Chapitre 29: Fluctuation et estimation

1 Intervalle de fluctuation

Dans une urne contenant 4 boules bleues et 6 boules rouges, on effectue 100 tirages avec remise et on s'intéresse à la fréquence f d'apparition d'une boule bleue. On a ainsi réalisé un échantillon de taille $n = 100$.

1.1 Loi binomiale

En classe de première, on a associé à l'échantillon de taille n la variable aléatoire X_n qui donne le nombre de fois où une boule bleue est apparue.

Les tirages étant indépendants, X_n suit une loi binomiale $\mathcal{B}(n, p)$ où $n = 100$ et $p = 0,4$ pour notre exemple.

En posant $F_n = \dfrac{X_n}{n}$ la fréquence d'apparition de la boule bleue, on va déterminer un intervalle de la forme $\left[\dfrac{a}{n}; \dfrac{b}{n}\right]$ avec a et b les plus petites entiers tels que $P(X_n \leq a) > 0,025$ et $P(X_n \leq b) \geq 0,975$. On a :

$$P\left(X_n \in [a; b]\right) \geq 0,95$$

soit

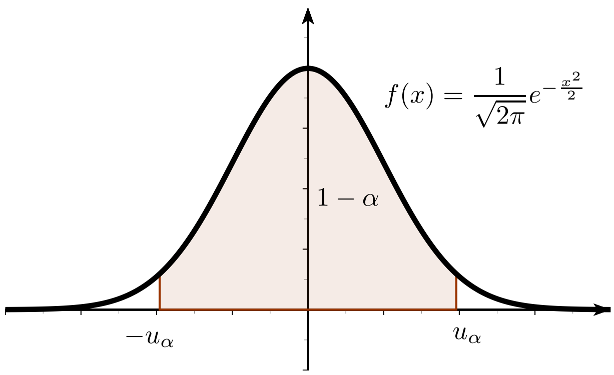
$$P\left(F_n \in \left[\frac{a}{n}; \frac{b}{n}\right]\right) \geq 0,95$$

et $\left[\dfrac{a}{n}; \dfrac{b}{n}\right]$ est alors appelé intervalle de fluctuation au seuil de $0,95$ de $F_n = \dfrac{X_n}{n}$.

A l'aide du tableur, on obtient pour notre exemple que $a = 31$ et $b = 50$ donc lorsque l'on réalise un échantillon de taille $n = 100$, la fréquence f d'apparition d'une boule bleue appartient à l'intervalle de fluctuation $[0,31; 0,5]$ avec une probabilité de $0,95$.

1.2 Théorème de Moivre-Laplace

Soit $\alpha \in]0; 1[$ et u_α le réel tel que $P\left(-u_\alpha \leq Z \leq u_\alpha\right) = 1 - \alpha$ où Z suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.



D'après le théorème de Moivre-Laplace, si X_n suit une loi binomiale $\mathcal{B}(n, p)$ alors :

$$\lim_{n \to +\infty} P\left(-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha\right) = P\left(-u_\alpha \leq Z \leq u_\alpha\right) = 1 - \alpha$$

or

$$-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha \iff -u_\alpha\sqrt{np(1-p)} \leq X_n - np \leq u_\alpha\sqrt{np(1-p)} \iff p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

On en déduit le théorème et la définition suivante :

Théorème:
Soit X_n une variable aléatoire suivant la loi binomiale $\mathcal{B}(n,p)$ et $F_n = \dfrac{X_n}{n}$.
Soit $\alpha \in]0;1[$ et u_α le réel tel que $P(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$ où Z suit la loi normale centrée réduite $\mathcal{N}(0,1)$ alors :

$$\lim_{n\to+\infty} P(F_n \in I_n) = 1-\alpha$$

où

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

Démonstration:
Soit $\alpha \in]0;1[$ et u_α le réel tel que $P(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$ où Z suit la loi normale centrée réduite $\mathcal{N}(0,1)$. D'après le théorème de Moivre-Laplace, si X_n suit une loi binomiale $\mathcal{B}(n,p)$ alors :

$$\lim_{n\to+\infty} P\left(-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha\right) = P(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$$

or

$$-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha \Longleftrightarrow -u_\alpha\sqrt{np(1-p)} \leq X_n - np \leq u_\alpha\sqrt{np(1-p)} \Longleftrightarrow p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

donc

$$\lim_{n\to+\infty} P(F_n \in I_n) = \lim_{n\to+\infty} P\left(-u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha\right) = 1-\alpha$$

Définition:
Sous les conditions énoncées dans le théorème ci-dessus, on dit que l'intervalle I_n est un intervalle de fluctuation asymptotique de la variable aléatoire fréquence $F_n = \dfrac{X_n}{n}$ au seuil de $1-\alpha$. Ainsi, pour n grand, la variable aléatoire F_n prend ses valeurs dans l'intervalle I_n avec une probabilité proche de $1-\alpha$. En pratique, on utilise cet intervalle de fluctuation de F_n dès que

$$n \geq 30 \quad ; \quad np \geq 5 \quad ; \quad n(1-p) \geq 5$$

Exemple:
Avec l'exemple initiale, si $\alpha = 0,05$, on a $u_{0,05} \simeq 1,96$ et donc :

$$I_{100} = \left[0,4 - 1,96\frac{\sqrt{0,4\times 0,6}}{\sqrt{100}} ; 0,4 + 1,96\frac{\sqrt{0,4\times 0,6}}{\sqrt{100}}\right] = [0,303;0,497]$$

$[0,303;0,497]$ *est intervalle de fluctuation asymptotique de la variable aléatoire fréquence au seuil de* $0,95$. *Lorsque l'on réalise un échantillon de taille $n = 100$, la fréquence f d'apparition d'une boule bleue appartient à cet intervalle avec une probabilité de* $0,95$.
Si $\alpha = 0,01$, on a $u_{0,01} \simeq 2,58$ et donc :

$$I_{100} = \left[0,4 - 2,58\frac{\sqrt{0,4\times 0,6}}{\sqrt{100}} ; 0,4 + 2,58\frac{\sqrt{0,4\times 0,6}}{\sqrt{100}}\right] = [0,273;0,527]$$

$[0,273;0,527]$ *est intervalle de fluctuation asymptotique de la variable aléatoire fréquence au seuil de* $0,99$. *Lorsque l'on réalise un échantillon de taille $n = 100$, la fréquence f d'apparition d'une boule bleue appartient à cet intervalle avec une probabilité de* $0,99$.

2 Prise de décision

On souhaite savoir si un pièce est équilibrée, au seuil de décision de 5%, sachant qu'en la lançant 100 fois on a obtenu 40 pile. La proportion de pile supposée (puisqu'on émet l'hypothèse « la pièce est équilibrée ») est $p = 0,5$. Comme $n \geq 30$; $np \geq 5$ et $n(1-p) \geq 5$, on peut appliquer la procédure suivante :

- On détermine $I_{100} = \left[0,5 - 1,96\dfrac{\sqrt{0,5 \times 0,5}}{\sqrt{100}} ; 0,5 - 1,96\dfrac{\sqrt{0,5 \times 0,5}}{\sqrt{100}}\right] = [0,402; 0,598]$;
- On calcule la fréquence f d'apparition du pile : $f = 0,4$;
- Comme $f \notin I_{100}$, on rejette l'hypothèse « la pièce est équilibrée »au seuil de 5%, c'est à dire que la probabilité de rejeter à tort l'hypothèse « la pièce est équilibrée »est de $0,05$. Le seuil de décision correspond à ce risque.
- Dans le cas où $f \in I_{100}$ on dit que l'on accepte l'hypothèse au seuil de 5%.

3 Intervalle de confiance

Dans une urne contenant un nombre inconnu de boules bleues et de boules rouges, on s'intéresse par exemple à la proportion p de boules bleues. Pour cela on effectue 100 tirages avec remise et on obtient 43 boules bleues soit une fréquence $f = 0,43$. Peut-on à partir de ce résultat estimer la proportion de boules bleues dans l'urne ?

Théorème:
Soit X_n une variable aléatoire suivant la loi binomiale $\mathcal{B}(n,p)$ et $F_n = \dfrac{X_n}{n}$ alors $P\left(F_n - \dfrac{1}{\sqrt{n}} \leq p \leq F_n + \dfrac{1}{\sqrt{n}}\right) \simeq 0,95$

Démonstration:
On sait que (vu en seconde) que $P\left(p - \dfrac{1}{\sqrt{n}} \leq f \leq p_n + \dfrac{1}{\sqrt{n}}\right) \simeq 0,95$. Or :

$$p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \Longleftrightarrow -F_n - \frac{1}{\sqrt{n}} \leq -p \leq -F_n + \frac{1}{\sqrt{n}} \Longleftrightarrow F_n + \frac{1}{\sqrt{n}} \geq p \geq F_n - \frac{1}{\sqrt{n}}$$

donc

$$P\left(F_n - \frac{1}{\sqrt{n}} \leq p \leq F_n + \frac{1}{\sqrt{n}}\right) \simeq 0,95$$

Définition:
Soit f la fréquence du caractère C sur un échantillon de taille n. L'intervalle $\left[f - \dfrac{1}{\sqrt{n}} ; f + \dfrac{1}{\sqrt{n}}\right]$ est un intervalle de confiance à 95% de la proportion p inconnue dans la population.

Exemple:
Avec notre problème initial, on a :

$$\left[0,43 - \frac{1}{\sqrt{100}} ; 0,43 + \frac{1}{\sqrt{100}}\right] = [0,33; 0,53]$$

donc la proportion p de boules bleues dans l'urne appartient à l'intervalle $[0,33; 0,53]$ avec une probabilité de $0,95$.