I. Principe de l'échantillonnage et de l'estimation

On illustre la situation ainsi : on dispose d'une grande urne où se trouvent un très grand nombre de boules rouges et bleues.

Cas 1 : la proportion p des boules rouges est connue	**Cas 2 : la proportion p des boules rouges n'est pas connue**
On tire au hasard avec remise n boules de l'urne. On obtient donc une fréquence f de boules rouges sur cet échantillon. On s'attend à ce que la fréquence f observée soit « proche » de p , fréquence théorique. On est ici dans le cadre d'un **échantillonnage**.	On veut donc estimer la proportion p de boules rouges dans l'urne. Comme il y en a beaucoup, on ne peut pas toutes les compter. On tire donc au hasard avec remise n boules de l'urne. On obtient donc une fréquence f de boules rouges sur cet échantillon. On s'attend à ce que la fréquence p théorique soit « proche » de f, fréquence observée. On est ici dans le cadre d'une **estimation**.

II. Intervalles de fluctuation et de confiance

II.1. Calcul des intervalles de fluctuation et de confiance

Cas 1 : Échantillonnage	**Cas 2 : Estimation**
On connait p, fréquence théorique d'un caractère sur une population. On a un échantillon de taille n. L'intervalle $\left[p-\dfrac{1}{\sqrt{n}};p+\dfrac{1}{\sqrt{n}}\right]$ est appelée intervalle de fluctuation au seuil 95% de la fréquence de ce caractère aléatoire de taille n issu de la population. <u>Condition de validité</u> : $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.	On connait f, fréquence observée d'un caractère sur un échantillon de taille n d'une population. $\left[f-\dfrac{1}{\sqrt{n}};f+\dfrac{1}{\sqrt{n}}\right]$ est appelé intervalle de confiance au seuil 95% de la proportion p de ce caractère aléatoire de la population.

II.2. Signification des intervalles

Cas 1 : Échantillonnage	**Cas 2 : Estimation**
La fréquence observée f sur un échantillon de taille n appartient à l'intervalle de fluctuation au seuil 95% dans 95% des cas.	Au moins 95% des intervalles de confiance au seuil 95% contiennent la fréquence théorique p .

Exemples :

— *On considère une pièce de monnaie équilibrée. La fréquence théorique (ou probabilité) de l'issue « Pile » est* $p = 0,5$.
Imaginons qu'on lance cette pièce 10000 *fois ; on obtiendrait un échantillon de taille* $n = 10000$.
Comme $n \geq 30$, $np = 5000 \geq 5$ *et* $n(1-p) = 5000 \geq 5$, *on peut calculer l'intervalle de fluctuation au seuil* 95% *du caractère « Pile » :*

$$\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right] = \left[0.5-\frac{1}{\sqrt{10000}};0.5+\frac{1}{\sqrt{10000}}\right] = [0.49;0.51]$$

Il y a donc 95% *de chance que la fréquence observée de l'issue « Pile » sur un échantillon de taille* 10000 *appartienne à cet intervalle.*

— *On dispose d'une pièce de monnaie que l'on sait truquée. On veut estimer la probabilité p de l'issue « Pile » pour cette pièce.*
On procède à 10000 *tirages. La fréquence observée sur cet échantillon de taille* $n = 10000$ *de l'issue « Pile » est* $f = 0,423$.
On calcule l'intervalle de confiance au seuil 95 % :

$$\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right] = \left[0.423 - \frac{1}{\sqrt{10000}}; 0.423 + \frac{1}{\sqrt{10000}}\right] = [0.413; 0.433]$$

Il y a donc 95% *de chance que la probabilité p de l'issue « Pile »appartienne à cet intervalle.*

II.3. Prise de décision à partir d'un échantillon

Propriété 1.

On considère une population dans laquelle on suppose que la proportion d'un caractère est p. On observe sur un échantillon de taille n une fréquence f du caractère.
On veut tester l'hypothèse : « La proportion de ce caractère dans la population est p ». Si I est l'intervalle de fluctuation au seuil 95% (en respectant les conditions de validité), la règle de décision est la suivante :

— Si $f \in \mathrm{I}$, on considère que l'hypothèse selon laquelle la proportion est p dans la population n'est pas remise en question (et on l'accepte au seuil de confiance 95%).

— Si $f \notin \mathrm{I}$, on rejette l'hypothèse selon laquelle cette proportion est p au seuil de confiance 95% (au risque de se tromper dans 5% des cas).

Exemple : L'entreprise A annonce que le pourcentage de moteurs défectueux dans la production est égal à 1%. Afin de vérifier cette affirmation 800 moteurs sont prélevés au hasard. On constate que 15 moteurs sont détectés défectueux. Le résultat de ce test remet-il en question l'annonce de l'entreprise A ? Justifier. On pourra s'aider d'un intervalle de fluctuation.

Solution : Pour une proportion p et un échantillon de taille n, l'intervalle de fluctuation asymptotique au seuil de 95 % est :

$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$$

sous les trois conditions : $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

L'échantillon de l'enquête est de taille $n = 800$ et l'entreprise annonce que le pourcentage de moteurs défectueux est égal à 1 % donc $p = 0,01$.

Regardons si les trois conditions sont vérifiées :
$n = 800 \geq 30$, $np = 800 \times 0,01 = 8 \geq 5$ et $n(1-p) = 800 \times 0,99 = 792 \geq 5$.

L'intervalle est : $\mathrm{I} = \left[0,01 - \frac{1}{\sqrt{800}}; 0,01 + \frac{1}{\sqrt{800}}\right] \approx [0,003; 0,017]$.

On constate que 15 moteurs sont détectés défectueux sur 800, ce qui fait une proportion de $\frac{15}{800} = 0,01875$; or $0,01875 \not\in \mathrm{I}$ donc le résultat de ce test remet en question l'annonce de l'entreprise A.