Chapitre 12

ÉCHANTILLONNAGE

On considèrera dans ce chapitre uniquement des **expériences aléatoires à deux issues**.
L'une est appelée **succès**, de probabilité $p \in [0;1]$ et l'autre, l'**échec**, est de probabilité $1-p$.

I - Fréquences observées

Définition : Échantillon

Un **échantillon** de taille n ($n \in \mathbb{N}^*$) est constitué des résultats obtenus par n répétitions indépendantes d'une même expérience aléatoire.

Remarque

Beaucoup de paramètres entrent en jeu dans la réalisation d'un échantillon : coût, temps, représentativité... En mathématiques, on procède virtuellement par **simulation informatique**. L'**échantillonnage** est utilisé, par exemple, pour les sondages : prévoir le résultat d'une élection, étudier les habitudes des consommateurs, tester la qualité d'objets ou de services produits par une entreprise...

Exemple

Un lancer de pièce de monnaie donne deux issues possibles : *pile* (**P**) ou *face* (**F**). **P**;**P**;**F**;**P**;**P**;**F**;**F**;**P** est un échantillon de taille 8 de cette expérience aléatoire. Cet échantillon possède les fréquences suivantes.

Issue	**P**	**F**
Fréquence	0,625	0,375

Définition : Fréquence de succès observée

La **fréquence de succès observée**, notée f_s, d'un échantillon de **taille** n avec k **succès** est :

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

Remarque

Considérons que *pile* soit le **succès** de l'expérience aléatoire du lancer de pièce. Ici, la pièce est truquée et $p = 0.2$.
On donne les fréquences de succès observées pour différents échantillons de taille 30 :

Échantillon de taille 30	numéro 1	numéro 2	numéro 3	numéro 4
Nombre de succès **P**	11	9	5	6
f_s	0,275	0,225	0,125	0,15

Les fréquences de succès ne sont pas les mêmes : elles varient selon l'échantillon. On appelle ce phénomène **fluctuation de l'échantillonnage**. Si on augmente les tailles des échantillons, nous pouvons obtenir des fréquences de succès plus proches les unes des autres mais aussi de p :

Échantillon de taille 500	numéro 1	numéro 2	numéro 3	numéro 4
Nombre de succès **P**	109	95	104	98
f_s	0,218	0,19	0,208	0,196

II - Estimation

Théorème : Loi des grands nombres

Quand on répète un **grand nombre** de fois une expérience aléatoire, sauf exception, la **fréquence de succès observée** de l'échantillon se stabilise autour de p.

Remarque : Estimation d'une probabilité

Si on ne connaît pas p, on peut essayer de le déterminer expérimentalement. La **loi des grands nombres** nous permet de l'**estimer** en calculant la fréquence de succès observée d'un échantillon de taille très grande.

Exemple

Une urne contient 8 boules colorées : 2 rouges, 3 bleues, 1 verte et 2 jaunes. On tire successivement et sans remise deux boules. Soit S l'événement : « On a tiré au moins une boule rouge ». La fonction suivante permet de simuler n fois l'expérience et renvoie la fréquence observée de S.

```python
from random import *
def simulation(n):
    k=0
    urne=['rouge','rouge','bleue','bleue','bleue','verte','jaune','jaune']
    for i in range(n):
        boule1=choice(urne)
        boule2=choice(urne)
        if boule1=='rouge' or boule2=='rouge':
            k=k+1
    return k/n
```

Voici plusieurs résultats de la commande `simulation(10000)` : 0,4393, 0,4378, 0,4339.
On donne ainsi une estimation de $\mathbb{P}(S)$:

$$\mathbb{P}(S) \simeq 0{,}44.$$

Méthode : Estimation d'une proportion

Dans une population, on note p la **proportion** théorique d'individus ayant un **caractère donné**. On considère un échantillon de taille n dans cette population et on calcule la fréquence f du caractère dans cet échantillon.

Théorème

Si $n \geqslant 25$ et $0{,}2 \leqslant p \leqslant 0{,}8$, alors f_s appartient à l'intervalle $I = \left[p - \dfrac{1}{\sqrt{n}}\,;p + \dfrac{1}{\sqrt{n}}\right]$ dans environ 95% des cas.

I est appelé **intervalle de fluctuation au seuil de** 95%.

Remarque

Si $f_s \in \left[p - \dfrac{1}{\sqrt{n}}\,;p + \dfrac{1}{\sqrt{n}}\right]$, alors $|f_s - p| \leqslant \dfrac{1}{\sqrt{n}}$.