Intervalles de fluctuation

- Définition générale

Soit X une variable alétoire et α un réel dans]0, 1[.
On appelle intervalle de fluctuation de X au seuil $1-\alpha$
tout intervalle $[a,b]$ tel que : $\mathrm{P}(X \in [a,b]) \geq 1-\alpha$

intervalle de fluctuation première.alg

Différents intervalles de fluctuation de la variable fréquence

- On peut chercher celui qui a l'amplitude minimale (IF1).
- On peut chercher le plus petit intervalle centré autour de l'espérance *p* (IF2)
- On peut chercher celui qui symétrise les probabilités que $F=X/n$ soit à l'extérieur (IF3)
- Enfin, il y a celui de seconde: (IF4)

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

Intervalle de fluctuation asymptotique

Si la variable aléatoire X_n suit la loi $\mathcal{B}(n, p)$

alors, pour tout réel α dans]0, 1[on a : $\lim\limits_{n\to+\infty} \mathrm{P}\left(\dfrac{X_n}{n} \in I_n\right) = 1-\alpha$,

où I_n désigne l'intervalle $\left[p-u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}, p+u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$

et u_α désigne l'unique réel tel que $\mathrm{P}\left(-u_\alpha \le Z \le u_\alpha\right) = 1-\alpha$

où Z suit la loi normale $\mathcal{N}(0,1)$.

Exploration de l'IF

Pour différentes valeurs de *p* et de α

exploration intervalle de fluctuation asymptotique.xls

IF vu en seconde

$$P\left(\frac{X_n}{n} \in I_n\right) \leq P\left(p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}}\right)$$

Théorème

Si la variable aléatoire X_n suit la loi $\mathcal{B}(n, p)$, alors, pour tout p dans]0, 1[,

il existe un entier n_0 tel que si $n \geq n_0$ alors $P\left(p - \frac{1}{\sqrt{n}} \leq X_n \leq p + \frac{1}{\sqrt{n}}\right) \geq 0,95$.

p	0,35	0,36	0,37	0,38	0,39	0,4	0,41	0,42	0,43	0,44	0,45	0,46	0,47	0,48	0,49	0,5
n_0	31	30	36	64	56	81	90	120	143	209	271	288	304	399	399	529

Intervalle de confiance

Pour n assez grand, l'intervalle aléatoire $\left[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}\right]$

a une probabilité supérieure à 0,95 de contenir p.

À partir de l'intervalle **aléatoire** $\left[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}\right]$

on obtient, en réalisant le tirage d'un échantillon,
une *réalisation* de cet intervalle qui fournit donc un intervalle numérique

de la forme $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$.

Niveau de confiance

Un intervalle de confiance pour une proportion p à un niveau de confiance $1-\alpha$ est la réalisation, à partir d'un échantillon, d'un intervalle aléatoire contenant la proportion p avec une probabilité supérieure ou égale à $1-\alpha$.

Cet intervalle aléatoire est déterminé à partir de la variable aléatoire

$F_n = \dfrac{X_n}{n}$ qui, à tout échantillon de taille n, associe la fréquence.

Niveau de confiance

Cette confiance est une « réminiscence » de la probabilité liée à la procédure : si on réalisait 100 fois l'échantillonnage on obtiendrait 100 intervalles de confiance réalisés différents et en principe environ 95 d'entre eux contiennent *p*.

Si on obtient [0,4 ;0,6] comme intervalle de confiance réalisé, on ne peut pas dire « la probabilité que *p* appartienne à [0,4 ;0,6] est 0,95 », on dit plutôt :

« **p appartient à [0,4 ; 0,6] avec un niveau de confiance de 0,95** »

Intervalles de confiance

- if_ic.ggb

$$f_1^{-1}(F_n) \approx F_n + \frac{u_\alpha}{\sqrt{n}}\sqrt{F_n(1-F_n)}$$

$$f_2^{-1}(F_n) \approx F_n - \frac{u_\alpha}{\sqrt{n}}\sqrt{F_n(1-F_n)}$$

IC ou IF : comment distinguer à partir d'un énoncé.

Exercice 1. On dispose d'un médicament efficace dans le traitement d'une maladie . On observe 36 guérisons sur 50 patients.

Question N°1 : la littérature dit que l'efficacité habituelle est de 80%. Les données sont-elles en accord avec cette affirmation ?

Question N°2 : comment estimer l'efficacité du médicament à partir des données ?

IC ou IF : comment distinguer à partir d'un énoncé.

Exercice : On dispose d'un médicament efficace dans le traitement d'une maladie . On observe 36 guérisons sur 50 patients. **(la fréquence observée est de 36/50=72%**

Question N°1 : la littérature dit que l'efficacité habituelle est **de 80%.** Les **données** sont-elles en accord avec cette affirmation ?

Réponse : on calcule un intervalle de fluctuation pour la **fréquence**, et on regarde si la **fréquence observée** est dans l'IF.

$$IF = \left[0,8 - \frac{1,96\sqrt{0,8\times 0,2}}{\sqrt{50}}, 0,8 + \frac{1,96\sqrt{0,8\times 0,2}}{\sqrt{50}}\right] \approx [0,69;0,91]$$

C'est une prise de décision par IF

Question N°2 : comment estimer **l'efficacité** du médicament à partir des **données** ?

Réponse : on estime **l'efficacité théorique p (inconnue),** à partir des données observées, on donne en plus un intervalle de confiance pour cette quantité inconnue p.

$$IC = \left[0,72 - \frac{1}{\sqrt{50}}, 0,72 + \frac{1}{\sqrt{50}}\right] \approx [0,58;0,96]$$

C'est une méthode d'estimation par intervalle

Formules pour IF et IC

	Intervalle de fluctuation ***p connue***	**Intervalle de confiance** ***p inconnue***
SECONDE	$n \geq 25$ et $0,2 \leq p \leq 0,8$, seuil 95% $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$	Sensibilisation
PREMIÈRE	Avec la loi binomiale	
TERMINALE	$n \geq 30$ et $np \geq 5$ et $n(1-p) \geq 5$ Asymptotique au seuil $1-\alpha$ $I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]^1$	Au niveau de confiance 95% $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$