Intervalle de fluctuation, estimation

Sommaire

1. Intervalle de fluctuation

Théorème (Intervalle de fluctuation asymptotique)

Si X_n suit une loi binomiale $\mathcal{B}(n;p)$, alors la fréquence de succès $F_n = \dfrac{X_n}{n}$ vérifie :

$$\lim_{n\to+\infty} P\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant F_n \leqslant p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1-\alpha$$

où u_α vérifie $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha$ pour Z suivant la loi normale centrée réduite $\mathcal{N}(0;1)$.

Démonstration

D'après le théorème de Moivre-Laplace :

$$\lim_{n\to+\infty} P\left(a \leqslant \frac{X_n - np}{\sqrt{np(1-p)}} \leqslant b\right) = P(a \leqslant Z \leqslant b)$$

En particulier, pour $a = -u_\alpha$ et $b = u_\alpha$, on a :

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

Or :

$$\begin{aligned}
-u_\alpha \leqslant \frac{X_n - np}{\sqrt{np(1-p)}} \leqslant u_\alpha &\iff -u_\alpha\sqrt{np(1-p)} \leqslant X_n - np \leqslant u_\alpha\sqrt{np(1-p)} \\
&\iff np - u_\alpha\sqrt{np(1-p)} \leqslant X_n \leqslant np + u_\alpha\sqrt{np(1-p)} \\
&\iff p - u_\alpha\frac{\sqrt{np(1-p)}}{n} \leqslant \frac{X_n}{n} \leqslant p + u_\alpha\frac{\sqrt{np(1-p)}}{n} \\
&\iff p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant F_n \leqslant p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}
\end{aligned}$$

Ainsi $P\left(-u_\alpha \leqslant \dfrac{X_n - np}{\sqrt{np(1-p)}} \leqslant u_\alpha\right) = P\left(p - u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant F_n \leqslant p + u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$

Et on a bien $\displaystyle\lim_{n\to+\infty} P\left(p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant F_n \leqslant p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1-\alpha$ ■

Définition

Pour tout nombre $\alpha \in]0;1[$, $\left[p - u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est appelé intervalle de fluctuation asymptotique de F_n au seuil $1-\alpha$.

En pratique

Pour $\alpha = 0{,}05$, $u_{0{,}05} = 1{,}96$. On en déduit que lorsque $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, la fréquence de succès F_n fluctue avec une probabilité de $0{,}95 = 95\ \%$ dans l'intervalle :

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

Cet intervalle est l'intervalle de fluctuation asymptotique de F_n à 95 %.

Intervalle de fluctuation, estimation

Explication (Intervalle de fluctuation vu en seconde)

Pour $p \in [0;1]$, on a $p(1-p) \leqslant 0,25$, donc $1,96\sqrt{p(1-p)} \leqslant 1,96\sqrt{0,25} \leqslant 1$

Ainsi $\left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subset \left[p - \dfrac{1}{\sqrt{n}}; p + \dfrac{1}{\sqrt{n}}\right]$.

Donc, $P\left(p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant F_n \leqslant p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \leqslant P\left(p - \dfrac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \dfrac{1}{\sqrt{n}}\right)$.

Sous les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, $P\left(p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant F_n \leqslant p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 0,95$

Par conséquent, on a pour n assez grand, $P\left(p - \dfrac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \dfrac{1}{\sqrt{n}}\right) \geqslant 0,95$.

Ainsi, on retrouve l'intervalle de fluctuation d'une fréquence au seuil de 95 % vu en seconde.

Exercice

On considère que la probabilité qu'une personne tirée au hasard soit une fille est de 50 %.

1. Dans quel intervalle fluctue la fréquence des filles dans un échantillon de 35 personnes tirées au hasard ?
2. Dans une classe de 35 élèves, on dénombre 13 filles. Peut-on considérer qu'il y a une sous-représentation des filles dans cette classe ?

2. Estimation

Explication

Considérons une expérience aléatoire à deux issues possibles (succès ou échec) dont **on ne connaît pas** la probabilité de succès p. On désire **estimer** au mieux p à partir de la réalisation de n expériences indépendantes. Pour cela, on va utiliser un **intervalle de confiance**.

Définition

Soit $\alpha \in [0;1]$. On appelle <u>intervalle de confiance</u> au niveau $1-\alpha$ pour l'<u>estimation</u> de p, noté I_C, un intervalle qui ne s'exprime qu'en fonction de F_n et n, et tel que $P(p \in I_c) \geqslant 1-\alpha$.

Théorème

Lorsque n est assez grand (en pratique $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$), l'intervalle :

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

est un intervalle de confiance au niveau de 95 % pour l'estimation de p.

Démonstration

On a vu que pour n assez grand, $P\left(p - \dfrac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \dfrac{1}{\sqrt{n}}\right) \geqslant 0,95$.

Or $p - \dfrac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \dfrac{1}{\sqrt{n}} \iff -\dfrac{1}{\sqrt{n}} \leqslant F_n - p \leqslant +\dfrac{1}{\sqrt{n}} \iff -\dfrac{1}{\sqrt{n}} \leqslant p - F_n \leqslant +\dfrac{1}{\sqrt{n}}$

$\iff F_n - \dfrac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \dfrac{1}{\sqrt{n}}$

Par conséquent $P\left(F_n - \dfrac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \dfrac{1}{\sqrt{n}}\right) \geqslant 0,95$ ■

Intervalle de fluctuation, estimation

En pratique

On dira que p est élément de l'intervalle $\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$ avec un niveau de confiance de plus de 95 %, où f désigne la fréquence observée sur un échantillon de taille n.
Avec les exigences usuelles de précision, on utilise cet intervalle dès que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

Exercice

Lors d'un scrutin électoral, on souhaite connaître la proportion p de français votant pour le candidat « A ».
L'institut IPSOS mène une enquête auprès de 1000 personnes tirées au hasard et avec remise (*i.e.* qu'une même personne peut éventuellement être choisie plusieurs fois). Le résultat indique qu'une proportion $f = 49$ % d'entre elles voteront pour le candidat « A ».

1. Quelle est la loi suivie par le nombre de personnes votant pour « A » dans une enquête avec remise effectuée auprès de 1000 personnes.
2. Étant donné le résultat de l'enquête, donner in intervalle de confiance à 95 % pour l'estimation de p.
3. Peut-on affirmer d'après l'enquête que le candidat « A »n'aura pas la majorité des votes ?