

# Échantillonnage

DURÉE ESTIMÉE

15 minutes

## OBJECTIFS DU CHAPITRE

Calculer la fréquence de succès dans un échantillon

Estimer une probabilité par une fréquence

Simuler des échantillons en Python

# 1 - Échantillon et fréquence

## Définition

On considère une expérience aléatoire à **deux issues** : l'une appelée **succès**, l'autre **échec**. On note  $p$  la **probabilité** du succès.

Répéter  $n$  fois cette expérience, de façon indépendante et dans les mêmes conditions, constitue un **échantillon de taille  $n$** .

## Définition

Pour un échantillon de taille  $n$ , si le succès est obtenu  $k$  fois, la **fréquence** du succès dans cet échantillon est :

$$f = \frac{k}{n}$$

## Exemple

On lance 50 fois une pièce équilibrée. Le succès est « obtenir Pile », de probabilité  $p = 0,5$ .

Si Pile apparaît 27 fois, la fréquence observée de Pile dans cet échantillon de taille  $n = 50$  est :

$$f = \frac{27}{50} = 0,54$$

La fréquence  $f = 0,54$  est proche de la probabilité  $p = 0,5$ , mais elle ne lui est pas exactement égale.

### Remarque

Il ne faut pas confondre  $p$  et  $f$  :

$p$  est une valeur **théorique**, fixée par l'expérience, toujours la même.

$f$  est une valeur **observée**, qui dépend de l'échantillon et change d'un échantillon à l'autre.

## 2 - Fluctuation d'échantillonnage

### Propriété

Deux échantillons de même taille  $n$  d'une même expérience donnent en général des fréquences de succès **différentes**. Ce phénomène s'appelle la **fluctuation d'échantillonnage**.

### Exemple

On simule 5 échantillons de taille  $n = 100$  pour le lancer d'une pièce équilibrée ( $p = 0,5$ ). On peut obtenir, par exemple, les fréquences de Pile suivantes :

0,48 ; 0,53 ; 0,47 ; 0,51 ; 0,50

Toutes ces fréquences sont proches de 0,5, mais elles varient d'un échantillon à l'autre : c'est la fluctuation d'échantillonnage.

### Propriété

Pour un échantillon de taille  $n$ , l'écart entre la fréquence observée  $f$  et la probabilité  $p$  est, le plus souvent, de l'ordre de :

$$\frac{1}{\sqrt{n}}$$

Autrement dit, dans la grande majorité des échantillons, on observe

$|f - p| \leq \frac{1}{\sqrt{n}}$ . Plus  $n$  est grand, plus ce seuil  $\frac{1}{\sqrt{n}}$  est petit, donc plus la

fréquence a de chances d'être proche de  $p$ .

### Exemple

Pour un échantillon de taille  $n = 100$ , le seuil vaut :

$$\frac{1}{\sqrt{100}} = \frac{1}{10} = 0,1$$

On s'attend donc à ce que la plupart des fréquences observées soient comprises entre  $0,5 - 0,1 = 0,4$  et  $0,5 + 0,1 = 0,6$ , ce qui est bien le cas dans l'exemple précédent.

Pour  $n = 10\,000$ , le seuil n'est plus que de  $\frac{1}{\sqrt{10\,000}} = \frac{1}{100} = 0,01$  : la fluctuation est dix fois plus faible.

### 3 - Loi des grands nombres (version vulgarisée)

#### Propriété

**Loi des grands nombres (énoncé vulgarisé).**

Lorsque la taille  $n$  de l'échantillon est **grande**, sauf exception, la fréquence observée  $f$  du succès est **proche** de sa probabilité  $p$ .

#### Remarque

La loi des grands nombres ne garantit pas que  $f$  sera exactement égale à  $p$  : il restera toujours une petite fluctuation. Elle affirme seulement que cette fluctuation devient de plus en plus faible à mesure que  $n$  augmente. Le seuil  $\frac{1}{\sqrt{n}}$  donne un ordre de grandeur de cette fluctuation.

#### Exemple

Pour le lancer d'une pièce équilibrée ( $p = 0,5$ ), on relève la fréquence de Pile pour des échantillons de tailles croissantes :

Plus  $n$  augmente, plus la fréquence observée se rapproche de  $p = 0,5$ .

## 4 - Estimer une probabilité par une fréquence

Dans de nombreuses situations, on ne connaît pas la probabilité  $p$  d'un succès (pièce truquée, proportion de pièces défectueuses dans une usine, intention de vote...). La loi des grands nombres permet alors d'en obtenir une **estimation**.

### Propriété

Si la taille  $n$  de l'échantillon est grande, la fréquence observée  $f$  donne une **estimation** de la probabilité (ou de la proportion)  $p$  inconnue :

$$p \approx f$$

L'estimation est d'autant plus fiable que  $n$  est grand.

### Exemple

Une usine veut estimer la proportion  $p$  de pièces défectueuses dans sa production. Elle prélève un échantillon de  $n = 500$  pièces et en trouve 15 défectueuses.

La fréquence observée de pièces défectueuses est :

$$f = \frac{15}{500} = 0,03$$

On estime donc la proportion de pièces défectueuses à environ 3 % de la production.

### Remarque

Cette estimation reste soumise à la fluctuation d'échantillonnage : un autre échantillon de 500 pièces donnerait probablement une fréquence un peu différente. Pour réduire l'incertitude, il faut augmenter la taille de l'échantillon.

## 5 - Simuler sur Python

On peut observer la loi des grands nombres en **simulant** une expérience aléatoire sur ordinateur. Le module *random* fournit la fonction *randint(a, b)* qui renvoie un entier au hasard entre *a* et *b* (bornes incluses).

Simulons le lancer d'un dé équilibré, où le succès est « obtenir un 6 », de probabilité  $p = \frac{1}{6} \approx 0,167$ . La fonction suivante simule un échantillon de taille *n* et renvoie la fréquence de succès :

```
from random import randint

def frequence_six(n):
    succes = 0                # on compte les succès
    for i in range(n):       # on répète l'expérience n fois
        if randint(1, 6) == 6: # un lancer du dé
            succes = succes + 1
    return succes / n        # fréquence observée
```

En appelant cette fonction pour des tailles d'échantillon de plus en plus grandes, on constate que la fréquence renvoyée se rapproche de  $p \approx 0,167$  :

```
print(frequence_six(10))      # par exemple 0.1
print(frequence_six(1000))   # par exemple 0.172
print(frequence_six(100000)) # par exemple 0.1668
```

### 📌 Remarque

Les résultats affichés changent à chaque exécution (le hasard intervient) : c'est encore la fluctuation d'échantillonnage. Mais pour les grandes valeurs de  $n$ , la fréquence reste à chaque fois très proche de 0,167, ce qui illustre la loi des grands nombres.

On peut aussi simuler  $N$  échantillons de taille  $n$  et compter la proportion de ceux pour lesquels l'écart entre la fréquence  $f$  et la probabilité  $p$  est inférieur ou égal

au seuil  $\frac{1}{\sqrt{n}}$  :

```
p = 1 / 6
n = 100
N = 1000
seuil = 1 / n ** 0.5      # le seuil 1/racine(n)

dans_le_seuil = 0
for j in range(N):      # on génère N échantillons
    f = frequence_six(n) # fréquence du j-ième échantillon
    if abs(f - p) <= seuil: # l'écart est-il sous le seuil ?
        dans_le_seuil = dans_le_seuil + 1

print(dans_le_seuil / N) # proportion proche de 0.95
```

On observe que cette proportion est très élevée (de l'ordre de 0,95) : pour la grande majorité des échantillons, la fréquence  $f$  est bien à une distance d'au plus  $\frac{1}{\sqrt{n}}$  de la probabilité  $p$ .

### 1. Comment calculer la fréquence de succès dans un échantillon ?

On divise le nombre de succès  $k$  par la taille  $n$  de l'échantillon :

$$f = \frac{k}{n}.$$

Voir la fiche méthode : [Calculer la fréquence de succès dans un échantillon](#)

### 2. Comment estimer une probabilité inconnue ?

On prélève un grand échantillon et on calcule la fréquence observée  $f$  : la loi des grands nombres permet d'affirmer que  $p \approx f$ .

Voir la fiche méthode : [Estimer une probabilité par une fréquence](#)

### 3. Comment simuler des échantillons en Python ?

On utilise la fonction `randint` du module `random` dans une boucle pour répéter l'expérience  $n$  fois, on compte les succès, puis on calcule la fréquence.

Voir la fiche méthode : [Simuler des échantillons en Python](#)

↓ Télécharger en PDF