ESTIMATION EN TERMINALE S

I - INTERVALLE DE FLUCTUATION ASYMPTOTIQUE

Pour étudier un caractère présent dans une population, on prélève de façon aléatoire un échantillon dans cette population.

On suppose connues :

- la proportion p du caractère **dans la population**
- la taille n de l'échantillon

On cherche à évaluer :

- la fréquence f du caractère **dans l'échantillon**

EXEMPLE

On sait que 48% des élèves d'un lycée sont des garçons (et donc 52% sont des filles...).

Si l'on sélectionne au hasard 100 élèves dans l'établissement, on devrait obtenir ***environ*** 52 filles et 48 garçons mais il n'est pas du tout certain que l'on obtienne **exactement** ces chiffres.

Par contre, on pourra rechercher un intervalle dans lequel se situera *"probablement"* la proportion de garçons dans cet échantillon.

Si n est élevé, on peut assimiler la sélection de l'échantillon à un tirage avec remise. Le nombre d'individus présentant le caractère étudié suit alors une loi binomiale $\mathscr{B}\left(n, p\right)$. Pour n élevé, on peut approximer cette loi binomiale par une loi normale. On obtient alors le résultat suivant :

DÉFINITION ET PROPRIÉTÉ

Soit un réel $\alpha \in]0\,;\,1[$.

Un **intervalle de fluctuation asymptotique au seuil de** $1-\alpha$ est :

$$I = \left[p - u_{\alpha} \times \frac{\sqrt{p\left(1-p\right)}}{\sqrt{n}}\ ;\ p + u_{\alpha} \times \frac{\sqrt{p\left(1-p\right)}}{\sqrt{n}}\right]$$

Si on note f_n la fréquence du caractère étudié dans l'échantillon de taille n :

$$\lim_{n\to+\infty} p\left(f_n \in I\right) = 1-\alpha.$$

REMARQUES

- **Rappel :** La définition de u_{α} est donné dans le chapitre « loi normale ».
  On a en particulier $u_{0,05} = 1,96$ et $u_{0,01} = 2,58$
- Cette propriété est démontrée en exercice
- On considèrera que, lorsque n est suffisamment élevé ($n \geqslant 30$, $np \geqslant 5$ et $n\left(1-p\right) \geqslant 5$), $p\left(f \in I\right) \approx 1-\alpha$

CAS PARTICULIER (INTERVALLE DE FLUCTUATION AU SEUIL DE 95%)

L'**intervalle de fluctuation asymptotique au seuil de** 95% est l'intervalle :

$$I = \left[p - 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \ ; \ p + 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

EXEMPLE

Si l'on reprend l'exemple donné en introduction, on a $n = 100$ et $p = \dfrac{48}{100}$.

On trouve $I = [0,38 \ ; \ 0,58]$.

La proportion de garçons dans l'échantillon devrait être comprise entre 38% et 58% (avec une probabilité de 0,95)

REMARQUES

- L' intervalle de fluctuation peut être utilisé pour valider ou rejeter une hypothèse. On procède de la façon suivante :
  — On suppose que la proportion du caractère étudié est p.
  — On prélève un échantillon de taille n
  — On regarde si la fréquence f du caractère dans l'échantillon appartient à I. Si oui, l'hypothèse est validée, si non, elle est rejetée.
- Une étude de fonction montre que pour tout $p \in [0 \ ; \ 1]$, $1,96 \times \sqrt{p(1-p)} < 1$. On en déduit que :

  $$\left[p - 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \ ; \ p + 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right] \text{ est inclus dans } \left[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}} \right]$$

  qui est l'intervalle vu en Seconde.

II - INTERVALLE DE CONFIANCE

Dans cette partie (contrairement à la première partie), on suppose que l'**on connait la fréquence** f **du caractère dans l'échantillon** mais que l' **on ne connait pas la proportion** p **du caractère dans la population**.

On cherche alors à évaluer p.

DÉFINITION ET PROPRIÉTÉ

On appelle **intervalle de confiance avec un niveau de confiance de 95%** l'intervalle :

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

Pour n élevé, la proportion p du caractère dans la population appartiendra à I dans 95% des cas.

EXEMPLE

On recherche le pourcentage de truites femelles dans un élevage de truites.

Pour cela, on a prélevé un échantillon de 50 truites et on a comptabilisé 28 femelles dans cet échantillon.

Le pourcentage de truites femelles dans l'ensemble de l'élevage appartient donc à l'intervalle :

$$I = \left[\frac{28}{50} - \frac{1}{\sqrt{50}} ; \frac{28}{50} + \frac{1}{\sqrt{50}}\right] \approx [0,42 ; 0,70]$$

avec un risque d'erreur inférieur à 5%.

REMARQUE

La longueur de l'intervalle I est $\frac{2}{\sqrt{n}}$.

Si l'on souhaite obtenir un intervalle d'amplitude maximale a, il faut choisir n tel que

$\frac{2}{\sqrt{n}} \leqslant a$ c'est à dire $n \geqslant \frac{4}{a^2}$