Échantillonnage & Simulation

1 Simulation d'une expérience

- Une **expérience aléatoire** est une expérience dont on ne peut pas prévoir l'issue.

  Exemple. — Lancer de dé, lancer d'une pièce de monnaie...

- Lorsque l'on répète n fois, de façon indépendante, une expérience aléatoire, on obtient une série de n résultats que l'on appelle **échantillon** de **taille** n.

  Exemple. — Séance du lundi 29 mars : on a réalisé 11 échantillons de 30 lancers de pièce.

- Lorsqu'on souhaite répéter une expérience aléatoire un grand nombre de fois (5000 lancers de dé, par exemple), on peut la simuler, par exemple avec un tableur.

  La **simulation** obtenue remplace l'expérience et permet d'étudier des séries statistiques comportant un très grand nombre de données.

2 Fluctuation d'échantillonnage

Si on réalise plusieurs échantillons de même taille, la distribution des fréquences observée sur chaque expérience varie. C'est ce qu'on appelle la **fluctuation d'échantillonnage**.

Exemple. — On effectue des séries de 30 lancers d'une pièce de monnaie et on observe la fréquence d'obtention de PILE .

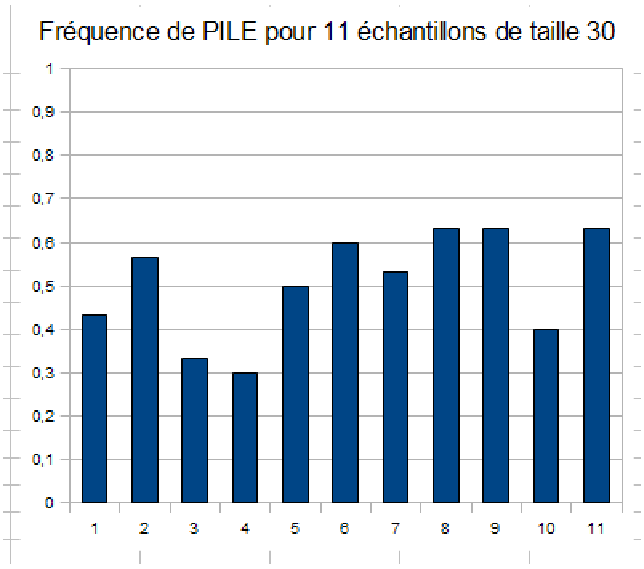

Fig. 1 – Expérience réalisée en classe.

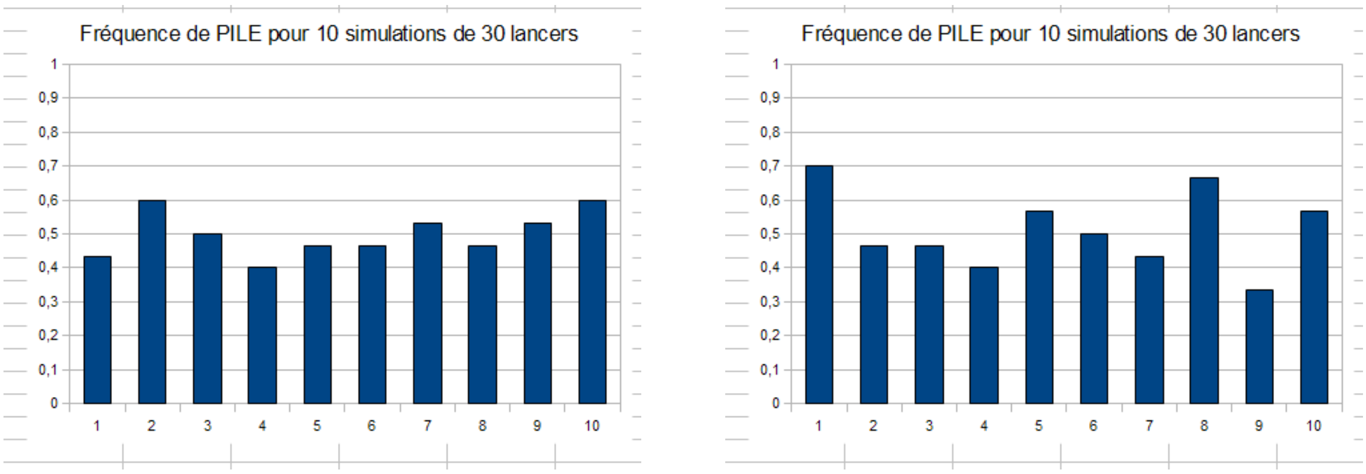

Fig. 2 – Deux exemples de simulation réalisées sur tableur.

3 Stabilisation des fréquences

On reprend l'exemple de la simulation des lancers de pièce.

En augmentant la taille des échantillons, on remarque que la fréquence d'obtention de PILE semble se stabiliser autour de 0,5 (qui est la probabilité théorique).

La fluctuation d'échantillonnage semble donc **moins importante** lorsque la taille n de l'échantillon étudié **augmente** : c'est le phénomène de **stabilisation des fréquences**.

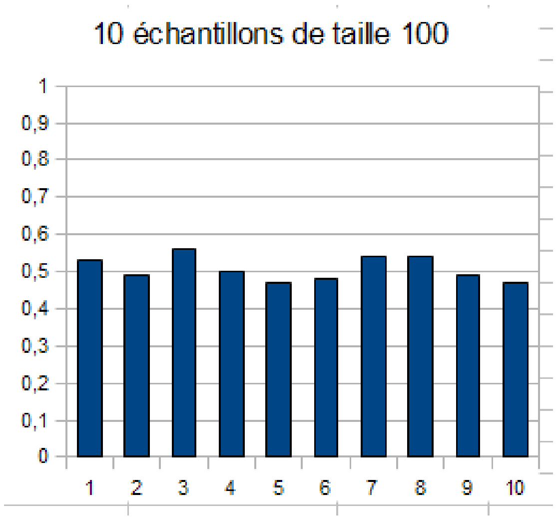

Fig. 3 – Simulation réalisée sur tableur.

Cette stabilisation des fréquences lorsque la taille de l'échantillon augmente a été **prouvée** : c'est l'objet de la partie suivante du chapitre.

4 Intervalle de fluctuation au seuil de 95 %

On souhaite étudier un caractère dans une population donnée.

Soit p la proportion du caractère dans la population.

On réalise un échantillon de taille n et on note f la fréquence observée du caractère dans cet échantillon.

Théorème

À condition que $n \geqslant 25$ et que $0,2 \leqslant p \leqslant 0,8$, il y a au moins 95 % de chances que $f \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$.

Ainsi, à condition que l'échantillon réalisé soit assez important, on est quasiment certain que la fréquence observée du caractère est proche de la proportion théorique. C'est par exemple le principe utilisé pour faire un sondage.

Exemple. — On sait que la probabilité théorique d'obtenir PILE en lançant une pièce de monnaie est 0,5.

On effectue 400 lancers de pièce de monnaie en comptant le nombre de PILE obtenus.

La taille de l'échantillon est $n = 400$, et on a : $p = 0,5$. Alors :

- $p - \frac{1}{\sqrt{n}} = 0,5 - \frac{1}{\sqrt{400}} = 0,5 - \frac{1}{20} = 0,5 - 0,05 = 0,45$.
- $p + \frac{1}{\sqrt{n}} = 0,5 + \frac{1}{\sqrt{400}} = 0,5 + \frac{1}{20} = 0,5 + 0,05 = 0,55$.

L'intervalle de fluctuation au seuil de 95 % est donc ici $[0,45 ; 0,55]$, ce qui confirme scientifiquement l'effet de « tube » perçu lors d'une simulation.

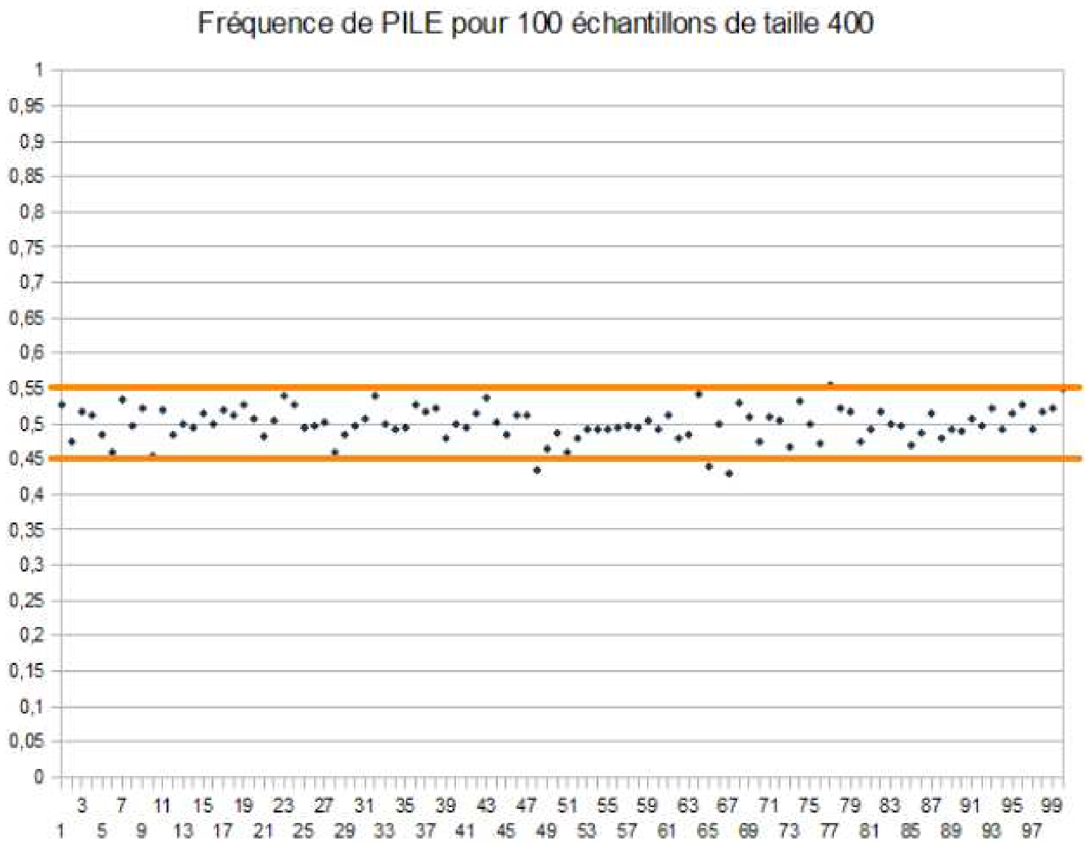

Fig. 4 – Simulation réalisée sur tableur.

Remarque. — Pour la même expérience aléatoire (lancer de pièce) mais avec un échantillon de taille 100 seulement, les calculs aboutissent à un intervalle de fluctuation au seuil de 95 % moins précis : $[0,4 ; 0,6]$ (voir la figure 3).

5 Estimation d'une proportion à partir d'un échantillon

On souhaite étudier un caractère dans une population donnée. Si on connaît la proportion « théorique » de ce caractère, le théorème évoqué dans la partie précédente nous permet d'estimer la fréquence f observée dans la pratique.

Cependant, il est souvent impossible de connaître exactement la proportion « théorique » p (par exemple, la proportion des yeux bleus dans la population française peut-elle être exactement connue ?).

Mais à partir d'un échantillon dont la fréquence observée f est connue, il est possible de trouver une estimation de la proportion « théorique » p.

En effet :

Théorème

À condition que $n \geqslant 25$ et que $0,2 \leqslant f \leqslant 0,8$, il y a au moins 95 % de chances que $p \in \left[f - \frac{1}{\sqrt{n}} \,;\, f + \frac{1}{\sqrt{n}}\right]$.

Exemple. — Sur 170 personnes âgées de plus de 60 ans, on a trouvé 34 personnes hyperglycémiques.

Donner un intervalle de fluctuation au seuil de 95 % de la proportion d'hyperglycémiques dans la population des personnes âgées de plus de 60 ans.