

Dans votre cours, commencez un nouveau chapitre : « Chapitre 8 — Échantillonnage ».

Introduction

Dans cette introduction, nous utilisons des formules qui seront données dans la suite du cours. Essayez donc de comprendre cette introduction sans vous attarder sur les formules.

Intervalle de confiance

Connaissant la fréquence d'un paramètre dans un échantillon, un *intervalle de confiance* permet d'estimer la proportion de ce paramètre dans l'ensemble d'une population.

Question. Pour connaître les intentions de vote des français lors des dernières élections présidentielles, on a interrogé un échantillon représentatif de 1 000 personnes annonçant aller voter, peu avant le second tour, et 59 % d'entre elles ont annoncé voter pour Emmanuel Macron.

Peut-on affirmer qu'Emmanuel Macron va être élu ?

Réponse. *Remarquons que l'on connaît la fréquence dans l'échantillon (parmi les 1 000 personnes interrogées), et on souhaite estimer la proportion dans l'ensemble de la population.*

L'intervalle de confiance à 95 % de la proportion de votants pour Emmanuel Macron est $\left[0, 59 - \frac{1}{\sqrt{1000}}; 0, 59 + \frac{1}{\sqrt{1000}}\right]$, soit $[0, 55; 0, 63]$. Cela signifie qu'on peut affirmer, sans trop de risques de se tromper, que dans l'ensemble de la population, la proportion de personnes allant voter pour Emmanuel Macron est comprise entre 55 % et 63 %. Donc Emmanuel Macron devrait avoir plus de la moitié des suffrages, et il devrait être élu.

Intervalle de fluctuation

Un *intervalle de fluctuation* permet de répondre au genre de questions inverse de la précédente : connaissant une probabilité d'un critère, il permet

d'estimer, dans un échantillon, la proportion d'individus satisfaisant ce critère.

Question. Un joueur malchanceux se demande si la pièce avec laquelle il joue est faussée ou non. Pour vérifier, il la lance 500 fois, et a obtenu 287 fois « Pile ». Peut-il affirmer que la pièce est faussée, ou le joueur a-t-il simplement manqué de chance ?

Réponse. Commençons par raisonner avec une pièce équilibrée (que nous comparerons ensuite à la pièce faussée).

L'intervalle de fluctuation à 95 % de la proportion de « Pile » sur 500 lancers d'une pièce équilibrée est $\left[0, 50 - \frac{1}{\sqrt{500}}; 0, 50 + \frac{1}{\sqrt{500}}\right]$, soit $[0, 45; 0, 55]$. Cela signifie qu'en lançant une pièce équilibrée 500 fois, il y a 95 % de chances d'obtenir une proportion de « Pile » comprise entre 0,45 et 0,55. La proportion de « Pile » obtenue par le joueur est $287/500$ soit environ 0,57, ce qui est à l'extérieur de l'intervalle de fluctuation. On peut, avec une probabilité d'erreur inférieure à 5 %, rejeter l'hypothèse que la pièce est équilibrée (en d'autres termes, il est probable qu'elle soit faussée).

Pour comprendre la différence entre intervalle de confiance et intervalle de fluctuation, vous pouvez regarder la vidéo suivante :

<http://youtu.be/97vzxWsyie8>



Cours

Recopiez (ou imprimez et collez) les deux parties « Intervalle de confiance » et « Intervalle de fluctuation ».

1 Intervalle de confiance

On cherche à estimer la proportion p , inconnue, d'un caractère d'une population (la probabilité d'obtenir Pile pour une pièce truquée ; la proportion de votants pour un candidat à une élection ; la proportion de pièces défectueuses produites par une machine ; etc.).

Définition et Propriété. On considère un échantillon de taille n dans une population, et on calcule la fréquence observée f du caractère dans cet échantillon.

Alors on appelle *intervalle de confiance* de la proportion p au niveau de 95 % l'intervalle :

$$\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}} \right]$$

Dans au moins 95 % des cas, la proportion p se trouve dans cet intervalle.

Exemple. *Pour mieux connaître sa clientèle, un magasin fait un sondage auprès de ses clients. Sur les 200 répondants, 163 ont déclaré avoir trouvé le personnel accueillant. Peut-on affirmer que plus des trois quarts des clients trouvent le personnel accueillant ?*

Sur un échantillon de taille 200 (donc $n = 200$), 163 individus trouvent le personnel accueillant (donc $f = 163/200 = 0,815$). Donc sur l'ensemble des clients, l'intervalle de confiance à 95% de la proportion de clients trouvant le personnel accueillant est :

$$\left[0,815 - \frac{1}{\sqrt{200}}; 0,815 + \frac{1}{\sqrt{200}} \right] = [0,74; 0,89]$$

Puisqu'une partie de l'intervalle est inférieure à 0,75 (soit trois quarts), on ne peut pas affirmer que plus des trois quarts des clients trouvent le personnel accueillant.

2 Intervalle de fluctuation

On connaît la probabilité d'apparition d'un caractère p dans une population (la probabilité d'obtenir Pile pour une pièce équilibrée ; la proportion de pièces défectueuses produites par une machine ; la proportion d'électeurs ayant voté pour un candidat à une élection ; etc.), et on souhaite estimer, pour un échantillon de taille n , quelle va être la fréquence observée.

Définition et Propriété. On connaît la proportion p d'apparition d'un caractère dans une population, et on considère un échantillon de taille n de cette population.

Alors on appelle *intervalle de fluctuation* à 95 % de la fréquence f l'intervalle :

$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$$

La probabilité que la fréquence observée f soit dans cet intervalle est supérieure à 95 %.

Propriété (Prise de décision). On suppose que la proportion d'un caractère dans une population est p .

- Si la fréquence observée f n'est pas dans l'intervalle de fluctuation, alors l'hypothèse que la proportion est p dans l'ensemble de la population est rejetée, avec un risque d'erreur de 5 %.
- Si la fréquence observée f est dans l'intervalle de fluctuation, alors l'hypothèse n'est pas rejetée.

Exemple. *On a observé qu'une certaine maladie guérit spontanément (c'est-à-dire sans traitement) au bout d'une semaine dans 60% des cas. Une équipe de chercheurs a développé un médicament qui, lors des tests effectués sur 132 personnes, en a guérit 85 en moins d'une semaine (soit légèrement plus de 60%).*

Peut-on affirmer que ce médicament est efficace ?

Remarquons que cette question est importante. En effet, en étudiant des patients au hasard, le taux de guérison spontanée ne sera pas exactement égal à 60% : il sera parfois un peu plus élevé (si on a de la chance), parfois un peu plus bas (si on n'a pas de chance). La question est donc de savoir si les résultats obtenus avec le médicament sont causés par l'efficacité du médicament, ou juste par la chance.

Que se serait-il passé si nous n'avions donné aucun médicament aux 132 patients de l'étude ? Parmi les 132 personnes (donc $n = 132$), environ

60% devrait guérir spontanément au bout d'une semaine (donc $p = 0,6$). L'intervalle de fluctuation à 95% est donc :

$$\left[0,6 - \frac{1}{\sqrt{132}}; 0,6 + \frac{1}{\sqrt{132}}\right] = [0,51; 0,69]$$

Cela signifie que la proportion de patients ayant guéri spontanément devrait être comprise entre 51% et 69%.

Or avec le médicament, 85 patients ont guéri, soit $85/132 = 64\%$. Cette proportion est à l'intérieur de l'intervalle de fluctuation, donc elle est parfaitement compatible avec le modèle où seul le hasard intervient : rien ne prouve que le médicament ait eu un quelconque effet.

Exercices

Essayez de faire les exercices, puis vérifiez vos résultats avec les corrigés à la fin de ce document.

1 Fluctuation d'échantillonnage

Vidéo d'un exercice corrigé. ATTENTION : La formule utilisée pour calculer l'intervalle n'est pas la même que celle vue dans le cours, mais à part cela, toute la méthode est la même.

<http://youtu.be/QZ0YFthGI0Y>



Exercice 1 (Polynésie, 7 juin 2016). Dans une population, on estime qu'il naît 51 % de garçons et 49 % de filles.

L'intervalle de fluctuation à au moins 95 % de la fréquence des filles dans un échantillon de 100 naissances choisies au hasard sera :

- a. $[0,48; 0,50]$ b. $[0,39; 0,59]$ c. $[0,41; 0,61]$ d. $[0,47; 0,51]$

2 Intervalle de confiance

Vidéo d'un exercice corrigé.
<http://youtu.be/cU5cJ1CVAM8>



Exercice 2 (Métropole La Réunion, 18 juin 2015). Pour le repas du midi, les visiteurs restant toute la journée dans un parc peuvent :

- soit déjeuner dans l'un des restaurants du parc ;
- soit consommer, sur une aire de pique-nique, un repas qu'ils ont apporté.

La direction souhaite estimer la proportion p de visiteurs déjeunant dans l'un des restaurants du parc.

Un sondage est effectué à la sortie du parc : 247 visiteurs parmi 625 ont déjeuné dans l'un des restaurants du parc.

Déterminer un intervalle de confiance au niveau de confiance de 95 % de la proportion p de visiteurs déjeunant dans l'un des restaurants du parc.

Exercice 3 (Pondichéry, 26 avril 2017). L'EFS affirme que dans une région donnée : « 23 % de la population donne son sang au moins une fois par an ».

On interroge au hasard un échantillon de 1 000 personnes habitant cette région. Parmi elles, 254 ont donné au moins une fois leur sang au cours de la dernière année.

Peut-on mettre en doute l'affirmation de l'EFS ? Justifier la réponse à l'aide d'un intervalle de fluctuation.

Bilan

Pas de bilan cette semaine à me rendre. Mais n'hésitez pas à me poser vos questions par courriel ou par l'ENT.

Corrigés

Exercice 1. On considère 100 naissances, donc la taille de l'échantillon est 100 (donc $n = 100$). La probabilité d'avoir une fille est 49% (donc $p = 0,49$). Donc l'intervalle de fluctuation à 95% de la fréquence de filles dans l'échantillon (c'est-à-dire la fréquence de filles que l'on peut espérer observer dans cet échantillon) est :

$$\left[0,49 - \frac{1}{\sqrt{100}}; 0,49 + \frac{1}{\sqrt{100}} \right] = [0,39; 0,59]$$

Exercice 2. On a interrogé 625 visiteurs (donc $n = 625$), et parmi ceux-ci, 247 ont déjeuné dans un des restaurants du parc (donc $f = 247/625 \approx 0,40$). L'intervalle de confiance à 95% de la proportion p de visiteurs déjeunant dans un des restaurants du parc est donc :

$$\left[0,40 - \frac{1}{\sqrt{625}}; 0,40 + \frac{1}{\sqrt{625}} \right] = [0,36; 0,44]$$

Exercice 3. On a interrogé 1 000 personnes (donc $n = 1000$), et parmi celles-ci, 245 ont donné au moins une fois leur sang au cours de l'année (donc $f = 245/1000 = 0,245$). Donc l'intervalle de fluctuation de la proportion de personnes ayant donné son sang au cours de l'année est :

$$\left[0,245 - \frac{1}{\sqrt{1000}}; 0,245 + \frac{1}{\sqrt{1000}} \right] = [0,213; 0,277]$$

L'EFS affirme que cette proportion est 23% (soit 0,230). Ce nombre est dans l'intervalle de fluctuation, donc il n'y a aucune raison de mettre en doute l'affirmation de l'EFS.