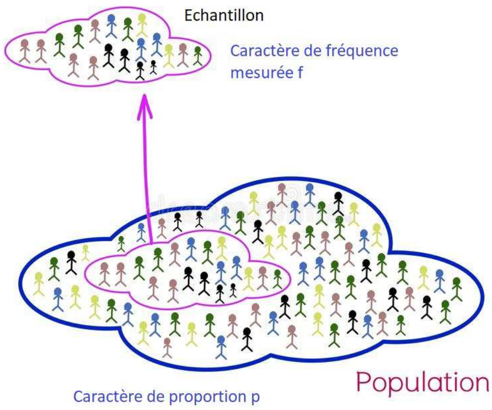
Math93.com

Statistiques et échantillonnage

Seconde

Introduction

L'échantillonnage est un domaine des mathématiques à la croisé entre statistiques et probabilités. L'idée est d'étudier une propriété d'une population dont l'effectif est trop grand pour être observée de façon exhaustive.
On prélève alors un échantillon de cette population, à partir duquel on souhaite obtenir des informations sur la population totale avec un certain degé de précision.
Un exemple connu est la réalisation de sondages.



I Fluctuation d'échantillonnage

I.1 Échantillons

Exemple

- Exemple 1 : On lance 10 fois un dé cubique et on note la liste des résultats obtenus, cela constitue un échantillon de taille 10. Par exemple :

$$\{1-5-3-1-6-3-4-3-6-2\}$$

- Exemple 2 : On interroge 100 personnes au hasard et on note à chaque fois leur couleur préférée, cela constitue un échantillon de taille 100. Par exemple :

$$\{vert\,;\,rouge\,;\,rouge\,;\,bleu\,;\,\cdots\,;\,jaune\,;\,noir\}$$

Définition 1 (Échantillon)

Un **échantillon (aléatoire) de taille n** est la liste de n résultats obtenus par **n répétitions indépendantes** d'une même expérience aléatoire.

Définition 2 (Distribution des fréquences)

La **distribution des fréquences** associées à un échantillon est la liste des fréquences des issues de cet échantillon.

Exemple

Dans l'exemple 1, on peut calculer les fréquences de chacune des issues $\{1-5-3-1-6-3-4-3-6-2\}$ de cet échantillon.

Numéro	1	2	3	4	5	6
Effectifs	2	1	3	1	1	2
Fréquences en %	20%	10%	30%	10%	10%	20%

I.2 Fluctuation d'échantillonnage

Dans une population, on s'intéresse à l'apparition d'un caractère. On note p la proportion de ce caractère dans la population.
On prélève un échantillon dans la population et on note f la fréquence d'apparition du caractère dans cet échantillon.
En observant plusieurs échantillons, on remarque que la fréquence observée fluctue autour de la proportion p.

Exemple

On s'intéresse au caractère : « *obtenir "Pile" en lançant une pièce* ».
La proportion est donc clairement $p = 0,5$.
On lance une pièce 100 fois et on note la fréquence de sortie des nombres pairs. On obtient alors un échantillon de taille 100. On répète 5 fois cette expérience, on obtient alors 5 échantillons de taille $n = 100$.

Échantillon n°	1	2	3	4	5
fréquences de "Pile" en %	45%	53%	51%	48%	49%

Propriété 1

Soit plusieurs échantillons de même taille d'une expérience aléatoire. La distribution des fréquences varie d'un échantillon à l'autre : c'est la **fluctuation d'échantillonnage**.

I.3 Réalisation d'une simulation

Définition 3

Simuler une expérience aléatoire, c'est la remplacer par un autre expérience qui conduit aux mêmes résultats dans les même conditions.

Par exemple l'expérience aléatoire associée à la naissance d'un enfant (fille ou garçon), peut être simulée par le jeu de pile ou face.

Consultez la fiche : simulation sur tableur, sur calculatrice ou en algorithmique.

Exemple

On cherche à modéliser un jeu de bonneteau. Un maître du jeu mélange 3 cartes, deux rois et une dame et le joueur doit trouver la dame. Il a donc 1 chance sur 3 de gagner.
La simulation d'une partie consiste à obtenir 1 (gagner) avec une probabilité $\frac{1}{3}$ et 0 avec une probabilité $\frac{2}{3}$. Il suffit d'entrer la commande :

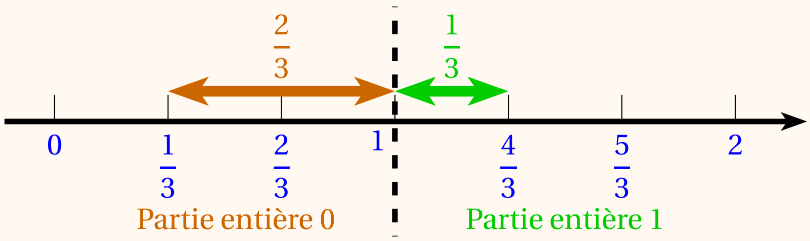
ent(NbrAléat+1/3) sur Texas ou Int(Rand#+1/3) sur Casio.

En effet :

- la fonction*NbrAléat* sur Texas ou *Rand#* sur Casio renvoie x avec $0 \leq x < 1$.
- Donc l'instruction *NbrAléat+1/3* sur Texas ou *Rand#+1/3* sur Casio renvoie $x+\frac{1}{3}$ tel que :

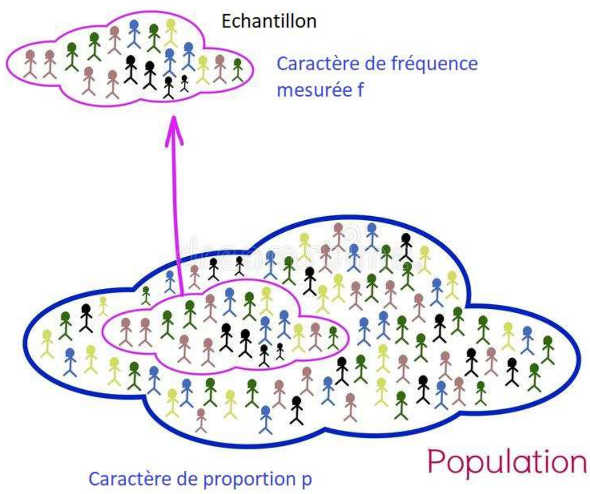
$$\frac{1}{3} \leq x+\frac{1}{3} < \frac{4}{3}$$

La partie entière de ce nombre est donc 0 avec une probabilité $\frac{2}{3}$ et 1 avec une probabilité $\frac{1}{3}$.



II Intervalle de fluctuation au seuil 95%

II.1 La notion d'intervalle de fluctuation



Dans une population, la proportion d'un caractère est p. Si l'on prélève des échantillons de cette population, le fréquence f d'apparition de ce caractère dans chaque échantillon fluctue autour de la proportion p.

Définition 4 (Intervalle de fluctuation au seuil 95%)

Un **intervalle de fluctuation au seuil 95%** est un intervalle I tel que, pour au moins 95% de l'ensemble des échantillons possibles, la fréquence observée appartient à I.
Remarque : pour une même situation, il en existe plusieurs.

II.2 Un intervalle particulier

Propriété 2 (Intervalle de fluctuation au seuil 95% (admis))

Soit un caractère dont la proportion dans une population est p.
Pour n assez grand et p ni proche de 0, ni proche de 1, il y a environ 95% des échantillons de taille n issus de cette population qui sont tels que la fréquence f observée appartienne à l'intervalle :

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

Les conditions sont : $\begin{cases} n \geq 25 \\ 0,2 \leq p \leq 0,8 \end{cases}$.

Exemple

On lance 2 500 fois un dé et on obtient 1 150 fois un nombre pair. Le dé est-il truqué ?

- On fait l'hypothèse que le dé n'est pas truqué, il y a donc 1 chance sur 2 d'obtenir un nombre pair soit
$$p = 0,5$$
- On assimile la situation à un échantillon de taille $n = 2\,500$, avec une probabilité $p = 0,5$ d'obtenir un nombre pair.
- **Les conditions sont vérifiées** :
$$\begin{cases} n = 2\,500 \geq 25 \\ 0,2 \leq p = 0,5 \leq 0,8 \end{cases}$$
- **Intervalle de fluctuation au seuil 95%** :
$$I = \left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right] = \left[0,5 - \frac{1}{\sqrt{2\,500}}\ ;\ 0,5 + \frac{1}{\sqrt{2\,500}}\right]$$
soit
$$\boxed{I = [0,48\ ;\ 0,52]}$$
- La fréquence observée f est $f = \dfrac{1\,150}{2\,500} = 0,46 \notin I = [0,48\ ;\ 0,52]$ donc on rejette l'hypothèse, le dé est sans doute mal équilibré (ou truqué).

Propriété 3 (Largeur de l'intervalle)

La largeur de l'intervalle de fluctuation au seuil 95% pour un échantillon de taille n est de l'ordre de grandeur de $\dfrac{2}{\sqrt{n}}$.