Échantillonnage, cours, classe de 2nde

Définition :

On considère une population d'individus. Lorsque l'on connaît ou lorsque l'on fait une hypothèse sur la proportion p d'individus ayant une caractéristique donnée dans une population et que l'on effectue un nombre n de tirages avec remise dans cette population, la fréquence observée appartient avec une certaine probabilité à un intervalle appelé *intervalle de fluctuation* de centre p et de longueur qui diminue lorsque n augmente. On parle alors de situation *d'échantillonnage*.

Propriété et définition :

Pour tout entier naturel n. Alors quand n « devient grand » l'intervalle

$$I_n = [p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$$

contient la fréquence f d'apparition du caractère étudié d'un échantillon avec une probabilité qui se rapproche de 0,95. On dit que c'est un *intervalle de fluctuation* au seuil de 95%.

Remarque :
En pratique, on utilise cette propriété dès que les conditions $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ sont vérifiées.

Test d'hypothèse :
On considère une population dans laquelle on suppose que la proportion d'un caractère est p.On fait l'hypothèse « La proportion dans la population est p ». On observe la fréquence f d'apparition de ce caractère sur un échantillon de taille n et on calcule l'intervalle de fluctuation I au seuil de 95%.

- Si $f \in I$, au risque de 5% d'erreur (ou au seuil de confiance de 95%), on accepte l'hypothèse que la proportion dans la population est f.
- Si $f \in I$, au risque de 5% d'erreur on rejette l'hypothèse que la proportion dans la population est p.

Exemple :
Un fournisseur d'accès à l'internet affirme que, sur sa hotline, seuls 20% des clients attendent plus de 5 minutes pour obtenir un interlocuteur. Une association de consommateurs interroge au hasard 200 personnes ayant eu à s'adresser à cette hotline. Parmi ces personnes, 53 ont dû attendre plus de 5 minutes. Peut-on mettre en doute l'affirmation du fournisseur d'accès ?

L'hypothèse à tester est « 20% des clients attendent plus de 5 minutes ».

$\frac{53}{200} = 0,265$ et $I = [0,2 - \frac{1}{\sqrt{200}}; 0,2 + \frac{1}{\sqrt{200}}] = [0,129; 0,271]$.

Or $0,265 \in I$ donc au seuil de confiance de 95%, on ne peut pas mettre en doute l'affirmation du fournisseur d'accès.

1 Estimation

Propriété et définition :

Soit X_n une variable aléatoire suivant la loi binomiale $\mathcal{B}(n;p)$ et $F_n = \frac{X_n}{n}$ où p est la proportion inconnue d'apparition d'un caractère.

- Alors, pour n assez grand, $P(p \in [F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}}]) \geq 0,95$.
- Si on appelle f la fréquence d'apparition du caractère sur un échantillon de taille n, Alors l'intervalle $[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}]$ est appelé *intervalle de confiance de la proportion p inconnue au niveau de confiance 0,95*.

Remarque :

Un intervalle de confiance au niveau de 95% a une amplitude de $\frac{2}{\sqrt{n}}$. L'amplitude diminue lorsque la taille n de l'échantillon augmente.

Exemple :

Un candidat à une élection municipale fait effectuer un sondage. Sur 100 personnes de la ville interrogées, 63 déclarent vouloir voter pour lui.

$I = [0,63 - \frac{1}{\sqrt{100}}; 0,63 + \frac{1}{\sqrt{100}}] = [0,53; 0,73]$

On peut donc estimer que la proportion de personnes dans la ville voulant voter pour lui est comprise dans l'intervalle I=[0,53 ;073].