# ECHANTILLONNAGE

## I. Fluctuation

Dans cette partie, on suppose qu'on connait la proportion de caractère sur l'ensemble de la population. On le note $p$. On choisit un échantillon d'individus de taille $n$. Quelle fréquence peut-on observer sur l'échantillon ?

### a) Lancer de pièces

On lance une pièce équilibrée. La probabilité d'obtenir Pile est $p = 0,5$. On lance la pièce 100 fois et on compte le nombre de Pile obtenus.

On reproduit l'expérience des 100 lancers 200 fois. (Voir simulateur geogebra).

On remarque que les 200 fois ou on répète l'expérience, les fréquences observées sur chaque lancer restent dans la plupart des cas assez proche de 0,5. On remarque qu'il est cohérant d'affirmer que dans 95 % des cas, on observe une fréquence comprise entre 0,4 et 0,6.

Lancer une pièce 100 fois correspond à choisir un échantillon de taille 100 parmi l'infinité de lancés de pièces. La fréquence que l'on observe sur cet échantillon de taille $n$ appartient dans 95 % des cas à l'intervalle

$$\left[0,5 - \frac{1}{\sqrt{n}} ; 0,5 + \frac{1}{\sqrt{n}}\right]$$

Si la fréquence observée sur l'échantillon fluctue, plus la taille de l'échantillon est grande, plus la fréquence se rapproche de la probabilité.

* $n = 100$, dans 95 % des cas, la fréquence observée sur l'échantillon appartiendra à

$$\left[0,5 - \frac{1}{\sqrt{100}} ; 0,5 + \frac{1}{\sqrt{100}}\right] = [0,4 ; 0,6]$$

* $n = 1000$ ; dans 95 % des cas la fréquence observée sur l'échantillon appartiendra à

$$\left[0,5 - \frac{1}{\sqrt{1000}} \; ; \; 0,5 + \frac{1}{\sqrt{1000}}\right] = [0,46 ; 0,54]$$

## b) Intervalle de fluctuation

Définition : $p$ = proportion sur l'ensemble de la population d'un caractère (c'est donc aussi si on choisit une personne au hazard, la probabilité qu'elle est à le caractère).

On choisit au hazard un échantillon d'individus de taille $n$. La fréquence observée sur l'échantillon peut fluctuer (changer), mais dans 95 % des cas, elle appartiendra à l'intervalle de fluctuation qui se note

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

## c) Application

La directrice de l'établissement affirme que 30 % des élèves aiment les mathématiques. On choisit 50 élèves au hazard. $p = 0,3$ , $n = 50$

La fréquence des élèves aimant les mathétiques que nous pouvons observer sur notre échantillon appartient dans 95 % des cas à l'intervalle

$$\left[0,3 - \frac{1}{\sqrt{50}} \; ; \; 0,3 + \frac{1}{\sqrt{30}}\right]$$

$$[0,16 ; 0,44]$$

Donc, dans 95 % des cas sur les 50 élèves choisit au hazard, il y aura entre 8 et 22 élèves qui aiment les mathétiques.

Chose extraordinaire, sur les 50 élèves interrogés hier 32 affirment aimer les mathématiques.

La fréquence observée sur notre échantillon est $f = \frac{32}{50} = 0,64$. Cette fréquence n'appartient pas à l'intervalle de fluctuation.

* Soit notre échantillon fait partie des 5% des cas en dehors de l'intervalle de fluctuation.

* Soit l'affirmation de la directrice est fausse.

Le mathématicien décide de rejeter l'affirmation de la directrice.

Propriété : On affirme qu'une proportion $p$ connu est établie sur l'ensemble d'une population

On choisit un échantillon de taille $n$ sur lequel on calcule la fréquence $f$.

* Si $f \in I_f = \left[ p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$

alors on valide l'affirmation de la proportion $p$ sur la population

* Si $f \notin I_f = \left[ p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$

alors on rejete l'affirmation de la proportion $p$ sur la population

Exemple : Dans son magazin à New York, la société m&ns affirme que 25% de sa production sont sans colorants, donc de couleur chocolat. Pour vérifier cette affirmation, on choisit 1000 m&ns au hazard. Sur cet échantillon, on observe que 212 m&ns sont sans colorants (couleur chocolat)

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

$= [0,218 ; 0,282]$

La fréquence observé sur notre échantillon est $f = \frac{212}{1000} = 0,212$

$f$ n'appartient pas a l'échantillon, il est fort probable que l'annonce fait par la société m&ns est de la publicité mensongère.

## II. Intervalle de confiance

Dans cette partie, on considère qu'on ne connait pas la proportion $p$ sur l'ensemble de la population, on va chercher à l'encadrer en observant un échantillon, c'est le principe de sondage.

Deux candidats se présentent aux élections, candidat A et candidat B. Monsieur A interroge 1000 personnes parmis elles 642 lui affirment qu'elles voteront pour lui. En admettant que personne n'est mentit, Monsieur A peut-il penser qu'il va être élu ?

La fréquence observée sur l'échantillon est $f = \frac{642}{1000} = 0,642$.

On appelle intervalle de confiance, l'intervalle

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

On affirme que dans 95 % des cas, la proportion sur l'ensemble de la population $p$ appartient à l'intervalle de confiance.

$$I_c = \left[ 0,642 - \frac{1}{\sqrt{1000}} ; 0,642 + \frac{1}{\sqrt{1000}} \right] = [0,61 ; 0,674]$$

Donc Monsieur A peut considérer que sur l'ensemble de la population, plus de 60 % des gens vont voter pour lui. Donc il sera élu.

* Juste un petit peu plus loin

Combien de personnes Monsieur A doit-il interroger pour avoir une estimation à 0,02 près ?

La longueur de l'intervalle de confiance est de $\frac{2}{\sqrt{n}}$. On cherche $n$ pour que $\frac{2}{\sqrt{n}} = 0,02$.

Donc $n = 10\,000$.

Donc Monsieur A doit interroger 10 000 personnes, s'il veut une estimation du résultat des élections à 2 % près.