Échantillonnage – Activités

Activité 1 : Pièce de monnaie

1) On va s'intéresser à l'expérience suivante : On choisit une pièce de monnaie, on la lance et on note le côté visible obtenu : P pour « Pile » et F pour « Face »
   a. Quelle est la probabilité théorique d'apparition du côté « Pile » ?
   b. Prenez une pièce de monnaie et réaliser 10 fois cette expérience, en notant les résultats obtenus à l'aide de P et F. Quelle est la fréquence d'apparition du côté « Pile » sur vos 10 lancers ?
   c. Regrouper les résultats de vos camarades dans le tableau suivant. Que remarque-t-on ?

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Fréquence du côté «Pile»					

2) A l'aide des échantillons de toute la classe, former 5 échantillons de taille 50.
   a. Quelle est la fréquence d'apparition du côté « Pile » dans ce nouvel échantillon
   b. Regrouper les résultats de vos camarades dans le tableau suivant :

	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Fréquence du côté «Pile»					

   c. Comparer les fréquences obtenues avec celles de la question 1.

Activité 2 : Simulation

Il est possible de simuler une expérience aléatoire à l'aide d'une calculatrice ou de l'ordinateur.

1) A l'aide de votre calculatrice, réaliser plusieurs fois la tâche suivante :

   Casio : Menu « RUN », touche OPTN, sélectionner PROB, puis Ran#.

   TI : Appuyer sur math, sélectionner PRB, puis NbrAléat

   Que permet de faire cette fonction ?

2) Avec l'algorithme 1 et la calculatrice, simuler 10 fois le jeu Pile ou Face.
3) On considère maintenant l'algorithme 2 :
   a. Quels sont les valeurs possibles pour x ?
   b. Que permet de faire cet algorithme ?
   c. On suppose que x prend successivement les valeurs 0,1,1,1,0,0,1,0,1,1. Quelle valeur sera affichée en sortie de l'algorithme ?
4) On s'intéresse maintenant à une paire de dés bien équilibrés.
   a. Proposer une façon de simuler un dé avec la calculatrice.
   b. Ecrire un algorithme permettant de simuler le lancer de deux dés.
   c. Ecrire un algorithme permettant le nombre d'apparition du chiffre « 6 » sur 100 lancers de dé.

Algorithme 1

$x \leftarrow$ *Nombre aléatoire entre* 0 *et* 1
Si $x \leq 0.5$
    Alors Afficher « Pile ».
Sinon
    Afficher « Face »
Fin Si

Algorithme 2

$NbPiles = 0$
Pour i allant de 1 à 10 **Faire**
  $x \leftarrow$ *Entier aléatoire entre* 0 *et* 1
  Si $x = 1$
    Alors $NbPiles \leftarrow NbPiles + 1$
  Fin Si
Fin Pour
Afficher $NbPiles$

Chapitre 17 : Échantillonnage

1 – Fluctuation d'échantillonnage

● On considère une population, dont on suppose que la **proportion** d'un caractère est p.

Ex : Dans la population française, il y a 13 % de gauchers.

● On prélève un échantillon de **taille** n dans la population française et on s'intéresse à la **fréquence** f du caractère dans cet échantillon.

Ex : Votre classe est un échantillon de taille $n =$ …… et la fréquence de gauchers est $f =$ ………………………….

● Si l'on évalue la fréquence f de ce caractère dans plusieurs échantillons de même taille n, on obtiendra des résultats légèrement différents. Ce phénomène s'appelle **fluctuation d'échantillonnage.**

Ex : Dans une autre classe, la fréquence de gauchers pourra être différente.

● Plus la taille n des échantillons est grande, moins les variations de f seront importantes : La fréquence f observée dans les échantillons se rapproche de la proportion p.

Ex : Résultats obtenus sur 10 classes de 40 élèves puis sur 10 amphithéâtres de 250 élèves :

Classe	1	2	3	4	5	6	7	8	9	10
Gauchers	10,00%	32,50%	10,00%	7,50%	10,00%	15,00%	10,00%	2,50%	7,50%	12,50%

Amphi	1	2	3	4	5	6	7	8	9	10
Gauchers	13,60%	11,60%	10,80%	18,40%	12,00%	11,20%	13,40%	10,40%	17,50%	14,40%

2 – Intervalle de fluctuation

Afin de rendre compte des fluctuations naturelles de la fréquence observée dans les différents échantillons on va utiliser un intervalle :

Théorème 1 : Dans au moins 95 % des cas la fréquence f appartient à l'intervalle $I = [p - \frac{1}{\sqrt{n}}\ ; p + \frac{1}{\sqrt{n}}]$.

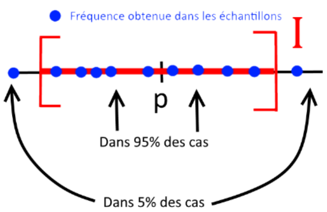
On dit que I est **un intervalle de fluctuation au seuil de 95%**

Condition d'application :

● Si $n \geq 30$, $np > 5$ et $n(1-p) > 5$, (p ni trop petit ni trop grand)

● Si $n \geq 25$ et $0.2 \leq p \leq 0.8$

Remarque : Plus n est grand, plus l'intervalle sera précis.



Exemple 1 : Calculer l'intervalle de fluctuation de la fréquence de gauchers dans des échantillons de taille n

● Comme $p = 0.13$ et que $np > 5$ on doit avoir $n \times 0.13 > 5$ c'est-à-dire $n > \frac{5}{0.13} \approx 38{,}49$ donc $n \geq 39$

● $n = 40$: $I = \left[0.13 - \frac{1}{\sqrt{40}}; 0.13 + \frac{1}{\sqrt{40}}\right] = [0.13 - 0.158\ ; 0.13 + 0.158] = [0; 0.288]$

● $n = 250$: $I = \left[0.13 - \frac{1}{\sqrt{250}}; 0.13 + \frac{1}{\sqrt{250}}\right] = [0.13 - 0.063\ ; 0.13 + 0.063] = [0.067; 0.193]$

3 – Prise de décision

On considère une population dans laquelle on suppose l'hypothèse que la proportion d'un caractère est p.

On prélève un échantillon de taille n, et on détermine la fréquence f de ce caractère dans l'échantillon.

Règle de décision : Soit $I = [p - \frac{1}{\sqrt{n}} \,; p + \frac{1}{\sqrt{n}}]$ l'intervalle de fluctuation au seuil de 95%.

- Si $f \in I$ alors on accepte l'hypothèse avec un **seuil de confiance** de 95 %.
- Si $f \notin I$ alors on rejette l'hypothèse avec un **risque** de 5 %.

Exemple 2 : On suppose que la proportion d'hommes dans la population française est de $50\ \%$. On prélève un échantillon de 5000 personnes où on compte 2400 hommes. Doit-on accepter ou rejeter l'hypothèse ?

$I = [0.5 - \frac{1}{\sqrt{5000}} \,; 0.5 + \frac{1}{\sqrt{5000}}] = [0.485 \,; 0.515]$

$$f = \frac{2400}{5000} = 0.48$$

Donc $f \notin I$, on rejette l'hypothèse avec un risque de $5\ \%$.

Echantillonnage - Exercices

1 **(Table de nombre aléatoires)**

Une urne contient 10 boules : cinq rouges, trois noires et deux blanches. On tire une boule et on note sa couleur et on la remet dans l'urne.

1) Avec la table de nombres aléatoires entiers de 0 à 9 donnée ci-dessous, simuler 25 tirages en expliquant votre méthode.

3	6	3	6	7	9	5	8	0	1	8	9	0	7	4	0	8	1	6	6
0	5	6	6	6	9	4	4	1	4	5	9	3	0	7	6	5	7	5	8
7	7	1	6	4	4	4	5	8	0	8	0	8	8	4	7	1	0	2	3
1	3	7	4	4	5	1	7	9	4	5	8	6	0	6	0	0	3	8	8
3	0	4	8	3	6	9	9	8	3	6	5	9	3	1	9	2	1	8	4

2) Calculer les fréquences obtenues pour chaque couleur.
3) Déterminer pour chacune des couleurs l'intervalle de fluctuation pour un échantillon de taille 25.
4) Vos fréquences sont-elles dans ces intervalles ? Conclure

2 **(Discrimination)**

L'ensemble des faits évoqués ci-dessous est réel. En novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison pour cambriolage d'une résidence et tentative de viol. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population de comté était d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine.

1) Déterminer l'intervalle de fluctuation correspondant à la proportion d'origine mexicaine pour un échantillon de taille 870.
2) La fréquence des personnes d'origine mexicaine dans les personnes convoquées est-elle dans cet intervalle ? Qu'en conclure ?

3 **(Réussite au baccalauréat)**

À Gustave Eiffel, pour la session 2009 du baccalauréat, il y a eu 154 reçus pour 170 candidats se présentant à l'épreuve. Les fréquences des reçus en Série STI, STL et S étaient, respectivement, 0,766, 0,896 et 0,963.

Après avoir vérifié que toutes les conditions sont remplies pour appliquer les intervalles de fluctuation, déterminer si les différences de réussite entre les filières peuvent être dues aux fluctuations d'échantillonnage.

4 **(Election présidentielle 2007)**

Au premier tour de l'élection présidentielle française de Mai 2007, parmi les suffrages exprimés, les proportions, en pourcentage, pour les candidats ayant obtenu plus de 2% des suffrages, étaient les suivantes :

Bayrou	Besancenot	De Villiers	Le Pen	Royal	Sarkozy
18,57	4,08	2,23	10,44	25,87	31,18

Cinq mois plus tôt, le 13 Décembre 2006, l'institut de sondage BVA faisait paraître un sondage effectué sur un échantillon de 797 personnes dont voici les résultats, en pourcentage, concernant les candidats précédemment cités :

Bayrou	Besancenot	De Villiers	Le Pen	Royal	Sarkozy
7	4	2	10	34	32

1) Pour quels candidats peut-on appliquer les intervalles de fluctuation parmi ceux présents au premier tour ?
2) Pour les quatre candidats arrivés en tête déterminer les intervalles de fluctuation pour un échantillon de taille 797.
3) Les résultats du sondage donnent-ils des fréquences appartenant à ces intervalles ?
4) Qu'en conclure ?

5 **(Assemblée nationale)**

On considère que la proportion de femmes dans la population française est $1/2$. À l'assemblée nationale, il y a 577 députés, dont 108 femmes. Peut-on considérer que cette répartition est un effet de la fluctuation d'échantillonage ou bien dire que la parité des sexes n'est pas respectée à l'assemblée nationale ?

6 **(Parité)**

Dans une région où il y a autant de femmes que d'hommes, les entreprises sont tenues de respecter la parité. L'entreprise A a un effectif de 100 personnes dont 43 femmes. L'entreprise B a un effectif de 2 500 personnes dont 1 150 femmes.

1) Calculer le pourcentage de femmes dans ces deux entreprises. Qu'en conclure ?
2) Si respecter la parité revient à ne pas tenir compte du caractère homme-femme, on peut alors considérer l'ensemble des salariés d'une entreprise comme un échantillon prélevé au hasard dans la population de la région.
   a. Déterminer les intervalles de fluctuation relatifs aux deux échantillons.
   b. Les résultats confirment-ils la conclusion de la première question ?