intervalle de fluctuation - intervalle de confiance

I) Intervalle de fluctuation asymptotique :

utile dans le cas où la proportion d'un caractère est connue (ou supposée) dans une population !

a) intervalle de fluctuation asymptotique au seuil de $1-\alpha$:

propriété :

Si X_n est une variable aléatoire suivant une loi binomiale $\mathscr{B}(n\ ;\ p)$ avec $p \in]0;1[$, alors pour tout nombre réel α appartenant à $]0;1[$,

$$\lim_{n \to +\infty} P\left(\frac{X_n}{n} \in I_n\right) = 1-\alpha \quad \text{avec } I_n = \left[p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ,\ p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

et u_α est le réel tel que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha$ quand Z suit $\mathcal{N}(0\ ;\ 1)$.

I_n est appelé intervalle de fluctuation asymptotique au seuil de $1-\alpha$

$\frac{X_n}{n} = F_n$ correspond à la fréquence des succès lors de la répétition de **n** épreuves de Bernoulli indépendantes de paramètre **p** !

► ***démonstration*** - *exigible* ROC -

Posons $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$

D'après le théorème de Moivre-Laplace, $\lim\limits_{n \to +\infty} P(-u_\alpha \leqslant Z_n \leqslant u_\alpha) = P(-u_\alpha \leqslant Z \leqslant u_\alpha)$

avec Z suivant $\mathcal{N}(0\ ;\ 1)$.

$$\begin{aligned}
\text{Or, } P(-u_\alpha \leqslant Z_n \leqslant u_\alpha) &= P(-u_\alpha\sqrt{np(1-p)} \leqslant Z_n\sqrt{np(1-p)} \leqslant u_\alpha\sqrt{np(1-p)}) \\
&= P\left(np - u_\alpha\sqrt{np(1-p)} \leqslant np + Z_n\sqrt{np(1-p)} \leqslant np + u_\alpha\sqrt{np(1-p)}\right) \\
&= P\left(np - u_\alpha\sqrt{np(1-p)} \leqslant X_n \leqslant np + u_\alpha\sqrt{np(1-p)}\right) \\
&= P\left(\frac{np - u_\alpha\sqrt{np(1-p)}}{n} \leqslant \frac{X_n}{n} \leqslant \frac{np + u_\alpha\sqrt{np(1-p)}}{n}\right) \\
&= P\left(\frac{p - u_\alpha\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{X_n}{n} \leqslant \frac{p + u_\alpha\sqrt{p(1-p)}}{\sqrt{n}}\right)
\end{aligned}$$

or, nous avons établi précédemment que $P(-u_\alpha \leqslant Z \leqslant u_\alpha) = 1-\alpha$

Donc, $\lim\limits_{n \to +\infty} P(-u_\alpha \leqslant Z_n \leqslant u_\alpha) = \lim\limits_{n \to +\infty} P\left(\frac{X_n}{n} \in I_n\right) = 1-\alpha$

b) cas particulier : intervalle de fluctuation asymptotique au seuil de 95%

définition : Soit X_n une variable aléatoire suivant une loi binomiale $\mathscr{B}(n\ ;\ p)$ avec $p \in]0;1[$. L'intervalle de fluctuation asymptotique de F_n $\left(\frac{X_n}{n}\right)$ au seuil de 95% est :

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

remarquons que l'intervalle de fluctuation asymptotique est inclus dans l'intervalle de fluctuation établi en seconde soit $\left[p - \frac{1}{\sqrt{n}}\ ,\ p + \frac{1}{\sqrt{n}}\right]$!

c) prise de décision :

On suppose que la proportion du caractère étudié dans la population est connu.
Supposons que cette proportion soit égale à p.
La prise de décision consiste à évaluer la pertinence de cette hypothèse.
Peut-on valider ou non, la proportion p ?

On effectue la procédure suivante :

► On prélève dans la population au hasard un échantillon de taille n

► On note la fréquence observée f_{obs} du caractère

► Si les 3 conditions sont vérifiées ($n \geqslant 30$, $np \geqslant 5$, $np(1-p) \geqslant 5$), on calcule l'intervalle de fluctuation asymptotique $\mathrm{I} = \left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ,\ p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ au seuil de 95%.

règle de décision :

le risque de se tromper !

- Si $f_{obs} \in \mathrm{I}$, on accepte l'hypothèse faite sur la proportion p
- Si $f_{obs} \notin \mathrm{I}$, on rejette l'hypothèse faite sur la proportion p avec un risque d'erreur de 5%.

la probabilité du risque de rejet à tort (5%) est évaluée mais dans le cas où on accepte l'hypothèse le risque d'erreur n'est pas déterminé !!

Ex : On considère que la probabilité qu'une personne choisie au hasard soit de sexe masculin est de 0,5 soit 50%.
Dans une classe de lycée de 35 élèves, on compte 11 garçons.
Peut on affirmer au seuil de 5% que les résultats constatés sur cet échantillon de 32 élèves confirment l'hypothèse de l'énoncé ?

L'hypothèse est que la proportion de garçons est $p = 0{,}5$.
On a $n = 35$, $np = 53 \times 0{,}5 = 17{,}5$, et $np(1-p) = 8{,}75$
donc les conditions ($n \geqslant 30$, $np \geqslant 5$, $np(1-p) \geqslant 5$) sont vérifiées.
On peut donc calculer l'intervalle de fluctuation asymptotique I à 95%.

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

$$= \left[0{,}5 - 1{,}96\frac{\sqrt{0{,}5(1-0{,}5)}}{\sqrt{35}}\ ,\ 0{,}5 + 1{,}96\frac{\sqrt{0{,}5(1-0{,}5)}}{\sqrt{35}}\right] \simeq [0{,}334\ ,\ 0{,}902] \text{ à } 10^{-3} \text{ près}$$

I est un intervalle de fluctuation de la fréquence de garçons dans un échantillon de taille 35.

Sur l'échantillon considéré, la fréquence observée est $f_{obs} = \frac{11}{35} \simeq 0{,}314$

Donc $f_{obs} \notin \mathrm{I}$ et on peut affirmer que les résultats constatés sur l'échantillon ne confirment pas l'hypothèse annoncée. On rejette l'hypothèse au seuil de 5%
(la probabilité de rejeter à tort l'hypothèse est de 5%)

II) Intervalle de confiance :

utile dans le cas où la proportion d'un caractère est inconnue dans une population !

Dans une population, la fréquence d'individus p (proportion d'individus) présentant un certain caractère est inconnue. On cherche à connaître cette proportion à partir d'un échantillon de n individus (tirage effectué au hasard et avec remise).
On suppose que les conditions d'approximation ($n \geqslant 30$, $np \geqslant 5$, $np(1-p) \geqslant 5$) sont vérifiées.
Soit F_n la variable aléatoire qui associe à chaque échantillon de taille n la fréquence du caractère étudié.

propriété : Soit une population pour laquelle la proportion p d'un certain caractère est inconnue. Soit F_n la variable aléatoire qui associe à chaque échantillon de taille n la fréquence du caractère étudié. On suppose que $n \geqslant 30$, $np \geqslant 5$, $np(1-p) \geqslant 5$
La proportion inconnue p vérifie $P\left(F_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \frac{1}{\sqrt{n}}\right) \simeq 0,95$

► ***démonstration*** - *exigible* ROC -

Nous avons vu en seconde que, dans environ 95% des cas, Fn appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}} , p + \frac{1}{\sqrt{n}}\right]$, ce qui revient à dire que $P\left(p - \frac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \frac{1}{\sqrt{n}}\right) \simeq 0,95$

Or, $p - \frac{1}{\sqrt{n}} \leqslant F_n \leqslant p + \frac{1}{\sqrt{n}}$ équivaut à $-\frac{1}{\sqrt{n}} \leqslant F_n - p \leqslant \frac{1}{\sqrt{n}}$

équivaut à $-\frac{1}{\sqrt{n}} - F_n \leqslant -p \leqslant \frac{1}{\sqrt{n}} - F_n$

équivaut à $F_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \frac{1}{\sqrt{n}}$

Il en résulte que $P\left(F_n - \frac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \frac{1}{\sqrt{n}}\right) \simeq 0,95$

définition :
Soit f, la fréquence observée d'un caractère sur un échantillon de taille n.
L'intervalle $\left[f - \frac{1}{\sqrt{n}} , f + \frac{1}{\sqrt{n}}\right]$ est un intervalle de confiance au niveau de confiance de 95% de la proportion inconnue p du caractère dans la population.

il y a 95% de chances que la véritable valeur de la proportion p soit comprise entre $f - \frac{1}{\sqrt{n}}$ et $f + \frac{1}{\sqrt{n}}$!

remarques :

► pour estimer une proportion inconnue on utilise en général un intervalle de confiance au niveau de confiance de 95%
► l'intervalle de confiance dépend de la taille de l'échantillon, pas de celle de la population
► l'amplitude de l'intervalle de confiance est $\frac{2}{\sqrt{n}}$
► dans certaines disciplines,on utilise comme intervalle de confiance au niveau de confiance de 95% celui -ci : $\left[f - 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}} , f + 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right]$

on l'admettra, la justification n'est pas au programme !

<u>Ex :</u> Dans un lycée, 440 élèves sont demi-pensionnaires. Un sondage est réalisé sur un échantillon de 54 élèves demi-pensionnaires. 67% se déclarent satisfaits des repas pris à la cantine.
Déterminez à partir de ce sondage l'intervalle de confiance au niveau de confiance de 95% de la proportion d'élèves demi-pensionnaires du lycée satisfaits de leur repas.

La taille de l'échantillon est n = 54

La fréquence observée du caractère «demi-pensionnaire satisfait» est f = 0,67.
L'intervalle de confiance au niveau de confiance de 95% de la proportion p de demi-pensionnaires satisfaits est :

$$\left[f-\frac{1}{\sqrt{n}}\,,\,f+\frac{1}{\sqrt{n}}\right]=\left[0{,}67-\frac{1}{\sqrt{54}}\,,\,0{,}67+\frac{1}{\sqrt{54}}\right]\simeq[0{,}533\,,\,0{,}806]\ \text{à } 10^{-3}\text{ près}$$

On peut donc affirmer avec un niveau de confiance de 95% que la proportion p des demi-pensionnaires satisfaits dans ce lycée est comprise entre 53% et 80%.