Intervalle de fluctuation

I- Un exemple

Dans la classe de Seconde E pour l'année scolaire 2011–2012, il y avait 24 garçons et 10 filles, ce qui paraît disproportionné.
On peut se demander toutefois si, lorsqu'on choisit 34 élèves au hasard dans une population constituée d'une moitié de filles et d'une moitié de garçons, cette distribution est rare.

1) Quelle était la fréquence des garçons dans la classe de Seconde E ?

2) Expliquer comment simuler le choix de 34 élèves au hasard dans une population d'une moitié de filles et d'une moitié de garçons à l'aide de la fonction *random* de la calculatrice.

3) Procéder à cette simulation en notant le nombre de filles et de garçons obtenus et calculer la fréquence des garçons dans votre simulation (arrondie au centième).

4) Écrire cette fréquence au tableau et noter les résultats des simulations de la classe dans le tableau ci-dessous :

5) Déterminer, pour cette série statistique :

- **a.** les valeurs extrêmes, les premier et troisième quartiles, les premier et neuvième déciles, la médiane et la moyenne ;
- **b.** représenter le diagramme en boite correspondant ;
- **c.** déterminer l'intervalle interquartile et interpréter le résultat ;

6) D'après ces résultats, peut-il arriver que le hasard produise une distribution comparable à celle de la Seconde E ? Si oui, est-ce fréquent ?

II- Loi des grands nombres et intervalle de fluctuation

Nous avons vu dans le chapitre de probabilité que, lorsque qu'on répète une expérience aléatoire un grand nombre de fois, les différentes fréquences d'apparition ont tendance à se stabiliser.
Ce constat est un résultat mathématique appelé *La loi des grands nombres* :

Théorème — *Loi des grands nombres*

Pour une expérience donnée, dans le modèle défini par une loi de probabilité, les distributions des fréquences calculées sur des séries de taille n se rapprochent de la loi de probabilité quand n devient grand.

Nous l'admettrons.
Les mathématiciens ont obtenu des règles assez précises sur la façon dont les fréquences se rapprochent de la probabilité et une première approximation de ces règles, la seule au programme de la Seconde, est la suivante, qu'on admettra :

Propriété — *Intervalle de fluctuation en statistiques*

Dans une population, la proportion d'un caractère est p.
On produit un échantillon de taille n de cette population et on détermine la fréquence f du caractère dans cet échantillon.
Si p est compris entre 0,2 et 0,8 et si n est supérieur ou égal à 25, alors, dans environ 95 % des cas, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$, que l'on appelle *intervalle de fluctuation au seuil de 95 %* (ou *au risque de 5 %*)

On peut aussi reformuler la propriété en termes de probabilités :

Propriété — *Intervalle de fluctuation en probabilité*

Soit une expérience aléatoire où la probabilité d'un évènement A est p. On reproduit cette expérience n fois et on détermine la fréquence f d'apparition de l'évènement A.
Si p est compris entre 0,2 et 0,8 et si n est supérieur ou égal à 25, alors, dans environ 95 % des cas, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$, que l'on appelle *intervalle de fluctuation au seuil de 95 %* (ou *au risque de 5 %*)

Remarque On remarquera que plus n est grand et plus l'intervalle de fluctuation est petit. En effet :

- avec $n = 25$, l'intervalle de fluctuation est de la forme $[p - 0,2\,;\, p + 0,2]$ (soit $p \pm 20\%$)
- avec $n = 100$, l'intervalle de fluctuation est de la forme $[p - 0,1\,;\, p + 0,1]$ (soit $p \pm 10\%$)
- avec $n = 400$, l'intervalle de fluctuation est de la forme $[p - 0,05\,;\, p + 0,05]$ (soit $p \pm 5\%$)
- avec $n = 10000$, l'intervalle de fluctuation est de la forme $[p - 0,01\,;\, p + 0,01]$ (soit $p \pm 1\%$)
- etc.

Cela est cohérent avec la loi de grands nombres : plus n est grand et plus la fréquence d'un évènement tend vers la probabilité de cet événement.

III- Retour à notre exemple d'introduction

Essayons de répondre à la question suivante :
« Dans le cas de la classe de Seconde E, peut-on avancer, au risque de 5 % de se tromper, que l'échantillon (la classe) est représentatif d'une population (le lycée) comportant une moitié de filles et d'une moitié de garçons ? Et si ce n'est pas le cas, quelles peuvent être les raisons ? »

1) **a.** Dans notre population de référence, quelle est la valeur de p qu'on a supposée ?
   b. Quelle est la valeur de n ?
   c. Déterminer alors l'intervalle de fluctuation correspondant à cette éxpérience.
   d. Quel pourcentage des fréquences obtenues par les simulations de la classe appartient à cet intervalle ?
   e. Répondre à la question.

2) Et si notre supposition, pour p, était fausse ?
   À l'administration du lycée, on pouvait obtenir l'information suivante : « Au Lycée Gustave Eiffel, pour l'année scolaire 2011–2012, il y a en Seconde 214 élèves, dont 80 filles et 134 garçons ».
   a. Déterminer l'intervalle de fluctuation (toujours pour un échantillon de taille 34).
   b. La fréquence des garçons de la Seconde E appartient-elle à cet intervalle ? Qu'en conclure ?

IV- Exercices

Exercice 1

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.
Vos fréquences sont-elles dans ces intervalles ?
Conclure.

Exercice 2

D'après le site de l'IREM de Paris 13.
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 ?

3) Qu'en conclure ?

Exercice 3

Les questions 1 et 2 sont indépendantes.

1) 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.
Déterminer si les différences de réussite entre les filières peuvent être dues aux fluctuations d'échantillonage.

2) Dans le village chinois de Xicun en 2000, il est né 20 enfants dont 16 garçons. On suppose que la proportion de garçons et de filles est la même à la naissance dans toute l'espèce humaine.
Déterminer si la fréquence des naissances de garçons dans le village de Xicun en 2009 peut être due aux fluctuations d'échantillonage.

3) Avez-vous vérifié que toutes les conditions étaient remplies pour appliquer les intervalles de fluctuation dans les deux questions précédentes ?

Exercice 4

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 ces candidats 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 ?

Exercice 5

Les questions 1 et 2 sont indépendantes.

1) On considère que la proportion de femmes dans la population française est $\frac{1}{2}$. 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 ?

2) En 1990, les employés et ouvriers constituaient 58,7 % de la population française (d'après le recensement de l'INSEE). Suite à l'élection législative de 1993 on recensait 1,6 % de députés dont l'ancien métier était employé ou ouvrier.
Peut-on considérer que cette répartition est un effet de la fluctuation d'échantillonage ?

Exercice 6

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 ?