Séquence 12 : Échantillonnage

I) Fluctuation d'échantillonnage

Définition :
Un échantillon de taille n est la liste de n résultats obtenus par n répétitions indépendantes de la même expérience aléatoires.

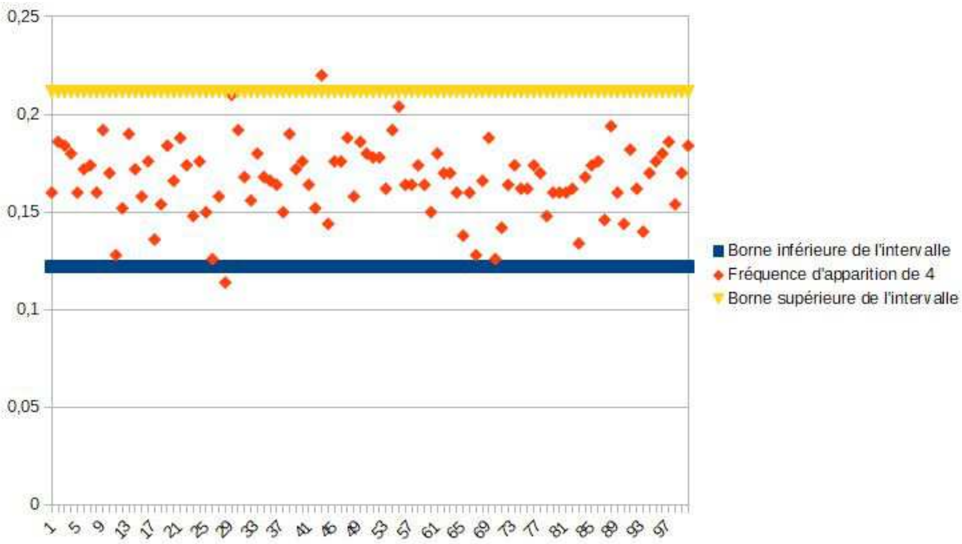
Propriété :
Si on analyse un grand nombre d'échantillons de taille n ($n \geqslant 25$) et que l'on observe à chaque fois la fréquence f d'un caractère dans la population, on s'aperçoit que pour une proportion théorique p du caractère comprise entre 0.2 et 0.8 environ 95% des fréquences observées se situent dans l'intervalle :

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

Cet intervalle est appelé intervalle de fluctuation au seuil de 95 %.

Exemple :

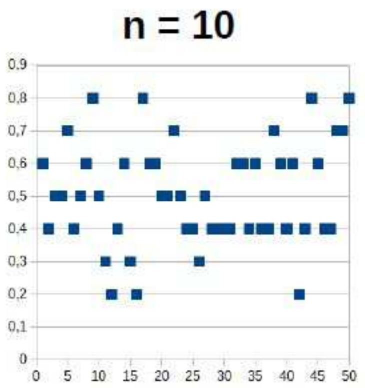
100 élèves ont lancés 150 fois un dé cubique équilibré. (100 échantillons de taille 150).

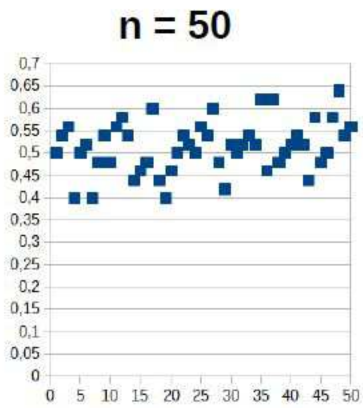


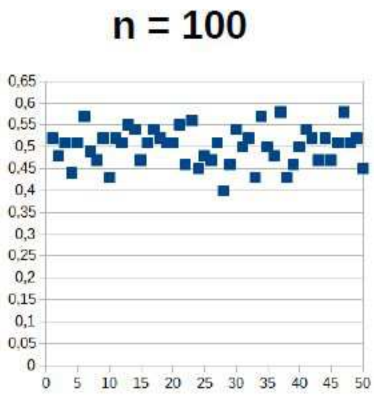
Déterminons l'intervalle de fluctuation au seuil de 95 %.

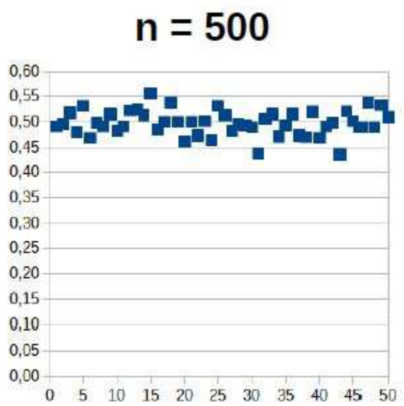
Remarques :

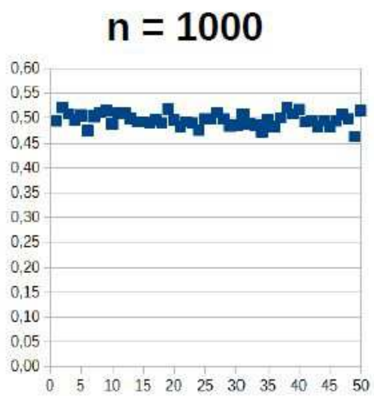
- La fréquence observée f d'un caractère varie d'un échantillon à l'autre : c'est la fluctuation d'échantillonnage.
- Les fluctuations diminuent lorsque la taille n des échantillons augmente.
- On peut changer le seuil en 90 % , 99 %.











Propriété :
On considère une population dans laquelle on **suppose** que la proportion d'un caractère est p.
On observe f comme fréquence de ce caractère dans un échantillon de taille n.

Soit l'hypothèse : "La proportion de ce caractère est p."
On considère I, un intervalle de fluctuation de la fréquence au seuil de 95% dans les échantillons de taille n.

- Soit $f \in I$, on considère que l'hypothèse selon laquelle la proportion est p dans la population n'est pas remise en question et on l'accepte.
- Soit $f \notin I$, on rejette l'hypothèse selon laquelle cette proportion vaut p avec un risque d'erreur de 5 %.

Propriété :
On considère une population dans laquelle la proportion d'un caractère est p est connue.
On observe f comme fréquence de ce caractère dans un échantillon de taille n.

On considère I, un intervalle de fluctuation de la fréquence au seuil de 95% dans les échantillons de taille n.

- Soit $f \in I$, on n'a pas de raison de considérer que l'échantillon n'est pas représentatif.
- Soit $f \notin I$, on considère l'échantillon non représentatif.

Exemple 1 :

Un laboratoire annonce un taux d'efficacité de 75% pour un médicament. Un test clinique sur un échantillon de 200 malades a mis en évidence la guérison de 138 patients. Que penser de l'annonce du laboratoire ?

Exemple 2 :

En France, 22 % des adultes ont de l'eczéma. Dans une petite ville, un médecin constate que, sur ses 330 patients adultes, 17 ont de l'eczéma. Que peut-on penser de ce résultat ?

II) Estimation

On considère dans cette partie que la proportion p est inconnue.

> Propriété :
> f est une estimation ponctuelle de la proportion inconnue p.

Exemple :

Un élève lance 150 fois une pièce équilibrée. Il obtient 70 fois Face.
Donner une estimation de la proportion p d'obtenir Face.

> Définition :
> Un intervalle de confiance (ou fourchette de sondage) de niveau de confiance supérieur à 95 % pour une proportion p inconnue basé sur un échantillon de taille n est :
>
> $$I = \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$$
>
> où f est la fréquence observée dans l'échantillon.

Remarque importante :

La valeur p n'appartient pas obligatoirement à un intervalle $I = \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ calculé à partir de la fréquence f d'un échantillon. En revanche, parmi tous les échantillons de taille n possibles , ayant comme fréquence observée f, environ 95 % des intervalles $I = \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ associés, contiennent le nombre p.

Exemple :

Lors de sondage, on interroge un panel dit représentatif d'une population pour obtenir une estimation du pourcentage de votants pour un candidat A à une future élection. Si, parmi 1000 personnes interrogées, le candidat A obtient 53 % dans les intentions de vote, peut-il être sûr d'être élu le jour de l'élection ?