***<u>Remarques initiales</u>* :**

---

**Echantillonnage – Intervalle de fluctuation :**

On utilise un *intervalle de fluctuation* quand :

- On <u>connaît la proportion</u> $p$ de présence du caractère dans la population

  Ou bien si

- On fait une *<u>hypothèse sur la valeur de cette proportion</u>* (situation de « prise de décision »)

**Estimation – Intervalle de confiance :**

On utilise un *intervalle de confiance* quand :

- On <u>ignore la proportion</u> $p$ de présence du caractère dans la population, *<u>sans faire d'hypothèse sur cette valeur</u>*.

---

*<u>Exemple</u>* 1 :

On dispose d'une pièce de monnaie, dont on cherche à savoir si elle est équilibrée ou pas.

On fait l'*<u>hypothèse</u>* d'une apparition de « Pile » égale à 0,5 , que l'on va ensuite tester.

*On est dans une situation d'échantillonnage.*

*<u>Exemple</u>* 2 :

Une usine fabrique des appareils photographiques. Sur 1.000 appareils fabriqués, on constate que 15 d'entre eux ne fonctionnent pas. On cherche à évaluer la proportion d'appareils défectueux dans l'ensemble de la production.

On ne connait pas au départ cette proportion, et on n'en fait pas d'hypothèse.

*On est en situation d'estimation.*

**<u>Echantillonnage</u> – Intervalle de fluctuation de $f$ au seuil de 95% :**

Soit $f$ la fréquence observée $f$ dans l'échantillon :

- Si $f$ appartient à l'intervalle de fluctuation au seuil de 95% , on accepte l'hypothèse,
- Si $f$ n'appartient pas à l'intervalle de fluctuation, on rejette l'hypothèse, <u>avec</u> 5% <u>de risque de rejeter une hypothèse pourtant vraie</u>.

*<u>Connaissances niveau Seconde</u>* :

**Si des échantillons de taille individuelle $n \geq 25$ relèvent du schéma de Bernoulli (loi binomiale) de probabilité $p$ , telle que $0{,}2 \leq p \leq 0{,}8$ , alors pour environ 95% des échantillons, la *<u>fréquence</u>* d'apparition de l'évènement $X$ appartient à l'intervalle $[\, p-\frac{1}{\sqrt{n}} \,;\, p+\frac{1}{\sqrt{n}} \,]$ .**

*<u>Connaissances niveau Première</u>* :

**Si la variable aléatoire $X$ suit la loi binomiale $B(n\,;p)$ , alors environ 95% des valeurs possibles de la <u>fréquence</u> de $X$ , $f=\frac{X}{n}$ , appartiennent à l'intervalle $[\, \frac{k_1}{n} \,;\, \frac{k_2}{n} \,]$ , avec $\begin{cases} k_1 \text{ plus petit entier } k \text{ tel que } p(X \leq k) > 0{,}025 \\ k_2 \text{ plus petit entier } k \text{ tel que } p(X \leq k) \geq 0{,}975 \end{cases}$ .**

*Connaissances niveau Terminale (Intervalle de fluctuation asymptotique au seuil de 95%)* :

**Si la variable aléatoire $X_n$ suit la loi binomiale $B(n\,;p)$, l'intervalle $I_n = [\,p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\,p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,]$ est appelé *intervalle de fluctuation asymptotique au seuil de 95%*. C'est-à-dire que : $\lim\limits_{n \to +\infty} p(\dfrac{X_n}{n} \in I_n) = 0{,}95$ .**

**(Conditions nécessaires : $n \geq 30$ , $np \geq 5$ , $n(1-p) \geq 5$ ) . On pose $Z = \dfrac{X_n - \mu}{\sigma} = \dfrac{X_n - np}{np(1-p)}$ , avec $Z = N(0\,;1)$.**

*On considère que la limite est atteinte dès que les conditions nécessaires sont satisfaites.*

***On peut alors écrire : $Prob\left( p - 1{,}96.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \dfrac{X_n}{n} \leq p + 1{,}96.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \right) = 0{,}95$ ( $u_{0,05} = 1{,}96$ ).***

Une population présente un caractère avec une probabilité $p$, et tout échantillon extrait de cette population présente ce caractère avec une fréquence observée $f_{obs} = \dfrac{X_n}{n}$, qui *fluctue* autour de $p$.

95% des échantillons fluctuent dans l'intervalle précédent, mais 5% d'entre eux, bien que de cette population, peuvent fluctuer de plus. On peut alors douter de l'appartenance de l'échantillon à la population, mais avec un risque d'erreur de 5%.

Plus généralement : *Intervalle de fluctuation asymptotique au seuil de* $1 - \alpha$ :

**Si la variable aléatoire $X_n$ suit la loi binomiale $B(n\,;p)$, pour tout réel $\alpha \in ]0\,;1[$ :**

**$u_\alpha$ désigne le réel tel que $p(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$ lorsque $Z = N(0\,;1)$ .**

**L'intervalle $I_n = [\,p - u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\,p + u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,]$ est appelé *intervalle de fluctuation asymptotique au seuil de $1 - \alpha$* . C'est-à-dire que : $\lim\limits_{n \to +\infty} p(\dfrac{X_n}{n} \in I_n) = 1 - \alpha$ .**

*On considère que la limite est atteinte dès que les conditions nécessaires sont satisfaites.*

***On peut alors écrire : $Prob\left( p - u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \dfrac{X_n}{n} \leq p + u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \right) = 1 - \alpha$ .***

**Estimation – Intervalle de confiance de $p$ au niveau de confiance de 95% :**

On rappelle que si la variable aléatoire $X_n$ suit la loi binomiale $B(n\,;p)$ , et que $F_n = \dfrac{X_n}{n}$ est la fréquence constatée associée à $X_n$ , alors, pour $n$ assez grand, l'intervalle aléatoire $[F_n - \dfrac{1}{\sqrt{n}}\,;\,F_n + \dfrac{1}{\sqrt{n}}]$ contient $p$ avec une probabilité supérieure ou égale à 95%.

**On réalise une expérience aléatoire $n$ fois et on appelle $f$ la fréquence observée d'apparition du caractère étudié. L'intervalle $[f - \dfrac{1}{\sqrt{n}}\,;\,f + \dfrac{1}{\sqrt{n}}]$ est appelé *intervalle de confiance* de $p$ au *niveau de confiance* 0,95 , où $p$ est la proportion inconnue d'apparition du caractère dans la population.**(Conditions nécessaires : $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$ ).

**Taille minimale de l'échantillon pour obtenir une précision donnée :**

L'intervalle de confiance $[f - \dfrac{1}{\sqrt{n}}\,;\,f + \dfrac{1}{\sqrt{n}}]$ est de largeur $\dfrac{2}{\sqrt{n}}$. Plus $n$ est grand, plus cet intervalle est précis.

**Si l'on souhaite situer $p$ dans un intervalle de largeur donnée $a$ , il faut $\dfrac{2}{\sqrt{n}} \leq a$ , soit $n \geq \dfrac{4}{a^2}$ .**

*On remarquera que pour resserrer l'intervalle, il faut grandement augmenter la taille de l'échantillon (carré).*