Intervalles de fluctuation

Supposons que dans une population donnée, une proportion p a un trait particulier (les yeux bleus, un salaire supérieur à 3 000 euros, ...).

On considère un échantillon de taille n de cette population. Dans cet échantillon, une proportion f possède le trait. On souhaite déterminer si l'échantillon correspond à la population globale, ou s'il est particulier. Autrement dit, si f est « assez proche » de p selon un critère à déterminer.

Propriété **(Rappel de seconde)** Si p est la proportion d'un caractère dans une population (avec $0,2 \leqslant p \leqslant 0,8$), alors pour un échantillon de taille n avec $n \geqslant 25$, la fréquence f du caractère dans l'échantillon appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins $0,95$.
Cet intervalle est appelé intervalle de fluctuation au seuil de confiance de 95% (ou de risque de 5%).

On souhaite une propriété qui ne soit pas limitée sur les valeurs de p et de n. Pour cela nous allons utiliser la loi binomiale.

Soit X une variable aléatoire égale au nombre d'individus d'un échantillon de n personnes qui ont la probabilité p d'avoir le trait étudié. Alors $X \sim \mathscr{B}(n;p)$.
On détermine deux entiers a et b de la manière suivante :
- a est le plus petit entier tel que $P(X \leqslant a) \geqslant 0,025$
- b est le plus petit entier tel que $P(X \geqslant b) \leqslant 0,025$ (autrement dit que $P(X \leqslant b) \geqslant 0,975$)

On a alors $P(a \leqslant X \leqslant b) \geqslant 0,95$.
Par suite on observe que $P(a \leqslant X \leqslant b) = P\left(\frac{a}{n} \leqslant \frac{X}{n} \leqslant \frac{b}{n}\right)$, et $\frac{X}{n}$ est la variable aléatoire égale à la fréquence des individus ayant le trait étudié.
Ainsi, sur un grand nombre d'expériences sur des échantillons de taille n, 95% des échantillons environ auront une fréquence observée comprise entre $\frac{a}{n}$ et $\frac{b}{n}$.

Ainsi :

Propriété L'intervalle de fluctuation au seuil de confiance de 95% de la fréquence est

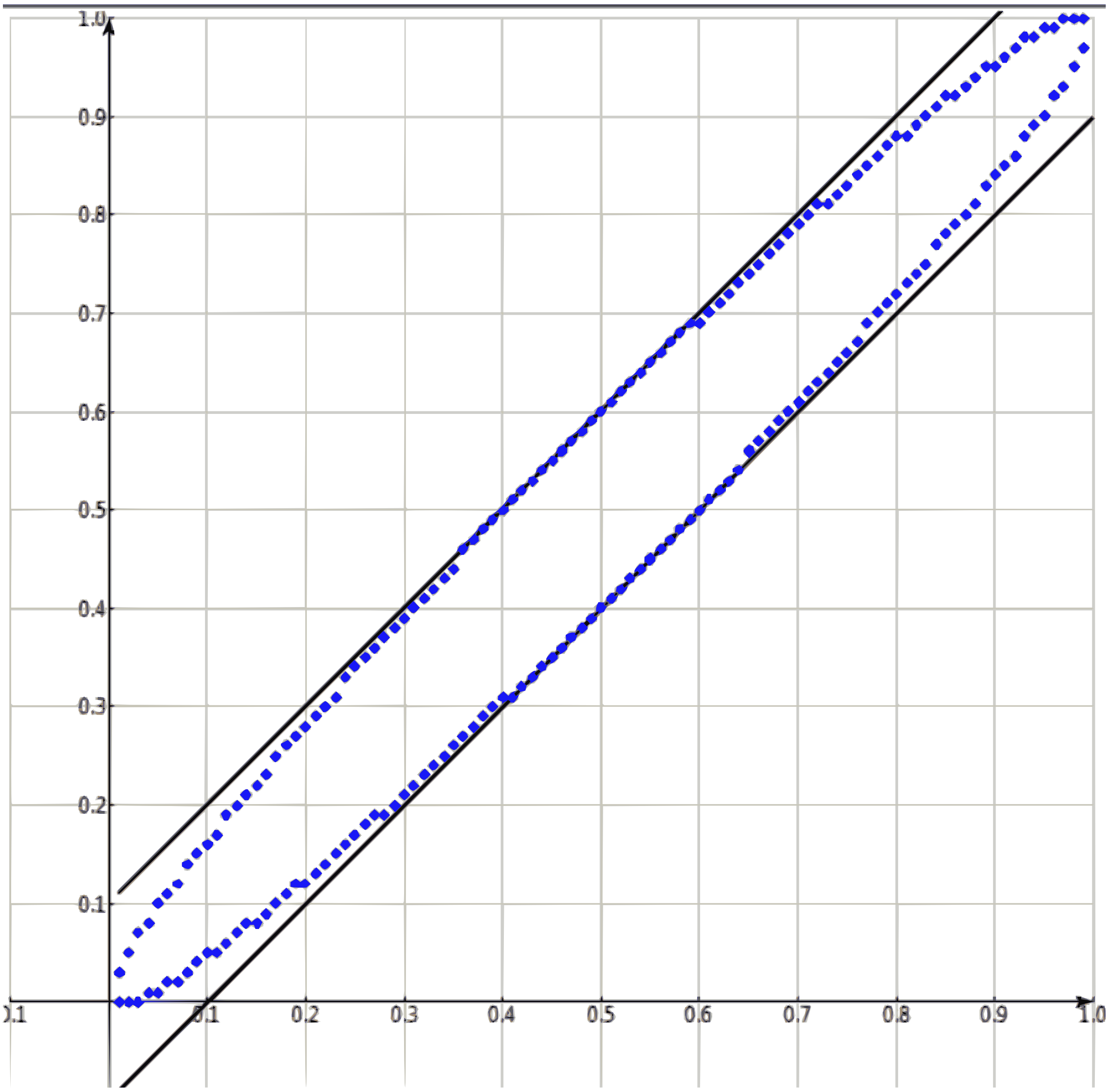
$$\left[\frac{a}{n}; \frac{b}{n}\right]$$

Pour revenir à la fréquence f observée, on adopte donc la règle de décision suivante :

Méthode On fait l'hypothèse que l'échantillon suit la même loi que la population totale, autrement dit que la proportion du trait étudié est p. Soit I l'intervalle de fluctuation de la fréquence à 95% dans des échantillons de taille n pour une proportion p.
- Si $f \notin I$, alors on rejette l'hypothèse au seuil de risque de 5%.
- Sinon, $(f \in I)$, on ne rejette pas l'hypothèse au seuil 5% (on dit parfois que l'on valide l'hypothèse).

On peut décider d'autres valeurs de seuils. Par exemple pour 99%, on cherche le plus petit a tel que $P(X \leqslant a) \geqslant 0,005$ et le plus petit b tel que $P(X \leqslant b) \geqslant 0,995$, pour que $P((a \leqslant X \leqslant b) \geqslant 0,99$.

Voici finalement une illustration de la taille des intervalles de fluctuation en fonction de la valeur de p (en abscisse) pour des échantillons de taille 100 selon la méthode de seconde (entre les deux droites puisque la taille de l'intervalle sera toujours constant égal à $\frac{2}{\sqrt{100}} = 0,2$) et la méthode utilisant la loi binomiale (entre deux points).



On observe qu'effectivement, plus p est petit ou au contraire grand, moins l'intervalle donné en seconde est précis par rapport à celui de la méthode de première. Ce dernier réduit ; en effet, si par exemple extrême $p = 0$, tous les échantillons devraient n'avoir qu'une fréquence nulle, il ne devrait pas y avoir de fluctuation.

Lire page 303 la méthode pour obtenir la table des valeurs $P(X \leqslant k)$ sur la calculatrice.

▶ **Exercices :** 68p311, 72,73p312 (intervalles)

▶ **Exercices :** 81,83,84p313 (prise de décision)