Chapitre 10: Échantillonnage

1 Simulation

Définition:
*Une **expérience aléatoire** est une expérience que l'on peut reproduire dans les mêmes conditions et dont on connait 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

Exemple:
On réalise deux fois l'expérience « Lancer cent fois de suite un dé à six faces ».

Issue	*1*	*2*	*3*	*4*	*5*	*6*
Échantillon 1						
Échantillon 2						

On observe ici que les fréquences des issues varient.

Définition:
Le résultat d'une expérience aléatoire est appelé un
Lorsque l'on répète plusieurs fois la même expérience aléatoire, on observe que les fréquences des issues varient pour chaque échantillon.
On parle alors de

3 Intervalle de fluctuation

Lors des deux expériences précédentes, on a obtenu deux échantillons de taille $n = \dots\dots\dots$ On s'intéresse à la fréquence d'apparition d'un nombre multiple de 3.
Les résultats obtenus sont consignés dans le tableau ci-dessous :

Issue	multiple de 3
Échantillon 1	
Échantillon 2	

La fréquence théorique d'apparition d'un nombre multiple de 3 est $p = \dots\dots\dots$

Comme $0,2 < p < 0,8$ et $n \geq 25$, les fréquences d'apparition d'un nombre multiple de 3 pour nos deux échantillons ont une probabilité de $0,95$ d'appartenir à l'intervalle

$$\left[\frac{1}{3} - \frac{1}{\sqrt{100}}; \frac{1}{3} + \frac{1}{\sqrt{100}}\right] =$$

Cet intervalle est appelé l'**intervalle de fluctuation au seuil de** 95% d'apparition d'un nombre multiple de 3 pour $n = 100$.

- Si la fréquence f d'un nombre multiple de 3 pour un échantillon est en dehors de l'intervalle de fluctuation, on rejette l'échantillon avec une erreur au seuil de 5%.
- Sinon on valide l'échantillon au seuil de 95%.

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

Pour notre exemple,

On réalise à présent deux fois l'expérience « Lancer 1000 fois de suite un dé à six faces ». On obtient deux échantillons de taille $n = \dots\dots\dots$ et on s'intéresse encore à la fréquence d'apparition d'un nombre multiple de 3.

Les résultats obtenus sont consignés dans le tableau ci-dessous :

Issue	multiple de 3
Échantillon 3	
Échantillon 4	

La fréquence théorique d'apparition d'un nombre multiple de 3 est encore $p = \dots\dots\dots$

Comme $\dots\dots < p < \dots\dots$ et $n \geq \dots$, les fréquences d'apparition d'un nombre multiple de 3 pour nos deux échantillons ont une probabilité de $0,95$ d'appartenir à l'intervalle

$$\left[\frac{1}{3} - \frac{1}{\sqrt{\dots\dots}}; \frac{1}{3} + \frac{1}{\sqrt{\dots\dots}}\right] =$$

Pour notre exemple,

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%.*

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 apparait 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.