Terminale S (2015-2016)
Statistiques

1 Echantillonnage

1.1 Intervalle de fluctuation asymptotique

On considère une variable aléatoire X_n suivant la loi binomiale $\mathcal{B}(n,p)$.

1.1.1 Propriété

Pour tout α dans $]0;1[$, on a :

$$\lim_{n\to+\infty} P\left(\frac{X_n}{n}\in I_n\right)=1-\alpha$$

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

u_α a été défini dans le chapitre sur les lois à densités par l'égalité $P(-u_\alpha\leq Y\leq u_\alpha)=1-\alpha$, où Y suit la loi normale centrée réduite.

Démonstration (exigible)

$$\frac{X_n}{n}\in I_n \Leftrightarrow p-u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\leq\frac{X_n}{n}\leq p+u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

En multipliant par n, on obtient :

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

puis, en retranchant np :

$$\frac{X_n}{n}\in I_n \Leftrightarrow -u_\alpha\sqrt{np(1-p)}\leq X_n-np\leq u_\alpha\sqrt{np(1-p)}$$

enfin, en divisant par $\sqrt{np(1-p)}$, $(\neq 0)$:

$$\frac{X_n}{n}\in I_n \Leftrightarrow -u_\alpha\leq\frac{X_n-np}{\sqrt{np(1-p)}}\leq u_\alpha$$

On pose $Z_n=\dfrac{X_n-np}{\sqrt{np(1-p)}}$; le théorème de Moivre-Laplace, (chapitre lois à densités), nous dit que, pour tous réels a et b,

$$\lim_{n\to+\infty} P\left(Z_n\in[a;b]\right)=\int_a^b\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x=P(a\leq Y\leq b)$$

où Y suit la loi normale centrée réduite.
Donc

$$\lim_{n\to+\infty} P\left(\frac{X_n}{n}\in I_n\right)=\lim_{n\to+\infty}P(-u_\alpha\leq Z_n\leq u_\alpha)=P(-u_\alpha\leq Y\leq u_\alpha)=1-\alpha.$$

On définit la variable aléatoire F_n par $F_n=\dfrac{X_n}{n}$, $(n>0)$; F_n représente la fréquence de succès.

1.1.2 Définition

L'intervalle $\left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est un intervalle de fluctuation asymptotique de F_n au seuil $1-\alpha$. Cet intervalle, déterminé à partir de p et de n, contient F_n avec une probabilité d'autant plus proche de $1-\alpha$ que n est grand.

1.1.3 Propriété

Au seuil de 95%, $\alpha = 0,05$ et $u_\alpha \simeq 1,96$. (Chapitre sur les lois à densités).

On obtient alors la propriété d'approximation suivante qui est valable dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$:

L'intervalle de fluctuation au seuil de 95% de F_n, la fréquence d'un caractère dans un échantillon de taille n, est $\left[p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ où p désigne la proportion de ce caractère dans la population.

Remarque

Si $p \in]0;1[$, alors $0 \leq p(1-p) \leq 0,25$, donc $\sqrt{p(1-p)} \leq 0,5$ et $1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}$.

On obtient alors $\left[p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subset \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ qui est l'intervalle de fluctuation présenté en classe de seconde.

1.2 Prise de décision

L'intervalle de fluctuation asymptotique défini plus haut est utilisé, (lorsque les conditions de validité sont vérifiées, soit $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$), dans l'élaboration d'un test permettant de vérifier une hypothèse. On considère une population dans laquelle on émet l'hypothèse H que la proportion d'un caractère est p. On appelle I l'intervalle de fluctuation asymptotique de la fréquence de ce caractère dans un échantillon aléatoire de taille n, au seuil de 95%, sous l'hypothèse H.

On observe alors la fréquence f du caractère dans un échantillon de taille n.

La règle de décision est :

- si $f \in I$, alors on accepte l'hypothèse H.
- si $f \notin I$, alors on rejette l'hypothèse H.

Remarques

- Au seuil de 95%, la probabilité que f, (obtenue dans un échantillon aléatoire), ne soit pas dans I est $\alpha = 0,05$. Cela signifie qu'il y a un risque de 5% de se tromper en rejetant l'hypothèse H alors qu'elle est vraie.
- Il est possible d'utiliser d'autres seuils, le plus courant après 95%, étant 99% ; il faut alors remplacer, dans l'intervalle de fluctuation, le nombre $u_{0,05} \simeq 1,96$ par le nombre $u_{0,01} \simeq 2,58$.

2 Estimation

On se propose ici d'estimer une proportion dans une population à partir de la fréquence observée sur un échantillon. C'est ce que l'on pratique par exemple lors d'un sondage.

2.1 Propriété

On suppose que la variable aléatoire X_n suit la loi binomiale $\mathcal{B}(n,p)$. Pour tout réel $p \in]0;1[$, il existe un entier n_0 tel que : si $n \geq n_0$, alors $P\left(p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}}\right) > 0,95$.

Démonstration

On reprend les notations de la démonstration page 1 ;

si $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$, d'après le théorème de Moivre-Laplace, on peut écrire :

$$\lim_{n\to+\infty} P\left(Z_n \in [-2;2]\right) = \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = P(-2 \leq Y \leq 2)$$

où Y suit la loi normale centrée réduite.

Or : $P(-2 \leq Y \leq 2) \simeq 0,954$

On obtient donc

$$\lim_{n\to+\infty} P\left(p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \simeq 0,954$$

D'après le cours sur les limites d'une suite, on peut déduire qu'il existe un entier n_0 tels que pour tout $n \geq n_0$,

$$P\left(p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) > 0,95$$

D'après la remarque page 2, puisque $2\sqrt{p(1-p)} \leq 1$, on a l'inclusion

$$\left[p - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subset \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$$

En conclusion, on obtient donc : $P\left(p - \frac{1}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{1}{\sqrt{n}}\right) > 0,95$

2.2 Propriété

Soit F_n la variable aléatoire qui à tout échantillon de taille n, extrait d'une population dans laquelle la proportion d'un caractère est p, associe la fréquence obtenue. Alors l'intervalle $\left[F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}}\right]$ contient, pour n assez grand, la proportion p avec une probabilté égale au moins à $0,95$.

Démonstration

Puisque $F_n = \dfrac{X_n}{n}$, il suffit d'appliquer le résultat précédent en remarquant que :

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

2.3 Définition

Soit f la fréquence observée d'un caractère dans un échantillon de taille n, extrait d'une population dans laquelle la proportion de ce caractère est p. Alors l'intervalle $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right]$ est appelé intervalle de confiance de la proportion p au niveau de confiance $0,95$.

Cet intervalle est la réalisation, à partir d'un échantillon, d'un intervalle aléatoire contenant la proportion p avec une probabilité supérieure à $0,95$.