

1 Fluctuations

1.1 Tirage dans une urne

Une urne contient 40 % de boules noires et 60 % de boules blanches. On prélève au hasard une boule (chaque boule a les mêmes chances d'être prélevée).

1. Que veut dire *random* en anglais ?
2. Pour simuler le tirage d'une boule dans cette urne, recopier le programme suivant (qui n'est pas encore correct).

```
from random import random

if random() < .3:
    print("Boule noire")
else:
    print("Boule blanche")
```

3. Exécuter le programme plusieurs fois, et vérifier qu'il affiche aléatoirement Boule noire et Boule blanche.
4. Modifier le programme pour respecter les probabilités de l'urne.
5. 🙌 Appeler le professeur pour vérifier votre travail.

1.2 Échantillon

On souhaite cette fois simuler le tirage de nombreuses boules, avec remise, dans cette urne, et compter le nombre de boules blanches obtenues.

1. En moyenne, en tirant 100 boules dans cette urne, combien de boules blanches devrait-on obtenir ?
2. Effacer le programme précédent, et recopier le programme suivant.

```
from random import random

def echantillon():
    blanches = 0
    for i in range(100):
        if random() < .6:
            blanches = blanches + 1
    return blanches

print(echantillon())
```

3. Exécuter ce programme plusieurs fois, et vérifier qu'il renvoie environ le nombre obtenu à la question 1.
4. Modifier ce programme pour que la fonction `echantillon()` prenne en argument le nombre de répétitions de l'expérience, et remplacer la dernière ligne par `print(echantillon(1000))` pour simuler 1000 répétitions.
5. Modifier ce programme pour qu'au lieu de renvoyer le *nombre* de boules blanches, il renvoie la *fréquence* de boules blanches.
6. 🙌 Appeler le professeur pour vérifier votre travail.

1.3 Fluctuation d'échantillonnage

On veut déterminer la probabilité que la fréquence de boules blanches obtenue à la partie précédente soit « éloignée » de la moyenne théorique 0,6.

1. Conserver le programme précédent, supprimer la ligne `print(...)`, et ajouter la fonction suivante.

```
def fluctuation(échantillon, répétitions, seuil):
    éloigné = 0
    for i in range(répétitions):
        if abs(.6-echantillon(échantillon)) > seuil:
            éloigné = éloigné + 1
    return éloigné / répétitions

print(fluctuation(100, 1000, .01))
```

2. Expliquez la ligne `if ...` du programme précédent.
3. Sur 50 échantillon de taille 1000, quelle est la fréquence d'échantillons éloignés de plus de 1% de la moyenne théorique ?
4. Même question avec 1%.
5. (a) Faites varier le nombre d'échantillon. Quand ce nombre augmente, la fréquence observée se rapproche-t-elle ou s'éloigne-t-elle de la fréquence théorique ?
(b) Faites varier la taille de chaque échantillon. Quand ce nombre augmente, la fréquence observée se rapproche-t-elle ou s'éloigne-t-elle de la fréquence théorique ?

2 Inquiétudes à Woburn

Woburn est une petite ville industrielle du Massachusetts, au Nord-Est des Etats-Unis. Du milieu à la fin des années 1970, la communauté locale s'émeut d'un grand nombre de leucémies infantiles survenant dans certains quartiers de la ville. Les familles se lancent alors dans l'exploration des causes et constatent la présence de décharges et de friches industrielles ainsi que l'existence de polluants.

Dans un premier temps, les experts gouvernementaux concluent qu'il n'y a rien d'étrange. Mais les familles s'obstinent et saisissent leurs propres experts.

L'objectif de ce paragraphe est que vous apportiez votre expertise statistique !

Les statistiques concernant les garçons de cette ville sont les suivantes : sur 5969 garçons, on a observé 9 leucémies. La fréquence des leucémies dans l'ensemble des États-Unis est 0,00052.

On peut considérer que les garçons de Woburn constituent un échantillon de taille $n = 5969$ extrait de la population des États-Unis.

1. On décide de simuler la naissance de 5969 garçons (ayant une probabilité 0,00052 d'avoir une leucémie). Utilisez le programme de la première partie pour faire cette simulation. Répétez-la plusieurs fois. Combien de cas de leucémie obtenez-vous ?
2. Quelle est la différence (en points) entre le nombre de leucémies *théorique* et le nombre de leucémies *observée* à Woburn ?
3. En utilisant le programme de la première partie, calculer la probabilité que le nombre de leucémies sur 5969 soit aussi écarté de la moyenne qu'à Woburn ?
4. Conclure : Le nombre de leucémies observées à Woburn est-il anormal ?