Chapitre 22 : Statistiques

1 Echantillonnage

1.1 Intervalle de fluctuation asymptotique

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

Propriété

Pour tout α dans $]0;1[$, on a : . . . .

où I_n désigne l'intervalle . . . .

u_α a été défini dans le chapitre sur les lois à densités par l'égalité . . . .

. . . .

Démonstration (exigible)

$$\frac{X_n}{n} \in I_n \Leftrightarrow \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

En multipliant par n, on obtient :

$$\frac{X_n}{n} \in I_n \Leftrightarrow \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

puis, en retranchant np :

$$\frac{X_n}{n} \in I_n \Leftrightarrow \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

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

$$\frac{X_n}{n} \in I_n \Leftrightarrow \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

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,

$$\lim_{n\to+\infty} P\left(Z_n \in [a;b]\right) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

où Y suit la loi normale centrée réduite.
Donc

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

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

Définition

L'intervalle . . . .

. . . .

. . . .

1.1.1 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 . . . .

. . . .

Remarque

Si $p \in]0;1[$, alors $0 \leq p(1-p) \leq \ldots\ldots$, donc $\sqrt{p(1-p)} \leq \ldots$ et $1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \ldots\ldots$

On obtient alors $\left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subset \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

. . . .

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 :

- . . . .
- . . . .

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 $u_{0,01} \simeq 2,58$.

Exemple 

Un joueur qui doit choisir au hasard une carte dans un jeu de 52 cartes obtient certains avantages s'il découvre un roi. On constate qu'il a retourné 22 fois un roi sur 100 essais.
Peut-on présumer, au risque de 5%, que ce joueur est un tricheur ?

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.

Propriété

On suppose que la variable aléatoire X_n suit la loi binomiale $\mathcal{B}(n,p)$. Pour tout réel $p \in]0;1[$,

. . . .

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 :

$$\lim_{n\to+\infty} P\left(Z_n \in [-2;2]\right) = \dots\dots\dots\dots\dots\dots$$

où Y suit la loi normale centrée réduite.

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

On obtient donc

. . . .

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

. . . .

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

. . . .

En conclusion, on obtient donc : . . . .

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

. . . .

. . . .

Démonstration

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

. . . .

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 . . . .

. . . .

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