Échantillonnage
2nd

Sacha Darthenucq

Prérequis:

- **Vocabulaire ensembliste et logique (2nd)**
- **Intervalles et distance (2nd)**
- **Statistiques (2nd)**
- **Probabilités (2nd)**

Le principe de l'échantillonnage est de dégager des tendances parmi une population de grande taille.

1 Échantillon

Définition: Un échantillon de taille n est une sélection de n individu parmi une population.

Exemple: Lors d'un sondage, on ne peut interroger toute la population, on va donc effectuer le sondage sur un échantillon représentatif de la population.

Remarque: La taille d'un échantillon doit être suffisamment grande pour que la tendance dégagée soit juste, mais pas trop car le travail à fournir pour récupérer les données risque d'être trop important.

2 Intervalle de fluctuation

En fonction de l'échantillon de population choisit, la fréquence d'apparition d'un caractère va changer, cela s'appelle la fluctuation d'échantillonnage.

Théorème: Soit p la proportion d'un caractère dans une population donnée. Soit f la fréquence d'apparition de ce caractère dans un échantillon de taille n.

Si $0.2 < p < 0.8$, $n \geq 25$, alors dans 95% des cas $f \in \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$. Cet intervalle, noté I s'appelle l'intervalle de fluctuation.

Remarque: Pour la borne inférieure de l'intervalle on arrondit toujours en dessous, tandis que pour la borne supérieure, on arrondit au dessus.

Remarque: Lorsque l'on applique ce théorème on doit connaître la proportion p du caractère dans la population. On peut aussi supposer la proportion p d'un caractère dans une population, et en fonction des fréquences obtenu dans divers échantillon valider ou non l'hypothèse.

Remarque: Attention, la fréquence obtenue n'appartient à l'intervalle que dans 95% des cas, donc lorsque l'on trouve un résultat qui invalide une affirmation, toujours émettre la possibilité de se trouer dans les 5% de cas où la fréquence trouvée n'appartient pas à l'intervalle.

Exemple: Deux candidats A et B s'affrontent lors d'une élection. Après un débat télévisé important, le sondage de 600 personnes donne le candidat B perdant avec seulement 232 voix. Sommes-nous sûrs que le candidat B va perdre l'élection ?
On suppose que 50% des personnes veulent voter réellement pour le candidat B (c'est à dire que le candidat B gagne l'élection). Nous allons voir si la fréquence obtenue lors du sondage se situe dans l'intervalle de fluctuation.
Ici $p = \frac{50}{100} = 0.50$, $n = 600$ et $f = \frac{232}{600} \simeq 0.387$.
$-\frac{1}{\sqrt{n}} \simeq -0.04$ et $+\frac{1}{\sqrt{n}} \simeq +0.041$
$I = [0.46; 0.541]$, $f \notin I$ donc nous sommes sûrs à 95% que moins de 51% des personnes voteront pour le candidat B, donc il ne sera pas élu.

3 Vulgarisation de la loi des grands nombres

Propriété: Considérons une expérience aléatoire, et un évènement E dont on ne connait pas la probabilité. La fréquence de réalisation de l'évènement E au cours de n expériences aléatoires, lorsque n est très grand, est proche de la probabilité de réalisation de l'évènement E.

Exemple: Lors d'un sondage, en interrogeant 10000 personnes représentative de la population, on est quasiment sûr de la probabilité de vote pour chaque candidat que l'on trouve (à moins qu'un grand bouleversement dans la campagne électorale survienne).