Chapitre 14 - FLUCTUATION D'ÉCHANTILLONNAGE - LOI DES GRANDS NOMBRES

1) ÉCHANTILLON ET DISTRIBUTION DE FRÉQUENCES

Définition :

Un **échantillon de taille** n est la série statistique formée des n résultats obtenus lorsqu'on répète n fois une expérience dans les mêmes conditions.

Exemple :
On lance un dé et on regarde si le nombre obtenu sur la face supérieure du dé est pair ou impair.
On répète 100 fois cette expérience.
On obtient successivement :

P P P I P I P I P P P I P P P I P I I P P P P I P I I P I P P I P I P I P I P I P I I P I P I P I P P P P I P I P I P P P I P P P I P P I P P I P I P I I P I P P I P I P I P I I I P I I P I P I P I I

Cette série constitue un échantillon de taille 100.

Remarque :
Pour obtenir un échantillon à l'aide d'un tirage celui-ci doit s'effectuer avec remise pour que la proportion du caractère considéré ne change pas.
Si la taille de l'échantillon est négligeable devant l'effectif total, on peut assimiler un tirage sans remise à un tirage avec remise.

Définition :

La distribution des fréquences associée à un échantillon est la liste des fréquences des issues de l'échantillon.

Exemple :
Dans l'exemple précédent, on peut calculer les fréquences de chacune des issues P et I.
Ce tableau donne la distribution des fréquences associée à l'échantillon.

Issue	P	I
f_i	0,48	0,52

2) FLUCTUATION D'ÉCHANTILLONNAGE

Définition :

Si on réalise plusieurs échantillons de même taille pour une même expérience, la distribution des fréquences varie.
C'est ce qu'on appelle **la fluctuation d'échantillonnage.**

Exemple :
On peut obtenir un nouvel échantillon de taille 100 en relançant 100 fois le dé.
On obtient successivement :

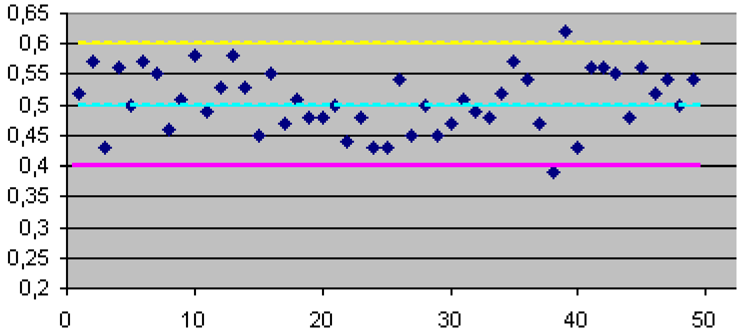
I I P I P I P I P I P I P I P I P I I P P P P I P I I P I P P I P I P I P I P I P IPI P I P I P I P I P P I P I P I P I P I P P P I P P I P P I P I P I I I I P P I P I P I P I P I P IPI P I P I P IPP

On obtient une nouvelle distribution des fréquences.

Issue	P	I
f_i	0,47	0,53

La distribution obtenue est différente.
On dit que les **fréquences fluctuent**.

On simule maintenant 50 échantillons de taille 100 de cette expérience et on s'intéresse à la fréquence d'apparition d'un nombre pair.
On se rend compte que les valeurs des fréquences obtenues sont assez proches de la valeur $\frac{1}{2}$.



Chaque point est associé à un échantillon.
On observe ici 2 points en dehors de l'intervalle $[0,4 ; 0,6]$

On constate que pour la plupart des échantillons (au moins 95%) la fréquence est comprise dans l'intervalle $[0,4 ; 0,6]$.

Propriété : (admise)

On considère une population dont la fréquence d'apparition d'un caractère donné est p et un échantillon de taille n, dont la fréquence d'apparition du caractère précédent est f.

Lorsque n est assez grand et p n'est ni proche de 0, ni de 1, il y a environ 95% des échantillons de taille n issus de cette population qui sont tels que la fréquence f du caractère appartient à un intervalle centré en p de la forme :

$$\left]p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}\right[$$

Remarque :

Quand la taille de l'échantillon augmente, l'ampleur des fluctuations des distributions de fréquences calculées sur ces échantillons de taille n diminue et les fréquences ont tendance à se stabiliser.

3) LOI DES GRANDS NOMBRES

On constate que si l'on répète un grand nombre de fois une expérience donnée, les différentes fréquences d'apparition ont tendance à se stabiliser, ce qui n'est parfois que théorique car répéter une expérience aléatoire dans les mêmes conditions n'est pas toujours envisageable.

Ce constat est un résultat mathématique appelé " **La loi des grands nombres** " ; Il se démontre, mais la démonstration est largement hors de vos compétences :

Propriété :

Pour une expérience donnée, dans le modèle défini par une loi de probabilité, les distributions des fréquences calculées sur des séries de taille n se rapprochent de la loi de probabilité quand n devient grand.

4) APPLICATIONS

- **Valider ou rejeter un modèle choisi**

A partir d'un échantillon pour lequel la fréquence relative à un caractère donné est f, on regarde si cette fréquence est dans l'intervalle $\left[p-\frac{1}{\sqrt{n}} ; p+\frac{1}{\sqrt{n}}\right]$. Si ce n'est pas le cas, on considère que l'observation n'est pas compatible avec le modèle.

- **Estimer la valeur d'une probabilité ou d'une proportion inconnue**

On a : $p-\frac{1}{\sqrt{n}} \leqslant f \leqslant p+\frac{1}{\sqrt{n}} \Leftrightarrow f-\frac{1}{\sqrt{n}} \leqslant p \leqslant f+\frac{1}{\sqrt{n}}$

L'équivalence précédente permet d'estimer, la valeur d'une proportion inconnue p dans la population à partir d'une fréquence observée f.
La précision de l'estimation vaut $\frac{1}{\sqrt{n}}$. Elle est d'autant meilleure que n est grand.