



Chapitre V : Echantillonnage

I Introduction

Exemple 1. On souhaite répondre à deux questions différentes :

Q1 Quel est le rappeur préféré des élèves de la classe ?

Q2 Quel est le rappeur préféré des français ?

Répondons à la question Q1 en remplissant le tableau suivant :

Rappeur	Cheu-B	Booba	Damso	Sans avis
Nombre de votes				

1. La réponse à la question Q1 est donc :

.....

2. Peut-on répondre à la question Q2 de la même façon ? Pourquoi ?

.....

.....

3. Comment pourrait-on procéder pour répondre à la question 2 ?

.....

.....

II Echantillonnage

On s'intéresse à un **caractère quantitatif** (ex : taille des vis produites par une usine, le nombre d'arbres par commune, durée de vie des tortues...) ou **qualitatif** (oui/non, le rappeur préféré des français, le pays d'origine d'une pièce européenne...) dans une **population** de taille N , où N est très grand (les vis, les communes, les tortues, les français, les pièces européennes...).

Définition II.1

Un **échantillon** de taille n est construit en tirant n fois aléatoirement, uniformément et avec remise un individu (une vis, une commune, une tortue, un français, une pièce....) parmi la population totale.



- Aléatoirement : au hasard.
- Uniformément : chaque individu a la même chance d'être tiré.
- Avec remise : chaque individu est remis dans la population pour le tirage suivant. Ceci est important pour que les chances de chaque individu d'être tiré soient les mêmes d'un tirage à l'autre. On admet que lorsque la population est très grande, l'absence de remise change tellement peu les probabilités qu'un tirage avec remise peut être approché par un tirage sans remise.



Exemple 2. On souhaite connaître la proportion de personnes à Montataire qui possède un téléphone.

1. Quelle est la population ?

.....

2. Quel est le caractère d'intérêt ? Est-il qualitatif ou quantitatif ?

.....

.....

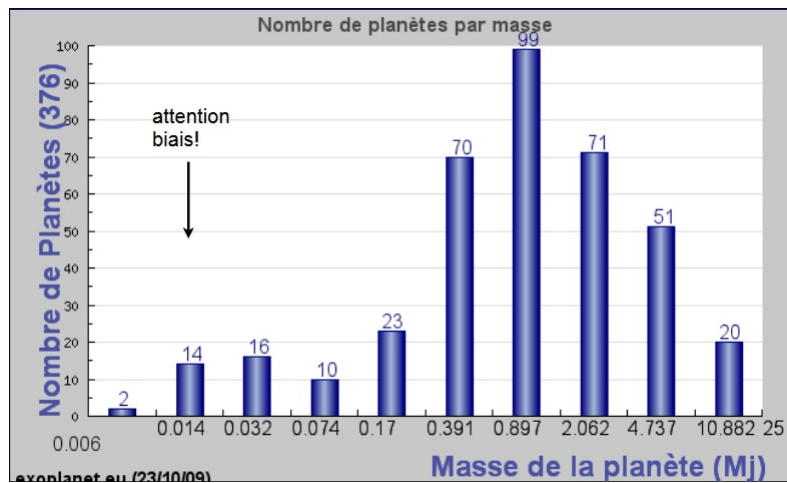
3. Comment pensez-vous procéder pour effectuer un échantillon ?

.....

.....

.....

Exemple 3. On donne ci-dessous l'histogramme du nombre d'exoplanètes découverte en 2009 en fonction de leur masse (en proportion de la masse de Jupiter).





A l'aide de l'extrait suivant de wikipédia, expliquer pourquoi ce relevé ne constitue pas un bon échantillon des exoplanètes de notre galaxie :

« Les méthodes de détection utilisées sont principalement la méthode des vitesses radiales qui déduit la présence d'une planète à travers l'incidence de son attraction gravitationnelle sur la vitesse radiale de son étoile et la méthode du transit qui identifie une planète lorsque celle-ci passe devant son étoile en mesurant l'affaiblissement de l'intensité lumineuse de l'astre. Un biais découlant de ces méthodes a abouti à la détection d'une majorité de planètes aux caractéristiques très différentes de celles présentes dans le Système solaire, en particulier de planètes gazeuses très proches de leur étoile hôte. »

.....

.....

.....

.....

.....

III Intervalle de fluctuations

On considère une population ayant une proportion p d'individus présentant un caractère donné. Un échantillon de cette population nous fournit une indication sur cette proportion p .

Exemple 4. Pièce équilibrée. On modélise un lancé d'une pièce équilibrée par le lancé d'un dé à dix faces. Lancer le dé 10 fois et remplir le tableau suivant en mettant un trait dans la colonne correspondante à chaque lancé :

pile : 1, 2, 3, 4 et 5	face : 6, 7, 8, 9, 0

En mettant les résultats en commun, nous obtenons les proportions de piles suivantes :

Echantillon	n°1	n°2	n°3	n°4	n°5	n°6
Nombre de piles : (sur 50 lancés)						
Fréquence f						

On observe que les échantillons ne sont pas tous égaux : ils fluctuent.

Exemple 5. Pièce truquée. On modélise un lancé d'une pièce déséquilibrée par le lancé d'un dé à dix faces. Lancer le dé 10 fois et remplir le tableau suivant en mettant un trait dans la colonne correspondante à chaque lancé :



pile : 1, 2 et 3	face : 4, 5, 6, 7, 8, 9, 0

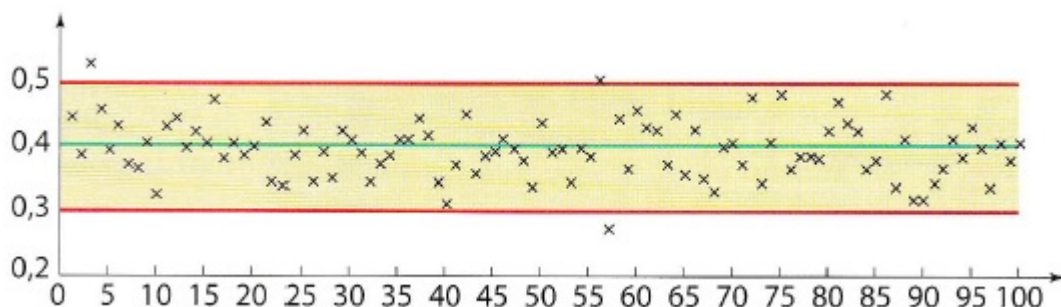
En mettant les résultats en commun, nous obtenons les proportions de piles suivantes :

Echantillon	n°1	n°2	n°3	n°4	n°5	n°6
Nombre de piles : (sur 50 lancés)						
Fréquence f						

Définition III.1

Un intervalle de fluctuation pour une fréquence f au seuil de 95% est un intervalle I tel que dans 95% au moins des échantillons, la fréquence observée appartient à cet intervalle I .

Exemple 6. En 2011, parmi les français de 15 ans ou plus ayant terminé leurs études la proportion p de titulaires du baccalauréat était égale à 40,4%, c'est-à-dire $p = 0,404$. Voici une simulation de 100 échantillons de taille 150.



L'intervalle $I = [0,3; 0,5]$ dessiné en rouge comprend 97% des échantillons et est donc un intervalle de fluctuation *empirique* (construit par l'expérience) au seuil de 95%.

Proposition III.2(admise)

On considère toujours une population ayant une proportion p d'individu exprimant un caractère. Soit un échantillon de taille n . Si

- $0,2 \leq p \leq 0,8$,
- $n \geq 25$,

alors un intervalle de fluctuation pour la fréquence observée f au seuil de 95% est

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

NB : cet intervalle est de taille $2/\sqrt{n}$. Donc plus l'échantillon est grand, meilleure est la précision.



Application 1.

1. Construire un intervalle de fluctuation à 95% pour l'exemple 4. Les fréquences observées sont-elles dans cet intervalle ?

.....

.....

.....

2. Rassembler les échantillons en un échantillon de taille plus grande. Calculer l'intervalle de fluctuation à 95% associé. La fréquence observée est-elle dans cet intervalle ?

.....

.....

.....

3. Faire le même travail pour l'exemple 5.

.....

.....

.....

IV Estimation et prise de décision

Proposition IV.1

Lorsque l'on connaît la proportion p , on peut déterminer si l'échantillon est représentatif de la population totale de la façon suivante :

1. On construit à l'aide de p et de la taille de l'échantillon n , l'intervalle I de fluctuation à 95%.
2. On regarde si la fréquence observé f appartient dans cet intervalle.
3. On conclut : si $f \in I$ alors à 95%, l'échantillon est conforme à la population totale, si $f \notin I$ alors à 95% l'échantillon n'est pas représentatif de la population globale.



Application 2. On s'intéresse à la diffusion des euros étrangers en France. Une étude en 2012 montre que 34% des euros circulant en France avaient été frappés à l'étranger. On souhaite savoir si cette proportion est toujours respectée dans la région de Montataire en 2017. Pour ce faire chaque élève apporte sa contribution en comptant le nombre d'euros (peu importe la valeur des pièces) français et étrangers chez lui.

1. Quelles critiques pouvons-nous faire sur cet échantillonnage ?

.....

.....

.....

2. Voici la fréquence que nous avons observée :

.....

Conclure.

.....

.....

.....

