FICHE n°3
Echantillonnage

Cette fiche présente des propriétés (admises) et applications correspondant à des situations d'échantillonnage. Elle doit être étudiée en lien étroit avec les illustrations et travaux pratiques effectués en classe.

I. Fluctuation d'échantillonnage

Définition
- ¤ Dans le sens commun des sondages, un ***échantillon*** est une partie obtenue par prélèvement aléatoire dans une population.
- ¤ En statistique, on dit qu'un ***échantillon de taille n*** relève du ***modèle de Bernoulli*** s'il est obtenu par n répétitions indépendantes d'une même expérience aléatoire à deux issues (0 ou 1).
- ¤ Les distributions des fréquences observées varient d'un échantillon à l'autre. Ce phénomène est appelé ***fluctuation d'échantillonnage***.

Propriété (admise)

Lorsque la taille de l'échantillon augmente, la fluctuation est plus faible et la distribution des fréquences tend à se rapprocher d'une distribution de fréquences théoriques.

Application ***Pour estimer la proportion p d'un caractère d'une population***

Pour estimer la proportion p d'un caractère d'une population, on peut **simuler des échantillons aléatoires de grande taille** ; la fréquence f du caractère dans ces échantillons tend vers p.

Remarque Voir la fiche 1 « Probabilités ».

II. Intervalle de fluctuation

Dans ce paragraphe, on considère un modèle de Bernoulli dont on connaît la probabilité p d'obtenir 1.

Propriété (admise)

Pour environ 95 % des échantillons de taille n relevant du modèle de Bernoulli avec une probabilité p d'obtenir 1, la fréquence d'apparition du 1 appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$

(à condition que $n \geq 25$ et $0{,}2 \leq p \leq 0{,}8$).

Cet intervalle s'appelle l'**intervalle de fluctuation au seuil de 95 %**.

Application ***Pour prendre une décision à partir d'un échantillon***

Pour déterminer si un échantillon est compatible avec un modèle donné, on détermine l'intervalle de fluctuation correspondant et on calcule la fréquence f d'apparition du 1 dans l'échantillon :

- ¤ si ***f* n'est pas dans l'intervalle de fluctuation**, alors **on peut rejeter** l'hypothèse que l'échantillon soit compatible avec le modèle ;
- ¤ si ***f* est dans l'intervalle de fluctuation**, alors **on ne peut pas rejeter** l'hypothèse que l'échantillon soit compatible avec le modèle… Cela ne vaut pas dire qu'il est sûr d'être compatible !

III. Fourchette de sondage

Dans ce paragraphe, on considère un modèle de Bernoulli dont on ne connaît pas la probabilité $\boldsymbol{p}$ d'obtenir 1.

<u>*Propriété*</u> *(admise)*

> Parmi tous les échantillons de taille $\boldsymbol{n}$ possibles, environ 95 % des intervalles associés $\left[\boldsymbol{f}-\frac{1}{\sqrt{\boldsymbol{n}}}\,,\,\boldsymbol{f}+\frac{1}{\sqrt{\boldsymbol{n}}}\right]$ contiennent le nombre $\boldsymbol{p}$.
> On dit alors que la **fourchette de sondage** $\left[\boldsymbol{f}-\frac{1}{\sqrt{\boldsymbol{n}}}\,,\,\boldsymbol{f}+\frac{1}{\sqrt{\boldsymbol{n}}}\right]$ est **l'intervalle de confiance au niveau 0,95** de $\boldsymbol{p}$.

<u>*Application*</u> ***Pour estimer une proportion inconnue à partir d'un échantillon***

> On considère une population dont la proportion $\boldsymbol{p}$ d'un caractère donné n'est pas connue.
> En réalisant un échantillon aléatoire de taille $\boldsymbol{n}$ dont on détermine la fréquence $\boldsymbol{f}$ du caractère étudié, on peut estimer la proportion $\boldsymbol{p}$ à l'aide de l'intervalle de confiance $\left[\boldsymbol{f}-\frac{1}{\sqrt{\boldsymbol{n}}}\,,\,\boldsymbol{f}+\frac{1}{\sqrt{\boldsymbol{n}}}\right]$ au niveau 0,95.