III - Fluctuation d'échantillonnage

Lorsqu'on étudie un caractère d'une population, la connaissance de la population entière n'est en général pas envisageable. On doit se contenter de la connaissance d'un échantillon de cette population.

Pour prendre des décisions, il est important que l'échantillon soit prélevé au hasard

Définition : On appelle **échantillon de taille n** une liste de n résultats obtenus par répétitions indépendantes d'une même **expérience aléatoire**.

Exemple : Simulation du lancer de 100 dés à l'aide d'un tableur (LibreOffice).

On pourrait utiliser la fonction =ALEA() du tableur qui renvoie un nombre réel de l'intervalle $[0\,;\,1]$.
Mais il y a une fonction plus pratique pour notre exemple : ALEA.ENTRE.BORNES qui renvoie un entier choisi aléatoirement entre deux valeurs données.

On écrit alors dans la case A1 : =ALEA.ENTRE.BORNES(1;6)
À l'aide de la poignée de recopie, on tire la formule vers le bas jusqu'en A100. On obtient alors l'équivalent sur tableur de 100 lancers d'un dé.

Définition : Pour une population donnée, des échantillons aléatoires produits suivant le même protocole peuvent avoir des compositions différentes : on dit qu'il y a une fluctuation d'échantillonnage

En reprenant l'exemple précédent conçu à l'aide du tableur, on obtient 100 nouveaux lancers en appuyant sur les touches $\boxed{\text{Ctrl}}\boxed{\Uparrow}$ (Shift) $\boxed{\text{F9}}$.

Remarque : La fréquence f d'un caractère dans l'échantillon n'est pas nécessairement égale à la proportion p de ce caractère dans la population.

Exemple : On compte le nombre de valeurs correspondant au critère recherché. Sur le tableur, on utilise la fonction =NB.SI( ) :
par exemple pour déterminer le nombre de fois où le 6 est apparu au cours des 100 lancers présents dans la plage de saisie A1:A100, on écrit en B1 :

=NB.SI(A1:A100;6)

En B2, on calcule alors la fréquence d'apparition du nombre 6 par la formule :

=B1/100

On constate que cette fréquence varie d'une expérience à l'autre. Par contre elle reste relativement proche de la valeur $\dfrac{1}{6}$.

Exercice :

1. Écrire un algorithme sous la forme d'une fonction simulant le lancer de 100 dés bien équilibrés et calculant la fréquence d'apparition de la face numéro 6.
2. Comment modifier la fonction précédente pour l'appliquer à un nombre quelconque de tirages ?

Loi des grands nombres :
Si on répète une expérience aléatoire un grand nombre de fois, la fréquence d'un caractère fini par se stabiliser autour de la probabilité d'apparition (ou proportion) de ce caractère.

Propriété :
Soit un caractère dont la proportion dans une population donnée est p.
Lorsque n est grand, la fréquence observée f d'individus présentant le caractère étudié dans cette population est telle que $-\dfrac{1}{\sqrt{n}} \leqslant f - p \leqslant \dfrac{1}{\sqrt{n}}$ dans une majorité des cas.

Repartons de la simulation de 100 lancers d'un dé bien équilibré.
Pour chaque simulation, la fréquence f obtenue devrait être proche de $\dfrac{1}{6}$.
D'après la propriété précédente, cette fréquence devrait vérifier :

Construire une fonction qui réalisera 500 fois cette expérience et qui déterminera le nombre de fois où f est dans l'intervalle $\left[\dfrac{1}{15}\,;\,\dfrac{4}{15}\right]$.