Donc:

$$(a < X < b) = (X < b) - P(X < a) = \int f(x)dx_{b-\infty} - \int f(x)dx_{a-\infty} = F(b) - F(a)$$

2.6. Loi de Laplace-Gauss ou loi normale

On parle de loi normale ou de loi de LAPLACE – GAUSS, lorsque l'on a affaire à une variable aléatoire continue dépendant d'un grand nombre de causes indépendantes, dont les effets s'additionnent et dont aucune n'est prépondérante.



***Définition :**

Une V.A continue X est dite distribuée selon une loi normale si sa densité de probabilité est :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]$$

La loi normale dépend de deux paramètres m et σ. On note : $X \longrightarrow N(m;\sigma)$.

2.6.1.Fonction de répartition

La fonction de répartition d'une variable normale est donnée par l'expression :

$$\prod(x) = p(X \leq x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]dx$$

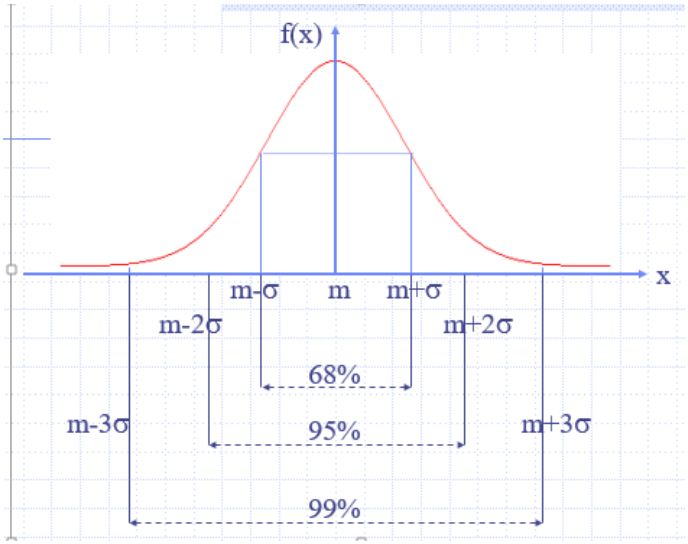
Caractéristiques :

$E(X) = m$

$V(X) = \sigma^2$

***Propriétés**

- Le graphique de la fonction de densité de probabilité de la Loi normale est une courbe en cloche symétrique par rapport au point d'abscisse x=*m*.
- La droite verticale x=m divise l'aire comprise entre la courbe et l'axe des abscisses en deux parties égales P(X<*m*) = 0,5 et P(X>*m*) = 0,5
- La grande partie des observations se situe dans l'intervalle [m-3σ ; m+3σ]



2.6.2.Intervalles remarquables

P[m-2/3 σ < X <m-2/3 σ] ≅ 50% ;

P[m - σ < X <m + σ] ≅ 68%

P[m - 2σ < X < m + 2σ] ≅ 95%;

P[m - 3σ < X < m + 3σ] ≅ 99,74%

2.6.3. Calcul des probabilités

Pour une VA continue, on s'intéresse surtout à une probabilité d'intervalle. La fonction de densité étant compliquée, des tables ont été prévues pour faciliter ce calcul.
Toutefois, étant donnée qu'il existe une infinité de lois normales distinctes par leurs paramètres, une seule variable normale est tabulée et sert de référence pour les autres : **il s'agit de la loi normale centrée réduite.**

***** Le passage de la loi normale à la loi normale centrée réduite s'effectue à l'aide du changement de variable suivant :

$$z = \frac{X - m}{\sigma}$$

La loi normale centrée réduite à pour paramètre : **m =0 et σ= 1**

Propriétés :

- Le graphique de la fonction de densité de probabilité de la LNCR est une courbe en cloche symétrique par rapport au point d'abscisse z= 0
- La droite verticale z= 0 divise l'aire comprise entre la courbe et l'axe des abscisses en deux parties égales P(Z<0) = 0,5 et P(Z>0) = 0,5.
- La grande partie des observations se situe dans l'intervalle -3 ;3.

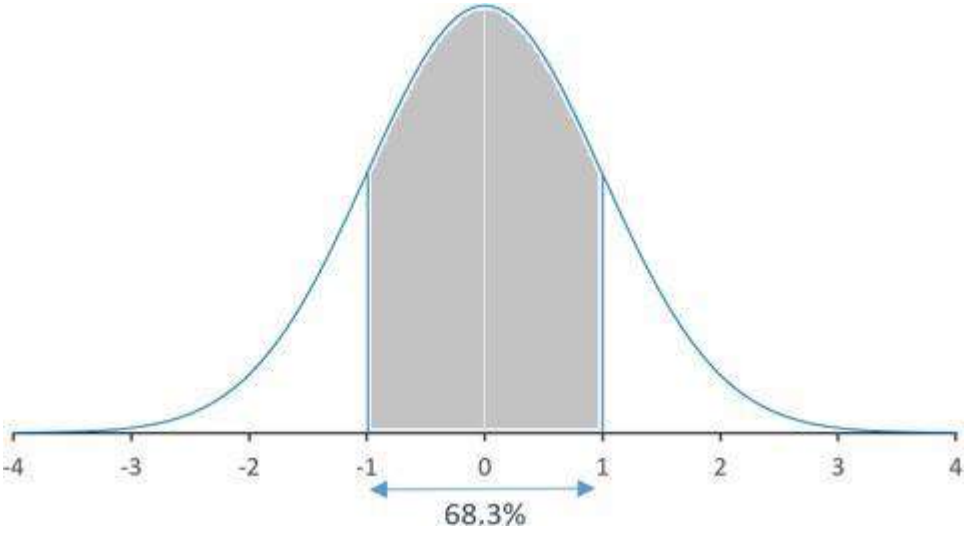
***Intervalles remarquables**

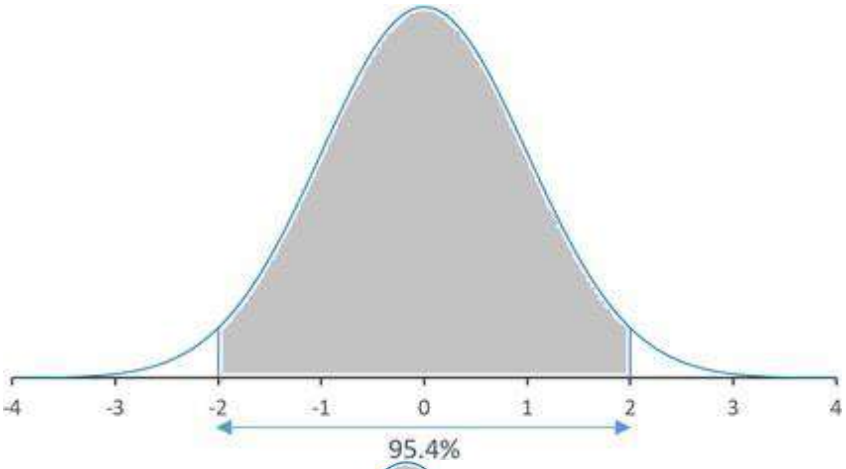
P[-2/3 <Z< 2/3] ≅ 50% ;

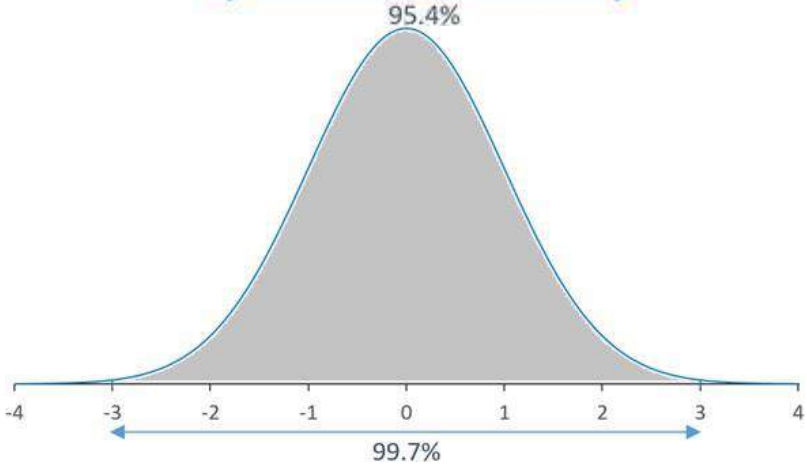
P[- 1 < Z < +1] ≅ *68%*

P[- 2 < Z < +2] ≅ 95%;

P[- 3 < Z < +3] ≅ 99,74%







2.6.4.Courbe de densité de la loi *N* (0*;* 1) :

La courbe de la densité de la loi normale N (0; 1) porte le nom de « courbe en cloche », qui est symétrique par rapport à l'axe de coordonnées. Elle admet donc un maximum au 0.

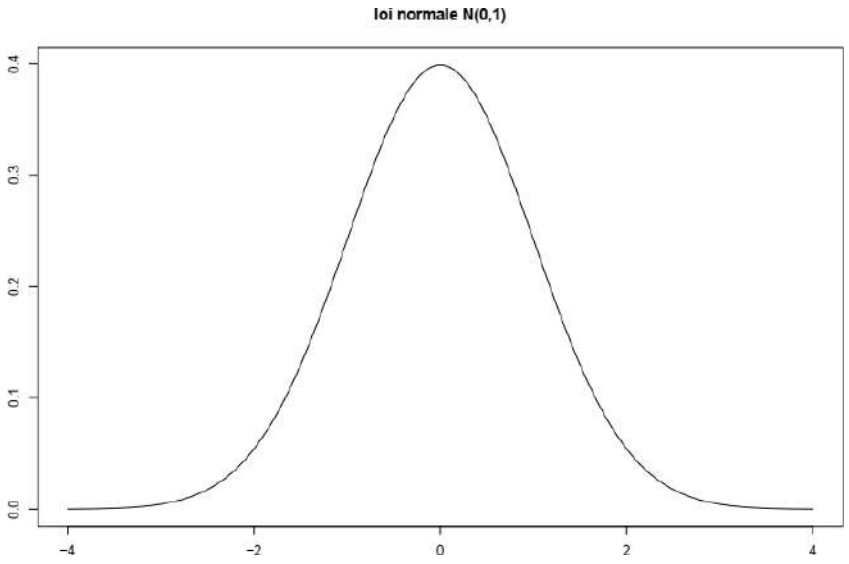


Figure 9. Courbe de densité de la loi N(0, 1)

La courbe de densité permet de calculer les probabilités. On peut noter :

*La probabilité $(a < X) = \int_{-\infty}(x)$ c'est l'air allant de l'infinie jusqu'as la valeur de à

- La probabilité $(-x < X) = (X > +x)$;
- La probabilité $(a > X) = 1$ - $(a < X)$

2.6.5. Utilisation de la table N(O; 1)

Cette table nous donne les probabilités de trouver une valeur inférieur à z

La Table comprenne deux zones. On retrouve la variable z qui est la donnée d'entrée et la probabilité. L'unité et le premier chiffre après la virgule sont dans la première colonne. Le second chiffre se trouve sur la première ligne.

EXEMPLE :

X suit une loi normale N(345; 167)
On souhaite connaître la probabilité pour que X soit inférieur à 500.

SOLUTION :

On effectue le changement de variable:

$$Z = \frac{X - \bar{x}}{\sigma} = \frac{X - 345}{167}$$

On cherche p(X < 500)=

$$p(X < 500) = p\left(Z \leq \frac{500-345}{167}\right) = p(Z \leq 0.93) = \pi(0.93) = 0.8238$$

L'erreur relative pour la loi normale est donnée par la formule :

$$\boldsymbol{P = \frac{tc * Cv}{\sqrt{N}}}$$

tc : est l'inverse de la probabilité de loi normale (ou coefficient de probabilité) ;

Cv : coefficient de variation ; $\boldsymbol{Cv = \frac{\sigma}{\bar{x}}}$

N : nombre d'individu

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} \, dx$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Chapitre III : Les méthodes d'interpolation spatiales

III.1. Introduction

L'interpolation est le procédé qui vise à cartographier une variable Z à des positions dans l'espace où aucun échantillon n'est disponible en utilisant un ensemble de données d'échantillons dont la position dans l'espace et la valeur de la variable Z sont connues (Fig. 11).

La plupart des techniques d'interpolation sont locales et déterminatives à l'exception du krigeage (qui sera étudier dans la partie géostatistique) qui est de nature stochastique. Ce dernier est généralement considéré comme un interpolateur exact (Arnaud and Emery 2000).

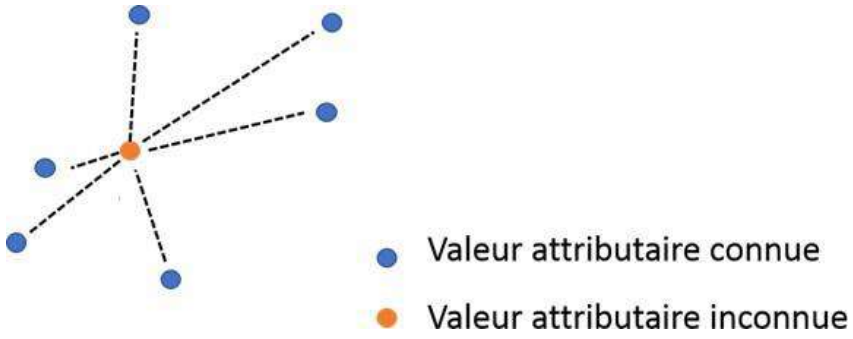


Figure 11. Principe des méthodes d'interpolation

Les méthodes d'interpolation spatiales peuvent être classées en deux principales catégories : les approches déterministes et géostatistiques ou stochastique. Ce sont des techniques utilisées pour estimer des valeurs inconnues à des endroits non échantillonnés, en se basant sur des points de données géoréférencées. Elles permettent de créer des surfaces continues ou des cartes de répartition d'une variable à partir d'un ensemble de points de données dispersés (Bosser 2011).

III.2. Méthodes d'interpolation barycentriques

Les méthodes d'interpolation barycentriques sont une classe de techniques d'interpolation spatiale qui calculent les valeurs d'un point inconnu comme une moyenne pondérée des valeurs aux points de données voisins. Elles sont basées sur le principe que la valeur interpolée en un point doit être une combinaison convexe (moyenne pondérée) des valeurs aux points de données environnants. Les poids sont déterminés en fonction de la géométrie des points de données autour du point cible (Bosser 2011).

III.3. La méthode du plus proche voisin

La méthode d'interpolation du plus proche voisin (Nearest Neighbor Interpolation) est une technique d'interpolation spatiale très simple. Prenant le cas de 2D. Si on divise notre espace en un grille et on attribue à chaque point de la grille de sortie la valeur du point de donnée le plus proche. En d'autres termes, pour chaque emplacement où on veut estimer une valeur, on cherche le point de donnée le plus proche dans l'espace et on lui assigne directement cette valeur (Despagne 2006).

Cette méthode est très rapide à calculer mais produit une surface en "marches d'escalier" avec des changements brusques entre les cellules voisines, correspondant à un manque de continuité. Elle convient lorsqu'on veut conserver les valeurs originales sans lissage, mais tend à créer un résultat découpé.

Les avantages sont sa simplicité de mise en œuvre et sa préservation exacte des valeurs d'entrée. Cependant, le résultat manque de lissage et n'est généralement utilisé que pour des données qualitatives.

Exemple :

Exemple de points de données : soit la une grille sur laquelle des données sont déterminées ou mesurées (indiquées en gras). Les valeurs estimées, qui sont indiquées en rouge, sont estimées à partir des valeurs mesurées les plus proches :

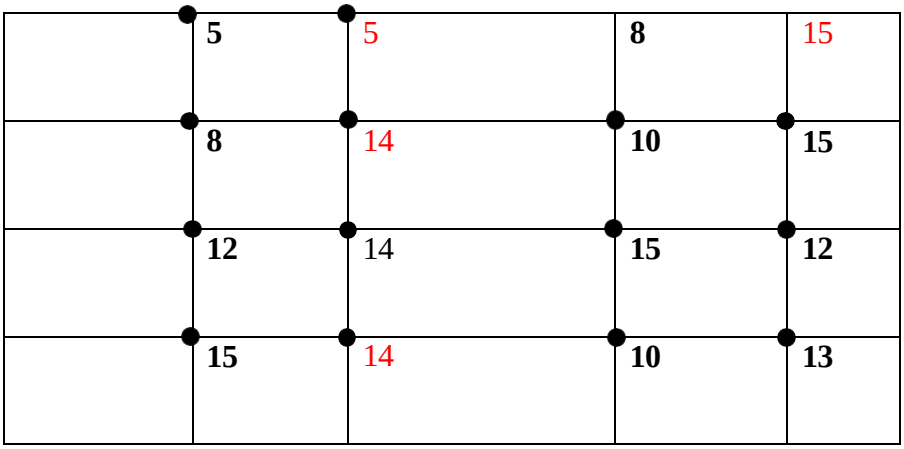


Figure 12. La méthode du plus proche voisin : en gras et noir les valeurs mesurées, et en rouge les valeurs estimées

III.4. Méthode de l'inverse des distances

La méthode d'interpolation par inverse des distances (Inverse Distance Weighting - IDW) est une technique d'estimation spatiale qui permet de calculer des valeurs inconnues à partir d'un ensemble de points de données géoréférencées dispersés (Mitas and Mitasova 1999).

Cette technique a pour but l'estimation de la teneur à un point donné (X_0) à partir des teneurs des autres points environnante, en tenant compte des distances séparant le point à estimer des autres points (Bosser 2011).

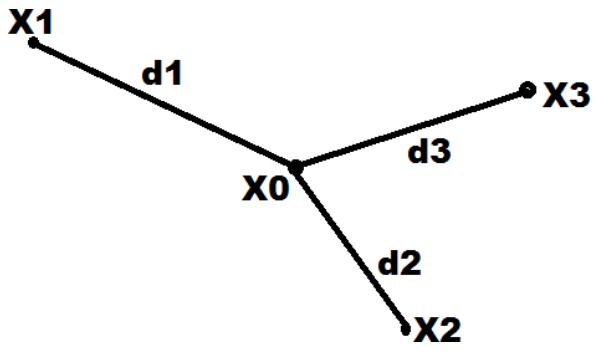


Figure 13. Principe de la méthode des inverses aux distances

La formule d'estimation est :

$$t = \frac{\sum_i \frac{t_i}{d_i^n}}{\frac{1}{\sum_i \frac{1}{d_i^n}}}$$

Où *ti* : sont les valeurs mesurées de la variable étudiée ;

t : la valeur estimé au point X0

Le principe de base est d'attribuer des poids plus importants aux points de données les plus proches de la position à estimer, et des poids plus faibles aux points plus éloignés. Les poids sont une fonction inverse de la distance. Les points plus proches ont un poids plus grand dans le calcul de la moyenne pondérée.

III.5. La méthode d'interpolation par triangulation

Cette technique est utilisée beaucoup plus pour des données géochimiques de surface. Dans un plan, on trace entre chacune de 3 échantillons un triangle chaque échantillon représente un sommet du triangle. La méthode la plus utilisée consiste à tracer des triangles les plus équilatéraux possibles (triangulation de Delaunay).

L'avantage de cette méthode est qu'elle attribue exactement les valeurs aux points de données. La surface interpolée est continue, formée de facettes triangulaires planes. Cependant, elle peut générer des artéfacts en "crêtes de toit" là où les triangles se rejoignent. En plus l'interpolation est limitée seulement au champ convexe du domaine des données qui est couvert par les triangles (Despagne 2006).

.

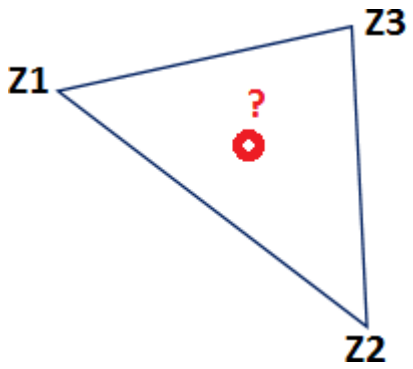


Figure 14. La méthode d'interpolation par triangulation

Il existe plusieurs méthodes pour l'interpolation de données à partir d'une triangulation.

A. Interpolation linéaire

On considère le triangle (x1, x2, x3) de surface S, contenant le point à estimer X de la variable régionalisée Z. donc les valeurs Z1, Z2 et Z3 sont attribuées aux point x1, x2, x3, successivement.
On divisant le triangle en 3 sous-triangles à partir du point X, ce qui définit donc trois surfaces (triangles) : la surface S1 (X1, X, X3), la surface S2 (X1, X, X2), et la surface S3 (X2, X3, X).

La valeur Z au point X :

$$Z(x) = (Z1 * S3 + Z2 * S1 + Z3 * S2)/S$$

Où chaque valeur est multipliée par la surface qui l'oppose.

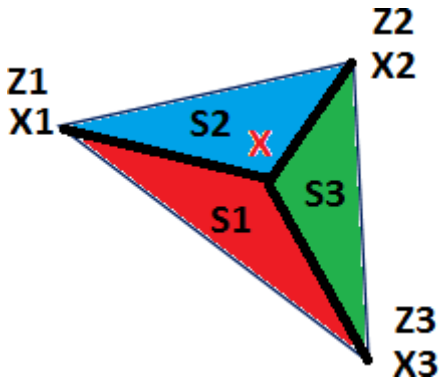


Figure 15. La méthode d'interpolation linéaire par triangulation

B. Interpolation par moyenne arithmétique

La teneur estimée pour le triangle est la teneur moyenne des trois sommets.

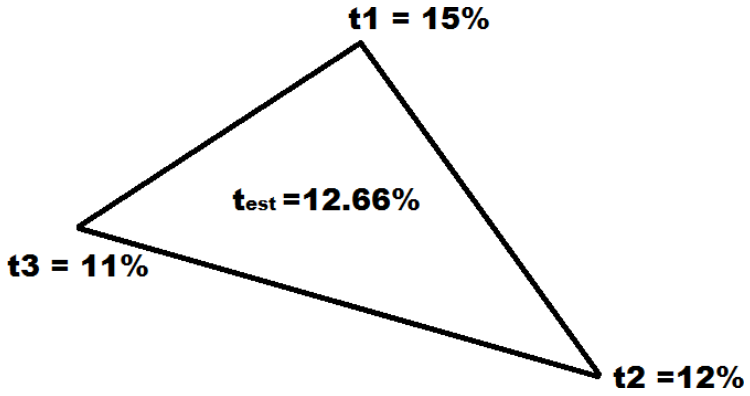


Figure 16. Principe de la méthode d'interpolation de triangulation par moyenne

Chapitre IV : Géostatistique linéaire

IV.1. Les Variables régionalisées

On appelle variable le caractère sur lequel porte une étude d'un ensemble d'individus et qui change de l'un a l'autre. Si le changement de ce caractère est imprévisible, la variable est dite *variable aléatoire*. Si cette variable aléatoire est répartie dans l'espace, dite *variable régionalisée*.

L'ensemble des variables aléatoires (teneurs mesurées sur des échantillons géologiques ou dans des sondages) implantées aux points Xi de coordonnées X_{1i}, X_{2i}, X_{3i} et notées $z(x_i)$ forme la fonction aléatoire Z(X). Mais les teneurs mesurées ne sont pas forcément les teneurs vraies. La teneur $z(x_i)$ mesurée en xi est une réalisation particulière de la variable aléatoire Z(x) et l'ensemble des teneurs mesurées en différents points est interprété comme une réalisation particulière de la fonction aléatoire Z(X).

IV.1.1. DEFINITION DES MOMENTS

En géostatistique appliquée d'estimation, on s'intéresse essentiellement aux deux premiers moments de la variable régionalisée Z (x) (Journel, 1978).

- moment d'ordre 1 - $E[Z(x)] = m(x)$ qui est l'Espérance mathématique

- moment d'ordre 2 - $2\gamma(x,h) = E\{[z(x) - z(x+h)]^2\}$ appelé variogramme

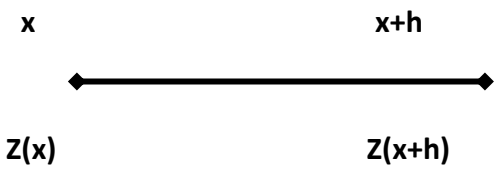
ou

$$C(x,h) = E\{[Z(x+h) - m(x+h)] * [Z(x) - m(x)]\}$$ appelé covariance

A partir de la covariance, on définit le corrélogramme qui exprime les variations de corrélations spatiales entre les valeurs (teneurs) mesurées au point x et celles observées au point (x+ h). Il est généralement noté ρ et est égale aux valeurs de la covariance au point x+h divisées par celle de la covariance au point x (h=0.)

$$\rho^h = \frac{C(h)}{C(0)}$$

Les trois moments quantifient chacun l'autocorrélation entre les valeurs Z(x) au point x et Z(x+h) au point x+h.



IV.1. 2. LA STATIONNARITE

La F.A. Z(x) est dite stationnaire d'ordre 2 si ses deux premiers moments sont invariants par translation sur l'espace de définition et par conséquent :

$$E[Z(x)] = m$$

$$E[Z(x+h) - Z(x)]^2 = 2\gamma(h) = 2\gamma(-h)$$

$$E[Z(x+h).Z(x)] - m^2 = C(h) = C(-h)$$

- Le variogramme est toujours positif mais la covariance peut présenter des valeurs négatives ;

- La relation entre le variogramme et la covariance est donnée par cette formule :

$$\gamma(h) = C(0) - C(h)$$

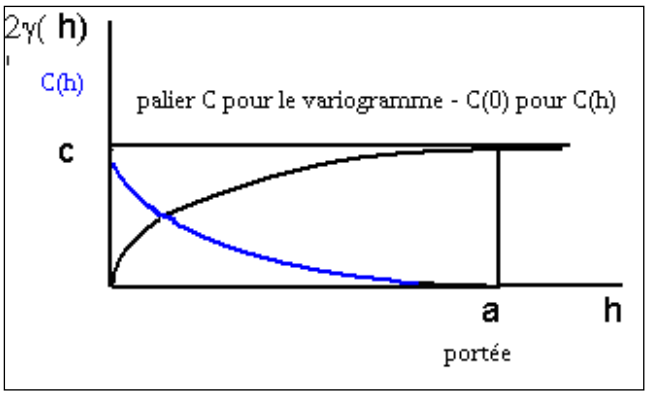
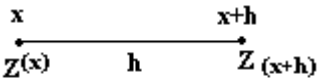


Fig. 2 - Schéma représentant la relation entre variogramme et covariance

« L'hypothèse d'existence du variogramme étant moins forte; en géostatistique appliquée au domaine des sciences de la terre et du génie civil, on préfère l'outil variogramme à la covariance »(Journel, 1978).

IV.1.3. Le variogramme théorique

Considérons deux valeurs numériques, Z(x) et Z(x+h), implantées en deux points distants du vecteur h,



on caractérise la variabilité entre ces deux mesures, par la fonction variogramme : $2\gamma(x,h)$, définie comme l'espérance de la variable aléatoire $[Z(x)-Z(x+h)]^2$

$$2\gamma(x,h) = E\{ [Z(x) - Z(x+h)]^2 \}$$; donc

Le variogramme est une fonction du vecteur h; il indique si les valeurs différent beaucoup au fur et à mesure que la distance augmente, il montre les particularités directionnelles du phénomène (si l'on examine dans différentes directions).

Le graphe de $\gamma(x,h)$ en fonction de h a les caractéristiques suivantes :

1- Il passe par l'origine (pour h=0 ; Z(x +h) = Z(x)) ;

2- C'est en général une fonction croissante de h ;

3- Dans la plupart des cas, il croît jusqu'à une certaine limite appelée **palier**, puis s'aplatit (fig.3)

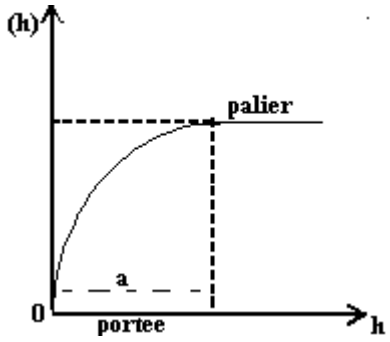


Fig. 3 - les caractéristiques du graphe $\gamma(x,h)$ en fonction de h

IV.3. - Portée et zone d'influence

Lorsque le variogramme a atteint sa limite supérieure c'est à dire son palier, il n'y a plus de corrélation entre les échantillons séparés par cette distance h : cette distance critique est appelée **portée** du variogramme (fig. 3), qui fournit une définition plus précise de la notion de ***zone d'influence.***

IV.1.4 - ESTIMATION DU VARIOGRAMME

Afin de pouvoir utiliser le variogramme dans la pratique, il est nécessaire de pouvoir l'estimer.

Considérons un champ S où la variable régionalisée (ex: teneur en CaO) est stationnaire. On peut alors considérer que le variogramme γ(x,h) ne dépend que du vecteur h (module et direction). Cette hypothèse rejoint en partie l'hypothèse de stationnarité et est appelée **Hypothèse intrinsèque**.

En pratique, on ne dispose que d'une seule réalisation [Z(x+h)-Z(x)] mais ces hypothèses permettent d'avoir plusieurs couples et l'on peut calculer le variogramme expérimental.

Un estimateur de 2 γ(h) c'est la moyenne arithmétique des différences aux carrées entre 2 mesures expérimentales implantées en 2 points distants de h.

$$2\gamma(h) = {}^{1}\!/_{N(h)} \cdot \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2$$

N(h) tant le nombre de couples expérimentaux [z(x)-z(x+h)]

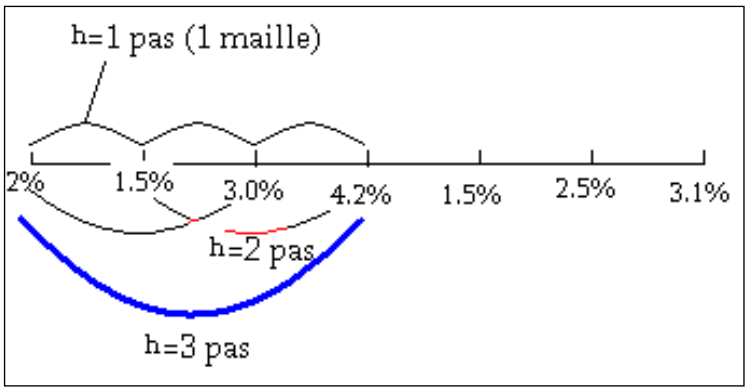


Fig. 4 - Schéma de calcul du variogramme expérimental

Les résultats sont aussi représentés sous forme graphique :

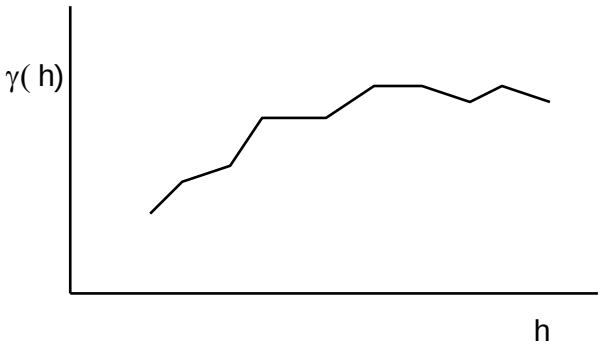


Fig. 5– Représentation graphique d'un variogramme

IV.2. INTERPRETATION DU VARIOGRAMME

Le variogramme caractérise la structure de la variabilité spatiale des variables régionalisées. Il quantifie la structure d'un phénomène «géologique – géotechnique ou agronomique» qui peut être utilisée par la suite pour l'évaluation des ressources par exemple. Il permet de distinguer entre différents types de parcelle, champ ou site - Par exemple, la variabilité des teneurs en sels dans différents sites.

Le variogramme, en général, croît avec le module du vecteur h. Le type de croissance du variogramme à l'origine caractérise la continuité de la variable étudiée. Au delà d'une certaine valeur de h appelée portée, le variogramme se stabilise c.a.d au delà de cette distance, les valeurs ne sont plus corrélées. Cette portée représente la zone d'influence d'un échantillon ou d'un sondage.

IV.2.1.REGLE PRATIQUE POUR LE CALCUL DU VARIOGRAMME EXPERIMENTAL

En règle pratique, afin que le variogramme expérimental soit un bon estimateur du variogramme local il faut que : N >30 couples et h < L/2 - moitié du champ.

IV.2.1.1. Variogramme à 1 Dimension -

Dans la pratique, il arrive souvent que l'on ait à caractériser la variabilité d'une variable dans une seule direction comme par exemple la variabilité des teneurs en sel, taux d'une fraction granulométrique,…dans

des sondages (puits) - les données sont dites jointives. Il arrive aussi que l'on ait à construire des variogrammes dans les directions horizontales à partir des données de sondages par exemple, les données dans ce cas ne sont pas jointives.

IV.2.1.2. Calcul du variogramme moyen à 1 dimension - 1D

Le variogramme à 1D ne peut être significatif que si le nombre de données est assez grand (25 et plus). Il est donc nécessaire de regrouper les variogrammes élémentaires à 1D en un seul variogramme moyen représentant la variabilité dans cette même direction. Cependant il faudra veiller à ne regrouper que les données homogènes et ayant même support.

Le variogramme moyen est estimé à partir de tous les couples de données distants de h. Il faut donc pondérer chaque variogramme élémentaire par le nombre de couples correspondant :

Soient 2 variogrammes élémentaires expérimentaux calculés pour un même h dans deux sondages différents par exemple:

$$2\gamma_1(h) = {}^{1}\!/_{N1}\sum_{i=1}^{N1}\left[Z(x)-Z(x+h)\right]^2 et 2\gamma_2(h) = {}^{1}\!/_{N2}\sum_{i=1}^{N2}\left[Z(x)-Z(x+h)\right]^2 et$$

Le variogramme moyen sera :

$$\gamma_{moy}(h) = \frac{N1.\gamma_1(h)+N2.\gamma_2(h)}{N1+N2} et..en..général : ...\gamma_{moy}(h) = \frac{\sum_{i=1}^{N} N_i.\gamma_i(h))}{\sum_{i=1}^{N} N_i}$$

IV.2.2. Variogramme à 2 Dimensions

Quand les données sont réparties suivant deux ou plusieurs directions à 2D, il est souvent nécessaire de calculer le variogramme moyen dans toutes ces directions. Si la structure de la variabilité est la même dans les différentes directions, les variogrammes expérimentaux de ces directions présenteront les mêmes allures (à peut prés même palier et même portée). On dira que le phénomène est **isotrope**, sinon le phénomène est **anisotrope**.

IV.2.2.1. Cas isotrope

Dans le cas où la variabilté est isotrope, le variogramme moyen à 2D est calculé en faisant la somme des variogrammes élémentaires pondérés par le nombre de couples correspondants (comme pour le variogramme moyen à 1D).

IV.2.2.2. Cas d'anisotropes

On distingue 2 types d'anisotropies : anisotropie géométrique et anisotropie zonale.

A.Anisotropie géométrique

Il y a anisotropie géométrique quand les variogrammes présentent la même variabilité globale et en particulier le palier mais ont des portées différentes.

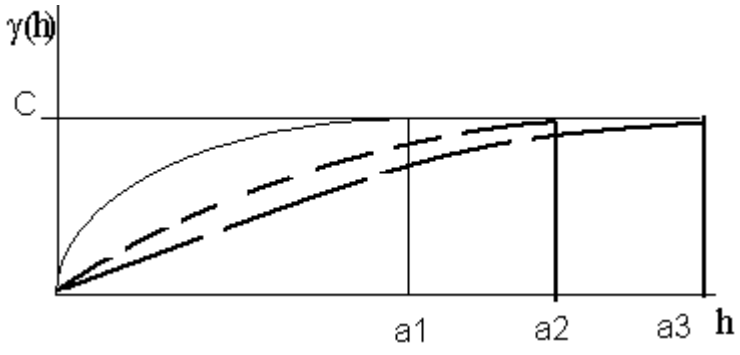


Fig. 10 - Schéma d'une anisotropie géométrique

L'étude de l'anisotropie est facilitée par l'établissement de roses de portées ou des inverses des pentes à l'origine.

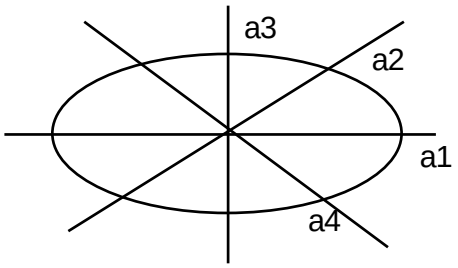


Fig. 11 - Rose des portées d'une anisotropie géométrique

Dans la rose des portée, la direction 3 (a3) - direction d'aplatissement de l'ellipse - est la direction de rapide variabilité du phénomène étudié. Dans ce cas la maille de reconnaissace la mieux adaptée est une maille rectangulaire dont les directions (profils) sont les directions principales de l'ellipse.

B.Anisotropie zonale

L'anisotropie zonale, cas le plus fréquent en pratique, affecte l'ensemble du variogramme - les portées et les paliers sont différents.

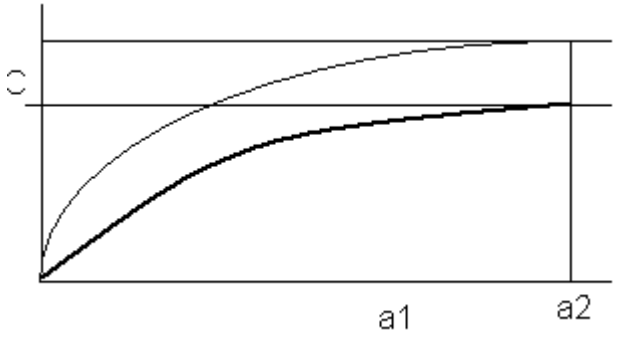


Fig. 12 - Schéma de variogrammes représentants une anisotropie zonale

IV.2.3. Variogramme de surface (multi et bivarié)

Le variogramme de surface permet l'identification d'un comportement anisotropique de la variable étudiée.

Les valeurs des variogrammes sont représentées dans les directions h_x et h_y.

La représentation de surface nécessite la segmentation de l'espace dans chacune des composantes h_x et h_y. en un nombre d'intervalles donnés. Ce ci abouti à une discrétisation de la surface en un ensemble de « maille » ou « panneau » de couleurs différentes et qui est fonction de la valeur du variogramme expérimental obtenu dans la direction Centre de la surface (de coordonnée relative 0.0) ->vers le centre du dit panneau. Le nombre de couple est inscrit à l'intérieur du panneau (fig.). On en déduit que la surface résultante est symétrique.

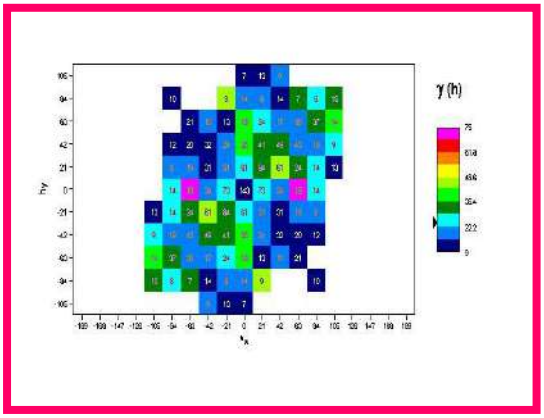


Fig. 13 – Graphe d'un variogramme de surface

IV.2.4.Effet de pépite pur

On dit qu'il y a effet de pépite pur lorsque le variogramme observé ne traduit que la seule constante de pépite (variogramme plat). $\gamma(h)=C0$ dès que $h>0$. Il y a alors indépendance spatiale et la géostatistique retrouve tous les résultats de la statistique des variables indépendantes.

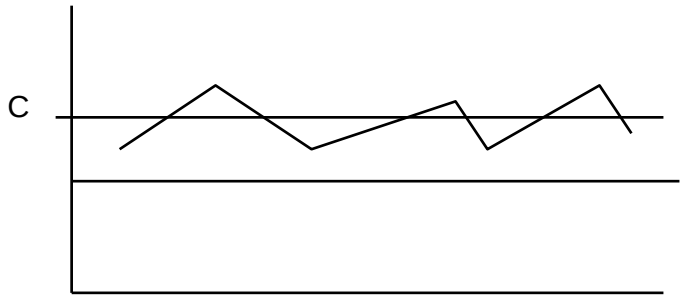


Fig. 15 - Schéma d'un variogramme représentant un effet de pépite pur

IV.4. SCHEMAS THEORIQUES ET AJUSTEMENT DES VARIOGRAMMES

Les variogrammes expérimentaux sont synthétisés dans un modèle théorique qui doit rendre compte des principales caractéristiques structurales de la régionalisation étudiée. Il doit être opérationnel et simple à l'emploi.

Les deux principales caractèristiques d'un variogramme stationnaire sont l'existence ou non d'un palier et le comportement à l'origine. L'élaboration d'un modèle synthétique se fait à l'aide de schémas théoriques de régionalisation. «Les modèles théoriques sont des expressions analytiques ».

Les schémas théoriques d'usage courant sont classés en Schémas à palier; schémas sans palier et Schémas à effet de trou.

IV.4.1. SCHEMAS A PALIER

Ce sont des variogrammes présentant un palier C. Le comportement des variogrammes à l'origine est soit linèaire soit parabolique.

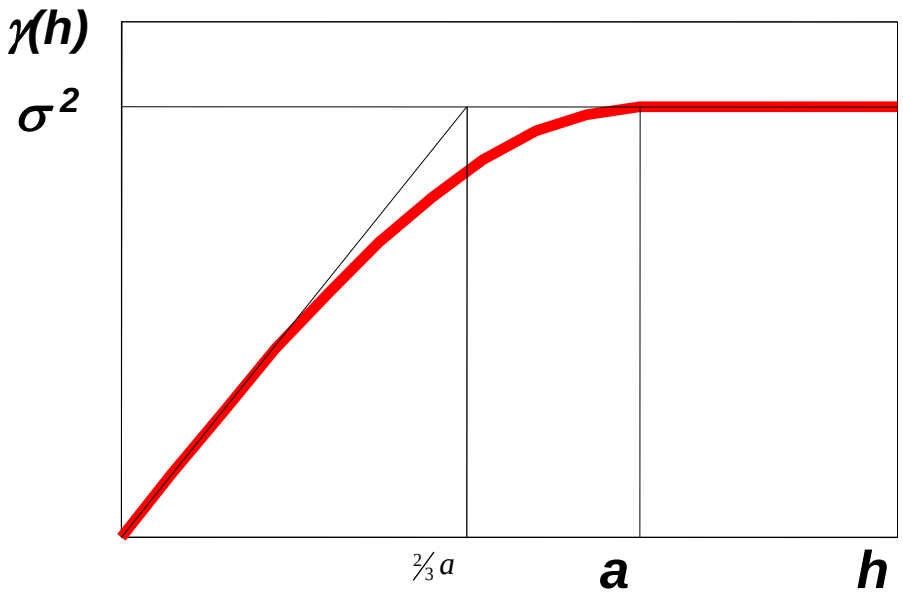
IV.4.1.1. Comportement linèaire à l'origine

On distingue principalement les schémas sphériques et les schémas exponentiels (Figs. 16 et 17).

- Schéma sphérique

Son expression mathématique est :

$$\gamma(h) = \frac{3}{2}.\frac{h}{a} - \frac{1}{2}.\frac{h^3}{a^3}; \forall h \in [0; a]$$
$$\gamma(h) = 1, pour, , h \geq a$$



a)

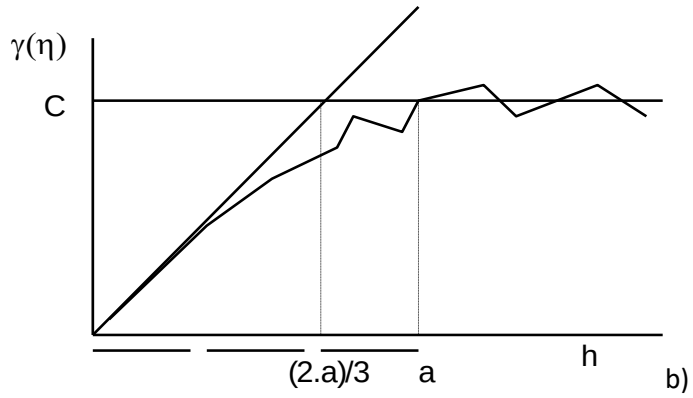
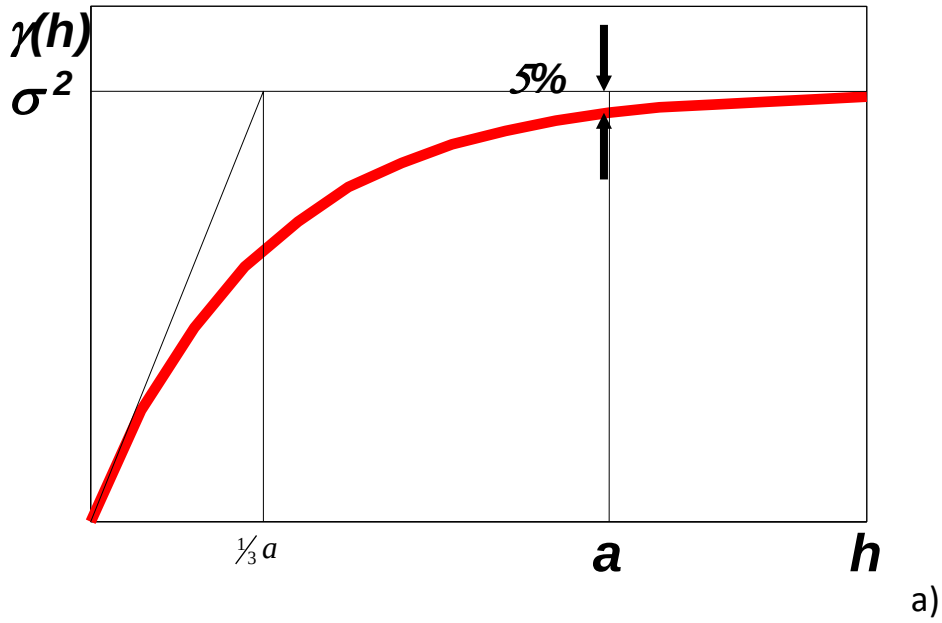


Fig. 16 - a) graphe d'un schéma théorique sphérique

b)graphe d'un variogramme expérimentale ajustable à l'aide d'un d'un schéma sphérique

B- Schéma exponentiel

$$\gamma(h) = 1 - e^{\frac{-h}{a}}, \forall h > 0$$



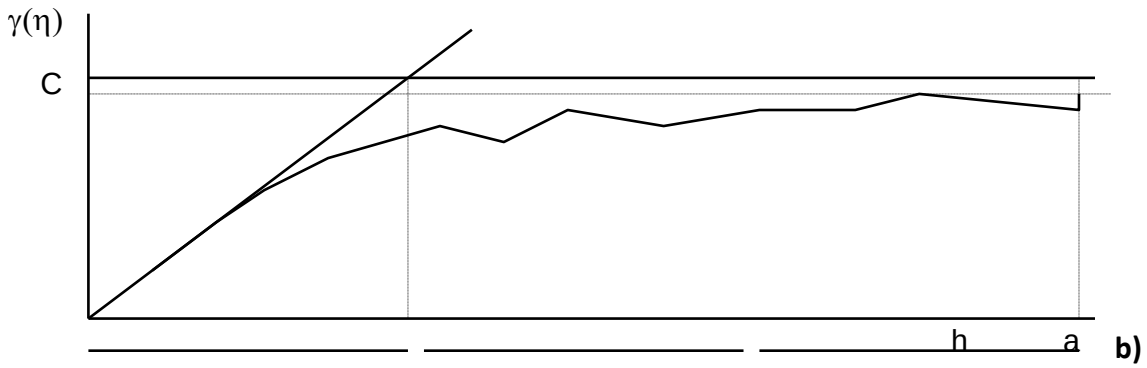


Fig. 17 - a) graphe d'un schéma théorique exponentiel

b) graphe d'un variogramme expérimentale ajustable à l'aide d'un schéma exponentiel

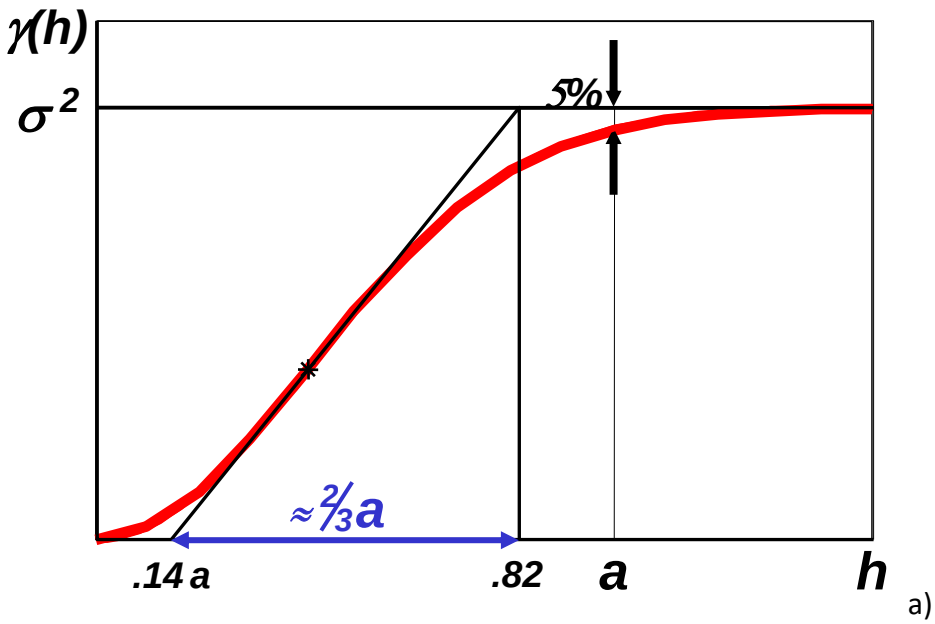
La différence entre schéma sphérique et schéma exponentiel réside dans les abscisses des intersections de leures tangentes à l'origine avec le palier :

- au deux tiers de la portée a pour le sphérique
- au un tiers de la portée pratique a' pour l'exponentiel.

IV.4.1 .2. Comportement parabolique à l'origine

En pratique le plus utilisé c'est le schéma gaussien :

$$\gamma(h) = 1 - e^{\frac{-h^2}{a^2}}, \forall h > 0$$



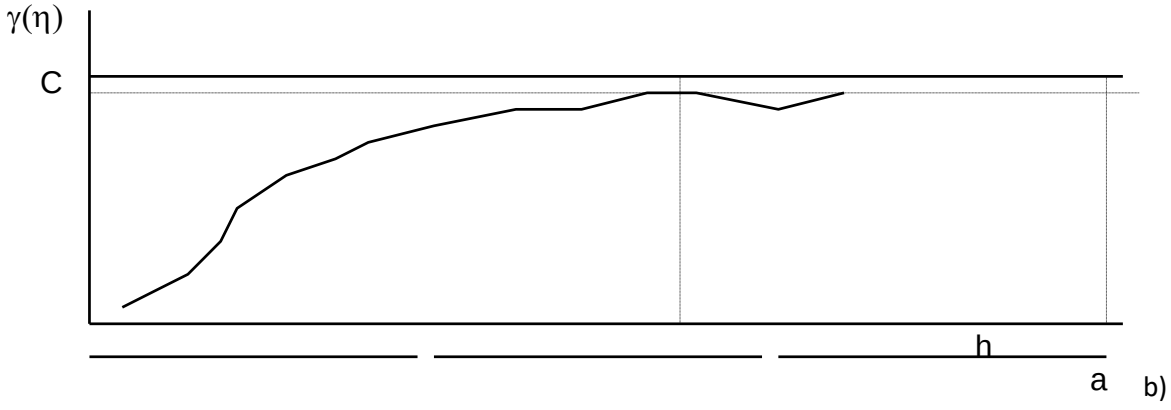


Fig. 18 - a) graphe d'un schéma théorique gaussien

b) graphe d'un variogramme expérimentale ajustable à l'aide d'un schéma gaussien

IV.4. 2. - SCHEMAS SANS PALIER

Ce sont des variogrammes théoriques qui correspondent à des variogrammes expérimentaux dont la croissance ne présente pas de palier dans les limites h < b où b est la limite de l'observation

. $\gamma(h)$ tend vers + ∞ quand h tend vers + ∞

Deux types de schémas sont assez souvent utilisés.

- les schémas en h^{λ} avec $0<\lambda<2$

IV.4.2.**1. Schémas en h^{λ}**

$\gamma(h) = h^{\lambda} \quad \forall h > 0$ avec $0 < \lambda < 2$

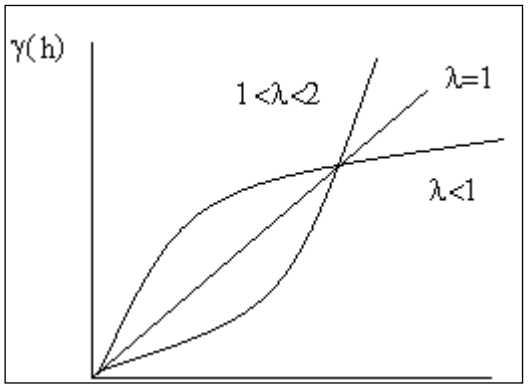


Fig. 19 - Graphe des schémas en h^{λ}

. IV.4.2.**2. SCHEMA A EFFET DE TROU**

On dit qu'un variogramme $\gamma(h)$ présente un effet de trou si sa croissance n'est pas monotone. Les schémas à effet de trou présentent une allure sinusoïdale au niveau du palier.

$$\gamma(h) = 1 - \frac{\sin(h)}{h}, \forall h > 0$$

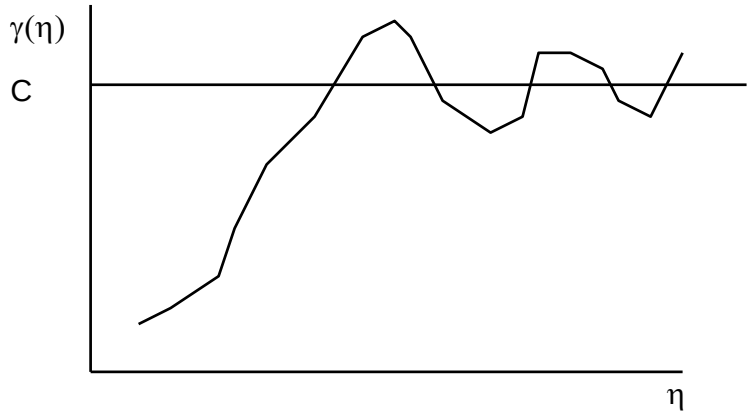


Fig. 20 - Graphe d'un schéma à effet de trou

Le schéma à effet de trou présente un comportement parabolique à l'origine :

$$\gamma(h) \cong \frac{h^2}{6}$$ quand h tend vers 0.

L'effet de trou reflète une pseudo-periodicité de la variable regionalisée. Ainsi la succession stationnaire dans un gisement de 2 types de sols bien differenciés provoque un effet de trou sur le variogramme experimental. Il peut être provoqué par l'hétérogeneité de l'information (2 campagnes d'échantillonnages par exemple).

IV.4.3. AJUSTEMENT D'UN VARIOGRAMME EXPERIMENTAL

Le variogramme représentant une structure gigogne est ajusté à l'aide d'une somme de deux ou plusieurs schémas théorique.

Dans la pratique, il existe plusieurs Methodes d'ajustement, cependant l'justement « à la main » est la méthode la plus simple et la plus juste.

Il faudra tenir compte de :

- L'existance ou non du palier ;
- De l'effet de pépite et du palier expérimental ;
- Du comportment à l'origine et de la tangente à l'origine dans le cas de comportement linéaire pour la proposition du schéma théorique.

Exemple : Si on veut ajuster un variogramme expérimental par un schéma sphérique, il faudra, à l'aide du graphe du variogramme expérimental tracé sur l'écran de l'ordinateur, choisir :

- un palier C

- un effet de pépite Co

- une portée a

- le choix d'un modèle en fonction du comportement à l'origine et de la tangente à l'origine

et le variogramme d'ajustement sera :

$\gamma(h) = Co + C.\gamma sphérique$ (pour le modèle choisi)

CHAPITRE V – VARIANCE D'ESTIMATION

V. 1.Définition d'estimation

Cela consiste à se servir des données d'un échantillon statistique pour attribuer certaines valeurs aux paramètres inconnus de la population. Cependant on peut se proposer d'attribuer une valeur unique aux paramètres inconnus et l'on aura alors une estimation dite ponctuelle comme on peut se proposer de d,terminer un intervalle de confiance dans lequel les paramètres se situeront et l'on aura alors l'estimation dite par intervalle. Dans ce dernier cas il sera encore opportun d'exprimer ou de chiffrer la crédibilité attachée à cet intervalle. Cette crédibilité est appelé niveau de confiance. Ces paramètres peuvent être estimés à l'aide de plusieurs méthodes qui ne donnent pas forcement le même résultat. Il est alors nécessaire de choisir une méthode d'estimation en fonction des qualités des estimations.

a. - QUALITE DES ESTIMATIONS

La teneur moyenne d'un bloc minier, par exemple, peut être estimée de plusieurs façons (moyenne arithmétique, krigeage ...). On peut donc obtenir plusieurs estimateurs de cette teneur moyenne. Il reste à savoir quelle est la meilleure estimation ou le meilleur estimateur.

Estimateur sans biais :

L'estimateur est dit sans biais si son espérance mathématique est égale au paramètre de la population.

$$E(x) = X$$

X étant le paramètre de la population et x l'estimateur de ce paramètre. Si on pose le biais égale à b alors :

$E(x) - X = b = 0$

Si $b \neq 0$ alors on dit que l'estimation est biaisé

Estimateur convergent :

Un estimateur est dit convergent si, étant sans biais, sa variance tend vers zéro, lorsque la taille de l'échantillon statistique n augmente indéfiniment.

Exemple : La moyenne arithmétique est un estimateur sans biais et convergent puisque E(m)- m = 0 et s (m) = S /n - donc quand n tend vers l'infinie s (m) tend vers 0

Estimateur efficace :

On dit qu'un estimateur est d'autant plus efficace que sa variance est plus petite.
Un estimateur sera donc d'autant meilleur qu'il sera sans biais, convergent et de variance aussi faible que possible.

V.1. ESTIMATION PONCTUELLE

Rappelons certaines estimations ponctuelles pour une loi de distribution normale :

V.1.1. Estimation d'une moyenne m :

$$m = \frac{\sum_{i=1}^{n} x_i}{n}$$

m étant la moyenne expérimentale, si μ est la variable aléatoire correspondante nous avons :

$E(m) = \mu$ et $\sigma_m^2 = \frac{\sigma_{pop}^2}{n}$; m est l'estimateur de μ

V.1.2. Estimation d'une variance σ^2

$$\sigma^2 = \frac{nS^2}{(n-1)}$$

V.1.3. ESTIMATION PAR INTERVALLE

L'estimation par intervalle donne un ensemble de valeurs susceptibles d'être prises par ce paramètre, avec une borne inférieure et une borne supérieure qui sont les limites de l'intervalle. Cet intervalle est appelé intervalle de confiance et on lui affecte un coefficient de crédibilité, appelé niveau de confiance. Exemple : La teneur moyenne tm d'un élément chimique dans un gisement est comprise entre 0.40 % et 0.50 % avec un niveau de confiance de 95 %.

0.40 % < tm < 0.50 %

avec un niveau de confiance (1-α) = 95 % ; α est appelé Risque d'erreur.

Ce coefficient de confiance "veut dire" que, par exemple, si l'on prélevait d'un même ouvrage minier et de la même façon un grand nombre d'échantillons (statistique) on trouverait pour chacun d'eux des teneurs moyennes différentes mais que 95% de ces valeurs moyennes seraient situées dans cet intervalle.

$$P(a < Z(x) < b) = 1-\alpha$$

En répartissant a/2 aux deux extrémités de la distribution, on calcule une valeur tm telle que :

$$P(tm1 < tm) = \alpha/2$$

et une autre valeur tm2 telle que :

$$P(tm2 > tm) = \alpha/2$$

Connaissant la loi de probabilité, on détermine tm1 et tm2 les limites de l'intervalle et l'on a l'intervalle de confiance pour tm.

$$tm1 < tm < tm2$$

V.1.4. Estimation par intervalle d'une moyenne m :

Il y deux cas à étudier séparément : le cas où l'effectif n de l'échantillon est inférieur à 30 et le 2ème cas où n est supérieur à 30.

- n < 30

Soit un échantillon statistique qui suit une loi normale N(m, s) où m est une variable aléatoire suivant aussi une loi normale N(m, s/√n).

Posons T = (μ-M)/ σ_{moy}

rappelons que $\sigma_{moy} = \sigma/\sqrt{n}$ et que $\sigma = \frac{S\sqrt{n}}{\sqrt{n-1}}$

Alors on peut écrire :

$$T = \frac{(\mu - M)}{\sqrt{S^2/(n-1)}}$$

T est, par définition, une variable de Student à n-1 d.l. que l'on note Tn-1.
L'on peut écrire :

$$P(-t_c < \frac{(\mu - M)}{\sqrt{S^2/(n-1)}} < +t_c) = 1 - \alpha$$

d'où on peut tirer : $\boxed{\mu = M \pm t_c . \frac{S}{\sqrt{n-1}}}$

où tc est pris de la table de Student pour n-1 d.l.

- n > 30

Dans le cas où n est supérieur à 30, en suivant le même raisonnement que pour n < 30, l'on abouti au résultat suivant :

$$P(-t_c < \frac{(\mu - M)}{\sqrt{S^2/n}} < +t_c) = 1 - \alpha \quad \text{et} \quad \mu = M \pm t_c . \frac{S}{\sqrt{n}}$$

où tc est pris de la table de la loi normale.

Exemple : Dans une galerie, on a prélevé 100 échantillons géologiques qui ont accusé une teneur moyenne tm en or de 50 g/t et une variance S de 400(g/t). La distribution des teneurs en or suit approximativement une loi Normale. On se pose alors les questions :
- avec un risque d'erreur de 5 %, quelle serait la teneur moyenne de tout le bloc géologique après exploitation ?

Solution :

1 - Estimation par intervalle de la moyenne

On pose

$$P(-t_c < \frac{(\mu - M)}{\sqrt{S^2/n}} < +t_c) = 1 - \alpha$$

et $$\mu = M \pm t_c . \frac{S}{\sqrt{n}}$$

$\alpha = 0.05$ alors $\alpha/2 = 0.025$

On obtient :

$1-(\alpha/2) = 0.975$ alors on lie sur la table normale $tc = 1.96$.

En remplaçant tc, S et m par leurs valeurs respectives, on obtient :

tm = 50 g/t + 3.92 g/t

46.08 g/t < tm < 53.92 g/t avec un risque d'erreur de 5 %.

Si Z(x) est la valeur inconnu que l'on cherche à estimer par la valeur mesurée ou calculée Z*(x), l'erreur commise est [Z - Z*]. Comme Z est une V.A. alors Z* et [Z - Z*] sont aussi des

réalisations particulières de V.A. L'erreur aléatoire [Z-Z*] est caractérisée par ses 2 moments :

- Moyenne , b = E{[Z - Z*]},quand cette moyenne est nulle (b =0), on dit que l'estimation est sans biais sinon l'estimation est biaisée.

- La variance de l ’erreur est appelée variance d'estimation et est égale à :

$$\sigma^2_{Est} = E\left\{\left[Z - Z^*\right]^2\right\} - b^2$$ **si b=0 alors** $$\sigma^2_{Est} = E\left\{\left[Z - Z^*\right]^2\right\}$$

V.1.5. ELABORATION D'UN ESTIMATEUR

L'estimateur Z* ne peut qu'être dépendant de l'information disponible. Si, par exemple,

L'information I est un ensemble discret de N teneurs :

$$I = \{Z(x_i), i = 1 à N\}$$

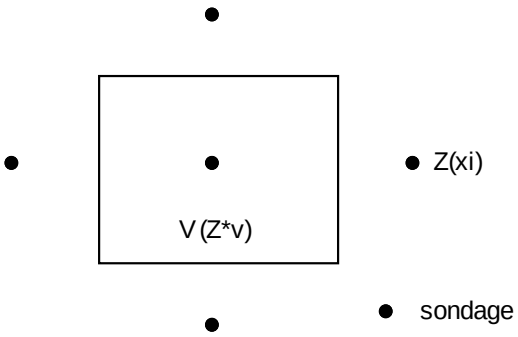


Fig. 21 - Schéma de répartition de l'information (sondages) par rapport à V - V n'a pas de dimension particulière – il peut être même assimilé en un point X : exemple :

L'estimateur Z* est fonction de ces données de sondages

$$Z^* = f\left[Z(x_1), Z(x_2), , , Z(x_n)\right] Z^* = f\left[Z(x_1), Z(x_2), , , Z(x_n)\right]$$

Cette fonction de n variables ne peut pas être quelconque :

- Elle doit vérifier le non-biais,

- Elle doit être telle que l'on puisse calculer la variance d'estimation, c'est à dire les termes du developpement suivant :

$$E\left[(Z - Z^*)^2\right] = E\left[(Z)^2\right] + E\left[Z^*)^2\right] - 2E\left[(Z.Z^*)\right]$$

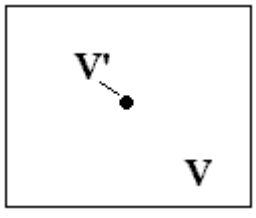
ce ci nous conduit à retenir, généralement, des estimateurs linèaires du type :

$$Z^* = \sum_{i=1}^{n} \lambda_i . Z(x_i)$$

Rappelons qu'un estimateur est dit optimal s'il minimise la variance d'estimation et s'il est sans biais.

V.2. ESTIMATION D'UNE MOYENNE PAR UNE AUTRE

Soit à estimer la moyenne Z_V sur un domaine V par la moyenne $Z_{V'}$ sur un domaine V'.



Rappelons que : $Z_V = \frac{1}{V}\int_v Z(x)dx$ **et** $z_{v'} = \frac{1}{V'}\int_{v'} z(x)dx$

Z(x) désigne la VR ponctuelle, et Z(X) la FA correspondante, stationnaire d'ordre 2

et de variogramme .

Dans les limites de cette hypothèse stationnaire, le biais est nul puisque :

$$E\{Z_v\} = E\{z_{v'}\}$$

Pour calculer la variance d'estimation, il suffit de calculer chacun des termes de l'égalité suivante

en remplaçant Z_V et $Z_{V'}$ par leurs valeurs respectives.:

$$\mathbf{E\{[Z_V - Z_{V'}]^2\} = E\{[Z_V]^2\} + E\{[Z_{V'}]^2\} - 2E\{Z_V.Z_{V'}\}}$$

- Calculons $E\{[Z_V]^2\}$:

$$E\{[\mathbf{Z_V}]^2\} = \frac{1}{V^2}\int_V dx \int_V E\{Z(x).Z(x')\}dx'$$

$$E\{Z(x).Z(x')\} = C[Z(x) \in V.Z(x') \in V] + m^2$$

or

ceci permet d'écrire :

$$E(Z_V^2) = \overline{C}(V,V) + m^2$$

Le symbole $\overline{C}(V,V)$ désignant la valeur moyenne de $C(V,V)$ lorsque les deux points d'appui x

et x' décrivent indépendamment l'un le domaine V, l'autre le même domaine V.

- Calculons $E(Z_{V'}^2)$:

or

$$E(Z_{V'}^2) = \frac{1}{V'^2}\int_{V'} dx \int_{V'} E\{Z(x).Z(x')\}dx'$$

$$E\{Z(x).Z(x')\} = C[Z(x) \in V'.Z(x') \in V'] + m^2$$

ceci permet d'écrire: $E(Z_{V'}^2) = \overline{C}(V',V') + m^2$

Le symbole désignant la valeur moyenne de $C(V',V')$ lorsque les deux points d'appui x et x' décrivent indépendamment l'un le domaine V', l'autre le même domaine V'.

- Calculons 2E(Z$_V$.Z$_{V'}$):

$$\mathrm{E(ZV.Zv')} = \frac{1}{V.V'}\int_V dx \int_{V'} E\{Z(x).Z(x')\}dx'$$

$$E\{Z(x).Z(x')\} = C[Z(x) \in V.Z(x') \in V'] + m^2$$

Or : $2E(Z_V.Z_V') = 2\overline{C}(V,V') + 2m^2$

ceci permet d'écrire:

Le symbole désignant la valeur moyenne de $\overline{C}(V,V')$ lorsque les deux points d'appui x et x' décrivent indépendamment l'un le domaine V, l'autre le domaine V'.

En remplaçant les différents termes par leur valeur alors on aura :

$$\sigma_{Est}^2 = E\left\{\left[Z_V - Z_{V'}^*\right]^2\right\} = \overline{C}(V,V) + m^2 + \overline{C}(V',V') + m^2 - 2\overline{C}(V,V') - 2m^2$$

$$\sigma_{Est}^2 = \overline{C}(V,V) + \overline{C}(V',V') - 2\overline{C}(V,V')$$

Si l'on préfère l'outilvariogramme à la covariance C(h), l'expression à l'aide du variogramme sera :

$$\sigma_{Est}^2 = 2\overline{\gamma}(V,V') - \overline{\gamma}(V,V) - \overline{\gamma}(V',V')$$

Cette notation symbolique s'étend à des domaines V et V' non forcément compacts ou continus

par exemple le domaine V à estimer peut être constitué de deux panneaux distincts, V = V1 + V2 ;

l'ensemble V' peut être constitue de plusieurs sondages ,

L'écriture symbolique précédente est générale quelles que soient les géométries des domaines V et V'. La simple écriture de cette formule rend compte des quatre faits essentiels et intuitifs que conditionnent toute estimation. La qualité d'une estimation de V par V' dépend:

1 - de la géométrie du domaine à estimer : terme $\bar{\gamma}(V,V)$

2 - des distances entre l'estimé et l'estimant : terme $\bar{\gamma}(V,V')$

3 - de la géométrie interne de l'estimant : terme $\bar{\gamma}(V',V')$

4 - du degré de régularité du phénomène étudié : utilisation de la caractéristique structurale . $\gamma(h)$

V.3. ESTIMATION D'UNE MOYENNE PAR UNE MOYENNE PONDEREE

La formule générale précèdente s'étend à la variance d'estimation de la teneur moyenne Zs d'un panneau S par une combinaison linèaire Zs des informations disponibles.

Par exemple si l'on dispose de N informations Si de teneurs moyennes z(xi), λi étant le pondérateur associé à l'information Si. L'estimateur Z* est égale à :

$$Z^* = \sum_{i=1}^{n} \lambda_i . Z(x_i)$$

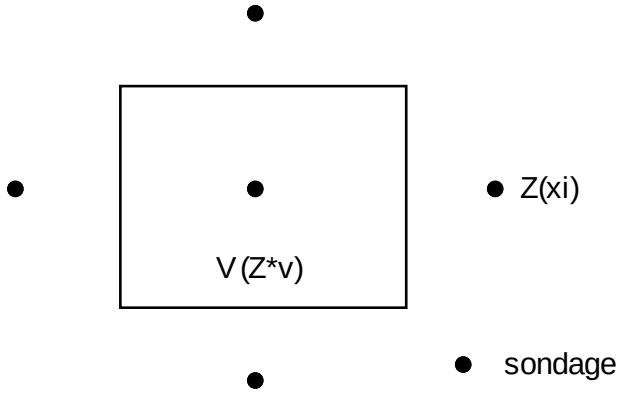


Fig. 24 - Schéma de configuration de reconnaissance d'un volume V par un certain nombre N informations de volume vi ou Si

La variance d'estimation s'écrit alors, en notation symbolique :

$$\sigma^2_{est} = 2\sum_{i=1}^{n} \lambda_i \, \gamma(V, S_i) - \bar{\gamma}(V,V) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i . \lambda_j \, \bar{\gamma}(S_i, S_j) \quad \text{(A)}$$

Cette formule (A) est générale quelles que soient les géomètries du panneau v et des informations Si, et quels que soient les pondérateurs λ_i . Le non-biais doit cependant être assuré:

$E(Z_V - Z^*) = 0$. Pour cela il suffit d'imposer la condition suivante:

$$\sum_{i=1}^{n} \lambda_i = 1$$

- Cette formule (A) peut donc servir à calculer la variance d'estimation d'estimateurs linèaires $Z^* = \sum_{i=1}^{n} \lambda_i . Z(x_i)$ du type pondérateur par moyenne arithmétique, par l'inverse de la distance, ou par le polygone d'influence et autres. Il y a donc une infinité de solutions possibles.

- Cependant en Géostatistique il exise une procedure de construction d'estimateur dite procédure de krigeage et qui consiste donc à déterminer les pondérateurs λi tels que l'on ait :

- non-biais E(Zv-Z*)=0

- Variance d'estimation minimale

V.4.CALCUL DES VALEURS MOYENNES DE γ(V,V') ou γ(S,S')

En géostatistique, il est souvent fait appel à des valeurs moyennes $\bar{\gamma}(v,v')$ (calcul des différentes variances, régularisations, krigeage ...) du variogramme ponctuel $\bar{\gamma}(h)$ quand les deux points d'appuis M et M' du vecteur h=MM' décrivent indépendamment les volumes v et v' (S,S') où :

$$\bar{\gamma}(v,v') = {}^{1}\!/_{v.v'} . \int_{v} dx . \int_{v'} \gamma(x - x') . dx'$$

dx désignant en réalité une intégrale triple :

$$\iiint_{v} dx_1 . dx_2 . dx_3$$

si v est à trois dimensions X = {$x_1.x_2.x_3$}; $\bar{\gamma}(v,v')$ désigne une intégrale sextuple. Dans la pratique, deux solutions se présentent :

- soit calculer numériquement à l'aide de calculatrice programmable la valeur moyenne $\bar{\gamma}(v,v')$ recherchée,

- soit décomposer la résolution analytique des intégrales sextuples en étapes successives dont certaines auront été résolues à l'avance et une fois pour toutes. Ces étapes intermédiaires sont définies comme fonctions auxiliaires.

- CALCUL NUMERIQUE

C'est souvent la solution la plus rapide si l'on dispose de calculatrice programmable.

On implante une maille régulière (x_i, i=1 à N) dans le volume v, une autre (x_j, j=1 … N') dans le volume v' et l'on assimile l'intégrale sextuple à une somme discrète :

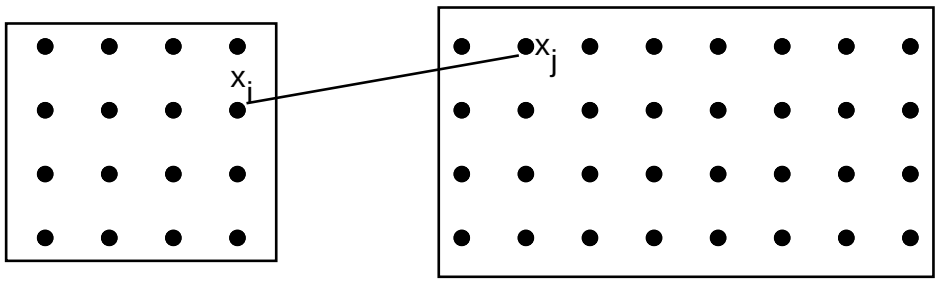


Fig. - Schéma de discrétisation de deux "volumes" V et V'

Cependant il faut noter que l'erreur est liée à la densité de discrétisation à l'interieur des volumes v et v'; elle décroît quand N et N' augmentent. Il est donc nécessaire de choisir N et N' telle que l'erreur soit pratiquement nulle et ne masque pas la variabilité étudiée.

En pratique, il y a 2 principales régles à suivre :

- La discrétisation doit rester la même pour toutes les estimations numériques γ^* des valeurs $\bar{\gamma}(h)$ d'une même formule (variance d'estimation, variance de dispersion, ..).

- La densité de discrétisation peut être choisie par itération, en s'arrêtant dès que le supplément de discrétisation n'apporte pas d'amélioration notable à la réalisation de l'objectif visé.

En pratique pour un domaine à 1D, on prend 10 points, pour 2D, 6x6 et pour 3D, N=4x4x4.

Dans de nombreux cas, le choix d'une approximation indique la discrétisation nécessaire. Par exemple, pour le calcul de $\bar{\gamma}(v,v')$,

- avec v et v' distants de plus de la portée : $\bar{\gamma}(v,v')$ = palier

- avec v et v' distants de h et petits vis-à-vis de la portée : $\bar{\gamma}(v,v') = \bar{\gamma}(h)$

- avec v' petit vis-à-vis de la portée; v' peut être assimilé à son centre de gravité (ponctuel) et on adoptera une maille N = 4 x 4 x 4 points pour discréditer v.

- FONCTIONS AUXILIAIRES

Une fonction auxiliaire est une valeur moyenne $\bar{\gamma}(v,v')$ correspondante à des géométries relativement simples et souvent rencontrées de v et v'. 4 fonctions auxiliaires essentielles sont utilisées: α, χ, F et H. Elles sont définies dans l'espace à 1, 2 ou 3 dimensions.

Ces fonctions auxiliaires sont présentées sous forme d'abaques.

CHAPITRE VI Estimation par krigeage

C'est une méthode d'interpolation spatiale. Elle porte le nom de son précurseur, l'ingénieur minier sud-africain, D.G. Krige ; c'est le Professeur George Matheron qui a baptisé la méthode « krigeage » (Gratton 2002).

La procédure de krigeage consiste à trouver *la meilleure estimation linaire* possible d'une caractéristique inconnue à partir de l'information disponible (expérimentale) et l'information structurale (variogramme, covariance ou corrélogramme) de F.A. représentative de la régionalisation des variables étudiées.

Il existe au moins trois types de krigeage (Baillargeon 2005): simple, ordinaire et universel,

La différence entre ces types d'estimation réside dans la connaissance de la statistique de la variable à interpoler (Bostan 2017):

1- Krigeage simple : Variable stationnaire de moyenne connue ;

2- Krigeage ordinaire : variable stationnaire de moyenne inconnue :

1 - Krigeage universel : variable non stationnaire.

V.1.Système du krigeage ordinaire

Elle consiste à trouver le meilleur estimateur linéaire possible d'une variable régionalisée d'un volume V implantée à l'intérieur ou à l'extérieur de V'. Pour cela on utilise le formalisme mathématique de Lagrange qui a permet d'aboutir un système de N+1 équations à N+1 inconnus.

Le système de krigeage ordinaire est donné par le système d'équation suivant :

$$\begin{cases} \sum_{i=1}^{N} \lambda i \gamma(v_i, v_j) + \mu = \gamma(v_j, V) \\ \\ \sum_{i=1}^{N} \lambda i = 1 \end{cases} \quad \forall i = 1.à.N...et..j = 1..à..N$$

La variance d'estimation de krigeage est donnée par la formule suivante :

$$\sigma_k^2 = \sum_{i=1}^{N} \lambda i \gamma(v, V)_i + \mu - \gamma(V, V)$$

Propriétés et remarques à propos du krigeage ordinaire :

- Le système de krigeage ordinaire est un système à **N+1** équations à **N+1** inconnues qui sont les N pondérateur λ_i et μ qui est le paramètre de Lagrange.
- Le krigeage est un estimateur linéaire sans biais. C'est un interpolateur exact.

Pour l'écriture matricielle du système de krigeage ordinaire :

$$\begin{vmatrix} \gamma_{11} & \gamma_{12} & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \gamma_{2n} & 1 \\ \gamma_{n1} & \gamma_{n2} & \gamma_{nn} & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} \begin{vmatrix} h_1 \\ h_2 \\ h_n \\ u \end{vmatrix} = \begin{vmatrix} \gamma_{10} \\ \gamma_{20} \\ \gamma_{n0} \\ 1 \end{vmatrix}$$

V.5.1. Système de Krigeage simple

Le krigeage le moins complexe est celui dans lequel la stationnarité postulée est de deuxième ordre et l'espérance de la fonction aléatoire étudiée est supposée connue et constante sur tout le champ. Il s'agit du krigeage simple. Donc, quand on connaît la moyenne "m" d'un champ à estimer, on utilise le Krigeage simple comme un estimateur sans biais minimisant la variance d'estimation (Matheron 1978).

Le système de krigeage simple est :

$$Zo = \sum \lambda i \, Zi + \left(1 - \sum \lambda i\right) m$$

Et son écriture matricielle est :

$$\begin{vmatrix} \gamma_{11} & \gamma_{12} & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \gamma_{2n} \\ \gamma_{n1} & \gamma_{n2} & \gamma_{nn} \end{vmatrix} \begin{vmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_n \end{vmatrix} = \begin{vmatrix} \gamma_{1\,0} \\ \gamma_{2\,0} \\ \gamma_{n\,0} \end{vmatrix}$$

V.5.2.Système de Krigeage universel

C'est une méthode de krigeage souvent utilisée sur les données présentant une tendance spatiale significative, comme une surface en pente. L'hypothèse de stationnarité sur laquelle repose les deux types de krigeage présentés précédemment peut souvent être mise en doute. En krigeage universel, les valeurs attendues des points échantillonnés sont modélisées en tant que tendance polynomiale (Bostan 2017).

Le modèle supposé pour la variable régionalisée est :

Z(x)=Y(x)+m(x)

Comportant une dérive m(x) déterministe et un résidu Y(x) stationnaire d'espérance nulle.

On modélise alors la tendance déterministe sous forme d'une somme de fonctions de base :

$$m(x) = \sum_{p=1}^{l} a_p f^p(x)$$

où les *ap* sont des coefficients réels et où chaque fonction de base f^p ne dépend que d'une seule coordonnée. En règle générale, on prend des fonctions classiques comme des exponentielles ou des monômes *x*, *x*2, *x*3,... et on fixe par convention la première fonction de base $f^1(x) = 1$.

Le krigeage universel consiste alors à estimer simultanément la tendance *m* et la fluctuation aléatoire
Y en *x*0. Pour la résolution du problème on obtient un système de $n + l$ inconnues :

$$\begin{bmatrix} C(0) & C(x_1-x_2) & \dots & C(x_1-x_n) & f^1(x_1) & \dots & f^l(x_1) \\ C(x_2-x_1) & C(0) & \dots & C(x_2-x_n) & f^1(x_2) & \dots & f^l(x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ C(x_n-x_1) & C(x_n-x_2) & \dots & C(x_n-x_n) & f^1(x_n) & \dots & f^l(x_n) \\ f^1(x_1) & f^1(x_2) & \dots & f^1(x_n) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ f^l(x_1) & f^l(x_2) & \dots & f^l(x_n) & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda_1 \\ \vdots \\ \lambda_l \end{bmatrix} = \begin{bmatrix} \overline{\gamma}(x_1-x_0) \\ \overline{\gamma}(x_2-x_0) \\ \vdots \\ \overline{\gamma}(x_n-x_0) \\ f^1(x_0) \\ \vdots \\ f^l(x_0) \end{bmatrix}$$

qui se généralise sans problème au cas intrinsèque en remplaçant comme à l'accoutumée *C* par *-γ*.

La variance de krigeage s'´écrit :

$$\mathrm{Var}[\widehat{Z} - Z] = C(0) - \sum_{i=1}^{n} w_i C(x_i - x_0) - \sum_{p=1}^{l} \lambda_p f^p(x_0)$$

V.5.3.Le cokrigeage

En géostatistique, le cokrigeage est une extension du krigeage au cas multivarié. qui prend en compte plusieurs variables (Rivoirard 2003).

VI.5.3.1. Définition

Le cokrigeage, une extension du krigeage, est applicable lorsque deux variables spatiales ou plus sont en jeu. Initialement développé dans le but d'améliorer la prédiction d'une variable pour laquelle seuls quelques échantillons sont disponibles, il exploite la corrélation spatiale avec d'autres variables plus facilement mesurables. Une distinction essentielle entre le cokrigeage et le krigeage avec dérive externe réside dans le fait que les variables explicatives ne servent pas à identifier une tendance dans la variable principale, mais sont en elles-mêmes des éléments de prédiction. Cela nécessite de définir covariogramme croisé.

VI.6.Covariogramme croisé

$$\gamma_{ZY}(h) = \frac{1}{2p(h)} \sum_{i=1}^{p(h)} (z(s_i) - z(s_i + h))(y(s_i) - y(s_i + h))$$

$$\text{avec } p(h) = Card\left\{(s_i, s_j) \middle| |s_i - s_j| \approx h\right\}$$

Comme pour le krigeage, il y aura plusieurs versions pour le cokrigeage. On ne présentera que le cokrigeage ordinaire. On se limitera au cas où l'on n'introduit qu'une variable auxiliaire, que l'on notera Y. L'estimateur que l'on calcule est de la forme :

$$Z(s_0) = \sum_{i=1}^{n_Z} \lambda_i Z(s_i) + \sum_{i=1}^{n_Y} \alpha_i Y(s_i)$$

avec les contraintes d'absence de biais :

$$\sum_{i=1}^{nZ} \lambda_i = 1$$

$$\sum_{i=1}^{ny} \alpha_i = 0.$$

Les équations de cokrigeage s'écrivent :

$$\sum_{j=1}^{nz} \lambda_j Cov(Z_i, Z_j) + \sum_{j=1}^{ny} \alpha_j Cov(Z_i, Y_j) + \mu_z = Cov(Z_0, Z_i) \qquad \forall i = 1...nz$$

$$\sum_{j=1}^{nz} \lambda_j Cov(Y_i, Z_j) + \sum_{j=1}^{ny} \alpha_j Cov(Y_i, Y_j) + \mu_y = Cov(Z_0, Y_i) \qquad \forall i = 1...ny$$

$$\sum_{i=1}^{nz} \lambda_i = 1$$

$$\sum_{i=1}^{ny} \alpha_i = 0$$

Et sa variance est

$$\sigma_{ck}^2 = Var(Z_0) - \sum_{i=1}^{nz} \lambda_i Cov(Z_0, Z_i) - \sum_{i=1}^{ny} \alpha_i Cov(Z_0, Y_i) - \mu_z$$

Toutes les propriétés du krigeage sont valides pour le cokrigeage. En plus, Si l'on estime directement par cokrigeage une combinaison linéaire des variables, la valeur cokrigée sera égale à la même combinaison linéaire appliquée aux valeurs cokrigées de chaque variable. (Ce ne serait pas le cas pour le krigeage).