ÉCHANTILLONNAGE

I Échantillon

Définition n°1. ***Notion d'échantillon***

Soit n un entier naturel non nul. On considère une expérience aléatoire à deux issues, que l'on peut répéter de manière indépendante (c'est-à-dire que la probabilité de chaque issue ne dépend pas des résultats précédemment obtenus).
Un échantillon de taille n est constitué des résultats obtenus par n répétitions de cette expérience aléatoire.

Exemple n°1.

On considère l'expérience aléatoire suivante « On lance une pièce de monnaie et on regarde si elle tombe sur Pile ou Face. »
Cette expérience peut bien être répétée de manière indépendante car le résultat d'un lancer ne dépend pas des lancers précédents.
Si on répète dix fois cette expérience, les résultats obtenus constituent un échantillon de taille 10.
Par exemple, $(P\,;F\,;P\,;P\,;P\,;F\,;P\,;F\,;P\,;P)$ est un échantillon de taille 10.

Remarque n°1.

Lorsqu'on s'intéresse à un caractère d'une population, celle-ci est souvent trop grande pour pouvoir être étudiée dans sa totalité. On observe alors ce caractère sur une partie de cette population, choisie de manière aléatoire, en considérant que la population est suffisamment grande pour que ce choix puisse être assimilé à un tirage avec remise. Les résultats obtenus constituent un échantillon.

Propriété n°1. ***Influence de la taille de l'échantillon (admise)***

On considère une expérience aléatoire à deux issues, que l'on peut répéter de manière indépendante.
Soit p la probabilité d'une issue ω .
Soit n un entier naturel non nul.
On considère un échantillon de taille n et on note f la fréquence de l'issue ω dans cet échantillon.
Lorsque n est grand, sauf exception, la fréquence observée f est proche de la probabilité p .

Exemple n°2.

On reprend l'expérience aléatoire de l'exemple précédent et on constitue un échantillon de taille 1000.
Dans cet échantillon, on observe que l'on obtient 531 fois l'issue « Pile ».
La fréquence observée de « Pile» pour cet échantillon est $f=\frac{531}{1000}=0{,}531$.
Si la pièce est bien équilibrée, la probabilité d'obtenir « Pile » est $p=0{,}5$.
On constate que, pour cet échantillon, la fréquence observée f est proche de la probabilité p .

Remarque n°2.

▪ La fréquence observée varie entre deux échantillons de même taille. Il est donc naturel que la fréquence observée f pour un échantillon ne soit pas exactement égale à la probabilité p .
▪ Plus la taille de l'échantillon est grande, plus il y a de chances que la fréquence observée soit proche de la probabilité.

II Principe de l'estimation

Propriété n°2. Fluctuation d'échantillonnage (admise)

On considère une expérience aléatoire à deux issues, que l'on peut répéter de manière indépendante et dont on connaît la probabilité p d'une issue ω.
On constitue un grand nombre d'échantillons de taille n sur lesquels on observe la fréquence f de réalisation de l'issue ω.
Plus la taille n des échantillons est grande, moins il y a de fluctuation de la fréquence observée f autour de la valeur de p.

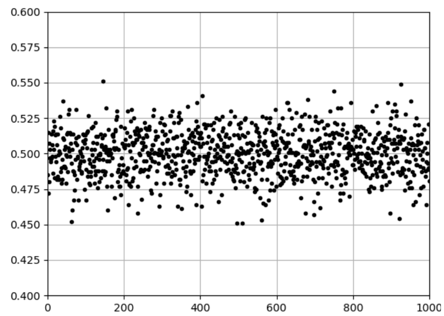
Exemple n°3.

On considère l'expérience aléatoire consistant à lancer une pièce équilibrée et à noter si on retombe sur « Pile » ou « Face ».
On sait que la probabilité p d'obtenir « Face» est égale à 0,5 .

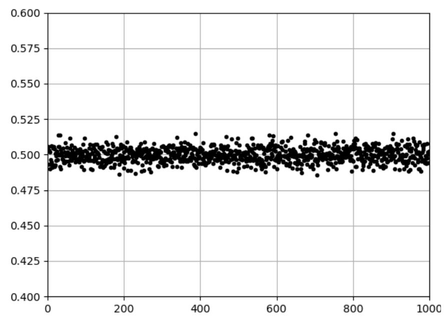
À l'aide d'un ordinateur, on simule 1 000 échantillons de taille 1 000 et, pour chaque échantillon on calcule la fréquence d'apparition de « Face ».
On peut observer que, sur ces échantillons environ 95% de ces fréquences appartiennent à l'intervalle $[0{,}47\ ;\ 0{,}53]$.

On simule à présent 1 000 échantillons de taille 10000 et, pour chaque échantillon, on calcule la fréquence d'apparition de « Face ».
On peut observer cette fois qu'environ 95 % de ces fréquences appartiennent à l'intervalle $[0{,}49\ ;\ 0{,}51]$ les fréquences observées sont plus proches de la valeur de p .

1000 échantillons de taille 1000



1000 échantillons de taille 10 000



Propriété n°3. Estimation d'une proportion (admise)

On considère une population dans laquelle on cherche la proportion des individus qui possèdent un certain caractère.
On prélève au hasard un échantillon de taille n dans la population et on observe la fréquence f du caractère dans cet échantillon. Cette fréquence f est une valeur approchée de p , appelée estimation ponctuelle de p .
Plus la taille de l'échantillon est grande, meilleure est l'estimation de p .

Exemple n°4.

Un candidat se présente à une élection dans une ville de 10000 habitants. Un sondage réalisé sur 1100 habitants montre que 517 personnes envisagent de voter pour ce candidat.

La fréquence observée sur cet échantillon est $f=\dfrac{517}{1100}=0{,}47$.

On peut estimer qu'une valeur approchée de la proportion des habitants souhaitant voter pour ce candidat est 47% . La qualité de cette estimation dépend de la taille de l'échantillon.

Remarque n°3.

De la même façon, pour une expérience aléatoire donnée, on peut estimer la probabilité d'une issue en observant sa fréquence dans un échantillon de taille suffisamment grande.

ÉCHANTILLONNAGE E01

EXERCICE N°1

L’étiquetage d’un sac de 100 kg de café précise qu’il contient 30 % de Robusta et 70 % d’Arabica. On prélève une poignée de 153 grains et on obtient 45 grains de Robusta.

1) Indiquer la population étudiée, le caractère étudié, sa proportion théorique et la taille de l’échantillon.

2) Calculer la fréquence du caractère dans cet échantillon.

EXERCICE N°2

On lance un dé équilibré et on observe si on obtient un résultat pair ou non.

1) Compléter la fonction python ci-dessous pour qu'elle simule cette expérience aléatoire et renvoie la fréquence de l'évènement « Obtenir un résultat pair » dans un échantillon de taille 5000.

```
from random import random
def de_pair():
    nombre = 0
    for compteur in range(...):
        if random() < ... :
            nombre = nombre + 1
    return ...
```

2) Exécuter 10 fois cette fonction et donner le plus petit intervalle auquel appartiennent tous les résultats renvoyés par la fonction.

ÉCHANTILLONNAGE E01

EXERCICE N°3

1) Que permet d'estimer l'algorithme ci-dessous ?

```
Somme ← 0
Pour compteur allant de 1 à 1000
    A ← nombre entier aléatoire entre 1 et 6
  Si A = 6
      Somme ← Somme + 1
Fréquence ← Somme / 1000
Afficher Fréquence
```

2) Comment pourrait-on améliorer la précision de cette estimation ?

ÉCHANTILLONNAGE E01

EXERCICE N°4 Des maths en justice !

En 1976, un accusé d'origine mexicaine condamné pour différents crimes au Texas attaqua le jugement sous le motif que la désignation des jurés dans l'état tu Texas était discriminatoire pour les Américains d'origine mexicaine.
Son argument était que ceux-ci n'était pas pas suffisamment représentés dans les jurys populaires.
À cette époque, au Texas, 79,1 % de la population était mexico-américaine, et parmi les 870 personnes convoquées en tant que « grands jurées » (sur une période de 11 ans), 339 étaient d'origine mexicaine.

La constitution des jurys pouvait-elle être considérée comme impartiale ?

À lire avant un éventuel débat ou commentaire avec la classe : ce lien