Intervalle de fluctuation – Intervalle de confiance – Prendre une décision – Donner une estimation

Intervalle de fluctuation: On réalise n fois une expérience aléatoire à deux issues, succès ou échec, (succès avec probabilité p). On note f la fréquence des succès.

- En seconde, on mettra en évidence que l'intervalle $\left[p-\frac{1}{\sqrt{n}}\ , p+\frac{1}{\sqrt{n}}\right]$ contient f dans plus de 95% des cas (à condition que $n \geq 25$ et $0{,}2 \leq p \leq 0{,}8$). On l'appellera intervalle de fluctuation au seuil de 95%.
- En classe de première, la loi binomiale étant connue, il est possible de déterminer un intervalle de fluctuation à l'aide d'une table, une calculatrice ou un logiciel. Cela peut être l'occasion de vérifier la validité du résultat vu en seconde.
- En terminale, on démontrera que si X_n suit une B(n,p) alors $P\left(\frac{X_n}{n} \in I_n\right) \approx 0{,}95$, avec $I_n = \left[p-\frac{1{,}96\sqrt{p(1-p)}}{\sqrt{n}}\ ; p+\frac{1{,}96\sqrt{p(1-p)}}{\sqrt{n}}\right]$.
- A noter que dans les situations où on ne connait pas p, il n'est pas possible de déterminer d'intervalle de fluctuation.

Prise de décision :
On ne connait pas p mais une valeur p_0 est proposée et un échantillon donné (et donc une fréquence de succès f).
Si l'expérience a une probabilité de succès égale à p_0, f appartient à l'intervalle de fluctuation (qui sera centré sur p_0) avec une probabilité supérieure ou égale à 95%, on convient d'accepter la valeur p_0 lorsque c'est le cas et de refuser p_0 dans le cas contraire.

- Avec cette méthode, la probabilité de refuser à tort la valeur p_0 est inférieure à 5%.

Estimation :
On ne connait pas p mais on cherche un intervalle dans lequel p doit raisonnablement se situer.
On sait que, sous certaines conditions ($n \geq 25$ et $0{,}2 \leq p \leq 0{,}8$), la probabilité qu'un échantillon aléatoire de taille n fournisse une fréquence f comprise dans l'intervalle $\left[p-\frac{1}{\sqrt{n}}\ , p+\frac{1}{\sqrt{n}}\right]$ est supérieure à 0,95.

$$\text{Or } f \in \left[p-\frac{1}{\sqrt{n}}\ , p+\frac{1}{\sqrt{n}}\right] \Leftrightarrow p \in \left[f-\frac{1}{\sqrt{n}}\ , f+\frac{1}{\sqrt{n}}\right].$$

On en déduit que, pour plus de 95 % des échantillons aléatoires de taille n, la fréquence inconnue p appartient à l'intervalle $\left[f-\frac{1}{\sqrt{n}}\ , f+\frac{1}{\sqrt{n}}\right]$. Cet intervalle, calculé à partir de l'échantillon, est appelé intervalle de confiance.

On conclut que avec un niveau de confiance égal à 95%, $\boldsymbol{p \in \left[f-\frac{1}{\sqrt{n}}\ , f+\frac{1}{\sqrt{n}}\right]}$.

- Il y a autant d'intervalles de confiance que d'échantillons.

- Dès que l'intervalle de confiance est déterminé, la situation de *p* par rapport à cet intervalle n'est plus aléatoire. (on ne peut donc plus parler de probabilité, c'est pourquoi on parle de niveau de confiance).

Deux exemples :

Prendre une décision :

Exercice 1 : En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population de comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine.

Déterminer l'intervalle de fluctuation au seuil de 5% du nombre d'américains d'origine mexicaine lors d'un tirage au sort de 870 personnes dans ce comté. Conclusion.

Donner une estimation :

Exercice 2 : Nous sommes en mars 2012. Est publié un sondage pour le second tour des présidentielles prévues en mai 2012.

Taille de l'échantillon : 1375
Abstention : 20%
Monsieur X : 52%
Monsieur Y : 48%

Dans les médias, on annonce donc que Monsieur X serait élu président de la république si les élections avaient lieu en mars 2012. Qu'en pensez-vous ?

Rappel : lorsque l'on veut estimer une proportion *p* au sein d'une population d'effectif important, on prélève un échantillon aléatoire de taille *n* pour lequel on obtient une fréquence *f* et on estime que *p* est compris dans l'intervalle de confiance $\left[f-\frac{1}{\sqrt{n}}\,, f+\frac{1}{\sqrt{n}}\right]$, avec un niveau de confiance de 0,95.

Eléments de réponses :

275 personnes se sont abstenues et 1100 se sont exprimées.

On peut considérer que l'on a un échantillon de 1100 personnes s'exprimant sur le deuxième tour. 52% d'entre elles se prononçant pour Monsieur X et 48% pour Monsieur Y. Les intervalles de confiance associés à ces deux fréquences sont $[0{,}489\,;0{,}551]$ et $[0{,}449\,;0{,}511]$. A priori ces résultats ne semblent pas garantir que la majorité de la population ait l'intention de voter pour X (au moment où le sondage a été réalisé).

<u>Attention</u> : L'intervalle de confiance (de X par exemple) est centré sur la fréquence calculée (ici 52%). La proportion *p* inconnue que l'on cherche à estimer est à l'intérieur ou à l'extérieur de cet intervalle (il n'est pas possible de dire que *p* a 95% de chance d'être dans cet intervalle). On dira plutôt : <u>*p* est dans l'intervalle $[0{,}489\,;0{,}551]$ avec un niveau de confiance de 95%.</u>