

Chapitre 8: Simulation et échantillonnage

1 Simulation

Définition:

Une **expérience aléatoire** est une expérience que l'on peut reproduire dans les mêmes conditions et dont on connaît à priori tous les résultats (ou issues) possibles, sans pouvoir dire avec certitude le résultat qui se produira.

Exemple:

Lancer un dé à six faces !

2 Échantillonnage

Définition:

Lorsque l'on répète plusieurs fois la même expérience aléatoire, la distribution des fréquences des issues varie. On parle de **fluctuation d'échantillonnage**.

Exemple:

Réaliser deux fois l'expérience « Lancer cent fois de suite un dé à six faces ». On observe que les fréquences des issues varient.

Issue	1	2	3	4	5	6
Expérience 1	0,13	0,15	0,18	0,21	0,19	0,11
Expérience 2	0,12	0,13	0,15	0,14	0,21	0,21

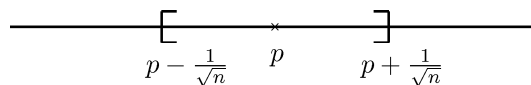
3 Intervalle de fluctuation

Propriété:

Au sein d'une population, on connaît la proportion p des individus ayant un caractère donné. Parmi les échantillons de taille n extraits de cette population, la fréquence d'apparition f du caractère varie avec l'échantillon prélevé. On admet que, pour un échantillon de taille $n \geq 25$ et pour p compris entre 0,2 et 0,8, la fréquence d'apparition f observé appartient à l'intervalle

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

avec une probabilité d'au moins 0,95. Cet intervalle est appelé **l'intervalle de fluctuation au seuil de 95%**.



- Si la fréquence observée f du caractère est en dehors de l'intervalle de fluctuation, on rejette l'échantillon avec une erreur au seuil de 5%.
- Si la fréquence observée f du caractère est dans l'intervalle de fluctuation, on valide l'échantillon au seuil de 95%.

⇒ Ainsi dans 5% des cas, la décision prise (rejet ou validation) risque d'être incorrect.

Exemple:

D'après l'exemple précédent :

Issue	multiple de 3
Expérience 1	0,29
Expérience 2	0,36

Dans l'exemple ci-dessus la fréquence théorique d'apparition d'un nombre multiple de 3 est de $\frac{1}{3} \simeq 0,33$. L'intervalle de fluctuation est

$$\left[0,33 - \frac{1}{\sqrt{100}}; 0,33 + \frac{1}{\sqrt{100}} \right] = [0,23; 0,43]$$

Dans les deux exemples du dessus, la fréquence d'apparition d'un nombre multiple de 3 est contenue dans l'intervalle $[0,23; 0,43]$. On peut donc dans les deux cas valider l'échantillon au seuil de 95%.

4 Intervalle de confiance

Définition:

Dans une population, on désire estimer la proportion inconnue p d'un caractère donné. On étudie un échantillon de taille n , $n \geq 25$. Le caractère étudié apparaît avec la fréquence f , $0,2 \leq f \leq 0,8$. On peut estimer que la proportion p du caractère dans la population totale est dans l'intervalle de confiance

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

avec une probabilité d'au moins 0,95. Dans 5% des cas, l'intervalle de confiance ne contient peut-être pas p .