Echantillonnage

Définition. Lorsque l'on réalise plusieurs fois une même expérience aléatoire de manière indépendante (les différentes réalisations n'ont pas d'influence les unes sur les autres), l'ensemble des résultats obtenus est appelé **échantillon**.

Exemple. Si l'on tire au sort 1 000 personnes dans la population française et que l'on observe si la personne est droitière ou non, on obtient un échantillon de taille 1 000.

Définition. Deux échantillons (obtenus par l'expérience ou simulés) de même taille associés à une même expérience ne sont a priori pas identiques. Ce phénomène s'appelle **la fluctuation d'échantillonnage**.

Propriété et définitions. On considère un échantillon de taille n associé à une expérience aléatoire dont l'une des issues (ou l'un des événements) a pour probabilité p.
La fréquence observée f de cette issue (ou événement) dans l'échantillon est généralement proche de sa probabilité p.

Propriété. Il y a environ 95% de chance que l'écart entre p et f soit $\leq \frac{1}{\sqrt{n}}$ à condition que :
n soit grand (en général $n \geq 30$) et p ne soit ni trop faible ni trop grand (en général $0{,}2 \leq p \leq 0{,}8$),
Définition. On dit que $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ est un **intervalle de fluctuation de f** au seuil de 95 %.
Définition. On dit que $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ est un **intervalle de confiance de p** au seuil de 95 %.

Exemple. On a lancé 1000 fois un dé à 6 faces et on a obtenu 159 fois le nombre 6. La fréquence observée de 6 est donc $f = \frac{159}{1000} = 0{,}159$, ce qui est assez proche de la probabilité d'obtenir 6 qui est $p = \frac{1}{6} \approx 0{,}167$. L'écart entre p et f est $\frac{1}{6} - 0{,}159 \approx 0{,}008$, ce qui est bien inférieur à $\frac{1}{\sqrt{1000}} \approx 0{,}032$.

Remarque. On peut simuler informatiquement une expérience aléatoire à deux issues x_1 et x_2 de probabilités respectives p et $1 - p$ en générant un nombre réel aléatoire entre 0 et 1 et en considérant que : x_1 est réalisée si ce nombre aléatoire est $\leq p$ et x_2 est réalisée sinon.
On peut simuler un échantillon en répétant simplement la simulation de l'expérience dans une boucle for.

```
import random
for i in range (1,1001):
  if random.random() <= 0.88:
      print("Droitière")
  else:
      print("Non droitière")
```