Intervalle de fluctuation, estimation

I) Intervalle de fluctuation

1) Théorème

On rappelle le théorème concernant la loi normale :

Si X est une variable aléatoire suivant la loi $\mathcal{N}(0 ; 1)$, alors pour tout $\alpha \in]0 ; 1[$ il existe un unique réel positif u_α tel que

$P(-u_\alpha \leq X \leq u_\alpha) = 1 - \alpha$

et en utilisant les notations de ce théorème :

Si la variable aléatoire X_n suit la loi binomiale $\mathcal{B}(n,p)$, avec p dans l'intervalle] 0 ;1[alors pour tout α dans] 0 ;1[, et pour

$I_n = \left[p - u_\alpha\sqrt{\frac{p(1-p)}{n}} ; p + u_\alpha\sqrt{\frac{p(1-p)}{n}}\right]$ **on a** $\lim_{n\to+\infty} P\left(\frac{X_n}{n} \in I_n\right) = 1 - \alpha$

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

La variable $F_n = \frac{X_n}{n}$ correspondant à la fréquence du succès lors de la répétition des n épreuves de Bernoulli indépendantes de paramètre p

Démonstration :

On pose : $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}} = \frac{\frac{X_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}$; $\frac{X_n}{n} \in I_n$ signifie $p - u_\alpha\sqrt{\frac{p(1-p)}{n}} \leq \frac{X_n}{n} \leq p + u_\alpha\sqrt{\frac{p(1-p)}{n}}$

soit $-u_\alpha \leq Z_n \leq u_\alpha$

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

où Z suit la loi normale $\mathcal{N}(0 ;1)$ et d'après le théorème énoncé en rappel on a

$$P(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$$

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

Cas particulier :

Dans de nombreux cas on utilise l'intervalle de fluctuation au seuil de 95 % c'est à dire le cas où $\alpha = 5\%$

Alors on a $u_\alpha = u_{0,05} \approx 1{,}95996$. Comme $u_{0,05} < 1{,}96$ on peut affirmer que pour n assez grand, la probabilité d'observer la fréquence F_n dans l'intervalle

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

est sensiblement égale à 95 %

Remarque :

La fonction définie par $f(p) = p(1-p)$ définie sur [0 ; 1] est une fonction du second degré qui admet pour maximum $\frac{1}{4}$ lorsque $p = \frac{1}{2}$

L'expression $1{,}96\sqrt{p(1-p)}$ admet donc pour majorant $2 \times \frac{1}{2}$ = 1 et l'intervalle $I_n = \left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}} \,;\, p + 1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$ est donc inclus dans l'intervalle $J_n = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ intervalle de fluctuation vu en classe de seconde.

2) Définition

Sous les hypothèses et avec les notations du théorème précédent la suite $P(F_n \in I_n)$ converge vers $1-\alpha$. On admet que $P(F_n \in I_n) \approx 1 - \alpha$ peut être validée dès que $n \geq 30$; $np \geq 5$ et $n(1-p) \geq 5$

Exemple :

Dans un jeu de grattage, la société qui émet les tickets estime que les joueurs doivent avoir une fréquence de gain p =0,07. Une fréquence inférieure ferait fuir les clients, une fréquence supérieure est susceptible de mettre la société en difficulté.

Trois employés sont chargés de vérifier les tickets émis par les machines trois jours distincts.

Le premier choisit 50 tickets et obtient 2 gagnants.

Le second choisit 130 tickets et obtient 16 gagnants.

Le troisième choisit 500 tickets et obtient 40 gagnants.

En utilisant des intervalles de fluctuation au seuil de 95% quelle conclusion doit prendre chaque employé ? Peut-il accepter ou rejeter l'hypothèse p = 0,07 ?

Premier employé :

Le contrôle a porté sur 50 tickets donc $n = 50$ donc on a bien $n \geq 30$, mais $p = 0{,}07$ donc $np = 50 \times 0{,}07 = 3{,}5$ donc $np \leq 5$ on n'est donc pas dans des conditions où l'intervalle de fluctuation peut permettre une conclusion quelconque.

Deuxième employé :

Le contrôle a porté sur 130 tickets donc $n = 130$ donc on a bien $n \geq 30$, $p = 0{,}07$ donc $np = 130 \times 0{,}07 = 9{,}1$ donc $np \geq 5$ et $n(1-p) = 130 \times 0{,}93 = 120{,}9$ on est donc dans des conditions où l'intervalle de fluctuation peut permettre une conclusion.

L'intervalle de fluctuation est égal à $I_{130} = \left[0{,}07 - 1{,}96 \times \sqrt{\frac{0{,}07 \times 0{,}93}{130}} \,;\, 0{,}07 + 1{,}96 \times \sqrt{\frac{0{,}07 \times 0{,}93}{130}}\right]$

En arrondissant $I_{130} = [0{,}0259 ; 0{,}1141]$

Or la fréquence observée par l'employé est $f = \frac{16}{130} \approx 0{,}123$

Cette fréquence n'est pas dans l'intervalle de fluctuation donc l'employé doit rejeter l'hypothèse $? = 0{,}07$. Il y a trop de tickets gagnants

Troisième employé :

Le contrôle a porté sur 500 tickets donc $n = 500$ donc on a bien $n \geq 30$, $p = 0{,}07$ donc $np = 500 \times 0{,}07 = 35$ donc $np \geq 5$ et $n(1-p) = 130 \times 0{,}93 = 465$ on est donc dans des conditions où l'intervalle de fluctuation peut permettre une conclusion.

L'intervalle de fluctuation est égal à $I_{500} = \left[0{,}07 - 1{,}96 \times \sqrt{\frac{0{,}07 \times 0{,}93}{500}}\ ;\ 0{,}07 + 1{,}96 \times \sqrt{\frac{0{,}07 \times 0{,}93}{500}}\right]$

En arrondissant $I_{130} = [0{,}0476\,;0{,}0924]$

Or la fréquence observée par l'employé est $f = \frac{40}{500} \approx 0{,}08$

Cette fréquence est dans l'intervalle de fluctuation donc l'employé est donc amené à accepter l'hypothèse $p = 0{,}07$.

II) Intervalle de confiance, estimation

1) Propriété

Soit X_n une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n, p)$, soit $F_n = \frac{X_n}{n}$ la fréquence et $J_n = \left[F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}}\right]$.

- **Sous les conditions données dans la définition précédente ($n \geq 30$; $np \geq 5$ et $(1-p) \geq 5$) l'intervalle J_n contient p avec une probabilité d'environ 95 %**
- **Pour n assez grand l'intervalle J_n contient p avec une probabilité au moins égale à 95 %**

Démonstration

Les propositions $p \in J_n$ (c'est à dire $F_n - \frac{1}{\sqrt{n}} \leq p \leq F_n + \frac{1}{\sqrt{n}}$) et $p - \frac{1}{\sqrt{n}} \leq F_n \leq p - \frac{1}{\sqrt{n}}$ sont équivalentes. Cette propriété découle donc du théorème et de la propriété qui précédent

2) Définition

Dans une population un certain caractère possède une proportion p. On prélève au hasard et avec remise un échantillon de taille n on obtient une fréquence f de ce caractère dans cet échantillon.

L'intervalle $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right]$ est appelé l'intervalle de confiance de la proportion p.

- **Sous les conditions ($n \geq 30$; $np \geq 5$ et $(1-p) \geq 5$) le niveau de cette confiance est d'environ 95 %**
- **Pour n assez grand le niveau de cette confiance au moins égale à 95 %**

Exemple :

Lors d'une livraison de biscuits, on a deux cartons de 500 paquets de biscuits chacun. Dans chaque carton un certain nombre de paquets sont abîmés.

A l'aide d'un tirage avec remise on extrait 40 paquets de chaque carton. On obtient 6 paquets abimés dans le premier carton et 10 paquets abîmés dans le second.

a) Calculer la fréquence des paquets abîmés dans chaque carton

b) Déterminer pour chaque carton, l'intervalle de confiance de la fréquence au seuil de 95 %

c) Lorsqu'on vide les cartons, on trouve 28 paquets abîmés dans le premier carton et 45 dans le deuxième carton. Ces résultats sont-ils compatibles avec les estimations obtenues au **b)** ?

Réponses

Premier carton :

a) La fréquence est : $f_1 = \frac{6}{40} = 0,15$

b) L'intervalle de confiance est: $\left[0,15 - \frac{1}{\sqrt{40}}\ ;\ 0,15 + \frac{1}{\sqrt{40}}\right] \approx [-0,0081\,;0,3081]$

c) La fréquence réelle est $F = \frac{28}{500} = 0,056$ cette fréquence appartient à l'intervalle obtenu.

Second carton :

a) La fréquence est : $f_2 = \frac{10}{40} = 0,25$

b) L'intervalle de confiance est: $\left[0,25 - \frac{1}{\sqrt{40}}\ ;\ 0,25 + \frac{1}{\sqrt{40}}\right] \approx [0,0918\,;0,4081]$

c) La fréquence réelle est: $F = \frac{45}{500} = 0,09$ cette fréquence appartient aussi à l'intervalle obtenu.

III) Intervalle de fluctuation ou intervalle de confiance ?

Dans une population dont on étudie un caractère

On utilise un **intervalle de fluctuation** **quand :**	**On utilise un** **intervalle de confiance** **quand :**
• On **connaît la proportion** $\boldsymbol{p}$ de présence du caractère dans la population **OU** • On **fait une hypothèse sur la valeur de cette proportion** (on est dans le cas d'une prise de décision)	• On **ignore la valeur de la proportion** $\boldsymbol{p}$ de présence du caractère et on ne formule pas d'hypothèse sur cette valeur