Intervalles de fluctuations - Intervalles de confiance

I Distinction entre échantillonnage et estimation

Définition

- On considère une population d'individus. Lorsque l'on connaît ou lorsque l'on fait une hypothèse sur la proportion p d'individus ayant une caractéristique donnée dans une population et que l'on effectue un nombre n de tirages avec remise dans cette population, la fréquence observée appartient avec une certaine probabilité à un intervalle appelé intervalle de fluctuation de centre p et de longueur qui diminue lorsque n augmente. On parle alors de situation d'échantillonnage.
- Lorsque l'on ne connaît pas la proportion d'individus ayant une caractéristique donnée, en procédant à un nombre n de tirages avec remise on peut estimer à l'aide de la fréquence f obtenue la proportion p d'individus ayant cette caractéristique.
  Cette estimation se fait à l'aide d'un intervalle de confiance dont l'amplitude diminue lorsque le nombre n de tirages augmente.

II Échantillonnage

Propriété et définition

Pour tout entier naturel n, soit X_n une variable aléatoire suivant une loi binomiale $\mathscr{B}(n;p)$.. Alors quand n devient grand, l'intervalle $I_n = \left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$ contient la fréquence X_n avec une probabilité qui se rapproche de 0,95. On dit que c'est un intervalle de fluctuation asymptotique au seuil de 95 %.

Remarque :
En pratique, on utilise cette propriété dès que les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont vérifiées.

Test d'hypothèse :
On considère une population dans laquelle on suppose que la proportion d'un caractère est p. On fait l'hypothèse « La proportion dans la population est p »
On observe la fréquence f d'apparition de ce caractère sur un échantillon de taille n et on calcule l'intervalle de fluctuation I au seuil de 95 %.

- Si $f \in I$, au risque de 5 % d'erreur (ou au seuil de confiance de 95 %), on accepte l'hypothèse que la proportion dans la population est f.
- Si $f \notin I$, au risque de 5 % d'erreur on rejette l'hypothèse que la proportion dans la population est p.

Exemple : Un fournisseur d'accès à Internet affirme que, sur sa hotline, seuls 20 % des clients attendent plus de 5 minutes pour obtenir un interlocuteur.
Une association de consommateurs interroge au hasard 200 personnes ayant eu à s'adresser à cette hotline. Parmi ces personnes, 53 ont dû attendre plus de 5 minutes. Peut-on mettre en doute l'affirmation du fournisseur d'accès ?

On a $p : 0,2$; $n = 200$.
Les conditions sont vérifiées. L'intervalle de fluctuation est $I = \left[0,2 - \frac{1}{\sqrt{200}}\ ;\ 0,2 + \frac{1}{\sqrt{200}}\right] \approx [0,129\ ;\ 0,271]$.

L'hypothèse à tester est « 20 % des clients attendent plus de 5 minutes ». La fréquence observée est $f = \frac{53}{200} \approx 0,265$.
$f \in [0;129;0;271]$..
Donc au seuil de confiance de 95 %, on ne peut pas mettre en doute l'affirmation du fournisseur d'accès.

III Estimation

Propriété et définition

Soit p la proportion inconnue d'apparition d'un caractère. On appelle f la fréquence d'apparition du caractère sur un échantillon de taille n. Alors, l'intervalle $\left[f-\dfrac{1}{\sqrt{n}}\ ;\ f+\dfrac{1}{\sqrt{n}}\right]$ contient pour n assez grand la proportion p avec une probabilité supérieure ou égale à 0,95. L'intervalle $\left[f-\dfrac{1}{\sqrt{n}}\ ;\ f+\dfrac{1}{\sqrt{n}}\right]$ est appelé intervalle de confiance de p au niveau de confiance 0,95.

Remarque : Un intervalle de confiance au niveau de 95 % a une amplitude de $\dfrac{2}{\sqrt{n}}$.
L'amplitude diminue lorsque la taille n de l'échantillon augmente.

Exemple : Lors d'un scrutin électoral, on souhaite connaître la proportion p de français votant pour le candidat « A ».
L'institut SOFOS mène une enquête auprès de 1 000 personnes tirées au hasard et avec remise (c'est-à-dire qu'une même personne peut éventuellement être choisie plusieurs fois). Le résultat indique qu'une proportion $f = 49\,\%$ d'entre elles voteront pour le candidat « A ».

1. Quelle est la loi suivie par le nombre de personnes votant pour « A » dans une enquête avec remise effectuée auprès de 1 000 personnes ?
2. Étant donné le résultat de l'enquête SOFOS. donner un intervalle de confiance à 95 % pour l'estimation de p.
3. Peut-on affirmer d'après l'enquête que le candidat « A » n'aura pas la majorité des votes, c'est- à-dire que p < 0,5 ?

Solution

1. ¿ chaque étape de l'enquête, il y a une probabilité p de tirer une personne votant pour « A », car on procède à des tirages avec remise. De plus, les 1 000 tirages sont effectués de façon indépendante. L'enquête est donc un schéma de Bernoulli dont chaque succès correspond à tirer une personne votant pour « A », Le nombre de personnes votant pour « A » dans l'enquête suit donc une loi binomiale $\mathscr{B}(\ ;\ )$ où p est inconnu.
2. Notons F la proportion de personnes votant pour « A » parmi 1 000.
D'après le cours, si $n \geqslant 30$, $np > 5$,$n(1-p) > 5$, un intervalle de confiance au niveau 95
% pour p est $\left[F-\dfrac{1}{\sqrt{1\,000}}\ ;\ F+\dfrac{1}{\sqrt{1\,000}}\right]$.
Ici, $n = 1\,000 > 30$, $np \geqslant 5 \Leftrightarrow p \geqslant 0,005$ et $n(1-p) \geqslant 5 \Leftrightarrow p \leqslant 0,995$.
On ne connaît pas p, mais au vu de l'enquête, $f = 0,49$, donc on peut raisonnablement penser que p les conditions sont vérifiées.
L'intervalle de confiance est $\left[0,49-\dfrac{1}{\sqrt{1\,000}}\ ;\ 0,49+\dfrac{1}{\sqrt{1\,000}}\right] = [0,458\ ;\ 0,522]$.
3. L'intervalle de confiance précédent montre qu'il est tut à fait possible que $p \geqslant 0,5$.
Donc il n'est pas raisonnable d'affirmer suite à l'enquête que le candidat « A » n'obtiendra pas majorité des votes.