Centre Universitaire Abdelhafid-Boussouf –Mila-
Institut de science et technologie
Département de science et technologie

Cours analyse et modélisation hydrologique

Hydrologie statistique (partie 01)

Master 01 : hydraulique urbaine

Rappels de quelques notions de probabilité et d'analyse statistique

L'analyse statistique est opéré sur un échantillon ou une série d'observations (données).

Ces données (variables) peuvent être continues ou discrètes.

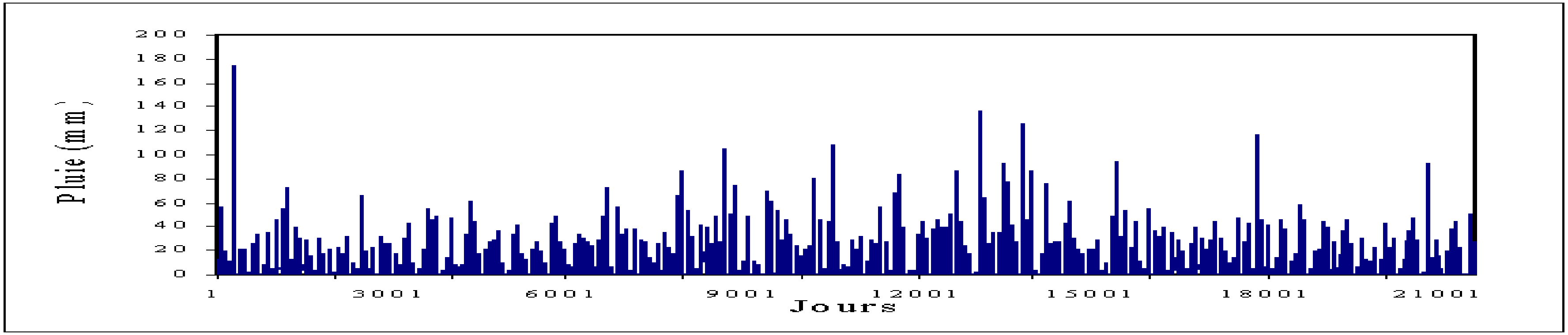
Variables continues : ce sont généralement des valeurs mesurées (précipitation, débit, …).

Variables discrètes : résultent habituellement d'un comptage (le fait de pleuvoir ou non est une variable discrète), (nombre de jours de pluie, de neige ou de gel…).

Elles proviennent d'une population, qui est un phénomène hydrologique dont une description est requise.

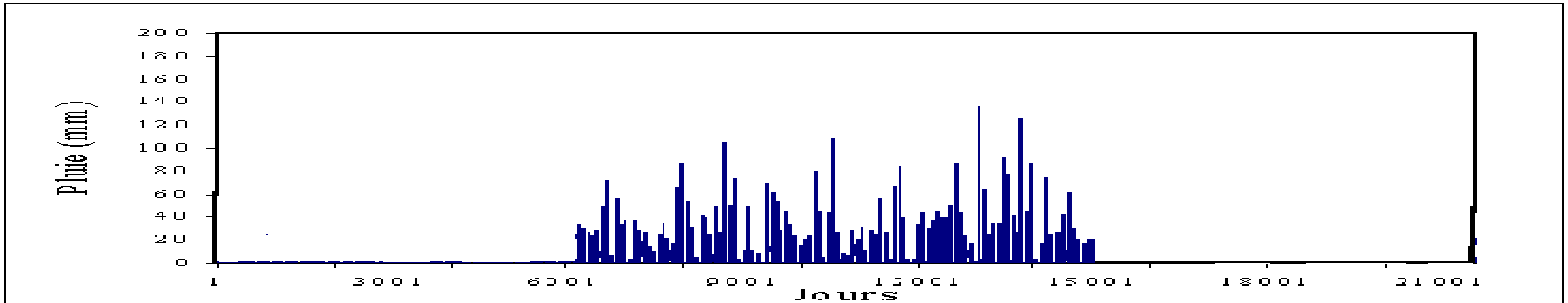
Rappels de quelques notions de probabilité et d'analyse statistique

On appelle **population** un ensemble fini ou infini d'éléments.



Population

Un **échantillon** est un sous ensemble de la population.



Echantillon continu

Rappels de quelques notions de probabilité et d'analyse statistique

On dira qu'un échantillon est exhaustif lorsque sa taille est celle de la population.

En hydrologie de surface, les populations sont généralement infinis, et par conséquent, les échantillons sont non exhaustifs.

Ces données doivent répondre à certains critères statistiques pour que ce genre d'analyse soit valable.

Elles doivent être, <u>indépendantes</u>, <u>homogènes</u>, <u>stationnaires</u> et <u>aléatoires.</u>

Rappels de quelques notions de probabilité et d'analyse statistique

Les variables que l'on manipule en hydrologie (précipitations, débits, mais aussi niveau de nappe phréatique, durée d'insolation, etc...), vont être considérées comme des Variables Aléatoires.

La Variable Aléatoire, parfois notée V.A., est une variable formelle, notée en majuscule, par exemple X :

X = "Précipitation annuelle à la station d'Alger"

Cette variable prendra une valeur x_k à chaque "tirage aléatoire", à chaque réalisation k.

Par exemple, en 1980, la variable X a pris la valeur **$x_{80} = 734$ mm.**

Rappels de quelques notions de probabilité et d'analyse statistique

Il n'en reste pas moins que, si l'on veut dimensionner un barrage pour compenser le manque d'eau nécessaire à l'AEP, à Alger, il faudra s'intéresser aux années futures (- par exemple de 1987, fin de la construction de l'ouvrage (Barrage de Keddara), à 2037, fin de la période d'amortissement -). Or on ne savait pas en 1987 ce que seraient les réalisations de la variable aléatoire X en 1988, 89 etc.., c'est à dire x_{88}, x_{89}, x_{90} etc...

On se trouve alors en avenir incertain : aucune approche déterministe, aucune mesure ou méthode déductive ne peut nous dire exactement, en 1987, ce que sera la réalisation x_{92} de X en 1992...

Rappels de quelques notions de probabilité et d’analyse statistique

Tout au plus pourra-t-on supposer que les phénomènes générateurs de la pluie seront les mêmes que dans le passé récent, et on fera l'hypothèse que les réalisations futures de la variable aléatoire X auront les mêmes caractéristiques, la même distribution statistique que par le passé….

Naturellement, cette hypothèse ne s’appliquera qu’à un futur relativement “ proche ” : sur une durée un peu supérieure à la durée d’amortissement de l’ouvrage, ou encore de l’ordre de grandeur de sa durée de vie utile, c’est à dire sur quelques dizaines ou centaines d’années…

Rappels de quelques notions de probabilité et d'analyse statistique

On va donc chercher bientôt à décrire et à résumer un échantillon, considéré comme un sous-ensemble d'une population qui sera souvent infinie.

Sur cette population, on peut définir une loi de probabilité : F(x) où x correspond à une valeur numérique.

Cette loi de probabilité, ou fonction de répartition, exprime la : "Probabilité que la Variable Aléatoire X reste inférieure ou égale à la valeur Numérique x."

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

Rappels de quelques notions de probabilité et d'analyse statistique

La fréquence (f) d'un événement est le nombre de fois que cet événement se répète divisé par le nombre total d'événements observés.

$$f = n/N$$

Lorsque le nombre total d'événements ou d'observations tend vers l'infini, cette mesure tend vers une valeur constante qui s'appelle la probabilité (p).

En hydrologie, on parle indistinctement de probabilité et de fréquence.

Le terme fréquence s'applique aussi à des intervalles fixes comprenant chacun un certain nombre d'observations.

Rappels de quelques notions de probabilité et d'analyse statistique

Nous avons vu au chapitre précédent que l'ensemble des couples (x_i, n_i) ou (x_i, f_i) définit la fonction de densité d'une variable, c'est à dire :

$$f_e(x_i) = {n_i}/{N}$$

Et que la fonction de répartition est :

$$FND = F_e(x_i) = \sum_{j=1}^{i} f_e(x_j) \text{ou } FD = F1_e(x_i) = \sum_{j=1}^{k} f_e(x_j)$$

Ces fonctions f_e et F_e (e pour échantillon) sont définies pour un échantillon.

Rappels de quelques notions de probabilité et d'analyse statistique

Pour une population ces fonctions sont f(x) et F(x) et sont définies comme suit :

$$f(x) = \lim_{\substack{N \to \infty \\ \Delta x \to 0}} \left\{ f_e(x) \big/ \Delta x \right\} \quad \text{et } F(x) = \lim_{\substack{N \to \infty \\ \Delta x \to 0}} F_e(x)$$

f(x) = fonction de densité de probabilité (f.d.p)

F(x) = Fonction de répartition des probabilités (f.p)

Si l'intervalle choisi est de plus en plus petit, le polygone des fréquences (ou histogrammes) devient une courbe continue et s'appelle fonction de densité de probabilité f(x) et le polygone cumulatif s'appelle alors fonction de distribution F(x).

Rappels de quelques notions de probabilité et d'analyse statistique

Variable aléatoire discrète (VAD)

Définition d'un VAD :

1. X peut prendre les valeurs $x_1, x_2, \ldots, x_n$

2. Les probabilités respectives sont $f(x_1), f(x_2), \ldots, f(x_n)$

3. Caractéristiques des probabilités :

$$\boxed{\sum_{i=1}^{n} f(x_i) = 1 \quad \text{avec} \quad 0 \leq f(x_i) \leq 1}$$

4. Fonction cumulative F :

$$F(x_i) = \sum_{j=1}^{i} f(x_j)$$

$$f(x_i) = F(x_i) - F(x_{i-1})$$

Rappels de quelques notions de probabilité et d'analyse statistique

Probabilité que X se trouve entre les valeurs a et b :

$$P\left(X \leq a = x_i\right) = F\left(x_i\right)$$

$$P\left(X \leq b = x_j\right) = F\left(x_j\right) \quad x_j > x_i$$

$$P\left(a < X \leq b\right) = F\left(x_j\right) - F\left(x_i\right) = \sum_{x_k > a}^{x_k \leq b} f\left(x_k\right)$$

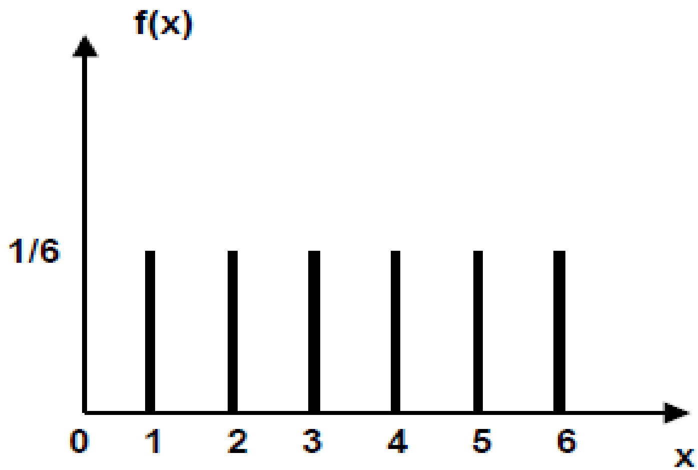
Probabilités discrètes :
x_i est exclus alors que x_j
est inclus dans la
probabilité cherchée

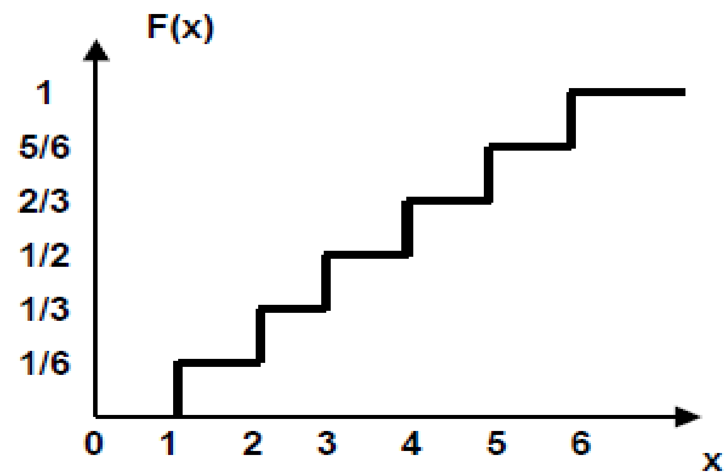
Rappels de quelques notions de probabilité et d'analyse statistique

Exemple 1 :

Dé à 6 faces classique (dé non lancé) :

Résultat n°	i	1	2	3	4	5	6
Valeurs possibles	x_i	1	2	3	4	5	6
Probabilité	$f(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6





$$P(2 \leq X \leq 5) = \sum_{2 \leq x_i \leq 5} f(x_i) = f(2) + f(3) + f(4) + f(5) = 2/3$$

Rappels de quelques notions de probabilité et d'analyse statistique

Variable aléatoire continue (VAC)

Variable aléatoire continue : variable qui peut prendre toutes les valeurs dans un intervalle fini ou infini et s'il existe une fonction f(x), appelée fonction de densité de probabilité telle que :

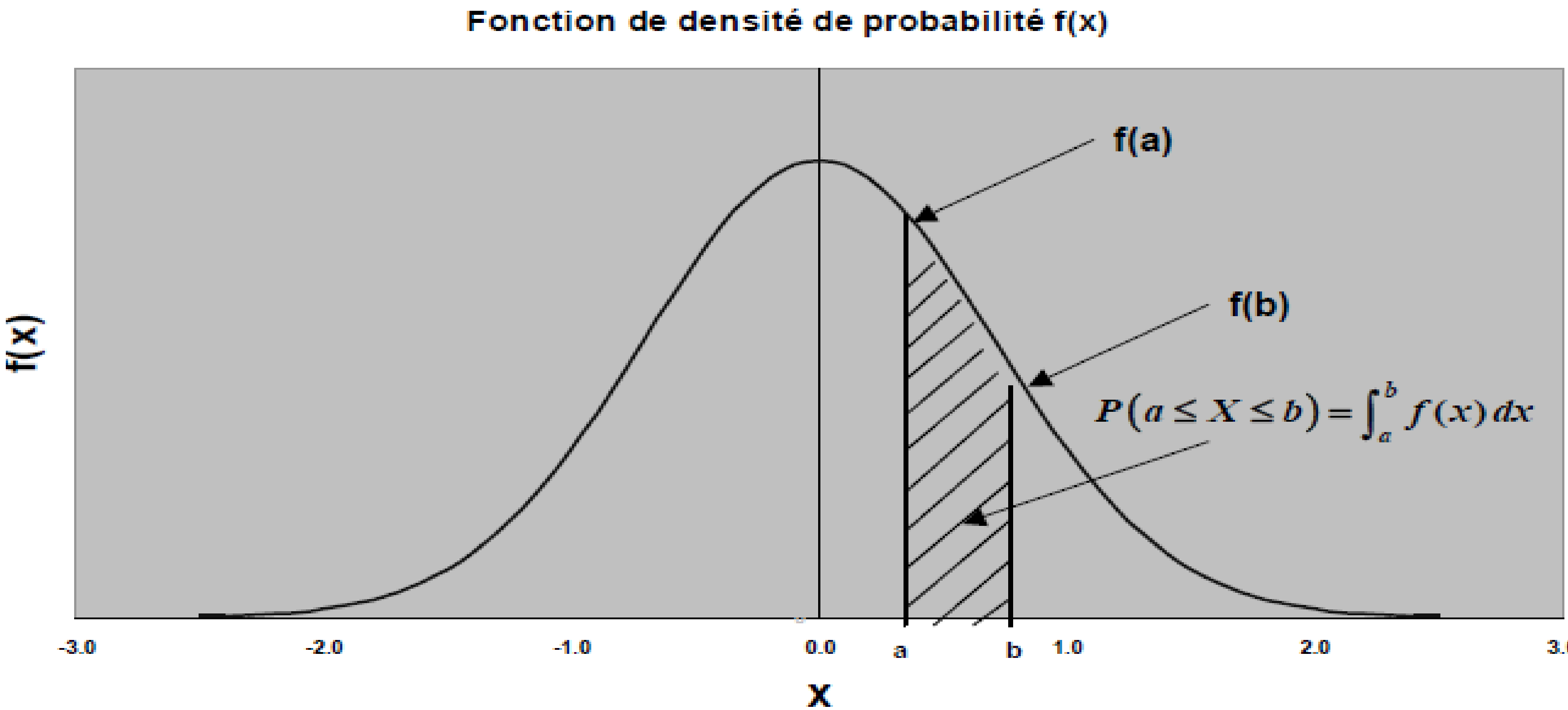
$$(1)\ f(x) \geq 0 \quad \forall x$$

$$(2) \int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$(3)\ P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$$

Rappels de quelques notions de probabilité et d'analyse statistique

Illustration de la fonction de densité de probabilité



Rappels de quelques notions de probabilité et d'analyse statistique

Fonction de la distribution cumulative :

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

Relation entre f(x) et F(x) :

$$f(x) = \frac{dF(x)}{dx}$$

Propriétés de F(x) :

$$\boxed{\begin{array}{l} 0 \leq F(x) \leq 1 \\ F(-\infty) = 0 \quad \text{et} \quad F(\infty) = 1 \\ P(a \leq X \leq b) = F(b) - F(a) \end{array}}$$

Remarque

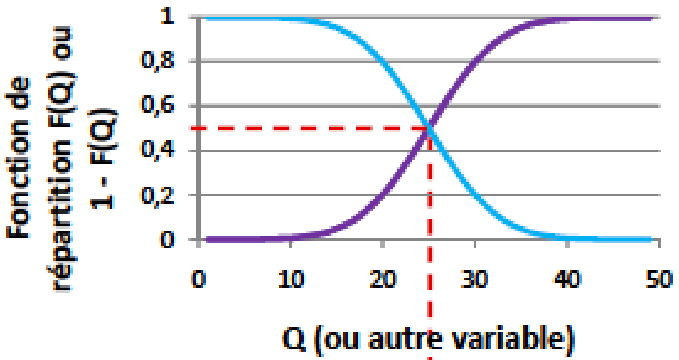
En génie, beaucoup de variables sont bornées à gauche, i.e. $\geq$ 0 (le débit, la vitesse du vent, ...). Dès lors,

$$F(x) = \int_{0}^{x} f(x)\,dx \quad \text{et} \quad \int_{0}^{\infty} f(x)\,dx = 1$$

Rappels de quelques notions de probabilité et d'analyse statistique

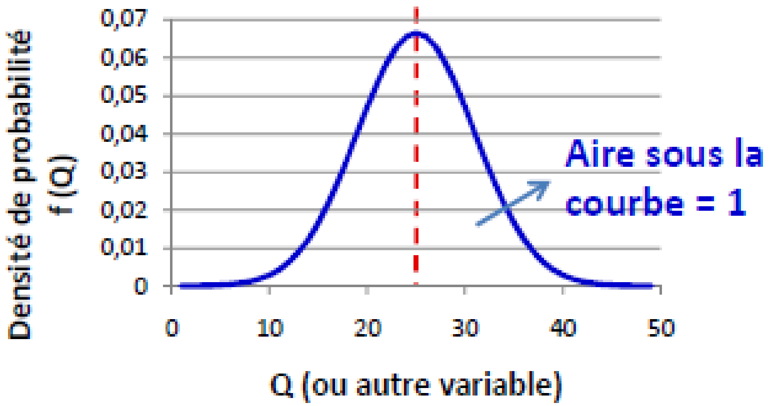
- Fonction de répartition (probabilités au non-dépassement théoriques)

  $F(x)$ et $1 - F(x)$ (dépassement)



- Fonction de densité de probabilité

$$f(x) = \frac{d}{dx}F(x) \text{ et donc } F(x) = \int_{-\infty}^{x} f(x)\,dx$$



Rappels de quelques notions de probabilité et d'analyse statistique

1.Les données hydrologiques doivent être traitées statistiquement selon trois étapes principales :

1.1. Description

Il est de pratique de condenser l'information hydrologique s'étendant sur de nombreuses années à une station fixe et de la remplacer par quelques caractéristiques bien choisies, à condition, toutefois, que ces dernières représentent la série chronologique de manière quasi exhaustive.

La statistique descriptive s'adapte parfaitement à ce type de problème. Elle définit certains paramètres types analysant fidèlement le phénomène à étudier.

Rappels de quelques notions de probabilité et d’analyse statistique

La série d'observations définie sera classée en ordre et décrite par trois types de paramètres :

- valeurs centrales
- paramètres de dispersion
- caractéristiques de forme des courbes de fréquence : histogrammes

1.2. L'analyse

Cette étape consiste à **formaliser** les données expérimentales par une expression mathématique tenant compte des valeurs types calculées à la première étape.

Rappels de quelques notions de probabilité et d'analyse statistique

Le problème à résoudre sera de choisir le modèle probabiliste adéquat qui représente au mieux la série expérimentale. On appelle cette phase la recherche d'un **ajustement théorique** adéquat.

1.3. La prévision

Dans cette étape, l'ingénieur aura à projeter dans l'avenir le modèle choisi pour pouvoir organiser l'avenir de la façon la plus avantageuse et pour pouvoir prendre des décisions optimales et sécuritaires.

Rappels de quelques notions de probabilité et d’analyse statistique

Remarque :

Ces 3 phases de l'analyse statistique supposent que les données utilisées sont homogènes, indépendantes et qu'elles ont subi la phase de **contrôle et de critique**, phase ultime pour la fiabilité des résultats qui est en fonction de la qualité des données.

Dans les exposés des méthodes statistiques qui suivront, nous supposerons que l'indépendance et l'homogénéité existent soit de prime d'abord, soit après correction des données .

Rappels de quelques notions de probabilité et d'analyse statistique

2. Probabilité d'apparition d'une valeur - Intervalle moyen de récurrence

Pour le dimensionnement des structures conditionnées par un phénomène naturel, il est important que l'ingénieur connaisse la probabilité d'apparition de ce phénomène et son importance afin d'établir des critères de dimensionnement adéquats. Des considérations socioéconomiques ont établi certaines règles (règles de I'Art) qui font que l'ingénieur assurera des capacités hydrauliques variables, selon le dommage qui peut résulter d'un événement dépassant cette capacité.

Fonctions de densité des probabilités et de Distribution

Un déversoir de barrage sera conçu pour évacuer une crue d'une **fréquence probable de 0.001** à cause des dommages à la propriété et des pertes de vie qui résulteraient d'une crue supérieure à la capacité du déversoir (danger pour la stabilité du barrage lui-même).

Par contre un égout pluvial peut être insuffisant de temps à autre, créant certains dommages à la propriété, dommages d'un montant inférieur toutefois au capital qu'il aurait fallu investir pour construire l'égout plus gros et éviter l'inondation. **La capacité de l'égout pluvial sera basée sur un débit de fréquence probable plus forte que le déversoir, de l'ordre de 0.2.**

Fonctions de densité des probabilités et de Distribution

On définit l'intervalle moyen de récurrence, dit période de retour, par l'équation :

$$T = 1/P$$

P : probabilité qu'une valeur au moins égale à une valeur donnée se produise. La probabilité est exprimée en fraction.

Un débit d'inondation dont la probabilité **d'apparition ou de dépassement** est 0.033 est appelé crue de 30 ans (**T=1/0.033=30 ans**) car la probabilité est établie à l'aide des crues annuelles. L'unité de temps de T est la même que celle de la variable qui a servi à déterminer la probabilité.

Fonctions de densité des probabilités et de Distribution

Il ne faut pas conclure qu'un débit de 30 ans se produira à intervalles fixes de 30 ans ou que, s'étant produit une fois, il ne se produira plus pendant 30 ans. On doit comprendre plutôt que sur une longue période, 300 ans par exemple, 10 crues au moins égales à ce débit se produiront. <u>Autrement, on peut dire qu'à chaque année, il ya 3.3% de chance qu'un tel débit soit atteint.</u>

Les intervalles de récurrence recommandés pour le dimensionnement de certaines structures sont comme suit :

a- déversoirs de barrages où un crue dépassant la capacité peut mettre le barrage en danger et créer des dommages considérables et des pertes de vie : 500 à 1000 ans ou 10 000 ans.

Fonctions de densité des probabilités et de Distribution

b- ponts sur routes importantes où l'exhaussement de l'eau crée par le pont peut entraîner des dommages importants ou la perte du pont : 50 à 100 ans.

c- ponts sur routes secondaires ou ponceaux sur routes importantes : 25 ans.

d- ponceaux sur routes secondaires, égouts pluviaux, fossés de drainage : 5 à 10 ans.

e- égouts pluviaux de moindre importance : 1 à 2 ans.

Le calcul des probabilités permet de calculer quelles sont les chances de non apparition d'une valeur égale ou supérieure à X connaissant sa probabilité d'apparition au cours de l'année calculé par : **p = 1/T.**

La probabilité de non apparition se calcule par : **q = 1- P =1-1/T**

Fonctions de densité des probabilités et de Distribution

3. Distribution d'une série statistique

Un échantillon hydrologique formé de n valeurs constitue une série continue dont l'effectif ou la fréquence d'apparition varie avec la valeur de la variable. On peut construire avec cet échantillon un histogramme ou **polygone des fréquences d'apparition, une courbe des valeurs classées, une courbe de distribution des fréquences.**

a- Polygone des fréquences d'apparition

Soit un échantillon de N observations décrivant une variable aléatoire X (x_1, x_2,.....x_N).

Fonctions de densité des probabilités et de Distribution

La construction d'un histogramme des fréquences de la variable X consiste à graduer l'axe des abscisses en valeur croissante de la variable étudiée et découpée en **intervalles de classes.** On porte alors en ordonnée le nombre d'apparitions constatées dans chaque intervalle. On obtient ainsi un graphique en "**escalier**".

b- Courbe des valeurs classées

Ces courbes sont obtenues en portant :

en ordonnée : les valeurs observées, classées en ordre décroissant

en abscisse : la fréquence d'apparition de l'ensemble des valeurs supérieures à la valeur portée en ordonnée.

Fonctions de densité des probabilités et de Distribution

Ce sont des courbes qui permettent de donner le pourcentage de probabilité où une valeur observée a été égalée ou dépassée. Elles ont généralement l'allure d'un S horizontale.

c- La courbe de distribution des fréquences

On peut construire soit la distribution des fréquences cumulées au non dépassement soit la distribution des fréquences cumulées au dépassement.

Si on calcule les fréquences cumulatives de toutes **les valeurs inférieures** ou égales à une valeur donnée xi, on obtient la fréquence cumulée au **non dépassement** de cette valeur. Le classement, dans ce cas, des valeurs observées, doit être fait en ordre croissant.

Fonctions de densité des probabilités et de Distribution

Inversement, le classement des données de l'échantillon par ordre décroissant et le calcul de la fréquence cumulée de toutes **les valeurs supérieures** à une valeur donnée x_i, permet de tracer la graphique des fréquences cumulées **au dépassement.**

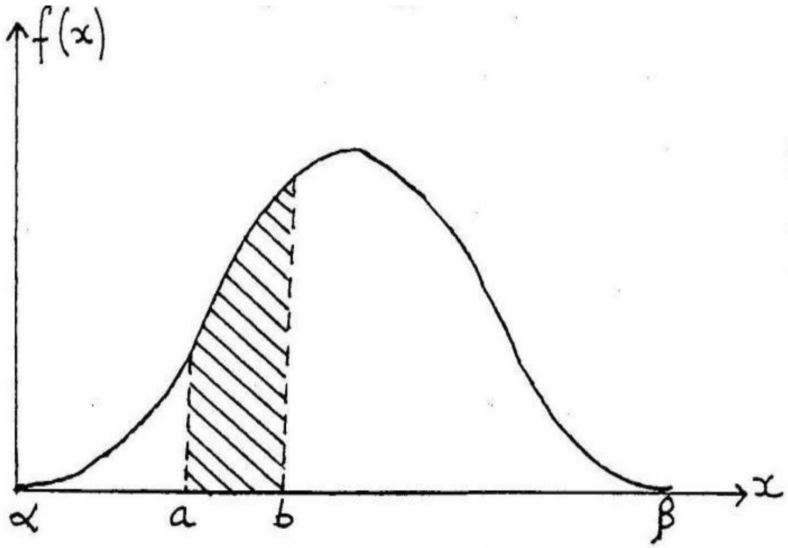
d- Fonction de densité de probabilité - Fonction de répartition

Si la taille de l'échantillon devient grande et l'intervalle de classe tend vers zéro, le polygone des fréquences relatives sera décrit par une courbe à laquelle correspond une certaine fonction de distribution continue f(x) dite **fonction de densité de probabilité,** notée **fdp.** Ainsi l'effectif ou la fréquence d'apparition d'une valeur x_i deviendra, la densité de probabilité $f(x_i)$.

Fonctions de densité des probabilités et de Distribution

Soit la courbe décrivant les fréquences d'apparition des événements d'une variable X.

Fonction de distribution



Si P(b) est la probabilité d'obtenir x < b et si P(a) est la probabilité d'obtenir x > a, alors la probabilité d'avoir a < x < b est :

Prob (a ≤ x ≤ b) = P(b) - P(a)

L'on peut démontrer que la probabilité d'avoir a < x < b

$$\operatorname{Prob}(a \leq x \leq b) = P(b) - P(a) = \int_a^b f(x)dx$$

C'est-à-dire égale à la surface hachurée sous la courbe.

Fonctions de densité des probabilités et de Distribution

Si α est la plus petite valeur possible de x et β la plus grande, on aura par définition :

$$\int_a^x f(x)dx = P(x) \qquad \text{et} \qquad \int_\alpha^\beta f(x)dx = 1$$

La dernière expression est évidente, puisqu'elle exprime que la probabilité pour que x soit situé entre les deux extrêmes possibles, est égale à 1 (ou 100% de chances).

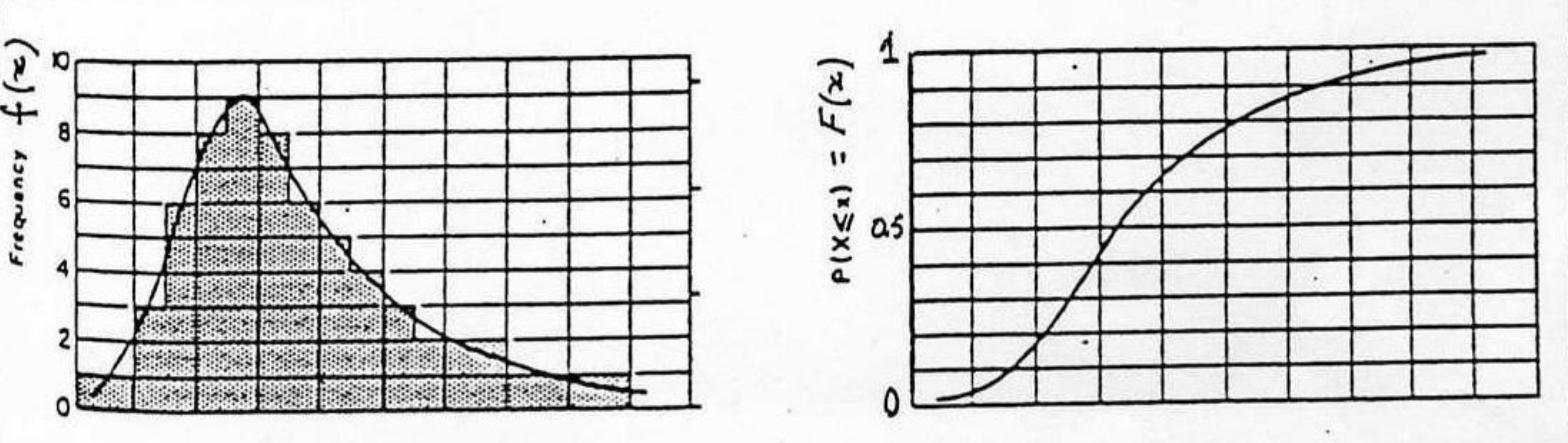
Elle exprime aussi que la totalité de l'aire sous la courbe de la fonction de distribution est égale à l'unité.

Une fonction de distribution f(x) est caractérisée par sa moyenne μ et son écart type σ.

De même, la limite du polygone des fréquences cumulées définit une fonction de répartition appelée loi de répartition de la population, notée F(x) telle que :

Fonctions de densité des probabilités et de Distribution

$$\frac{dF(x)}{dx} = f(x)$$



Fonction de répartition

Soit un échantillon X de valeurs observées (x_1, ...x_N), classées en ordre croissant.

$$F(x_i) = P(X \leq x_i) = \sum_{x_k \leq x_i} f(x_k)$$

Fonctions de densité des probabilités et de Distribution

C'est la probabilité qu'une valeur de la variable X soit inférieure ou égale à la valeur x_i.

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

Cette fonction est connue aussi sous le nom de la fonction de répartition au **Non Dépassement.** Elle prend des valeurs de 0 à 1.

Par opposition, le complément de cette fonction s'appelle la fonction de répartition au **Dépassement** et se définit par :

$$F_1(x_i) = P(X > x_i) = 1 - F(x_i)$$

Le calcul au dépassement nécessite le classement de l'échantillon en ordre décroissant.

Fonctions de densité des probabilités et de Distribution

Les lois mathématiques de distribution de probabilité peuvent être comparées à la distribution de l'échantillon disponible. Si la loi mathématique s'applique à l'échantillon, on peut déclarer que cette loi s'applique au phénomène et en déduire les probabilités de toutes les valeurs que peut prendre la variable étudiée.

Les échantillons étant petits en hydrologie, il sera parfois difficile de déclarer si une loi de probabilité s'applique d'une façon absolue à un phénomène. On devra souvent se contenter d'utiliser la loi qui semble le mieux s'appliquer au phénomène étudié, représentée d'une façon plus ou moins parfaite par l'échantillon disponible.

Type de fonction de densité des probabilités

5- Lois de distribution utilisées en hydrologie

5-1/ Présentation des lois

Vu la mauvaise qualité des données (échantillons trop réduits en particulier), l'hydrologue ne peut se montrer très difficile sur les lois théoriques qu'il utilise : aussi se contente-il d'un nombre limité de lois :

- **La loi Normale ou de Gauss** est utilisée pour des échantillons constitués de moyennes (en effet, la somme de variable aléatoires quelconques tend à devenir normale si le nombre de variables pris en compte augmente) : distribution de modules, de débits ou apports mensuels, de pluie annuelle ou mensuelle.

Type de fonction de densité des probabilités

- **La loi de Gumbel** est la seconde loi de base : sa dissymétrie constante semble très bien adaptée aux événements extrêmes habituellement rencontrés : maxima annuels de précipitations journalières en particulier, débits de pointe de crues etc.

- **La loi de Pearson III** complète ces deux lois fondamentales : à faible coefficient de variation, elle tend vers la loi de Gauss, à coefficient de variation plus fort, elle se rapproche de la loi de Gumbel.

- On utilise également **les lois de Galton,** où le logarithme de la variable suit une loi de Gauss et la loi de Fréchet, où le logarithme de la variable suit une loi de Gumbel.

Type de fonction de densité des probabilités

Ces lois présentent l'avantage de pouvoir être utilisées comme des lois à trois paramètres, ce qui leur confère une souplesse presque aussi grande que celle de la loi de Pearson III.

Enfin, on trouve dans la littérature d’autres lois, dont l'utilisation, beaucoup moins pratique, n'apporte rien de plus à l'hydrologue.

Nous traiterons, avec des étude de cas détaillées , les lois de Gauss, log normale et Gumbel.

La Loi Normale ou Loi de Gauss

<u>Distribution normale</u> : loi de Gauss, Notée N(μ, σ)

Lorsqu'une variable x subit l'influence de causes nombreuses très petites et indépendantes les unes des autres, les valeurs de cette variable se distribuent suivant une fonction de distribution, dite "Normale". C'est la loi la plus commune et la mieux étudiée. Sa fonction de densité de probabilité est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

dont les deux seuls paramètres sont la moyenne μ et l'écart type σ de la population étudiée.

Dans cette distribution, x peut varier de manière continue de - ∞ à + ∞ et

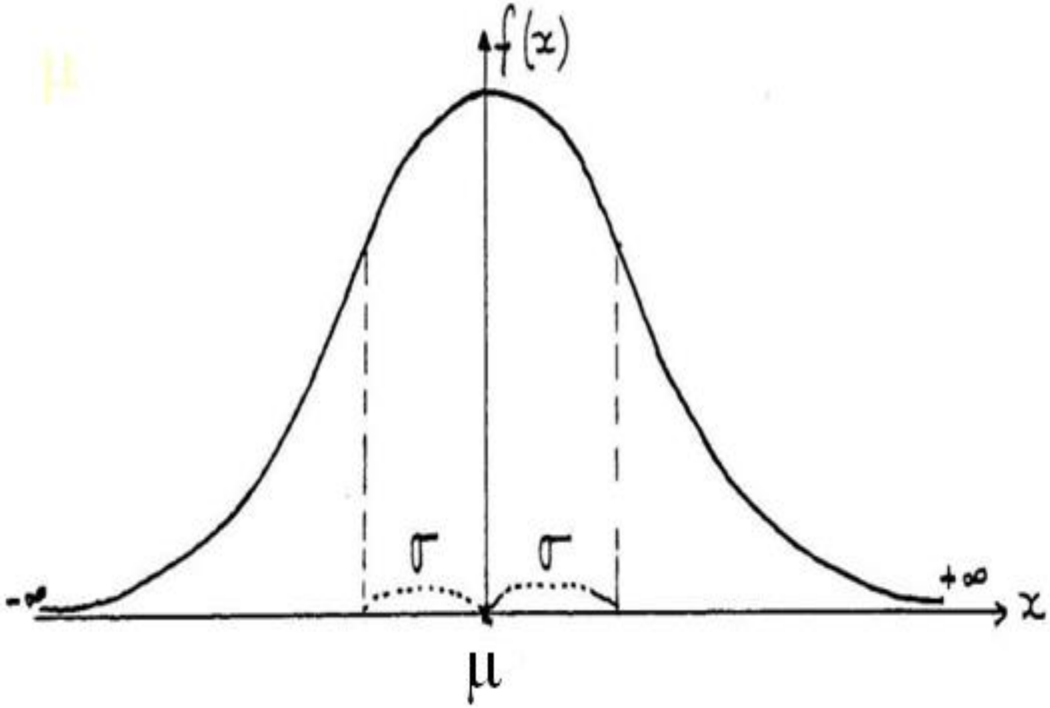
$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

La Loi Normale ou Loi de Gauss

La fonction de distribution ou de répartition est définie par :

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

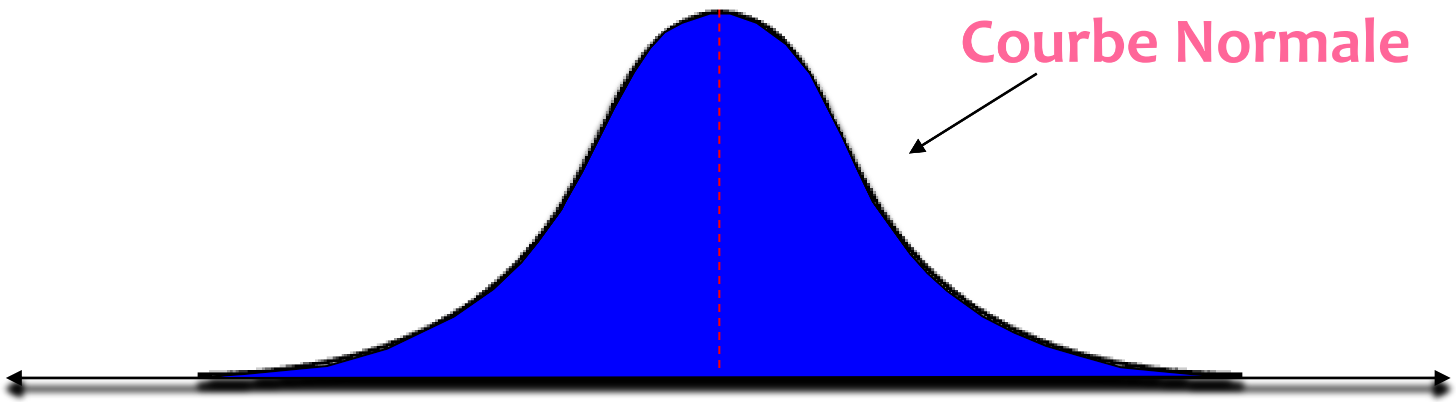
La fonction f(x) se représente par une courbe en forme de cloche, symétrique par rapport à x = μ, admettant un maximum pour x = μ et tendant asymptotiquement vers 0 pour x→-∞ et x→+∞.



Les points d'inflexion se situent à (μ - σ) et (μ +σ).

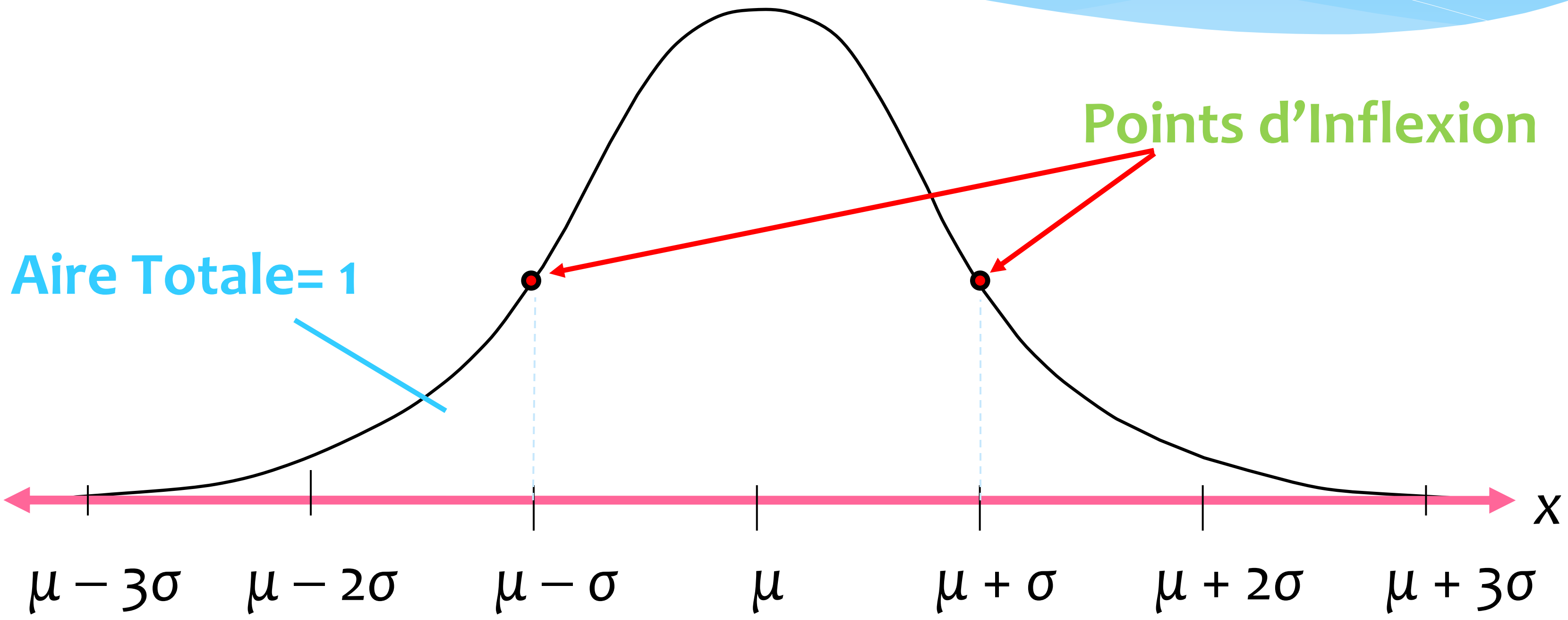
Propriétés de La Distribution Normale

La plus importante distribution des probabilités en statistiques est la **distribution normale**.



Une distribution normale est une distribution de probabilité continue d'une variable aléatoire, X. Le graphique d'une distribution normale est appelée la courbe normale.

Propriétés de La Distribution Normale



Il existe une famille entière de lois normales. Elles se différencient par leur moyenne et leur variance

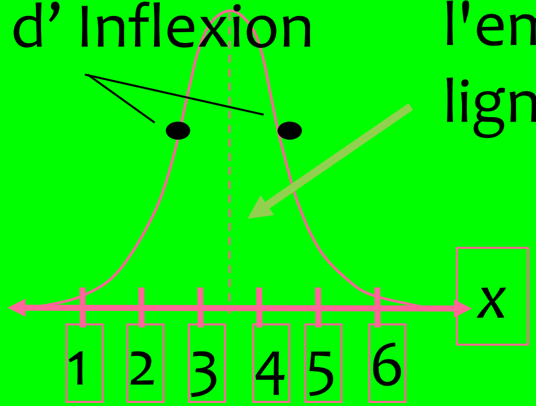
Moyennes et Ecarts Types

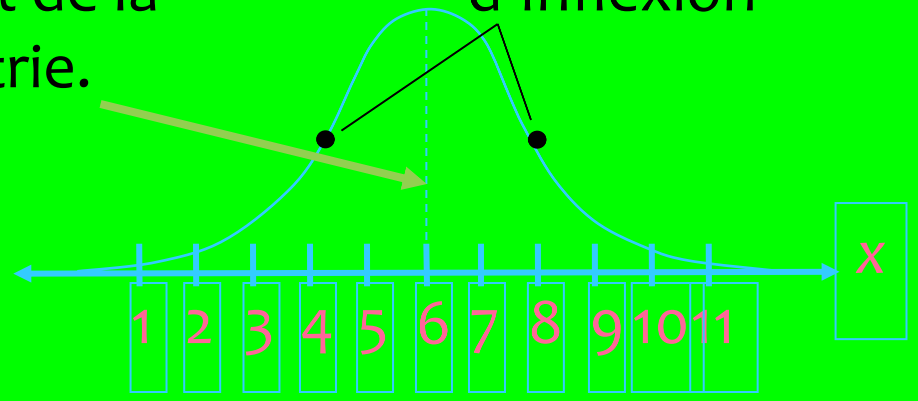
Une distribution normale peut avoir n'importe quelle moyenne et n'importe quel écart type positifs.

Points d' Inflexion

La moyenne donne l'emplacement de la ligne de symétrie.

Points d'Inflexion





Moyenne : μ = 3.5

Ecart-type: $\sigma \approx 1.3$

Moyenne : μ = 6

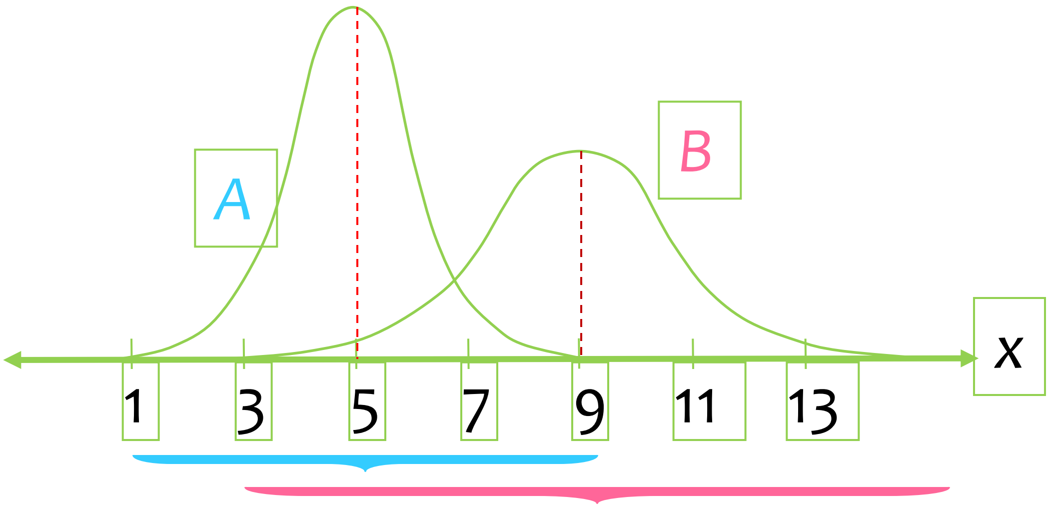
Ecart-type: $\sigma \approx 1.9$

L'écart-type décrit la propagation des données.

Moyennes et Ecarts Types

Exemple:

- Quelle est la courbe qui a la plus grande moyenne?
- Quelle est la courbe qui a le plus grand écart type?

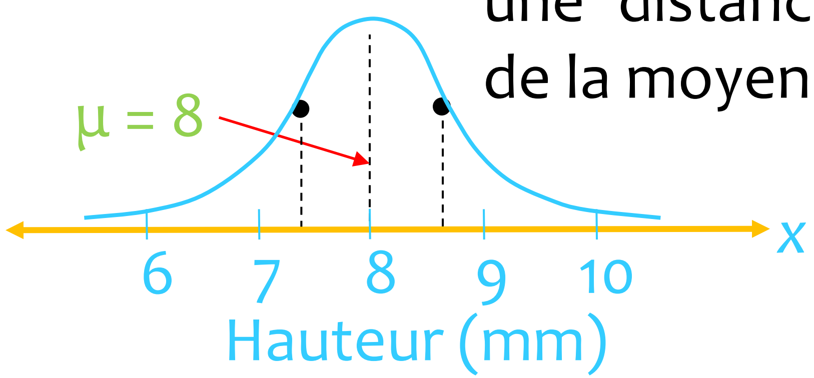


- L'axe de symétrie de la courbe A se produit à x = 5.
- L'axe de symétrie de la courbe B se produit à x = 9.
- La courbe B a la plus grande moyenne.
- La courbe B est plus étendue que la courbe A, de sorte que la courbe B a le plus grand écart-type.

Interprétation de Graphes

Exemple:

Les hauteurs de pluie journalières au printemps sont normalement distribuées. La courbe ci-dessous représente cette distribution. Quelle est la hauteur moyenne des pluies journalières? Estimer l'écart type.



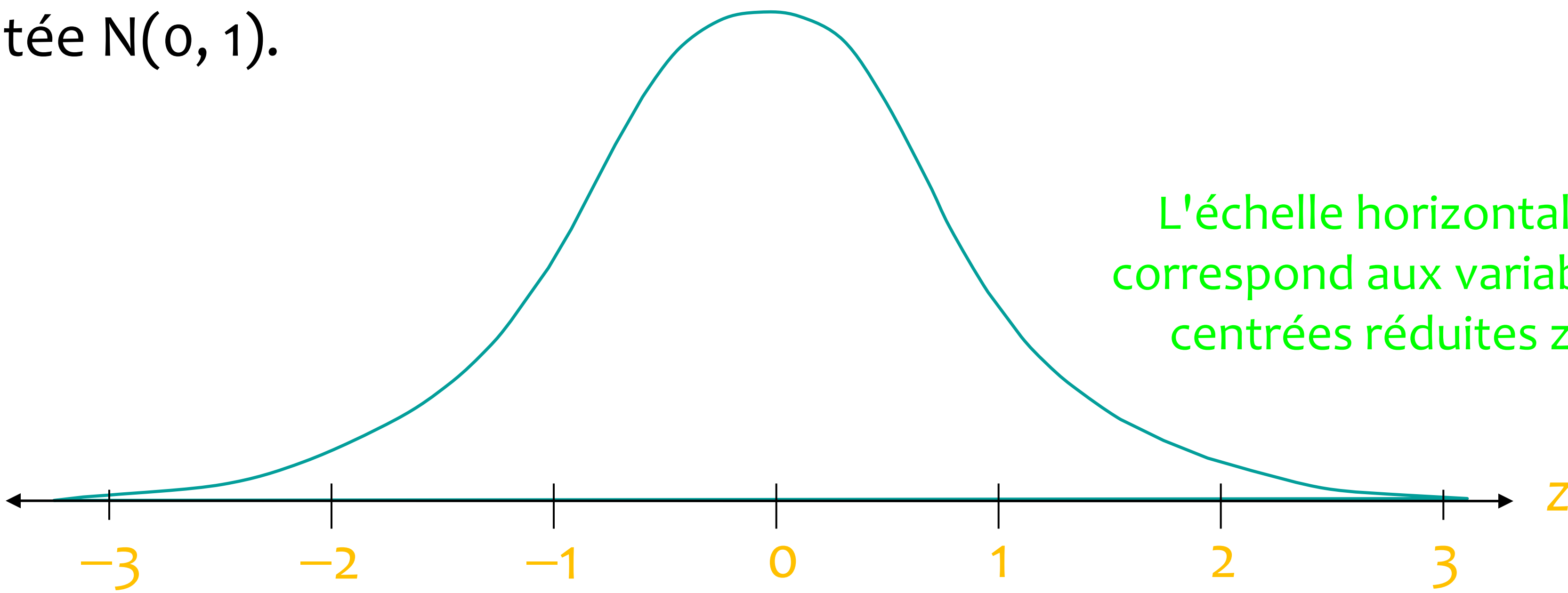
Les points d'inflexion sont à une distance d'un écart type de la moyenne.

$\sigma \approx 0.7$

Les hauteurs de pluies journalières sont normalement distribuées avec une hauteur moyenne d'environ 8 mm et un écart-type d'environ 0,7 mm.

La Distribution Normale Standard

La **Distribution Normale Standard** des probabilités est une distribution normale avec une **moyenne nulle** et un **écart type égal à 1,** notée N(0, 1).



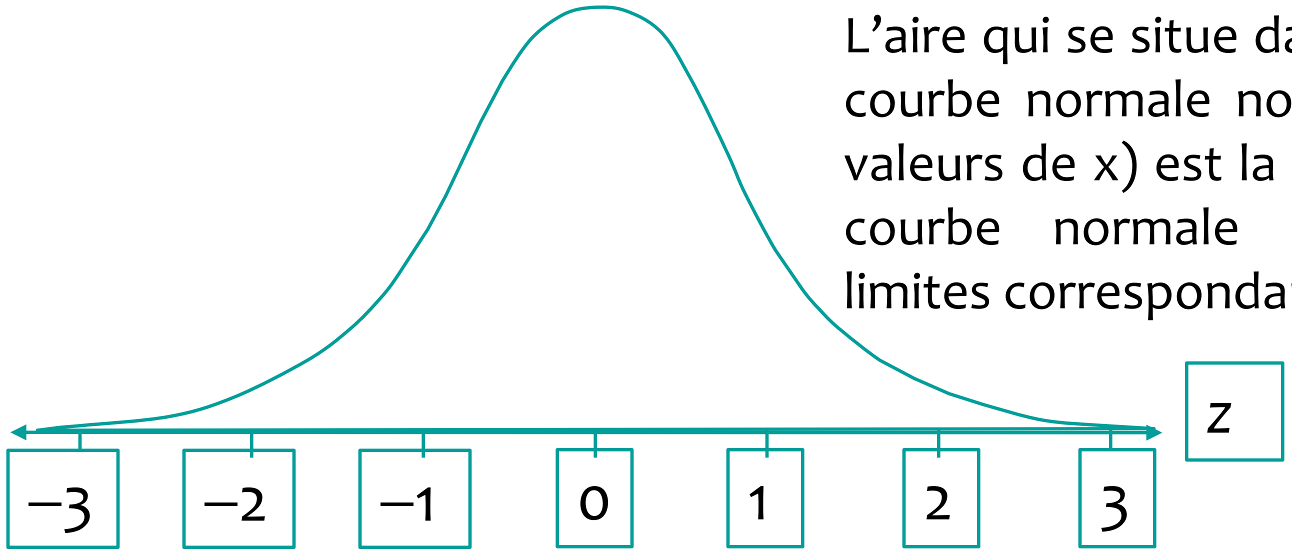
L'échelle horizontale correspond aux variables centrées réduites z.

N'importe quelle valeur observée peut être transformée en une Variable Centrée Réduite (VCR) z en utilisant la formule :

$$z = \frac{\text{Valeur} - \text{Moyenne}}{\text{Ecart type}} = \frac{x - \bar{x}}{s}$$

La distribution Normale Standard

Si chaque valeur d'une variable aléatoire x, distribuée normalement, est transformée en une variable centrée réduite z, le résultat sera la distribution normale standard.



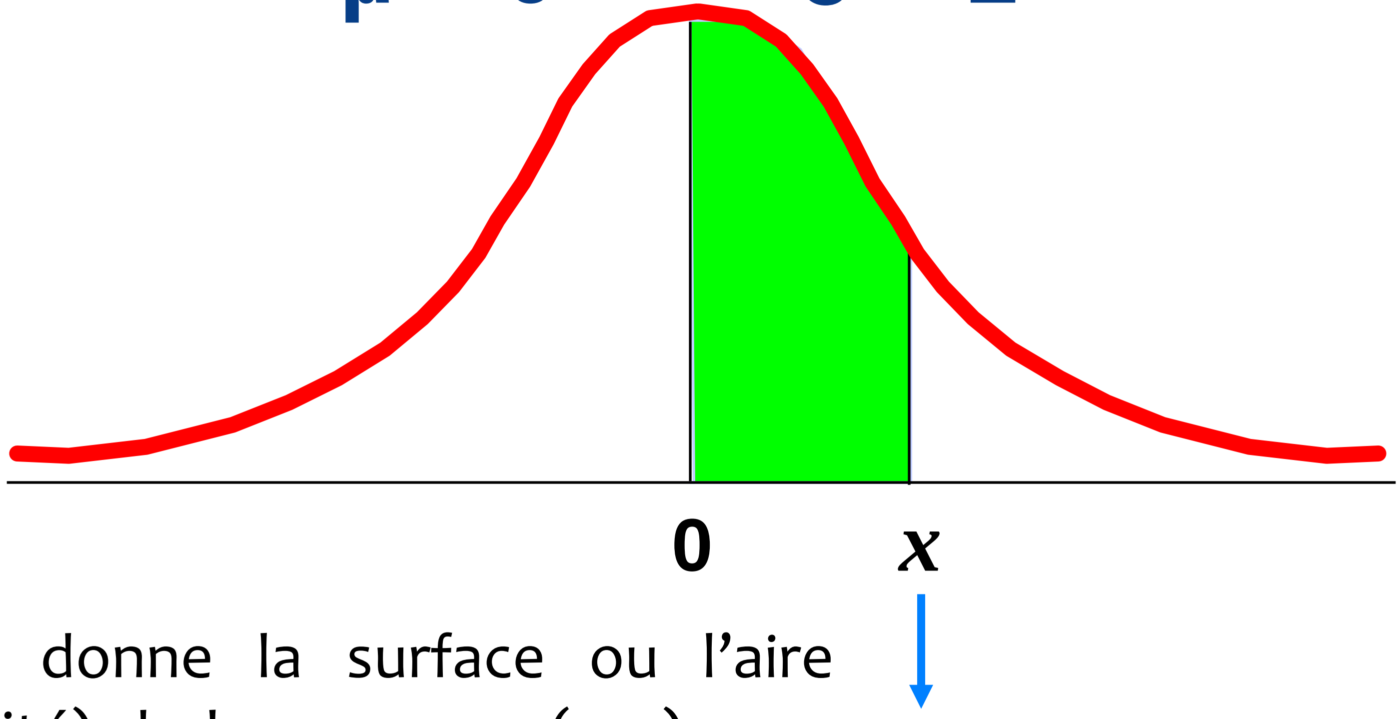
L'aire qui se situe dans l'intervalle sous la courbe normale non standard (pour les valeurs de x) est la même que celle de la courbe normale standard (dans les limites correspondantes à z).

La formule est utilisée pour transformer une valeur de x en une VCR z, alors que la table de la loi normale standard est utilisée pour calculer l'aire cumulée sous la courbe.

La distribution Normale Standard

La Table de la Distribution Normale Standard

$\mu = 0$ $\sigma = 1$



La table donne la surface ou l'aire (probabilité) de la moyenne (= 0) pour la VCR z seulement.

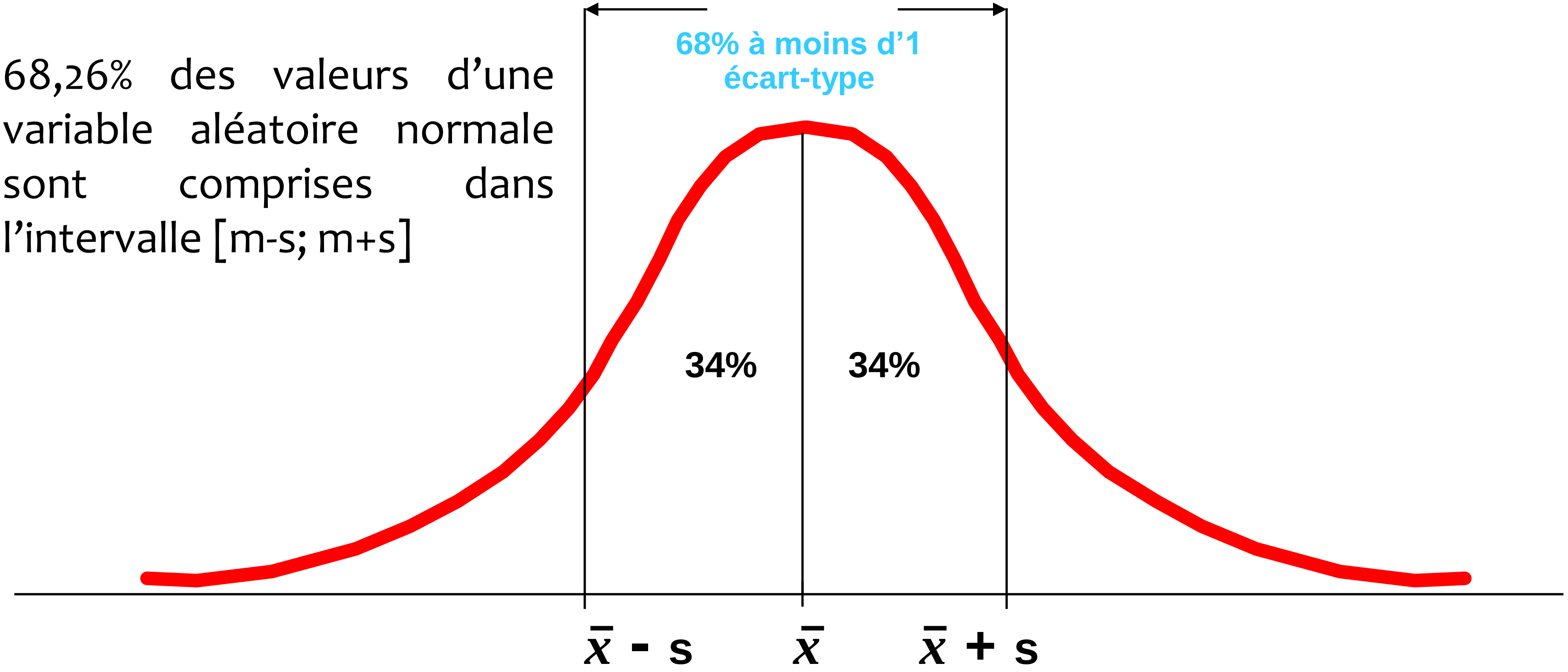
Table de la Distribution Normale Standard (z)

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	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Propriétés de La Distribution Normale

Règle empirique

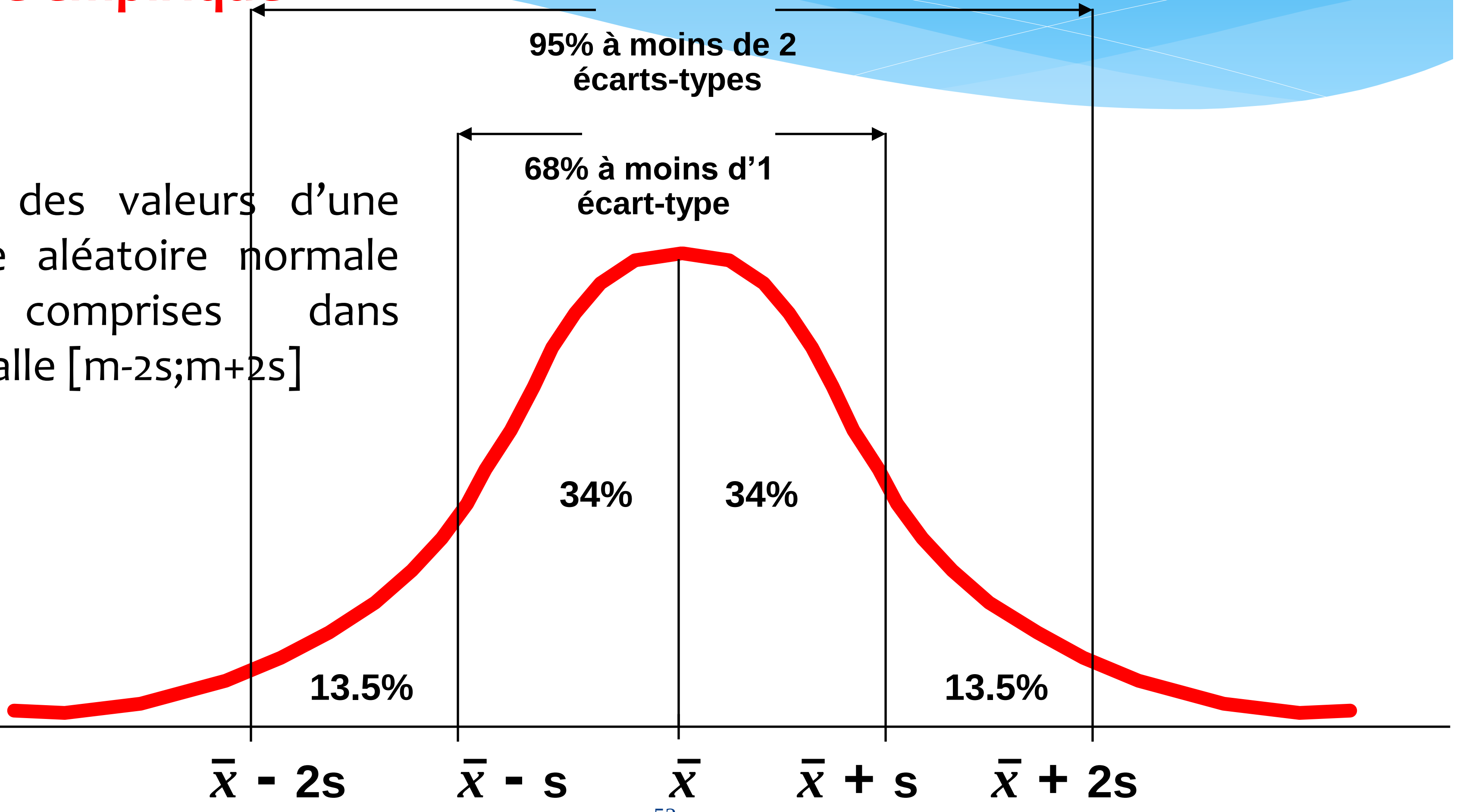
68,26% des valeurs d'une variable aléatoire normale sont comprises dans l'intervalle [m-s; m+s]



Propriétés de La Distribution Normale

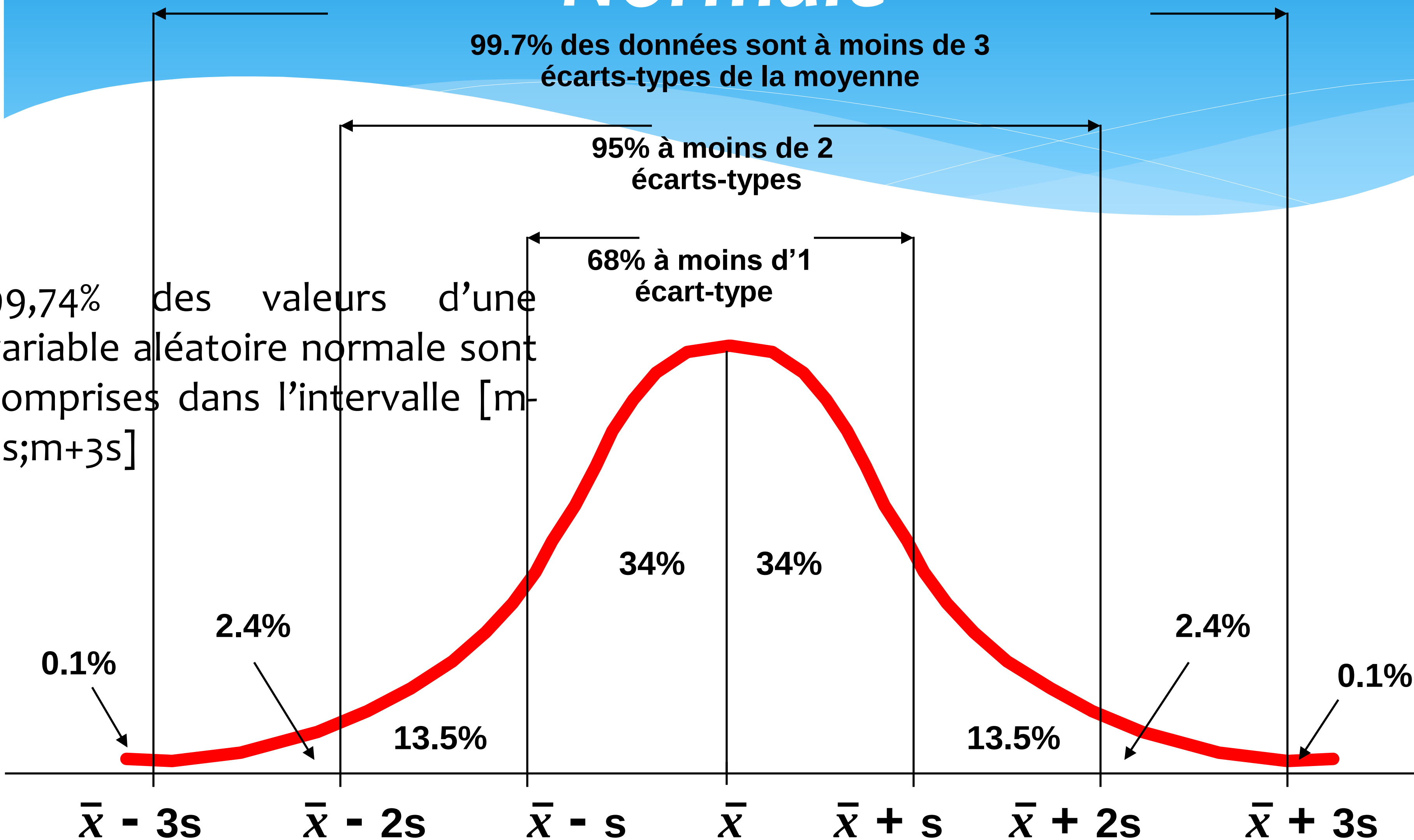
Règle empirique

95,44% des valeurs d'une variable aléatoire normale sont comprises dans l'intervalle [m-2s;m+2s]



Propriétés de La Distribution Normale

99,74% des valeurs d'une variable aléatoire normale sont comprises dans l'intervalle [m-3s;m+3s]



La distribution Normale Standard

Propriétés de la Distribution Normale Standard :

1. La surface cumulée est proche de 0 pour des VCR z proches de z = - 3,49.
2. La surface cumulée augmente à mesure que les VCR z augmentent.
3. La surface cumulée pour z = 0 est 0.5000.
4. La surface cumulée est proche de 1 pour des VCR z proches de z = 3,49.

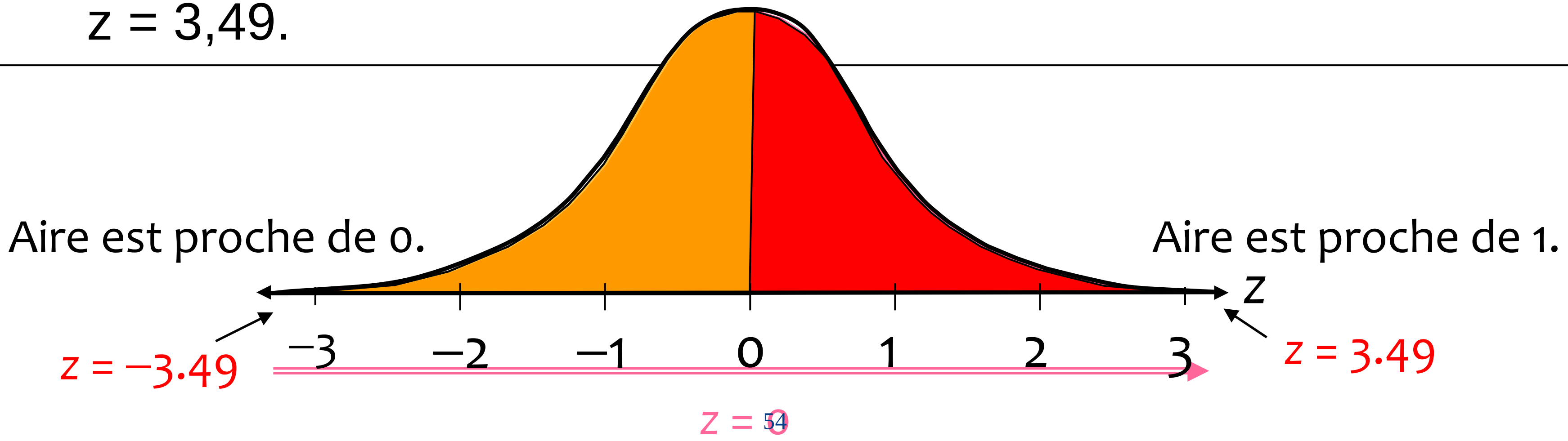


Table de la distribution Normale Standard

Exemple:

Trouvez l'aire cumulée qui correspond à une VCR z = 2,71.

Table de la Distribution Normale Standard

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981

Se déplacer dans la colonne de gauche jusqu'à trouver la valeur de z = 2.7, puis se déplacer en horizontale jusqu'à la colonne 0.01.

L'aire située à gauche de z = 2.71 est 0.9966.

Table de la distribution Normale Standard

Pour déterminer :

- **Coulissez le long de l'échelle horizontale de la table de la distribution normale standard, et reportez-vous à la valeur de la colonne la plus à gauche et à celle de la ligne du haut**

- l'Aire

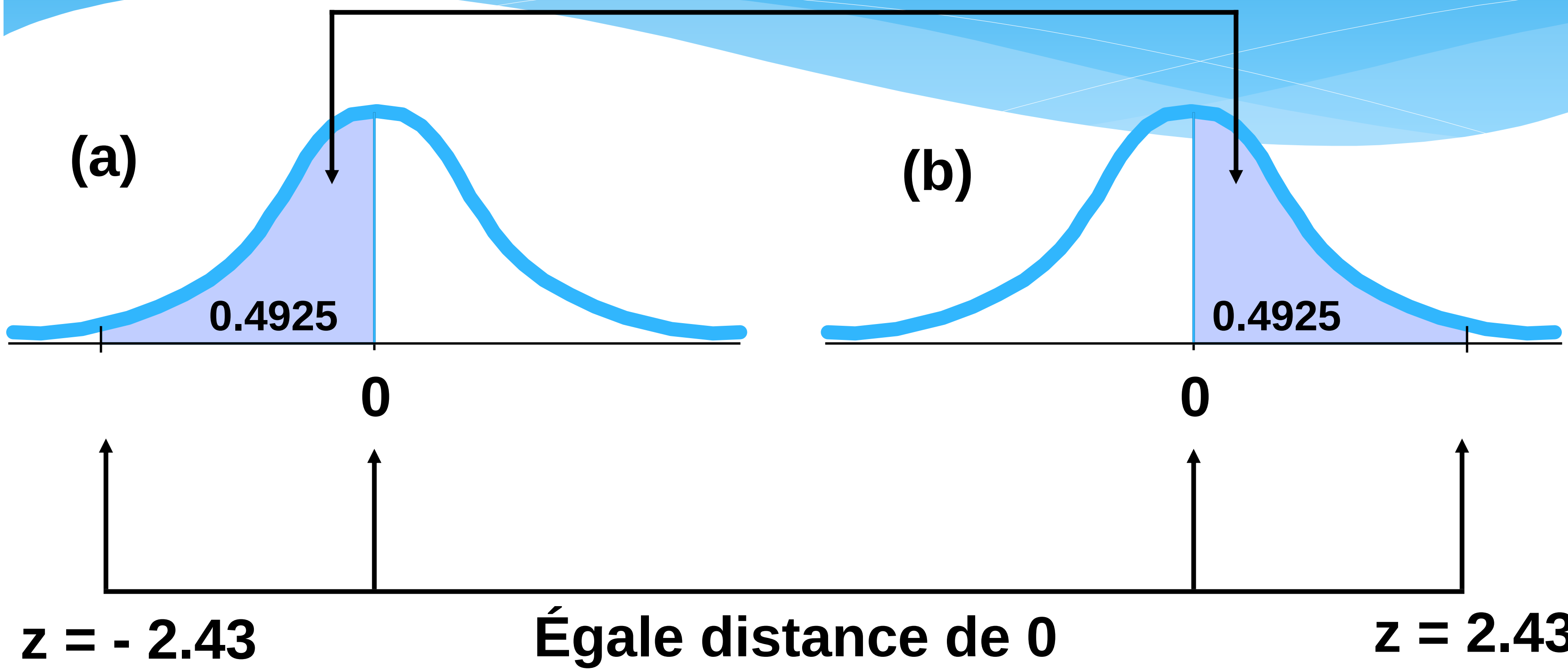
- l'aire sous la courbe, se référer aux valeurs dans la table.

Il sera important de faire la distinction entre la recherche d'une probabilité (l'aire) donnée d'une VCR z particulière de celle de trouver une VCR z donnée d'une probabilité particulière.

Être capable de reconnaître la bonne méthode pour chacun de ces processus sera essentielle.

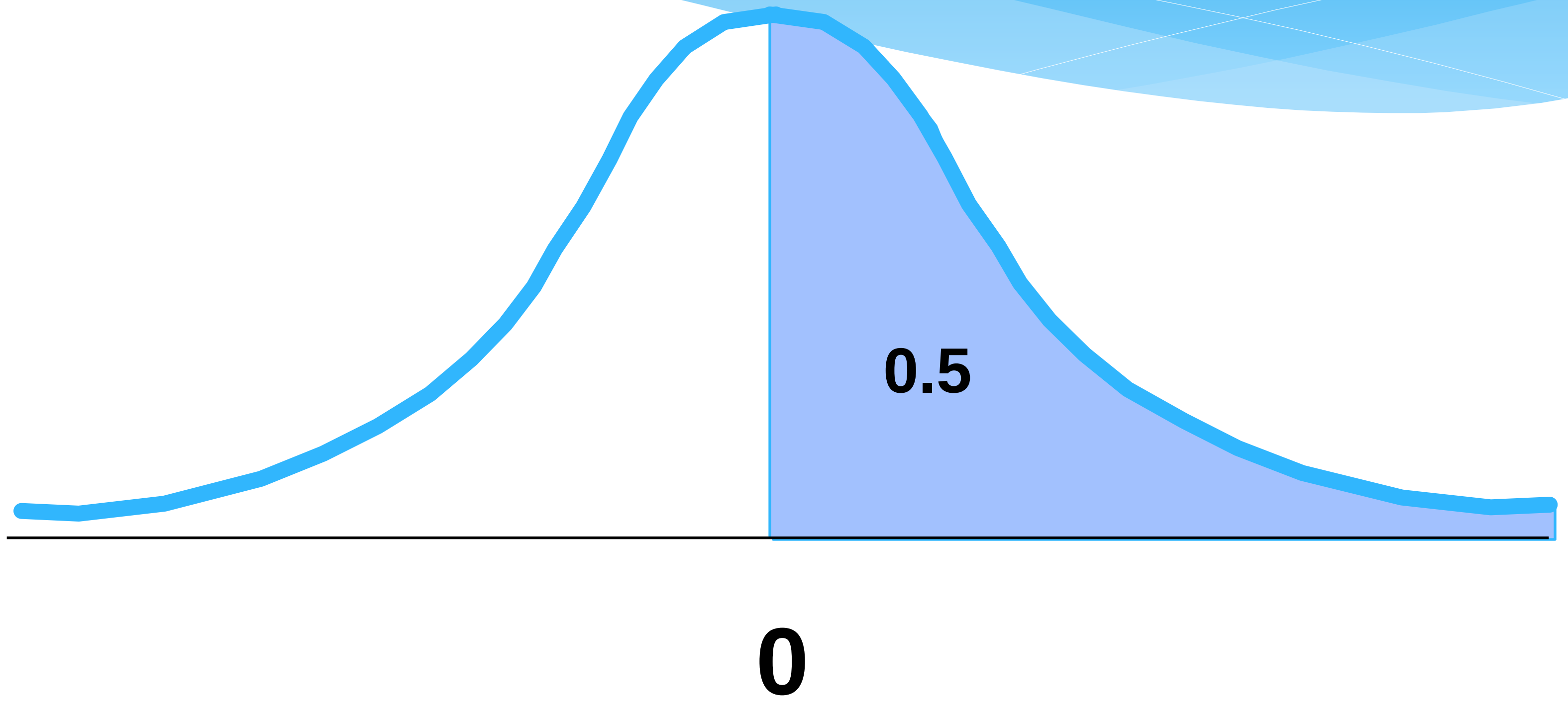
Utiliser la symétrie pour trouver l'aire à gauche de la moyenne

En raison de la symétrie, ces deux aires sont égales.



NOTE: Même si la VCR z peut être négative, l'aire sous la courbe (ou la probabilité correspondante) ne peut jamais l'être.

Probabilité de La moitié d'une distribution



Conversion de Distribution Normale Standard

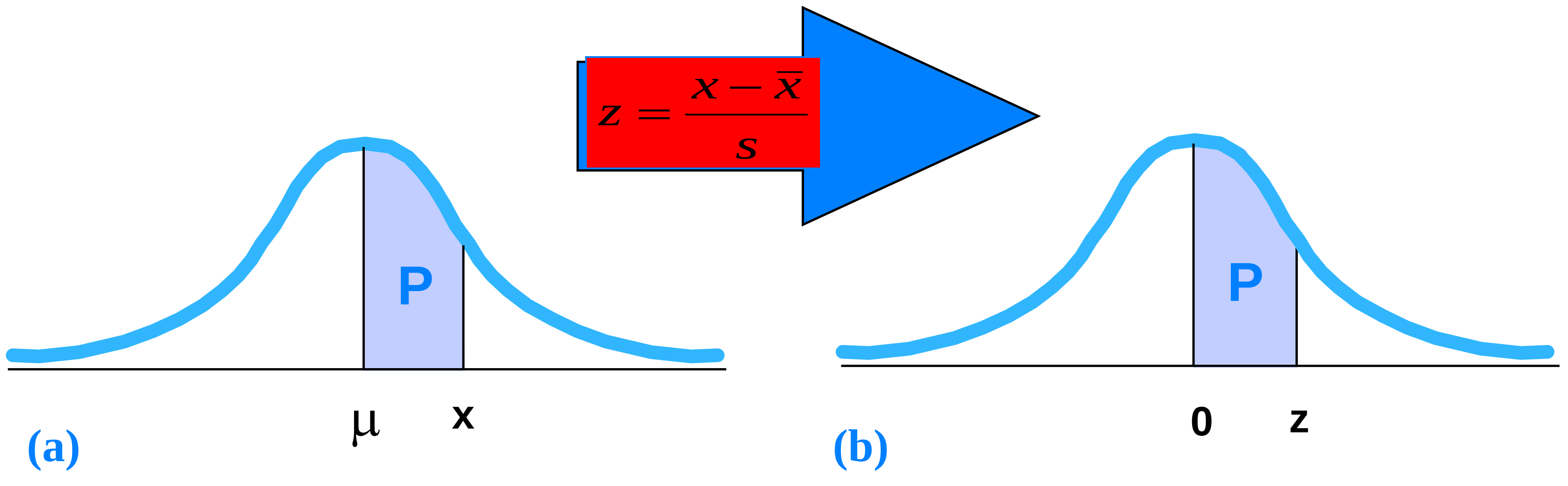


tableau de la loi Normale centrée réduite

Note :

$P(x<a)$ = lecture directe

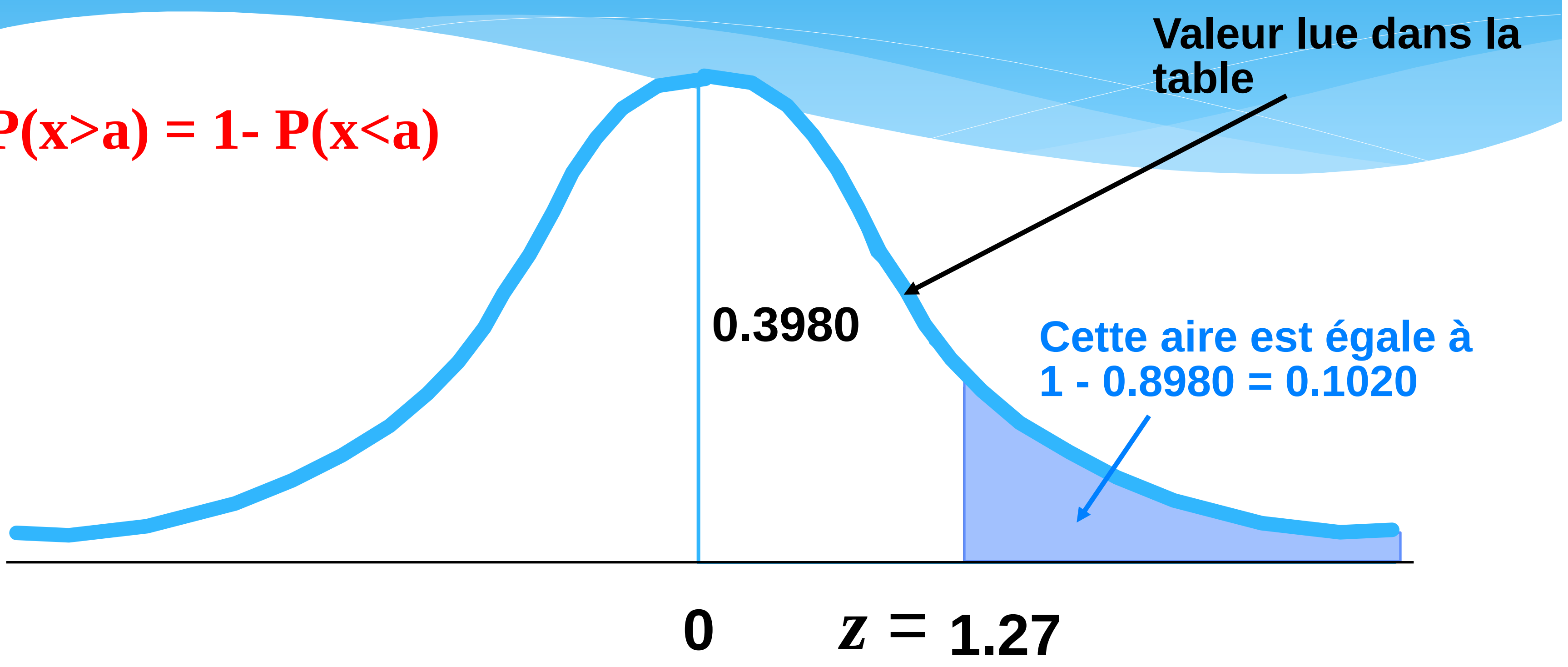
$P(x>a) = 1- P(x<a)$

$P(x<-a) = 1- P(x<a)$

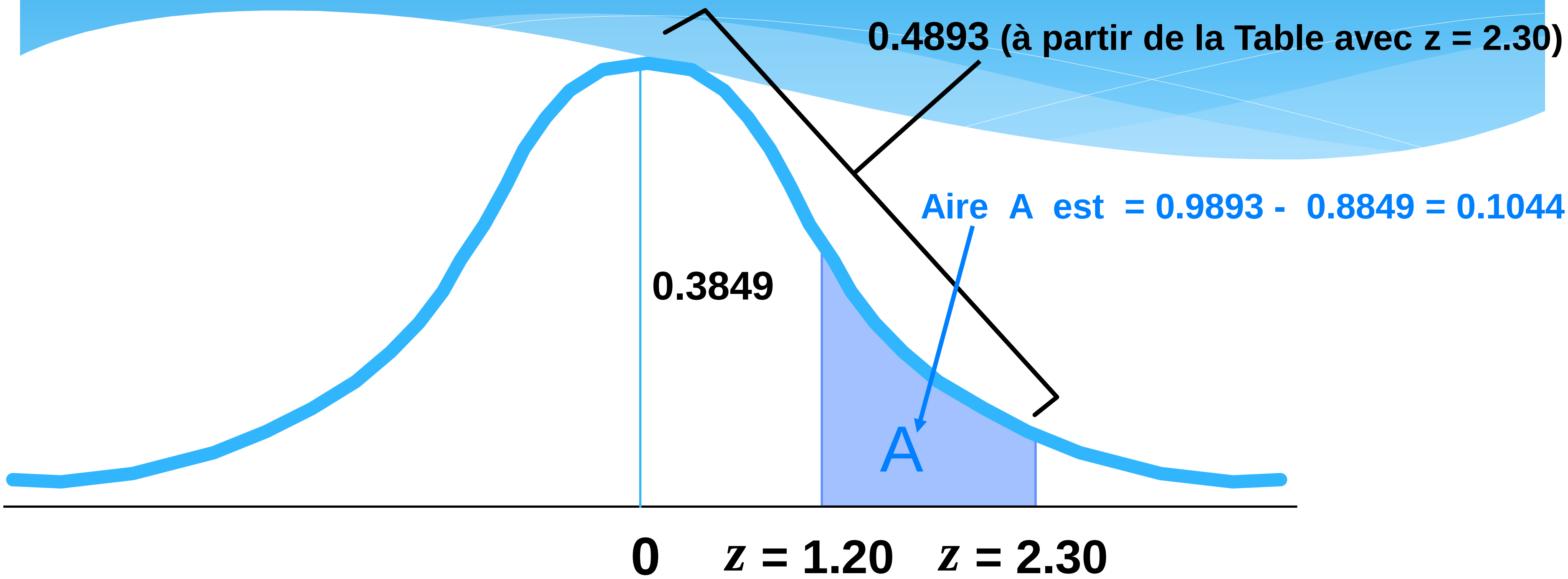
$P(a<x<b) = P(x<b) - P(x<a)$

Déterminer l'Aire à droite de *z* = 1.27

P(x>a) = 1- P(x<a)



Déterminer l'Aire comprise entre *z* = 1.20 et *z* = 2.30

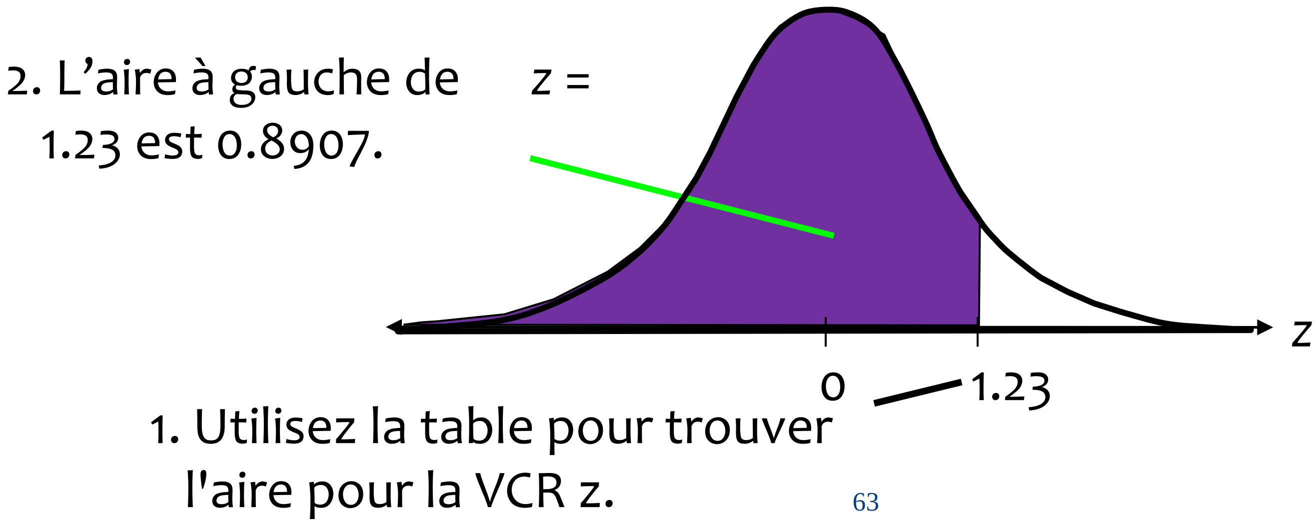


$$P(a<x<b) = P(x<b) - P(x<a)$$

Détermination des Aires

Détermination des aires sous la courbe Normale Standard

1. Dessinez la courbe normale standard et ombragez la zone appropriée sous la courbe.
2. Trouvez l’aire en suivant les indications suivante :
   a. Pour déterminer l’aire située à gauche de z, trouver l'aire qui correspond à z dans la table de la loi normale standard.

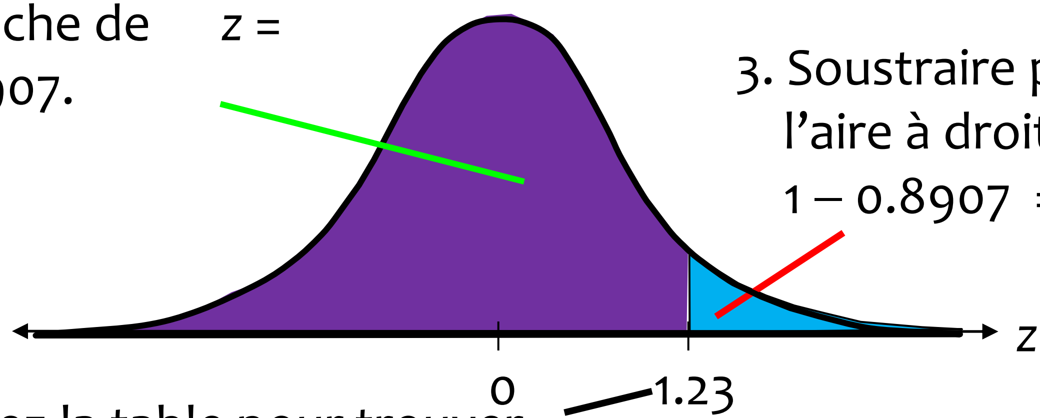


Détermination des Aires

Détermination des aires sous la courbe Normale Standard

b. Pour déterminer l'aire située à droite de z, utilisez la table de la loi normale standard pour trouver l'aire qui correspond à z. Puis soustraire cette aire ou ce résultat de 1.

2. L'aire à gauche de z = 1.23 est 0.8907.



3. Soustraire pour trouver l'aire à droite de z = 1.23:
$1 - 0.8907 = 0.1093$.

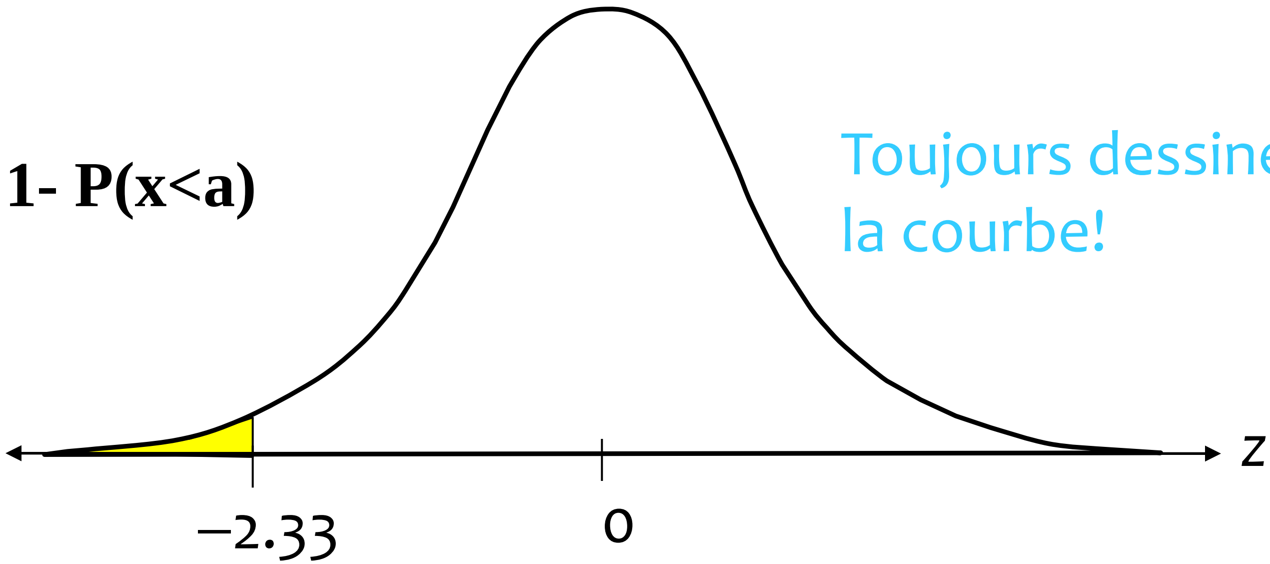
1. Utilisez la table pour trouver l'aire pour la VCR z.

Détermination des Aires

Exemple:

Trouvez l'aire sous la courbe normale standard à gauche de z = −2.33.

P(x<-a) = 1- P(x<a)



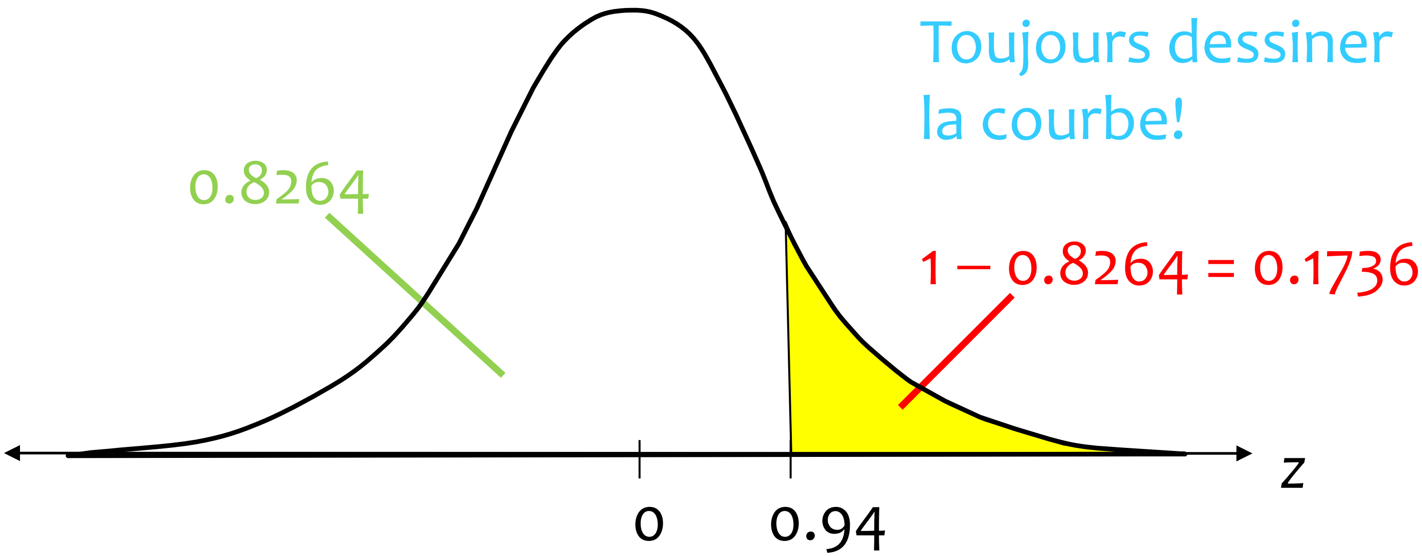
Toujours dessiner la courbe!

De la table de la loi normale standard, l’aire est égale à 0.0099.

Détermination des Aires

Exemple:

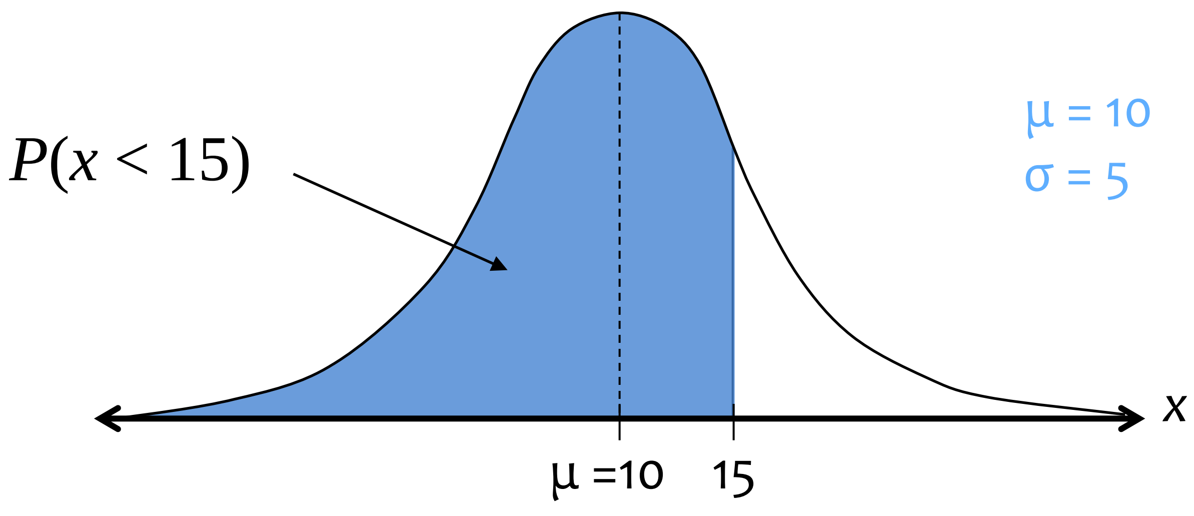
Trouvez l'aire sous la courbe normale standard à droite de $z = 0.94$.



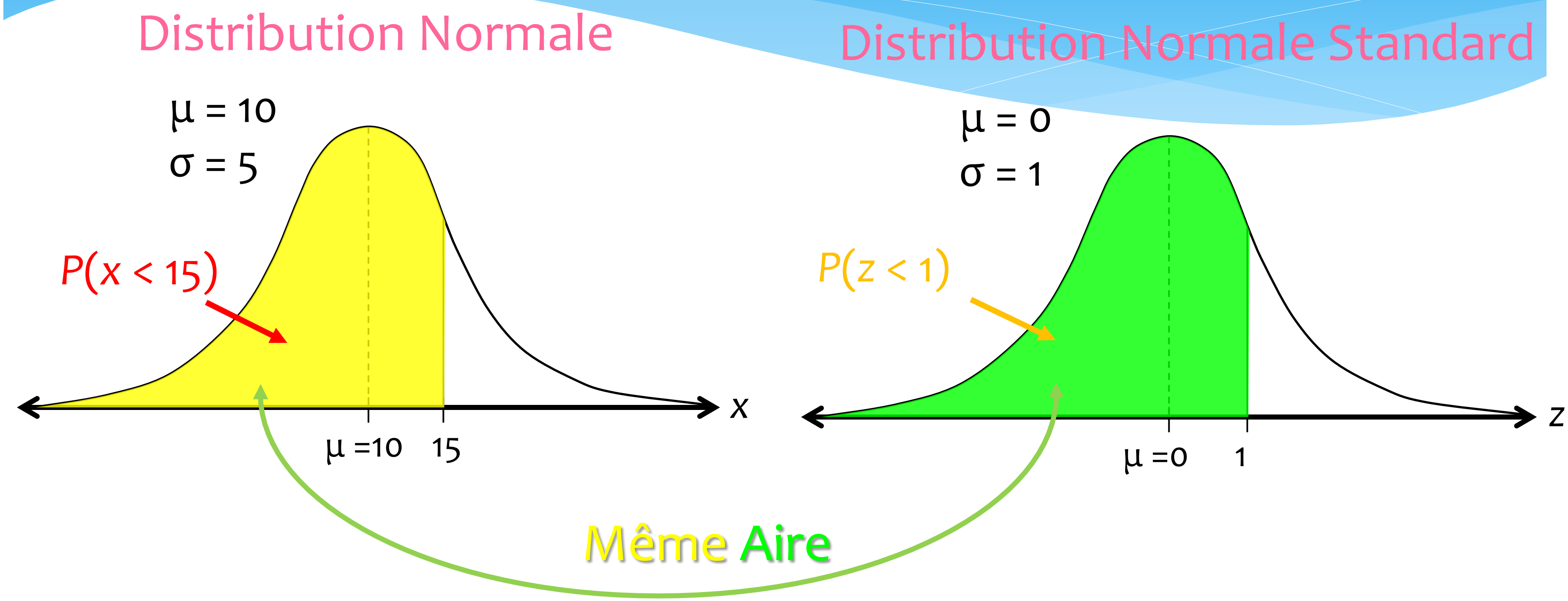
A partir de la table de la loi normale standard, l'aire est égale à 0.1736.

Probabilité et Distributions Normales

Si une variable aléatoire, x, est normalement distribuée, on peut déterminer la probabilité que x va se trouver dans un intervalle donné par le calcul de l'aire sous la courbe normale pour cet intervalle.



Probabilité et Distributions Normales



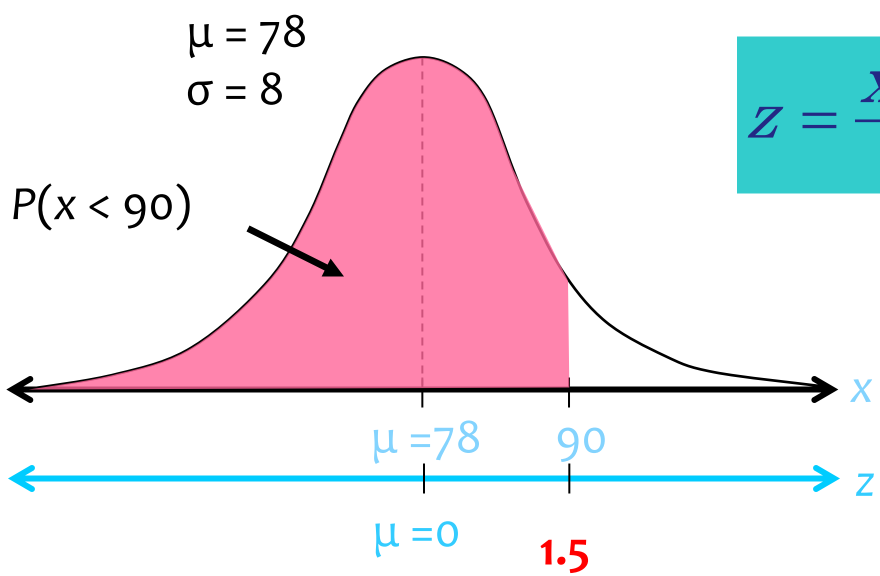
$P(x < 15) = P(z < 1) =$ L'aire ombragée sous la courbe

$= 0.8413$

Probabilité et Distributions Normales

Exemple:

La moyenne d'une série de pluies était de 78 mm avec un écart type de 8 mm. Si ces pluies sont normalement distribuées, trouver la probabilité que la pluie soit inférieure à 90 mm.



$$z = \frac{x - \mu}{\sigma} = \frac{90 - 78}{8}$$

$$= 1.5$$

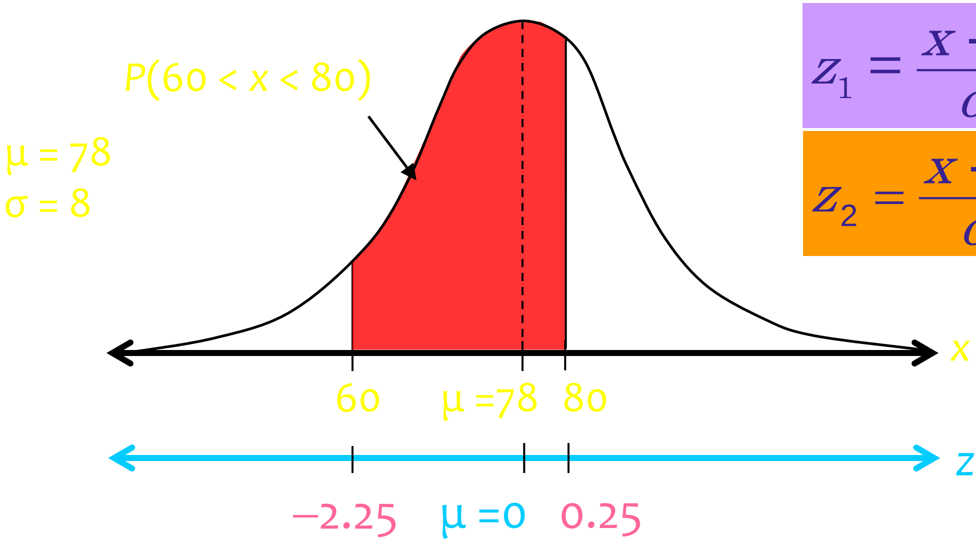
La probabilité qu'une pluie soit inférieure à 90 mm est 0.9332.

$$P(x < 90) = P(z < 1.5) = 0.9332$$

Probabilité et Distributions Normales

Exemple:

La moyenne d'une série de pluies était de 78 mm avec un écart type de 8 mm. Si ces pluies sont normalement distribuées, trouver la probabilité que la pluie soit comprise entre 60 et 80 mm.



$$z_1 = \frac{x - \mu}{\sigma} = \frac{60 - 78}{8} = -2.25$$

$$z_2 = \frac{x - \mu}{\sigma} = \frac{80 - 78}{8} = 0.25$$

La probabilité qu'une pluie soit comprise 60 et 80 mm est 0.5865.

$$P(60 < x < 80) = P(-2.25 < z < 0.25) = P(z < 0.25) - 1 + P(z < +2.25)$$

$$= 0.5987 - 1 + 0.9878 = 0.5865$$

Probabilité d'une Pluie Annuelle comprise entre 143 et 201 mm

$\bar{x} = 143$

$S = 29$

$$z = \frac{201 - 143}{29} = 2.00$$