

**Intitulé du Master : Ressources minérales et géomatériaux**

**Semestre : 03 - Intitulé de l'UE : UEM1**

**Intitulé de la matière : Géostatistiques**

**Crédits : 4 - Coefficients : 2**

### **Objectifs de l'enseignement**

En géologie le recours à la géostatistique a pour objectif de traiter les données afin d'estimer le potentiel et l'intérêt des éléments analysés (intérêt scientifique et/ou économique). Le traitement statistique des données conduit à une meilleure évaluation et quantification des objets géologiques étudiés. Donc cette matière est nécessaire dans la formation des étudiants de Master de la filière « Géologie ».

**Connaissances préalables recommandées :** Notions de statistiques

### **Contenu de la matière**

#### **Cours :**

Chap. I : Notions de statistiques

- Rappels de mathématiques - Les paramètres statistiques
- Définitions (échantillons, populations, variables...)
- Théorie élémentaire des probabilités.

Chap. II : Traitement statistique

- Analyse univariante (fonctions de distribution et histogrammes de fréquences)
- Analyse bivariante (diagrammes de dispersion, droite de régression et corrélation linéaire)
- Analyse multivariante (Analyse en Composante Principale, ACP)

Chap. III : Différents types de distributions - Distribution normale

- Distribution lognormale
- Distribution de poisson

**TD :**

Calcul des paramètres statistiques

Distribution de fréquences

Tests de distributions

Traitements graphiques des données  
d'analyses Cartographie.

---

---

## Chapitre I : Notions de statistiques

---

### I. Définitions

La statistique descriptive est l'ensemble de méthodes permettant de décrire et d'analyser, de façon quantitative, des observations pour un phénomène donné. Ces observations consistent généralement en la mesure d'une ou plusieurs caractéristiques communes sur un ensemble de personnes ou d'objets équivalents.

- A. **Population** : ensemble d'objets ou d'individus ayant des caractéristiques qui leurs sont propres.
- B. Chaque objet d'une population s'appelle un **individu** ou **unité statistique**.
- C. **Échantillon** : sous-ensemble d'une population. Lorsque la population est trop importante pour être reconnue entièrement, on prélève un échantillon. Un bon échantillon doit être représentatif de la population dont il est issu.

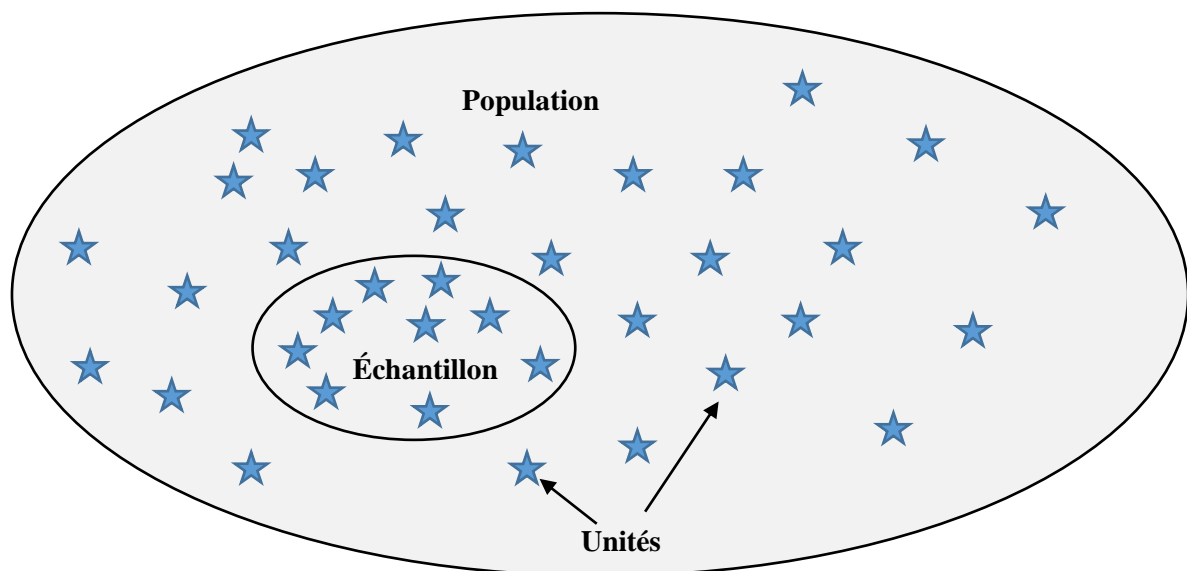


Figure 1. Schéma montrant la population, l'échantillon et les individus statistiques

### D. Caractère et variable statistique

On appelle variable statistique ou caractère, la caractéristique que l'on étudie et l'on mesure et qui est commune à tous les individus de la population de référence. L'ensemble des résultats s'appelle série statistique, alors que les mesures s'appellent

des observations. Un caractère peut être qualitatif comme la couleur des yeux et le sexe. Alors que d'autres sont quantitatifs comme l'âge et la taille.

**Exemple :**

\* Les étudiants de la faculté représentent la population, alors que l'étudiant représente l'individu statistique et une classe d'étudiants représente un échantillon statistique. Le caractère qu'on veut étudier (exemple : la couleur des yeux, le sexe)

\* Les teneurs d'un élément chimique dans un gisement représentent une population statistique. L'ensemble des échantillons prélevés correspondent à un échantillon statistique, alors qu'un seul échantillon géologique sur lequel on fait la mesure correspond à un individu ou unité statistique.

**E. Valeur discrète et valeur continue**

Une variable quantitative peut être discrète ou continue. Si le nombre de valeur possible est fini, la variable est discrète. Et si le nombre de valeurs possibles est infini, la variable est dite continue.

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

**F. 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.

**G. Effectifs et fréquences**

\* L'effectif ou la fréquence absolue  $n_i$  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  $N$ .

\* **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  $i$  :

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

\* **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 :

$$f_i = n_i/N$$

\* **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  $i$ .

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

\* **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  $i$ .

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

## 2. 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.

	$X_i$	$n_i$ (effectif)
	12	2
	16	13
	22	15
<b>mode</b>	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 :

$$\bar{X} = \sum_{i=1}^p fi * xi$$

### 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  $\text{var}(X)$  ou  $S^2$

calculé par la formule :  $\text{var}(X) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{X})^2$

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

### \* Quartiles

Soit  $X$  une série, on définit les quartiles  $Q_1$ ,  $Q_2$  et  $Q_3$  de la manière suivante :

- $Q_1$  est une valeur du caractère telle que 25% de la population a un caractère inférieur à  $Q_1$ .
- $Q_2$  est une valeur du caractère telle que 50% de la population a un caractère inférieur à  $Q_2$ :
- $Q_3$  est une valeur du caractère telle que 75% de la population a un caractère inférieur à  $Q_3$ .

#### **Remarque :**

- On note que  $Q_2 = Me$
- L'intervalle  $[Q_1-Q_3]$  s'appelle l'intervalle interquartile. Il contient 50% de la population.

### \* Fonctions de répartition

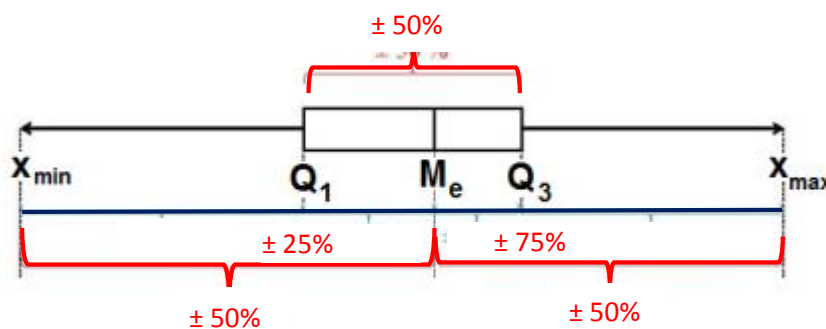
Soit  $X$  une série. On appelle fonction de répartition  $F$ , la fonction qui a une valeur du caractère  $x$  cumulée croissante jusqu'à  $x_i$

$$F : \{\text{caractères}\} \rightarrow [0;1] \text{ donc } : X \mapsto f_X^{c\uparrow} = f_{xi < X}$$

Avec cette définition, les quartiles sont simplement définis par :  
 $Q_1 = F^{-1}(0;25)$ ;  $Q_2 = F^{-1}(0;5)$  et  $Q_3 = F^{-1}(0;75)$

### \* 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'œil.



**Figure 2. La boîte à moustaches**

---

## Chapitre II : Les Probabilités

---

### 1. Définition

Une épreuve est une expérience dont l'issue est incertaine. Les résultats éventuels d'une épreuve sont généralement appelés au hasard. L'ensemble des résultats éventuels (les résultats possibles, les éventualités) s'appelle ensemble fondamental.

Exemple : Un dé possède six faces numérotées (éventualités ou cas possibles). Si le dé n'est pas trafiqué et lancé loyalement, une face d'un numéro donné a autant de chances d'apparaître que les autres faces. On dit alors que les six éventualités sont équiprobables.

Les probabilités vont nous servir à modéliser une **expérience aléatoire**, c'est-à-dire un phénomène dont on ne peut pas prédire l'issue avec certitude, et pour lequel on décide que le déroulement sera le fait du hasard.

L'étude d'une expérience aléatoire est basée sur l'analyse des résultats favorables par rapport au nombre de résultats possibles. L'**espace des possibles  $\Omega$**  ou **univers** décrit tous les résultats possibles de l'expérience  $A_i$ . Chacun de ces résultats est appelé **événement élémentaire  $\omega$** . Un **événement** est un sous-ensemble de  $\Omega$ , ou une réunion d'événements élémentaires.

**Exemple :** lancer d'un dé régulier est une expérience  $A_i$ , donc :

$$\Omega = \{1, 2, 3, 4, 5, 6\} \text{ et } \omega = 3 \text{ est un résultat possible.}$$

### 2. Opérations sur les événements

\* **Complémentaire de A :** événement constitué des résultats élémentaires de  $\Omega$  qui ne sont pas dans A. Soit  $\omega$  le résultat de l'expérience :

$$\bar{A} = \{\omega \in \Omega, \omega \notin A\}$$

$\bar{A}$  se réalise si A ne se réalise pas

\* **Réunion de A et B :** événement constitué des résultats élémentaires de  $\Omega$  qui appartiennent à A ou à B (ou aux deux). Soit  $\omega$  le résultat de l'expérience :

$$A \cup B = \{\omega \in \Omega, \omega \in A \text{ ou } \omega \in B\}$$

\* **Intersection de A et B :** événement constitué des résultats élémentaires de  $\Omega$



qui appartiennent à la fois à A et à B. Soit  $\omega$  le résultat de l'expérience :

$$A \cap B = \{\omega \in \Omega, \omega \in A \text{ et } \omega \in B\}$$

\* **Disjonction ou incompatibilité** : A et B sont disjoints si A et B n'ont pas d'éléments communs :

$$A \text{ et } B \text{ disjoints} \Leftrightarrow A \cap B = \{\phi\} \quad (\phi: \text{ensemble vide})$$

### 3. Probabilité

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

**3.1. Définition** : Soit  $(\Omega, A)$  un espace probablisable. Une probabilité sur  $(\Omega, A)$  est une application  $P \rightarrow [0,1]$  satisfaisant les 3 conditions suivantes :

$$0 \leq P(A) \leq 1$$

$$P(\Omega) = 1$$

$$P(\cup A_i) = \sum P(A_i)$$

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

### 3.2. Opérations sur les probabilités

$$* P(\phi) = 0$$

$$* P(\bar{A}) = 1 - P(A)$$

$$* 0 \leq P(A) \leq 1$$

### 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.

$$E(x) = \sum_{i=1}^k x_i \times P(X = x_i)$$

Si  $P(X = x_i)$  est remplacé par  $n_i/N$  (qui est  $f_i$ ), l'espérance mathématique peut alors s'écrire :

$$E(x) = \sum_{i=1}^k x_i \times f_i$$

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 :

$$* E(ax) = a x E(x)$$

$$* E(ax + b) = a x E(x) + b$$

$$* E(x + y) = E(x) + E(y)$$

$$E[XY] = E[X].E[Y] + COV(X; Y)$$

• **Moment d'ordre n:**

- **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 :**

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

La variance d'une variable aléatoire est donc le moment centré d'ordre 2.

$$\text{On peut donc écrire : } \sigma^2 = \sum_{i=1}^k [(x_i - E(x))^2 \cdot P(X = x_i)]$$

En remplaçant  $P(X=x_i)$  par  $f_i$  et  $E(x)$  par  $\bar{X}$  on obtient donc :

$$S^2 = \sigma^2 = \sum_{i=1}^k [(x_i - \bar{X})^2 \cdot f_i]$$

## 5. Loi d'une variable aléatoire

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

1. l'ensemble des valeurs qu'elle peut prendre (son domaine de définition  $D_X$ );
2. les probabilités attribuées à chacune de ses valeurs  $P(X = x_i) : 0 \leq P(X = x_i) \leq 1$

**Exemple** : on lance une pièce de monnaie trois fois, le nombre de fois possible qu'une face pile peut apparaisse est 0, 1, 2 et 3. Donc, L'ensemble fondamental est  $\Omega = \{\text{Face, pile}\}$  et  $X$  le nombre d'apparition de "pile", donc  $D_X = \{0, 1, 2, 3\}$ .

### 5.1. Fonction de répartition

La fonction de répartition n'est autre que le cumul des probabilités individuelles. La probabilité pour que la variable aléatoire  $X$  prenne une valeur inférieure à  $x$  est une fonction  $F(x)$  que l'on appelle fonction de répartition :

$$F(x) = P(X \leq x).$$

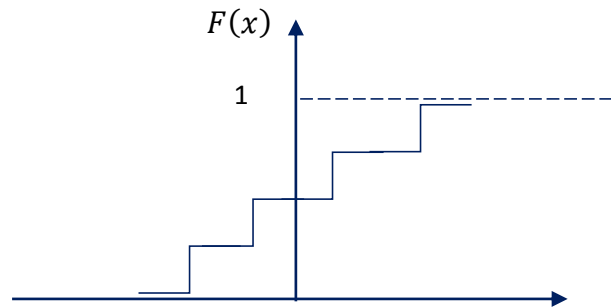
*Exemple* : on jette une pièce de monnaie  $N$  fois et on s'intéresse à la probabilité d'obtention la face pile pour au moins  $x$ -fois de jeter la pièce. Dans ce cas, on parle de probabilité cumulée ou encore de **fonction de répartition** qui est notée  $F(x)$ . Il s'agit de la somme des probabilités ponctuelles.

#### \* Propriétés de la fonction $F(x)$ :

- $F(x)$  est défini sur  $R \rightarrow [0,1]$  : il prend des valeurs allant de 0 à 1 ;
- La fonction  $F(x)$  est croissante ;
- $F(x)$  est continue à droite

\* **Pour une variable discrète** :  $F(x) = P(X \leq x_i) = \sum_{i=1}^n P(X = i)$

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

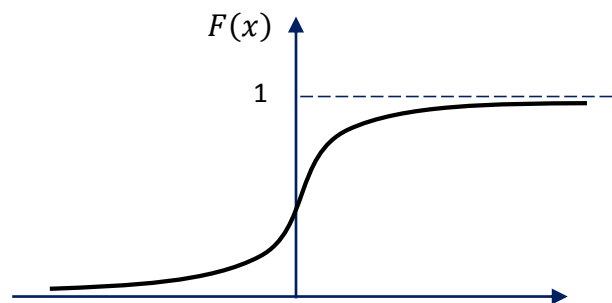


**Figure 3. Graphe de fonction de répartition d'une v.a. discrète**

\* **Pour une variable continue** : La probabilité ponctuelle  $P(X = x) = f(x)$  est appelée la fonction de densité. La fonction de répartition est :

$$F(x) = P(X \leq x_i) = \int_{-\infty}^{x_i} f(x) dx$$

$F(x)$  est une fonction continue.



**Figure 4. Graphe de fonction de répartition d'une v.a. continue**

## 5.2. Fonction de densité

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  $\mathbb{R}$  dans  $\mathbb{R}$ . Cette fonction est caractérisé par :

\*  $f(x) \geq 0$

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

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 5). La distribution de probabilité d'une v.a. continue est une distribution continue.

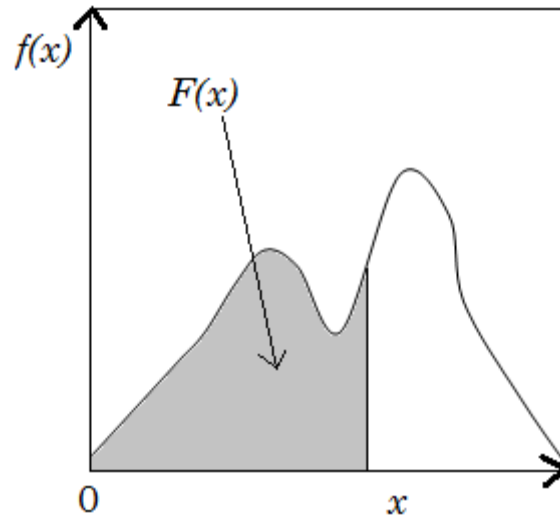


Figure 5. La fonction de densité

### 5.3. Probabilité d'un intervalle

Dans certains cas on s'intéresse à la probabilité d'un intervalle, comme exemple la probabilité d'avoir des teneurs comprises entre de valeurs  $P(a < X < b)$ . Graphiquement, elle correspond à la surface comprise entre  $a$  et  $b$  sur le graphe de la fonction de densité.

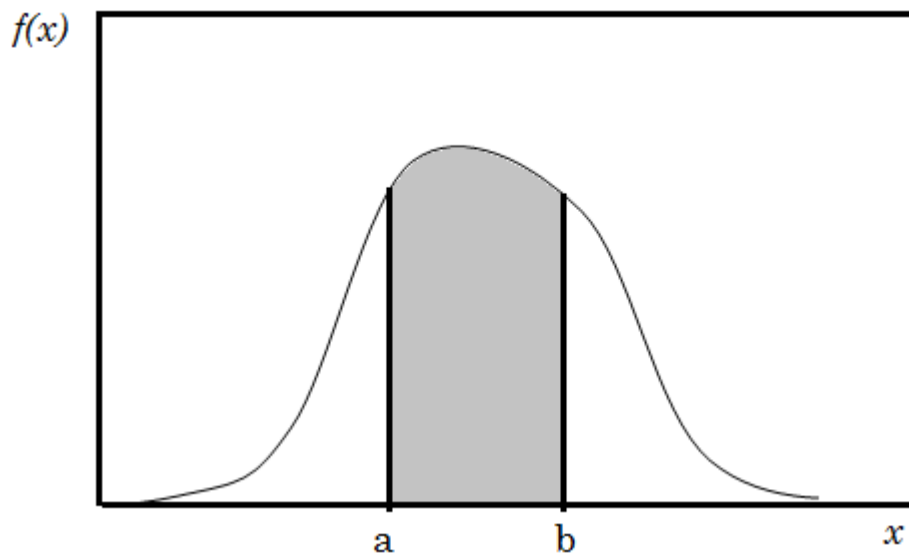


Figure 6. Probabilité d'un intervalle  $[a,b]$

Analytiquement, il s'agit de :

$$P(a < X < b) = \int_a^b f(x)dx$$

Donc:

$$P(a < X < b) = P(X < b) - P(X < a) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = F(b) - F(a)$$

#### 5.4. Loi de Laplace-Gauss ou loi normale

C'est la loi la plus importante. Son rôle est central dans de nombreux modèles probabilistes et dans toute la statistique. Elle possède des propriétés intéressantes qui la rendent très utile.

Une v.a.  $X$  suit la loi normale (Loi gaussienne, Loi de Laplace-Gauss), lorsqu'elle est formée à partir d'un grand nombre de facteurs s'additionnant les uns aux autres, aucune ne jouant un rôle prédominant, et agissant de manière indépendante.

**Exemple :** la teneur d'un élément chimique d'un échantillon géologique est le résultat de plusieurs paramètres géochimiques.

\* **Définition :** Une v.a.  $X$  suit une loi normale (ou loi gaussienne ou loi de Laplace-Gauss) et qui est notée  $\mathcal{N}(\bar{X}; \sigma)$  si sa densité  $f$  est donnée par :  $f(x) = \frac{1}{\sqrt{2\pi}} \exp - \frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2$ .

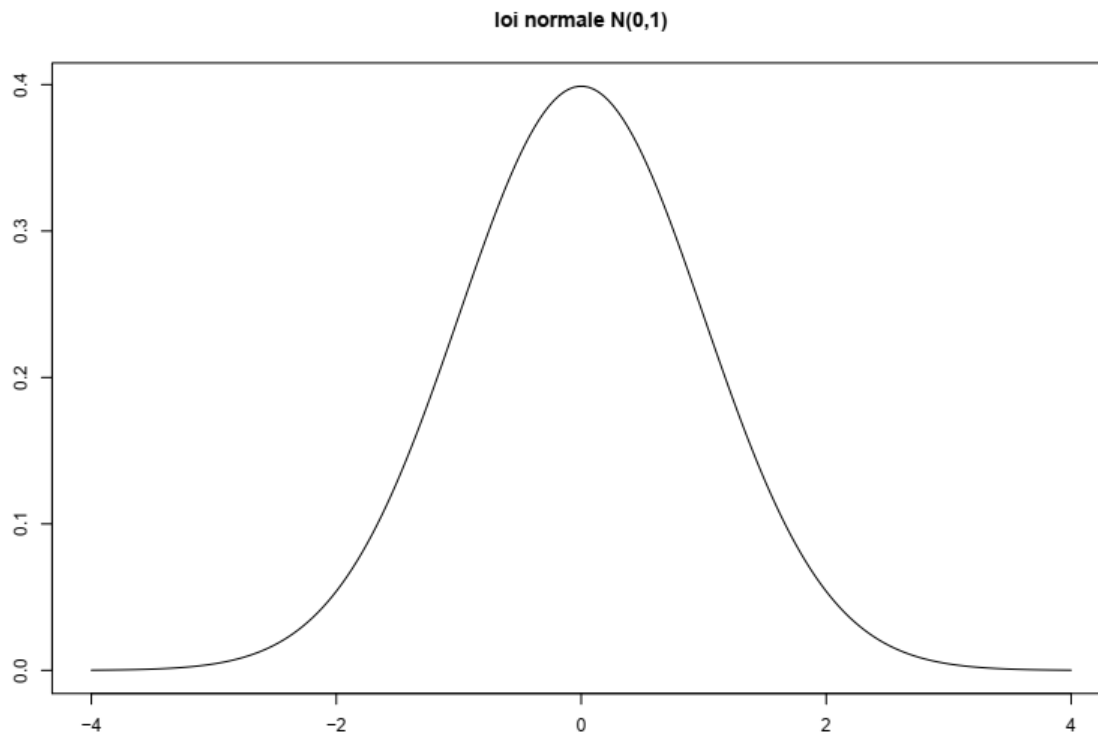
**Remarque :**

- $N(0; 1)$  : signifie une loi normale centrée et réduite de moyenne  $\mu = 0$  et d'un écart type  $\sigma = 1$ . Donc :  $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$
- Le passage d'une v.a.  $X$  de moyenne  $\mu$  et d'un écart type  $\sigma$  à une v.a.  $Z$  centrée et réduite s'effectué par la relation :

$$Z = (X - \mu) / \sigma$$

##### 5.4.1. 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 7. Courbe de densité de la loi  $N(0, 1)$**

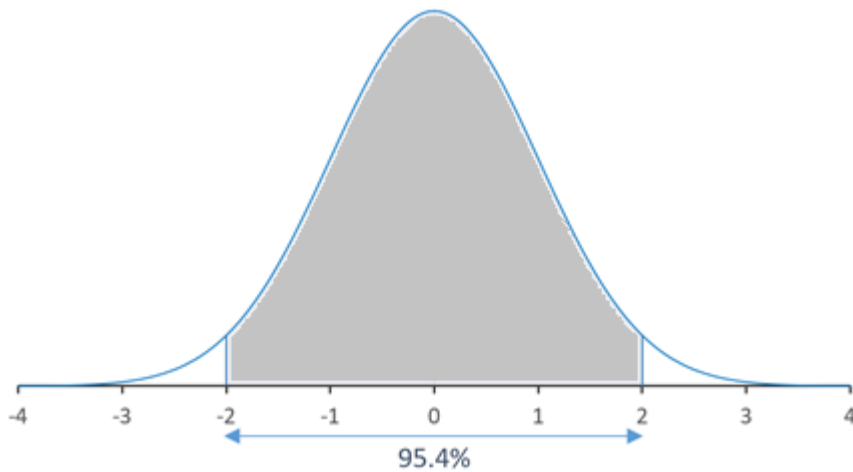
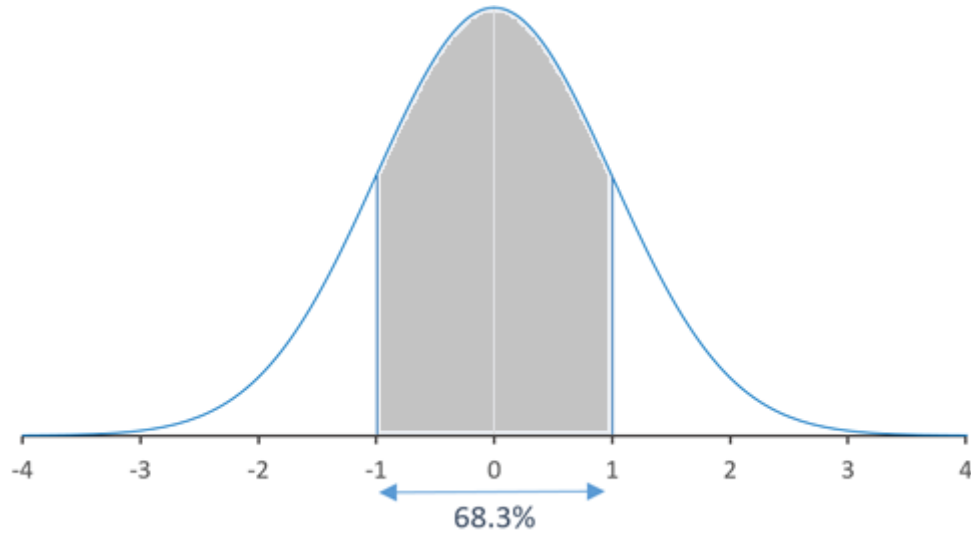
La courbe de densité permet de calculer les probabilités. On peut noter :

- La probabilité  $P(a < X) = \int_{-\infty}^a f(x)dx$  : c'est l'aire allant de l'infini jusqu'à la valeur  $a$  ;
- La probabilité  $P(-x < X) = P(X > +x)$  ;
- La probabilité  $P(a > X) = 1 - P(a < X)$

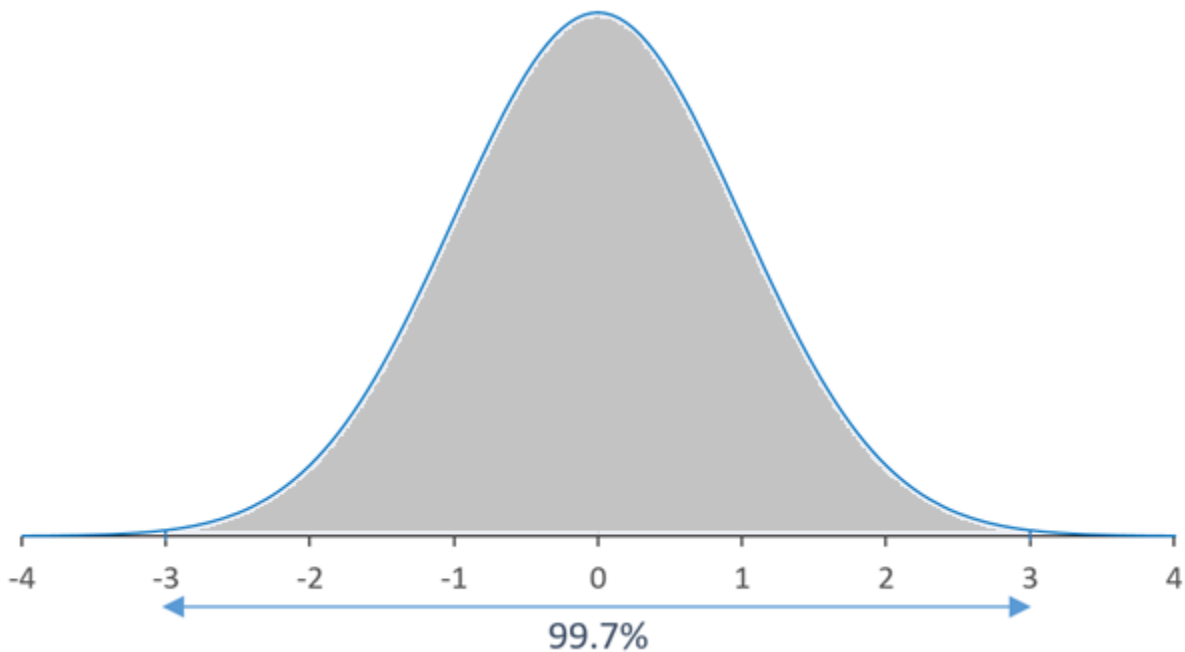
#### 5.4.2. Interprétation de l'écart-type de la loi normale centrée réduite

- Elle est spéciale car ses valeurs en abscisse représentent des unités d'écart-type.
- Sur l'axe des abscisses on trouve des valeurs allant de  $-6\sigma$  à  $+6\sigma$ . Le centre de la courbe est positionné au-dessus de 0.
- Toutes les distributions normales de moyenne 0 et d'écart-type de 1 sont appelées des **distributions normales standard**.
- La probabilité  $P(a < X < b)$  correspond à l'aire comprise entre  $a$  et  $b$ . En analysant la courbe de fonction de densité de la loi  $N(0, 1)$  pour certains cas. On remarque que (Figure 8) :
  - $P(-1 < X < 1)$  correspond à la surface comprise entre le 0 et  $\pm 1\sigma$  qui est de **68.3%** des observations.

- $P(-2 < X < 2)$  correspond à la surface comprise entre le 0 et  $\pm 2\sigma$  qui est de **95.4%** des observations.
- $P(-3 < X < 3)$  correspond à la surface comprise entre le 0 et  $\pm 3\sigma$  qui est de **99.7%** des observations.







**Figure 8. Probabilités des principaux intervalles de la loi  $N(0,1)$**

#### **5.4.3. Utilisation de la table de probabilités de la loi $N(0, 1)$**

La distribution normale est une distribution de probabilité continue et qui correspond à un intervalle et donc à l'aire donnée par la fonction de distribution. Il est important de savoir qu'il existe plusieurs types de tables de loi normale centrée réduite. La lecture de la probabilité varie d'une table à l'autre. Pour cela les tables possèdent généralement un schéma explicatif au sommet de celle-ci. Il faut donc être attentif à la **zone grisée** qui représente l'aire de la probabilité considérée.

La Table comprend 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.

