

**Mathématiques 1<sup>re</sup> année**

1. Analyse Combinatoire

2. Probabilités

3. Variables Aléatoires

4. Lois Discrètes

5. Lois Continues

6. Séries Statistiques Simples

**Mathématiques 2<sup>ème</sup> année**

1. Distribution, Estimateur, Estimation

2. Séries Statistiques Doubles

3. Tests d'Hypothèses

4. Test du Chi 2

5. Analyse de Variance à 1 Facteur

6. Tests Non Paramétriques

# 1. Introduction

## 2. Distribution d'échantillonnage

2.1 Approche empirique

2.2 Approche théorique

2.3 Loi de probabilité de la moyenne

2.4 Convergence

## 3. Estimateur

3.1 Convergence

3.2 Biais d'un estimateur

3.3 Variance d'un estimateur

## 4. Estimation ponctuelle et par intervalle

4.1 Estimation Ponctuelle: Espérance, Variance & Fréquence

4.2 Intervalle de confiance d'une moyenne

4.3 Intervalle de confiance d'une proportion

**Distribution**

**Estimateur**

**Estimation**

# 1. Introduction

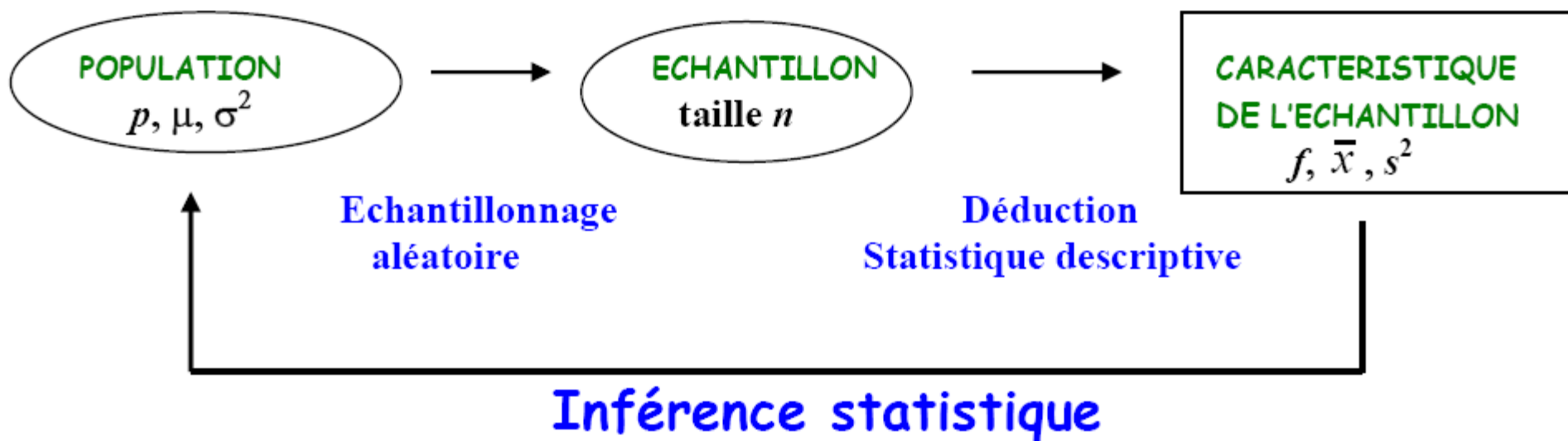
Un phénomène biologique sera entièrement déterminé si l'on connaît la loi de probabilité suivie par la variable aléatoire donnée dans la population. On a alors deux cas de figure :

soit la **loi de probabilité suivie par  $X$  est connue *a priori*** et on vérifie ***a posteriori*** que les observations faites à partir d'un échantillon sont en accord avec elle. C'est le cas par exemple de la répartition des génotypes attendus dans une population sous le modèle de Hardy-Weinberg. On effectue alors **un test d'ajustement entre la distribution théorique et la distribution observée**

soit la **loi de probabilité suivie par  $X$  est inconnue** mais suggérée par la description de l'échantillon (nature de la variable, forme de la distribution des fréquences, valeurs des paramètres descriptifs) (chapitre 5). Dans ce cas, il est nécessaire d'**estimer** les paramètres de la loi de probabilité à partir des paramètres établis sur l'échantillon.

L'**estimation** a pour objectif de déterminer les valeurs inconnues des paramètres de la population ( $p, \mu, \sigma^2$ ) ou (proportion, moyenne, variance) à partir des données de l'échantillon ( $f, \bar{x}, s^2$ ). Il est alors nécessaire de déterminer la précision de ces estimations en établissant **un intervalle de confiance** autour des valeurs prédites.

# 1. Introduction



L'**estimation** a pour objectif de déterminer les valeurs inconnues des paramètres de la population ( $p, \mu, \sigma^2$ ) ou (proportion, moyenne, variance) à partir des données de l'échantillon ( $f, \bar{x}, s^2$ ). Il est alors nécessaire de déterminer la précision de ces estimations en établissant un intervalle de confiance autour des valeurs prédites.

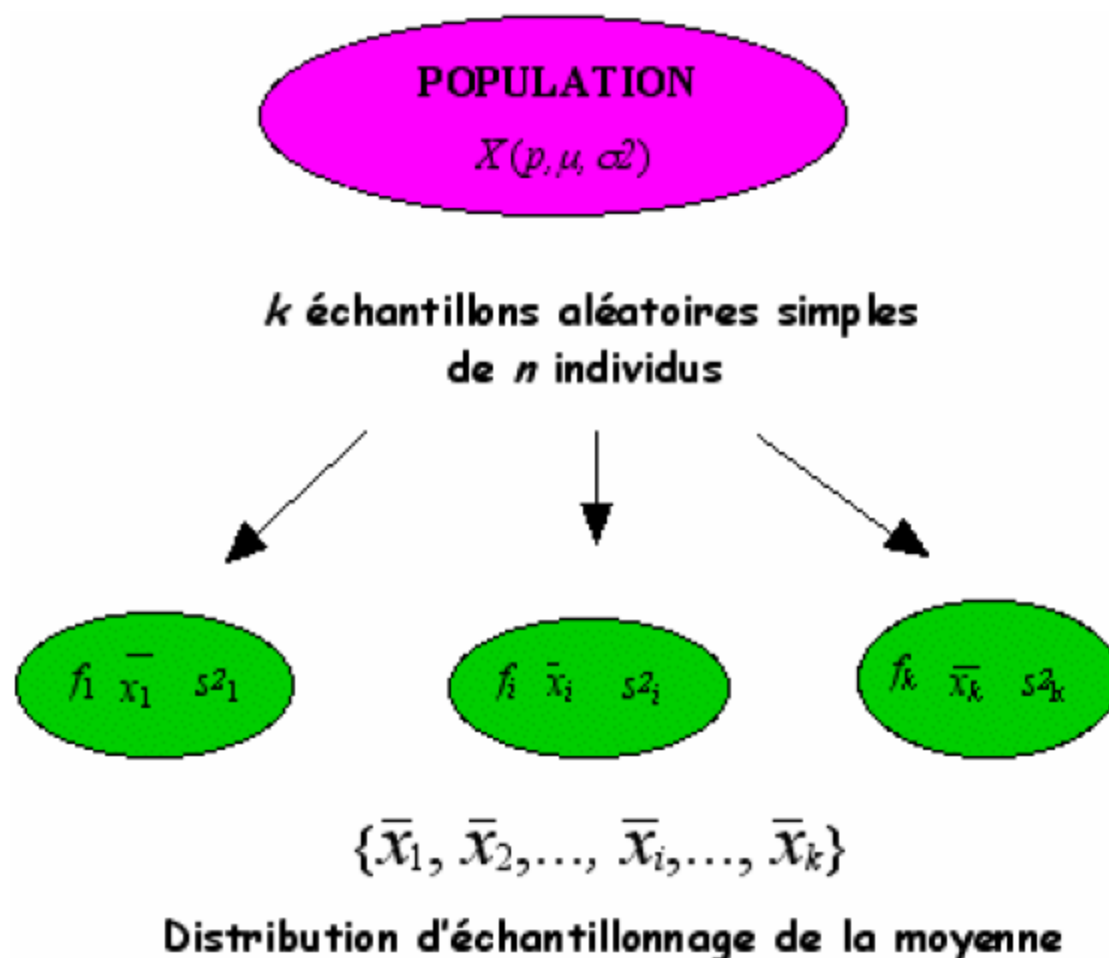
## 2. Distribution d'échantillonnage

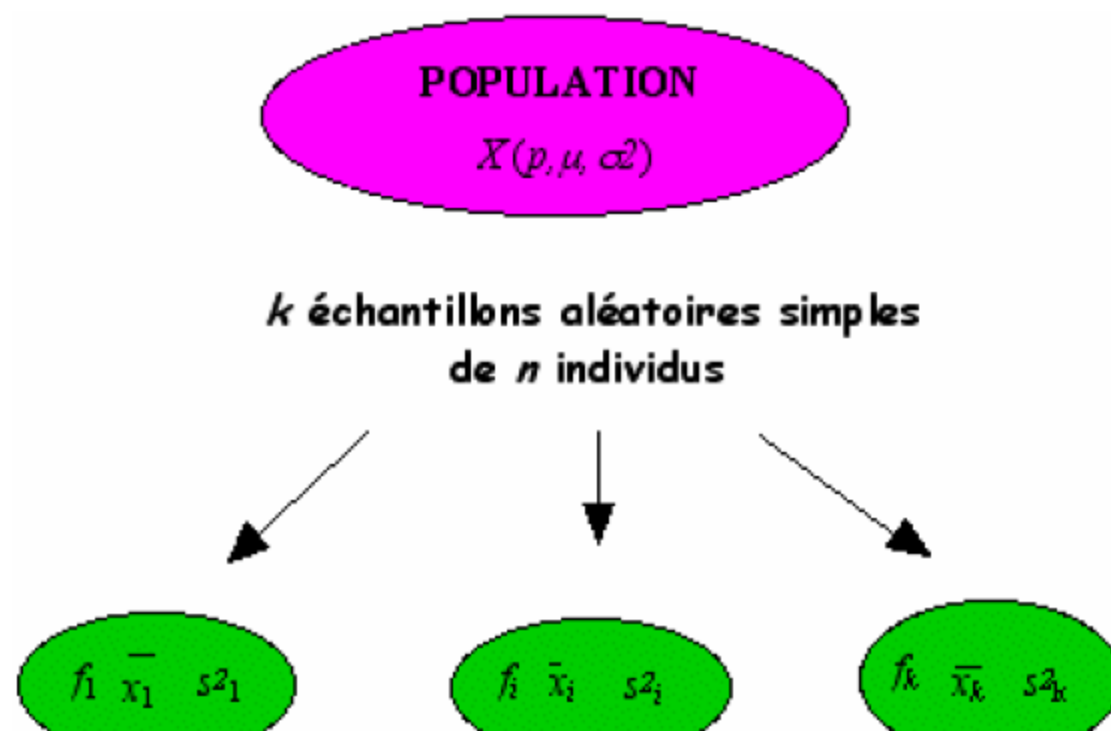
Pour résoudre les problèmes d'estimation de paramètres inconnus, il faut tout d'abord étudier les **distributions d'échantillonnage**, c'est à dire la loi de probabilité suivie par l'estimateur.

**Remarque :** En théorie de l'estimation, il s'agit de distinguer soigneusement trois concepts différents :

- ◆ les paramètres de la **population** comme la moyenne  $\mu$  dont la valeur est certaine mais inconnue symbolisés par des **lettres grecques**.
- ◆ les résultats de **l'échantillonnage** comme la moyenne  $\bar{x}$  dont la valeur est certaine mais connue symbolisés par des **minuscules**.
- ◆ les **variables aléatoires des paramètres**, comme la moyenne aléatoire  $\bar{X}$  dont la valeur est incertaine puisque aléatoire mais dont la loi de probabilité est souvent connue et symbolisées par des **majuscules**.

Il est possible d'extraire d'une population de paramètres  $p$ ,  $\mu$  ou  $\sigma^2$  pour une variable aléatoire  $X$ ,  $k$  échantillons aléatoires simples de même effectif,  $n$ . Sur chaque échantillon de taille  $n$ , on calcule les paramètres descriptifs ( $f$ ,  $\bar{x}$ ,  $s^2$ ).

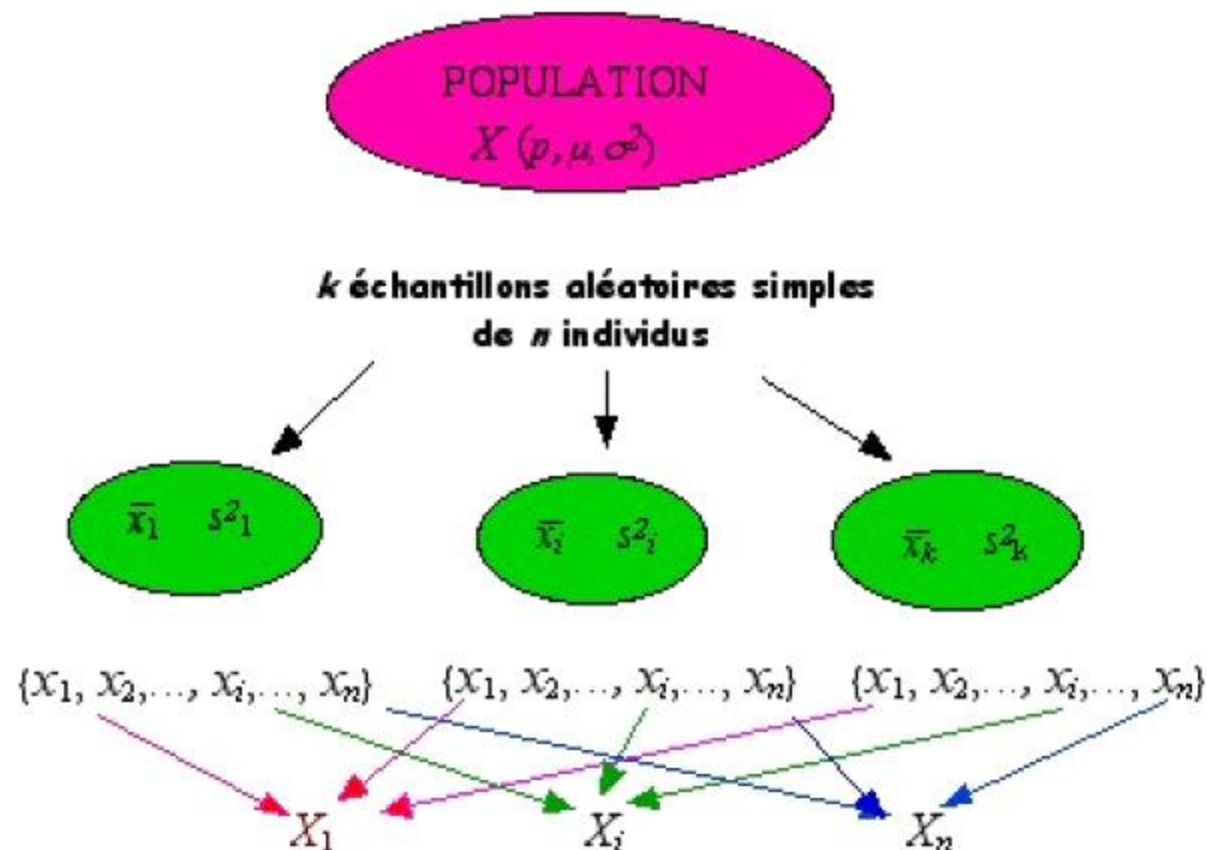


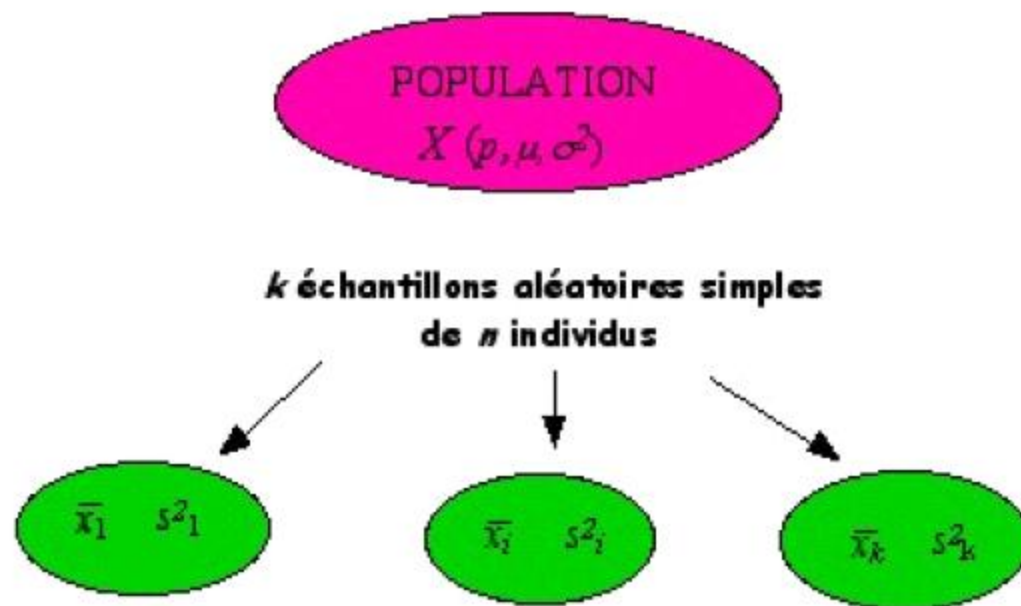


La distribution associée à ces  $k$  estimations constitue la **distribution d'échantillonnage du paramètre**. On peut alors associer une variable aléatoire à chacun des paramètres. La loi de probabilité suivie par cette variable aléatoire admet comme distribution, la distribution d'échantillonnage du paramètre auquel on pourra associer une espérance et une variance.



En pratique, les données étudiées sont relatives à **un seul échantillon**. C'est pourquoi, il faut rechercher les propriétés des échantillons susceptibles d'être prélevés de la population ou plus précisément les lois de probabilité de variables aléatoires associées à un échantillon aléatoire.





Ainsi les  $n$  observations  $x_1, x_2, \dots, x_i, \dots, x_n$ , faites sur un échantillon peuvent être considérées comme  $n$  variables aléatoires  $X_1, X_2, \dots, X_i, \dots, X_n$ . En effet, la valeur prise par le premier élément extrait de la population  $X_1$ , dépend de l'échantillon obtenu lors du tirage aléatoire. Cette valeur sera différente si l'on considère un autre échantillon. Il en est de même pour les  $n$  valeurs extraites de la population.

## 2. Distribution d'échantillonnage

## 3. Loi de Probabilité de la Moyenne

Soit  $X$  une variable aléatoire suivant une loi de probabilité  $\mathcal{L}$ . On tire  $n$  copies indépendantes  $X_1, X_2, \dots, X_i, \dots, X_n$ . On échantillonne avec  $E(X_i) = \mu$  et  $V(X_i) = \sigma^2$ .

On construit alors **la variable aléatoire**  $\bar{X}$ , telle que

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_i + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

avec pour **espérance** :

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n\mu$$

Propriétés de l'espérance

d'où  $E(\bar{X}) = \mu$

$E(\bar{X})$  est notée également  $\mu_{\bar{X}}$

et pour **variance** si  $V(X_i) = \sigma^2$ :

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} n\sigma^2$$

Propriétés de la variance

d'où  $V(\bar{X}) = \frac{\sigma^2}{n}$

$V(\bar{X})$  est notée également  $\sigma_{\bar{X}}^2$

Lorsque la variance  $\sigma^2$  **est connue** et  $n$  **grand** ( $n \geq 30$ ), on se trouve dans les conditions du théorème central limite et la loi suivie par :

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow \mathcal{N}(0,1) \quad \text{loi normale réduite}$$

Ceci reste vrai lorsque  $n \leq 30$  seulement si la loi suivie par  $X$  suit une loi normale.

Lorsque la variance  $\sigma^2$  **est inconnue** et  $X$  **suit une loi normale**, la loi suivie par la variable centrée réduite est alors :

$$\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \rightarrow T_{n-1} \quad \text{loi de student à } n-1 \text{ degrés de liberté}$$

Lorsque la variance  $\sigma^2$  **est inconnue** et  $X$  **ne suit pas une loi normale**, la loi suivie par

$$\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \quad \text{n'est pas connue.}$$

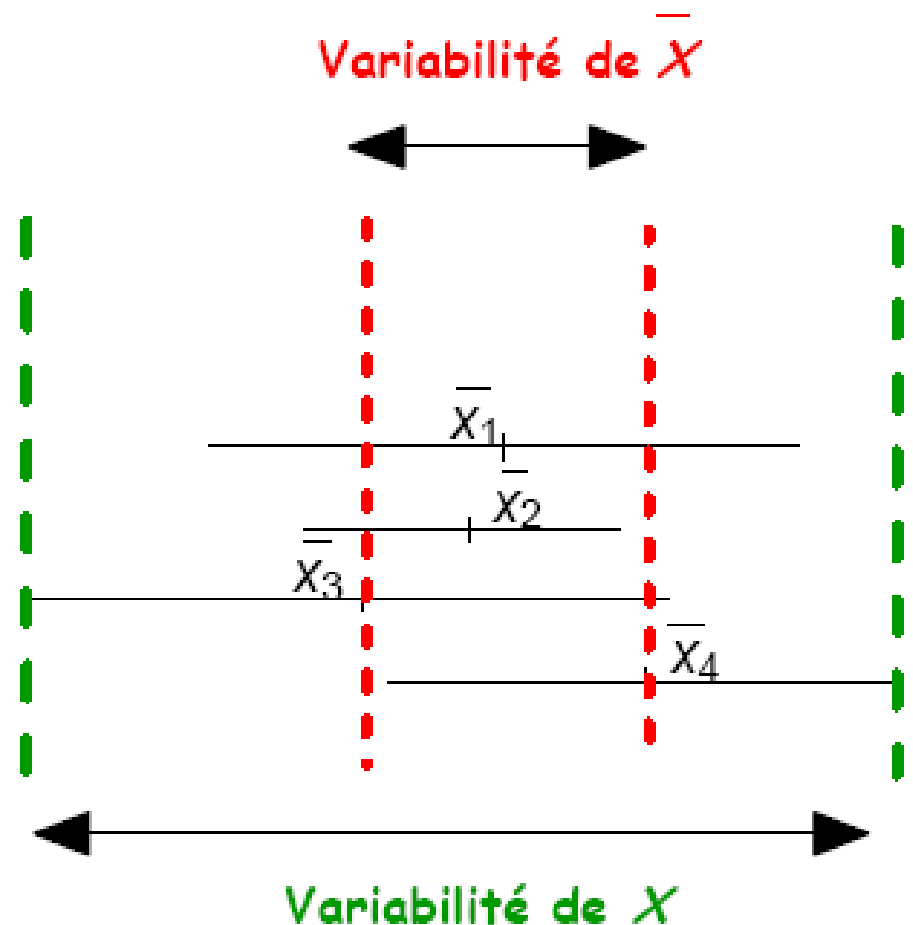
La **loi de probabilité** de la variable aléatoire  $\bar{X}$ , moyenne de  $n$  v.a.  $X$  de loi de probabilité  $\mathcal{N}(\mu, \sigma)$ , est une **loi normale**  $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ .

La **loi de probabilité** d'une fréquence  $\frac{K}{n}$ , suit une **loi normale**  $\mathcal{N}(p, \sqrt{\frac{pq}{n}})$   
vrai si  $np > 5$  et  $nq > 5$ .

**Remarque :** il est aisé de voir sur le graphe ci-dessous que la variance associée à une moyenne ( $\frac{\sigma^2}{n}$ ) est plus faible que la variance de la variable elle-même ( $\sigma^2$ ).

Soit l'étendue des valeurs observées d'une variable aléatoire  $X$  pour 4 échantillons de même taille d'une même population.

Les valeurs des moyennes arithmétiques sont indiquées ainsi que les limites relatives à l'étendue des valeurs de la **variable** observée et celle des **moyennes** observées.



L'**estimation** de  $\theta$  est une variable aléatoire  $\Theta$  dont la distribution de probabilité s'appelle la **distribution d'échantillonnage** du paramètre  $\theta$ .

L'estimateur  $\Theta$  admet donc une espérance  $E(\Theta)$  et une variance  $V(\Theta)$ .

L'estimateur  $\Theta$  doit tendre vers la valeur réelle du paramètre  $\theta$  lorsque le nombre d'individus étudié augmente. On dit que **l'estimateur est convergent**.

$$\text{Si } \forall \varepsilon > 0 \quad P(|\Theta - \theta| > \varepsilon) \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty$$

Le biais d'un estimateur noté  $B(\Theta)$  est la différence moyenne entre sa valeur et celle du paramètre qu'il estime. Le biais doit être égal à 0 pour avoir un bon estimateur.

$$B(\Theta) = E(\Theta - \theta) = E(\Theta) - E(\theta) = E(\Theta) - \theta = 0 \quad (\text{voir } \underline{\text{propriétés de l'espérance}})$$

$$\text{d'où } E(\Theta) = \theta$$

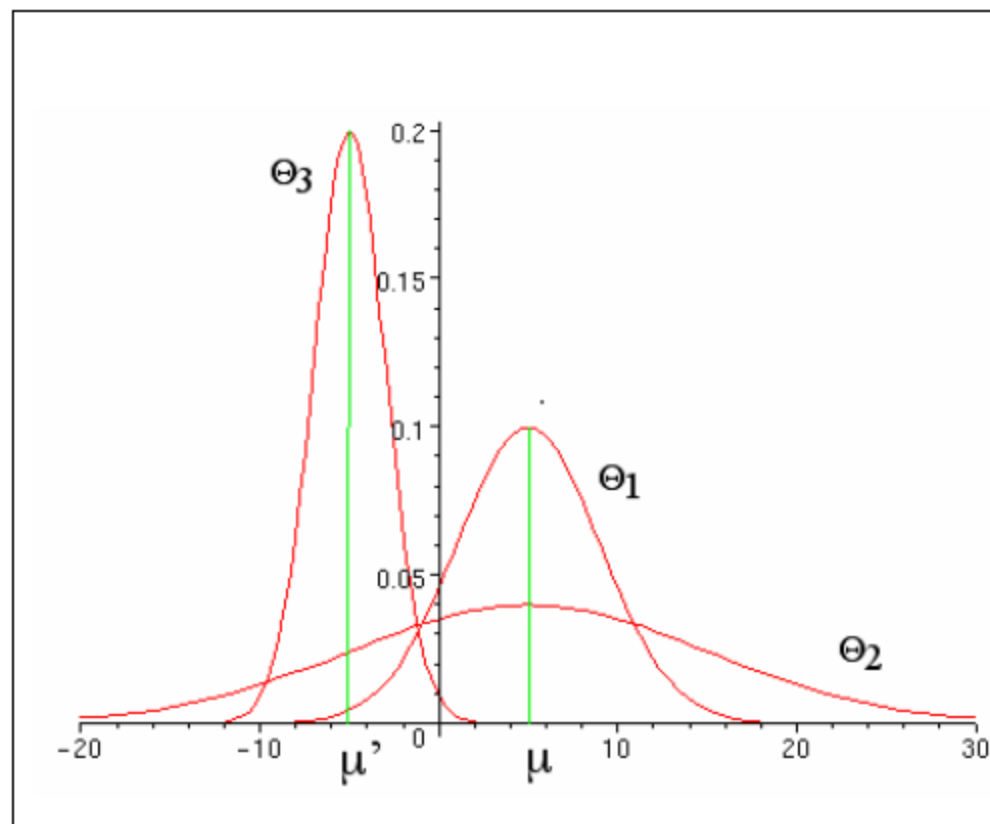
Ainsi l'estimateur sera **sans biais** si son espérance est égale à la valeur du paramètre de la population.

$$E(\Theta) = \theta$$

### Exemple :

Soit les densités de probabilité de 3 estimateurs d'une espérance  $\mu$ ,





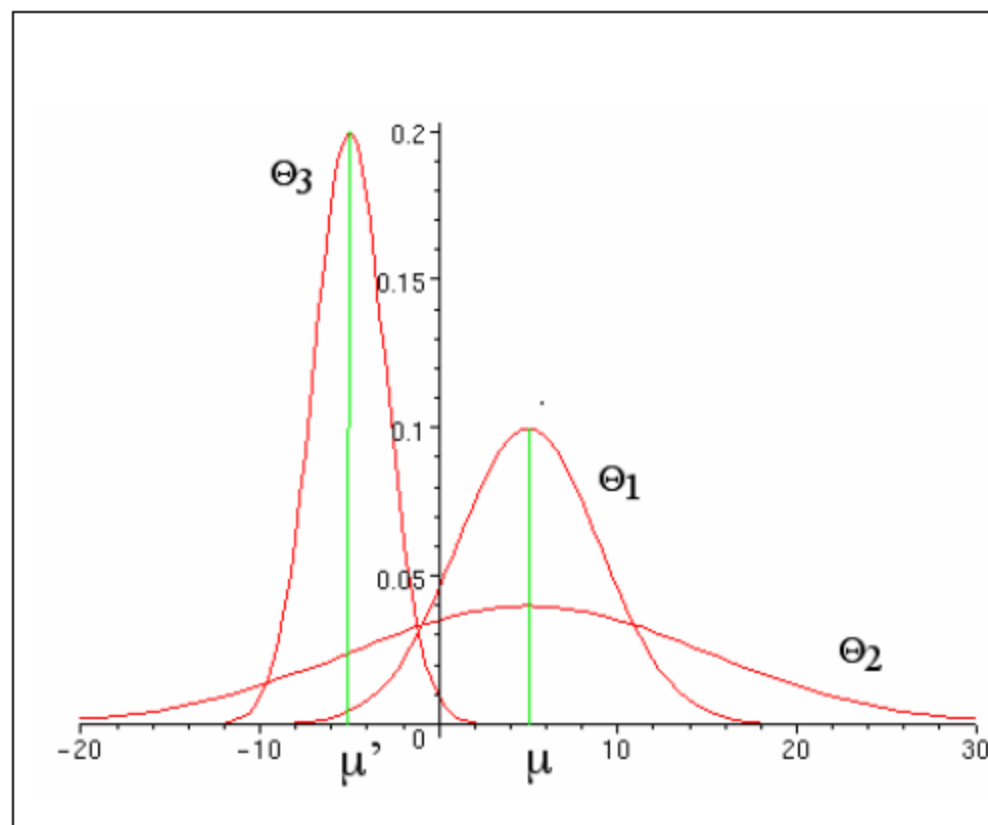
$\Theta_1$  et  $\Theta_2$  sont des **estimateurs sans biais**

de  $\mu$  car

$$E(\Theta_1) = E(\Theta_2) = \mu$$

$\Theta_3$  est un **estimateur biaisé** de  $\mu$  car

$$E(\Theta_3) - \mu = \mu' - \mu \neq 0$$



$\Theta_1$  et  $\Theta_2$  sont des **estimateurs sans biais**

de  $\mu$  car

$$E(\Theta_1) = E(\Theta_2) = \mu$$

$\Theta_3$  est un **estimateur biaisé** de  $\mu$  car

$$E(\Theta_3) - \mu = \mu' - \mu \neq 0$$

Dans l'exemple ci-dessus,  $\Theta_1$  et  $\Theta_2$  sont des **estimateurs sans biais** de  $\mu$  car  
 $B(\Theta_1) = E(\Theta_1 - \mu) = E(\Theta_1) - \mu = 0$  car  $E(\Theta_1) = \mu$ , de même pour  $B(\Theta_2)$

alors que  $\Theta_3$  est un **estimateur biaisé** de  $\mu$  car

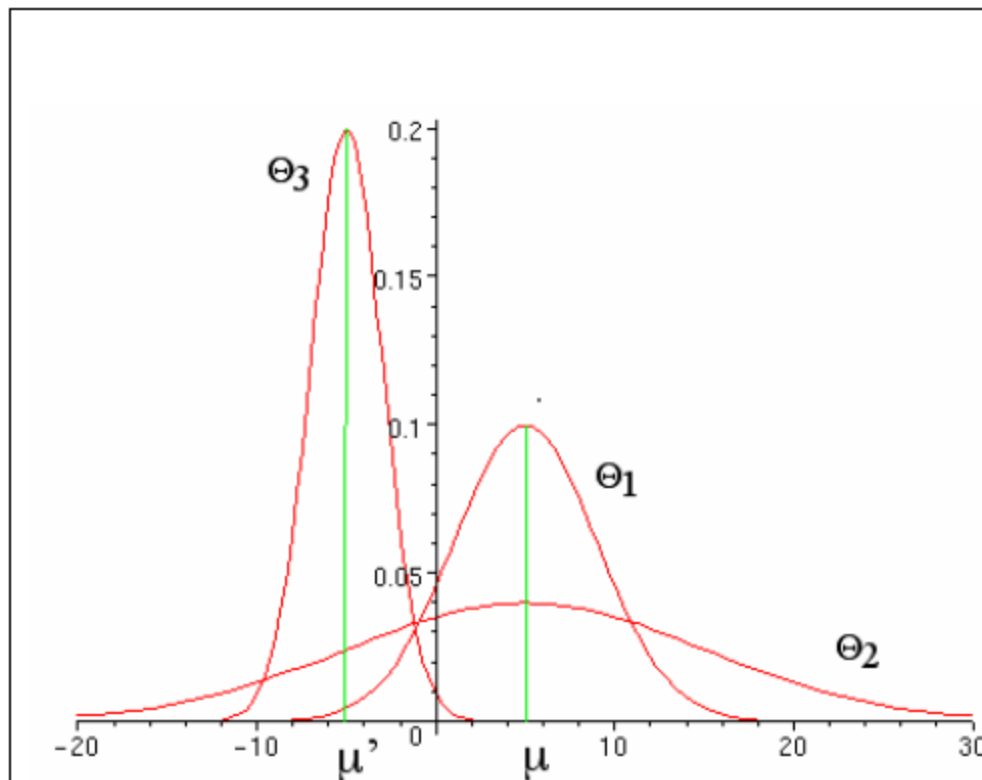
$$B(\Theta_3) = E(\Theta_3 - \mu) = E(\Theta_3) - \mu = \mu' - \mu \neq 0 \text{ car } E(\Theta_3) = \mu'$$

### 3. Estimateur

### 3. Variance d'un Estimateur

Si deux estimateurs sont convergents et sans biais, le plus efficace est celui qui a **la variance la plus faible** car ses valeurs sont en moyenne plus proches de la quantité estimée.

$$V(\Theta) = E(\Theta - E(\Theta))^2 \text{ minimale}$$



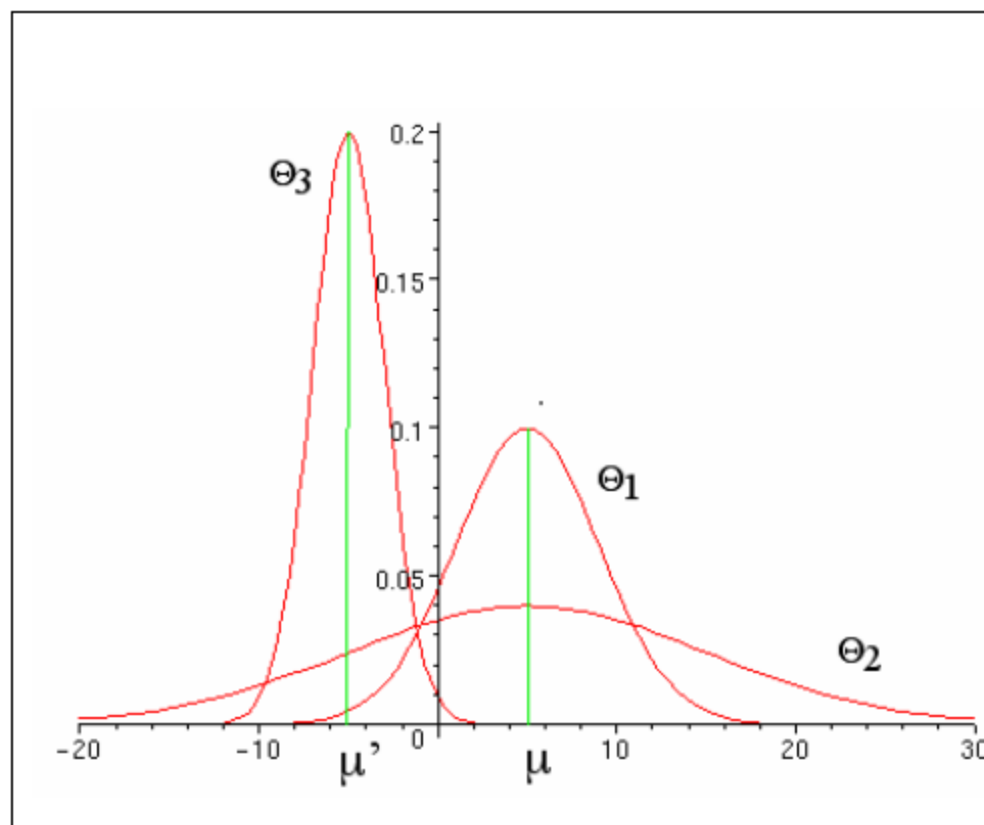
$\Theta_1$  et  $\Theta_2$  sont des **estimateurs sans biais** de  $\mu$  car  
 $E(\Theta_1) = E(\Theta_2) = \mu$

$\Theta_3$  est un **estimateur biaisé** de  $\mu$  car  
 $E(\Theta_3) - \mu = \mu' - \mu \neq 0$

Exemple

Dans l'exemple précédent, on voit que  $V(\Theta_1) < V(\Theta_2)$ . On peut donc conclure que  $\Theta_1$  est un meilleur estimateur de  $\mu$  que  $\Theta_2$ .

**Remarque :** Quand les estimateurs sont biaisés, en revanche, leur comparaison n'est pas simple. Ainsi un estimateur peu biaisé mais de variance très faible, pourrait même être préféré à un estimateur sans biais mais de grande variance.



$\Theta_1$  et  $\Theta_2$  sont des **estimateurs sans biais**

de  $\mu$  car

$$E(\Theta_1) = E(\Theta_2) = \mu$$

$\Theta_3$  est un **estimateur biaisé** de  $\mu$  car

$$E(\Theta_3) - \mu = \mu' - \mu \neq 0$$

Théorème :

Si un estimateur est asymptotiquement sans biais et si sa variance tend vers 0 lorsque  $n \rightarrow \infty$ , il est convergent.

$$P( |\Theta - \theta| \geq \varepsilon ) \leq \frac{V(\Theta)}{\varepsilon^2} \quad \text{avec } \varepsilon > 0$$

(Inégalité de Bienaymé-Tchébycheff)

Cette inégalité exprime que si  $|\Theta - \theta|$  tend vers 0 quand  $n$  augmente,  $V(\Theta)$  doit aussi tendre vers 0.

## 4. Estimation Ponctuelle et par Intervalle

---

L'**estimation** d'un paramètre quelconque  $\theta$  est **ponctuelle** si l'on associe **une seule valeur** à l'estimateur  $\hat{\theta}$  à partir des données observables sur un échantillon aléatoire. L'**estimation par intervalle** associe à un échantillon aléatoire, **un intervalle**  $[\hat{\theta}_1, \hat{\theta}_2]$  qui recouvre  $\theta$  avec une certaine probabilité.

Soit  $X$  une variable aléatoire continue suivant une loi normale  $\mathcal{N}(\mu, \sigma)$  dont la valeur des paramètres n'est pas connue et pour laquelle on souhaite estimer **l'espérance  $\mu$** .

Soient  $X_1, X_2, \dots, X_i, \dots, X_n$ ,  $n$  réalisations indépendantes de la variable aléatoire  $X$ , un estimateur du paramètre  $\mu$  est une suite de variable aléatoire  $\Theta$  fonctions des  $X_i$ :

$$\Theta = f(X_1, X_2, \dots, X_i, \dots, X_n)$$

La méthode des **moindres carrés** consiste à rechercher les coefficients de la combinaison linéaire

$$\Theta = a_1X_1 + a_2X_2 + \dots + a_iX_i + \dots + a_nX_n$$

telle que  $E(\Theta) = \mu$  et  $V(\Theta)$  soit minimale

La **moyenne arithmétique** constitue le meilleur estimateur de  $\mu$ , espérance de la loi de probabilité de la variable aléatoire  $X$ :

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

*Voici pourquoi :*

Estimateur *sans biais* :  $E(\bar{X}) = \mu$  (voir loi de la moyenne)

Estimateur *convergent* : si l'on pose l'inégalité de Bienaymé-Tchébycheff :

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{V(\bar{X})}{\varepsilon^2} \quad \text{avec } \varepsilon > 0$$

lorsque  $n \rightarrow \infty$   $\frac{V(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0$  et ceci  $\forall \varepsilon > 0$

ainsi en limite,  $P(|\bar{X} - \mu| \geq \varepsilon) = 0$ , ce qui indique que  $\bar{X} \rightarrow \mu$  en probabilité.



Soit  $X$  une variable aléatoire continue suivant une loi normale  $\mathcal{N}(\mu, \sigma)$  pour laquelle on souhaite estimer **la variance  $\sigma^2$** .

Soient  $X_1, X_2, \dots, X_i, \dots, X_n$ ,  $n$  réalisations indépendantes de la variable aléatoire  $X$ , un estimateur du paramètre  $\sigma^2$  est une suite de variable aléatoire  $\Theta$  fonctions des  $X_i$ .

$$\Theta = f(X_1, X_2, \dots, X_i, \dots, X_n)$$

- Cas où l'espérance  $\mu$  est connue

La méthode des moindres carrés consiste à rechercher les coefficients de la combinaison linéaire

$$\Theta = a_1(X_1 - \mu)^2 + a_2(X_2 - \mu)^2 + \dots + a_i(X_i - \mu)^2 + \dots + a_n(X_n - \mu)^2$$

telle que  $E(\Theta) = \sigma^2$  et  $V(\Theta)$  soit minimale

La **variance observée** constitue le meilleur estimateur de  $\sigma^2$ , variance de la loi de probabilité de la variable aléatoire  $X$  lorsque l'espérance  $\mu$  est connue :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

- Cas où l'espérance  $\mu$  est inconnue

Le meilleur estimateur de  $\sigma^2$ , variance de la loi de probabilité de la variable aléatoire  $X$  lorsque l'espérance  $\mu$  est inconnue est :

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

**Remarque :** Lorsque  $n$  augmente, la variance observée  $s^2$  tend vers la variance de la population  $\sigma^2$ .

$$\lim_{n \rightarrow +\infty} s^2 = \lim_{n \rightarrow +\infty} \frac{(n-1)}{n} \sigma^2 = \sigma^2$$

Soit le schéma de Bernoulli dans lequel le caractère A correspond au succès. On note  $p$  la fréquence des individus de la **population** possédant le caractère A. La valeur de ce paramètre étant inconnu, on cherche à estimer la fréquence  $p$  à partir des données observables sur un échantillon.

A chaque échantillon non exhaustif de taille  $n$ , on associe l'entier  $k$ , nombre d'individus possédant le caractère A.

La **fréquence observée** du nombre de succès observé dans un échantillon de taille  $n$  constitue le meilleur estimateur de  $p$  :

$$\hat{p} = \frac{K}{n}$$

L'**estimation par intervalle** associe à un échantillon aléatoire, **un intervalle**  $[\hat{\theta}_1, \hat{\theta}_2]$  qui recouvre  $\theta$  avec une certaine probabilité.

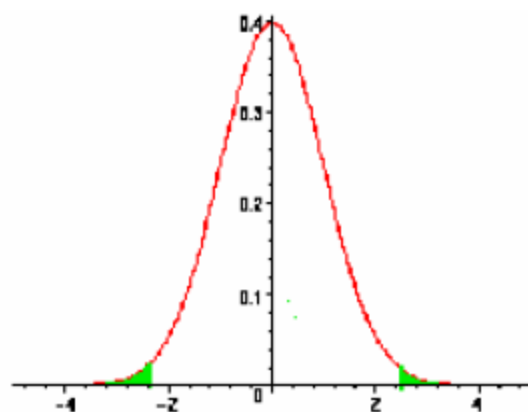
Cet intervalle est appelé l'**intervalle de confiance** du paramètre  $\theta$  car la probabilité que  $\theta$  dont la valeur est inconnue se trouve compris entre  $\hat{\theta}_1$  et  $\hat{\theta}_2$  est égale à  $1-\alpha$ , **le coefficient de confiance**

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$$

Son complément  $\alpha$  correspond au **coefficient de risque**.

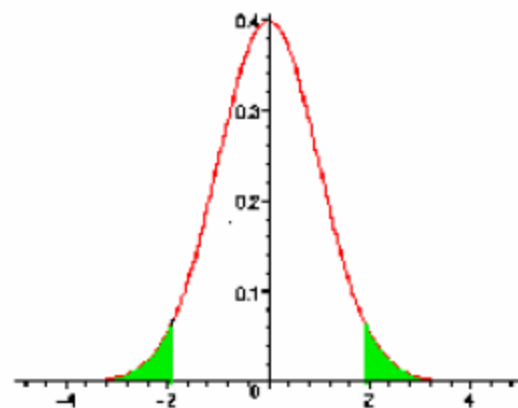
$$P(\theta \notin [\hat{\theta}_1, \hat{\theta}_2]) = \alpha$$

Un intervalle de confiance indique **la précision d'une estimation** car pour un risque  $\alpha$  donné, l'intervalle est d'autant plus grand que la précision est faible comme l'indiquent les graphes ci-dessous. Pour chaque graphe, **l'aire hachurée en vert** correspond au coefficient de risque  $\alpha$ . Ainsi de part et d'autre de la distribution, la valeur de l'aire hachurée vaut  $\frac{\alpha}{2}$ .



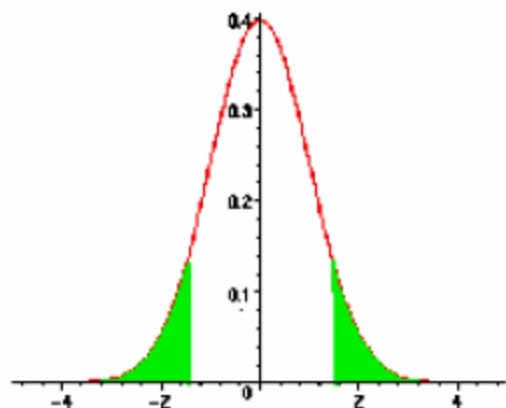
$$\alpha = 0,01$$

**99 chances sur 100** que la valeur du paramètre recherché se trouve dans l'intervalle de confiance mais la **précision** autour de la valeur prédite est **faible**



$$\alpha = 0,05$$

**95 chances sur 100** que la valeur du paramètre recherché se trouve dans l'intervalle de confiance et la **précision** autour de la valeur prédite est **correcte**.

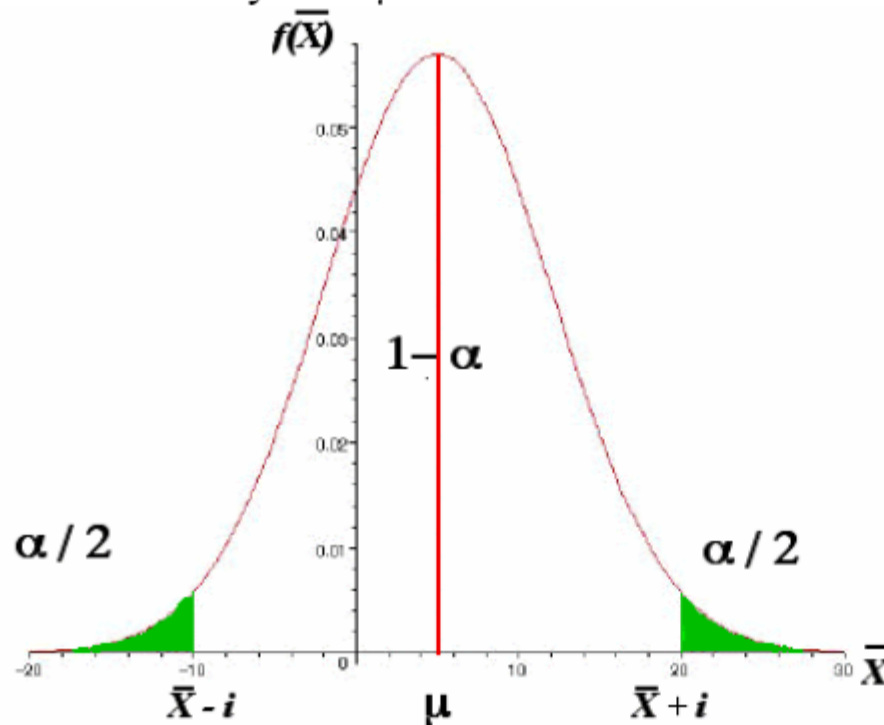


$$\alpha = 0,10$$

**90 chances sur 100** que la valeur du paramètre recherché se trouve dans l'intervalle de confiance mais la **précision** autour de la valeur prédite est **élevée**.

En fonction de la nature de la variable aléatoire continue  $X$ , de la taille de l'échantillon  $n$  et de la connaissance que nous avons sur le paramètre  $\sigma^2$ , l'établissement de l'intervalle de confiance autour de  $\mu$  sera différent.

- **Quelque soit la valeur de  $n$ , si  $X \rightarrow \mathcal{N}(\mu, \sigma)$  et  $\sigma^2$  est connue**, Etablir l'intervalle de confiance autour de la moyenne  $\mu$  revient à établir la valeur de  $i$  pour

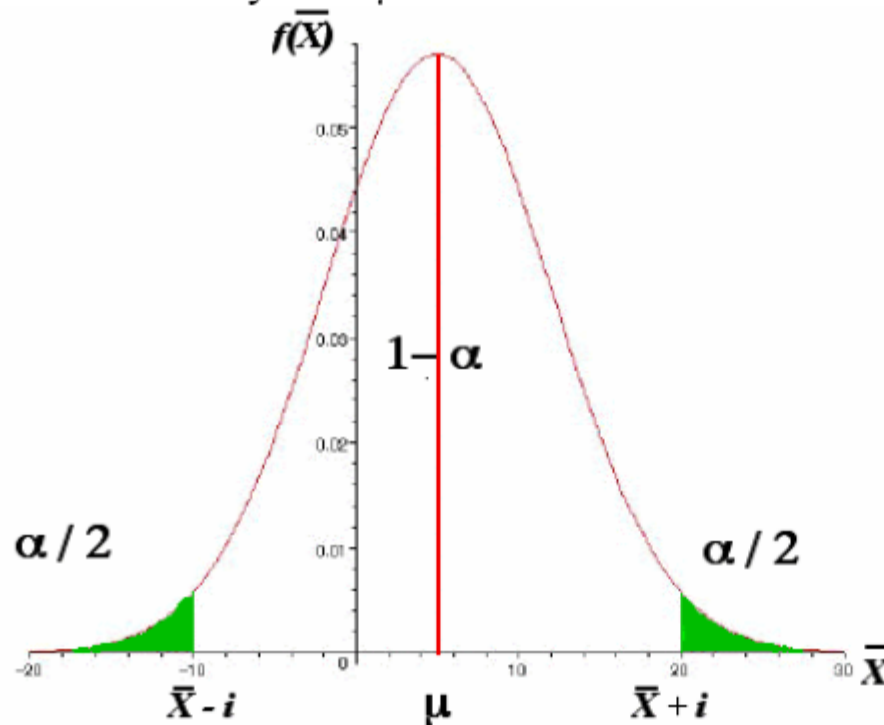


L'**intervalle de confiance de la moyenne  $\mu$**  pour un coefficient de risque  $\alpha$  est donc

$$\bar{X} - \varepsilon_{\alpha} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + \varepsilon_{\alpha} \frac{\sigma}{\sqrt{n}}$$

quelque soit la valeur de  $n$  si  $X \rightarrow \mathcal{N}(\mu, \sigma)$  et la variance  $\sigma^2$  est connue

• Quelque soit la valeur de  $n$ , si  $X \rightarrow \mathcal{N}(\mu, \sigma)$  et  $\sigma$  est connue, l'intervalle de confiance autour de la moyenne  $\mu$  revient à établir la valeur de  $i$  pour



Coefficient de risque	Ecart-réduit
$\alpha = 0,01$	$\varepsilon_{\alpha} = 2,576$
$\alpha = 0,05$	$\varepsilon_{\alpha} = 1,960$
$\alpha = 0,10$	$\varepsilon_{\alpha} = 1,645$



- Quelque soit la valeur de  $n$ , si  $X \rightarrow \mathcal{N}(\mu, \sigma)$  et  $\sigma^2$  est **inconnue**,

Le raisonnement reste le même mais la variance de la population  $\sigma^2$  doit être estimée par

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2 \quad (\text{voir } \underline{\text{estimation ponctuelle}})$$

L'**intervalle de confiance de l'espérance  $\mu$**  pour un coefficient de risque  $\alpha$  est donc

$$\bar{X} - t_{\alpha} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha} \frac{\hat{\sigma}}{\sqrt{n}}$$

quelque soit la valeur de  $n$  si  $X \rightarrow \mathcal{N}(\mu, \sigma)$  et  $\sigma^2$  est **inconnue**

**Remarque :** Lorsque  $n > 30$ , la loi de student converge vers une loi normale réduite. Ainsi la valeur de  $t_{\alpha}(n-1)$  est égale à  $\varepsilon_{\alpha}$ . Ci-dessous, un exemple pour un risque  $\alpha = 0,05$ .

Taille de l'échantillon	Ecart-réduit	Variable de student
$n = 10$	$\varepsilon_{\alpha} = 1,960$	$t_{\alpha} = 2,228$
$n = 20$	$\varepsilon_{\alpha} = 1,960$	$t_{\alpha} = 2,086$
$n = 30$	$\varepsilon_{\alpha} = 1,960$	$t_{\alpha} = 2,042$
$n = 40$	$\varepsilon_{\alpha} = 1,960$	$t_{\alpha} = 1,960$

- Si  $n > 30$  et  $X$  suit une loi inconnue,

La démarche est la même que pour le cas précédent puisque par définition la variance de la population est inconnue et doit être estimée avec la variance observée :

$$\hat{\sigma}^2 = \frac{n}{n-1} s^2 \quad (\text{voir } \underline{\text{estimation ponctuelle}})$$

Comme pour le cas 1, la loi suivie par la variable centrée réduite  $\frac{\bar{X} - \mu}{\hat{\sigma} / \sqrt{n}} \rightarrow \mathcal{N}(0,1)$

L'**intervalle de confiance de l'espérance  $\mu$**  pour un coefficient de risque  $\alpha$  est donc

$$\bar{X} - \varepsilon_{\alpha} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + \varepsilon_{\alpha} \frac{\hat{\sigma}}{\sqrt{n}}$$

vraie seulement si  **$n$  est grand.**

- Si  $n < 30$  et  $X$  suit une loi inconnue,

La loi de probabilité suivie par  $\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}$  n'est pas connue et l'on a recours aux statistiques non paramétriques.

Etablir l'intervalle de confiance autour de la fréquence  $p$  de la population à partir de son estimateur  $\frac{K}{n}$  revient à établir la valeur de  $i$  pour une valeur du coefficient de confiance

$(1 - \alpha)$  donnée par l'expérimentateur telle que :

$$P\left(\frac{K}{n} - i < p < \frac{K}{n} + i\right) = 1 - \alpha \quad \text{ou} \quad P\left(p - i < \frac{K}{n} < p + i\right) = 1 - \alpha$$

L'**intervalle de confiance de la fréquence**  $p$  pour un coefficient de risque  $\alpha$  est donc

$$\frac{K}{n} - \varepsilon_{\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \frac{K}{n} + \varepsilon_{\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

vraie seulement si  **$n$  est grand et  $np, nq > 5$**