❉ Chapitre 21 ❉

Échantillonnage

Objectif du chapitre :

- Connaitre le sens d'échantillon, taille d'un échantillon
- Comprendre la loi des grands nombres.

I. Échantillon

Définition 1:

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 de n répétitions indépendantes de la même expérience aléatoire.

Exemple 1:

On considère l'expérience suivante : « On lance une pièce de monnaie ». On peut la répéter de manière indépendante car le résultat d'un lancer ne dépend pas des lancers précédents.
Si on répète 10 fois cette expérience, les résultats 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 :

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 1 :

On considère une expérience aléatoire à deux issues, que l'on peut répéter de manière indépendante. Soit p la probabilité d'une issue. Soit n un entier naturel non nul.
On considère un échantillon de taille n et on note f la fréquence de l'issue dans cet échantillon.
Lorsque n est grand, sauf exception, la fréquence observée f est proche de la probabilité p.
La plupart du temps, l'écart entre p et f est inférieur ou égal à $\dfrac{1}{\sqrt{n}}$.

Exemple 2:

Reprenons la même expérience que dans l'exemple précédent et on constitue un échantillon de taille 1 000. Dans cette échantillon, on observe que l'on obtient 531 fois l'issue « Pile ».
La fréquence observée de « Pile » pour cet échantillon est $f = \dfrac{531}{1\,000} = 0,531$.
Si la pièce est bien équilibrée, la probabilité d'obtenir « Pile » est $0,5$.
Pour cet échantillon, la fréquence observée est proche de la probabilité p.

Remarque :

- 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 2 :

On considère une expérience aléatoire à deux issues, que l'on peut répéter de manière indépendante et dont on connaît la probabilité d'une issue ω (on lit « oméga »).
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 3:

Reprenons la même expérience que dans les exemples précédents.
On constitue 1 000 échantillons de taille 1 000, et pour chaque échantillon, on calcule la fréquence d'apparition de « Pile ». On observe alors que, sur ces échantillons, environ 95 % de ces fréquences appartiennent à l'intervalle $[0,47\,;\,0,53]$.
On a ensuite constitué 1 000 échantillons de taille 10 000. Sur ces échantillons, environ 95 % des fréquences appartiennent à l'intervalle $[0,49\,;\,0,51]$: les fréquences observées sont plus proches de la valeur de p.

Propriété 3 :

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 4:

Un candidat se présente à une élection dans une ville de 100 000 habitants. Un sondage réalisé sur 1 100 habitants montre que 517 d'entre eux envisagent de voter pour ce candidat.
La fréquence observée sur cet échantillon est $f = \dfrac{517}{1\,100} = 0,47$.
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 :

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.

III. Fluctuation d'échantillonnage

Définition 2:

Deux échantillons (obtenues par l'expérience ou simulés) de même taille associés à une expérience aléatoire ne sont a priori, pas identiques : ce phénomène s'appelle la fluctuation d'échantillonnage.

Exemple 5:

On lance un dé numéroté de 1 à 6, bien équilibré, et on repère le chiffre qui apparaît sur la face supérieure.
On répète ce lancer deux fois 100 fois et on obtient deux échantillons A et B de taille 100.
On a noté les fréquences d'apparition de chaque chiffre dans un tableau de distribution des fréquences :

Chiffre	1	2	3	4	5	6
Fréquence A	$0,14$	$0,17$	$0,19$	$0,18$	$0,17$	$0,15$
Fréquence B	$0,15$	$0,16$	$0,16$	$0,18$	$0,17$	$0,18$

Dans l'exemple précédent, on constate que les distributions des fréquences des deux échantillons ne sont pas les mêmes : c'est ce qu'on appelle la **fluctuation d'échantillonnage**.

La moyenne de l'échantillon A est de $3,52$ et celle de B est $3,60$.