Chapitre 16

Fluctuations d'échantillonage

Sommaire

16.1 Activités

Rappels :

- *l'effectif d'un résultat est le nombre de fois que ce résultat apparaît ;*
- *la fréquence d'un résultat est l'effectif de ce résultat divisé par l'effectif total.*

ACTIVITÉ 16.1 (Simulations de séries de lancers de dés)**.**
L'objectif de cette activité est de produire des séries de 50 lancers de dé à 6 faces et d'observer la distribution des fréquences de chacune des faces.
Pour éviter des lancers de dés qui peuvent être bruyants, on va simuler ces lancers à l'aide de la calcultrice.

1. (a)
    - Se rendre dans la console d'éxecution de Python de la calculatrice
    - Entrer *from random import **
    - Entrer *random()* à plusieurs reprises

   (b) Que fait cette fonction ?

   (c) Comment peut-on simuler le lancer d'un dé à 6 faces avec cette fonction ?

   Pour la suite de l'activité, on appelera *lancer de dé* la simulation d'un dé obtenu à la calculatrice.

2. **Par binôme (groupe de deux élèves)**

(a) Lancers.
On notera les résultats dans les tableaux 16.1 page suivante.

- l'un lance un dé 50 fois, l'autre note le résultat obtenu ;
- on recommence en permuttant les rôles ;
- chaque binôme (groupe de deux) obtient alors deux tableaux de cinquante résultats et complète les trois tableaux de fréquence.

(b) Graphiques.
Les graphiques sont à faire dans les repères de la page 142.
On note en abscisses les numéros des faces du dé et en ordonnées les fréquences de chacun des numéros.

- Faire les diagrammes des fréquences de vos résultats et de ceux de votre voisin sur un même graphique en utilisant deux couleurs différentes.
- Faire les diagrammes des fréquences de votre binôme sur le graphique suivant.

3. **Par groupe puis pour la classe**

(a) Lancers.
On notera les résultats dans les tableaux 16.2 page suivante.

- relever les résultats de tous les binômes de votre groupe et compléter le quatrième tableau de fréquence ;
- relever enfin les résultats des deux groupes et compléter le dernier tableau.

(b) Graphiques.
Les graphiques sont à faire dans les repères de la page 142.

- Faire les diagrammes des fréquences de votre groupe.
- Faire les diagrammes des fréquences de votre classe sur le graphique suivant.

4. **Comparaison des graphiques**

(a) Comparer le diagramme de vos fréquences à celui de votre voisin.

(b) Comparer le diagramme des fréquences de votre binôme à celui d'autres binômes.

(c) Comparer le diagramme des fréquences de votre groupe à celui de l'autre groupe puis à celui de la classe.

(d) Que constate-t-on ?
Ce phénomène s'appelle *fluctuation d'échantillonage sur des séries de taille 50.*

TABLE 16.1: Binôme

Mes 50 lancers

Tableau de fréquence de mes résultats

Face	Effectif	Fréquence
1		
2		
3		
4		
5		
6		
Total	50	

Ceux de mon voisin

Tableau de fréquence des résultats de mon voisin

Face	Effectif	Fréquence
1		
2		
3		
4		
5		
6		
Total	50	

Tableau de fréquence de mon binôme

Face	Effectif	Fréquence
1		
2		
3		
4		
5		
6		
Total	100	

TABLE 16.2: Pour mon groupe puis pour la classe

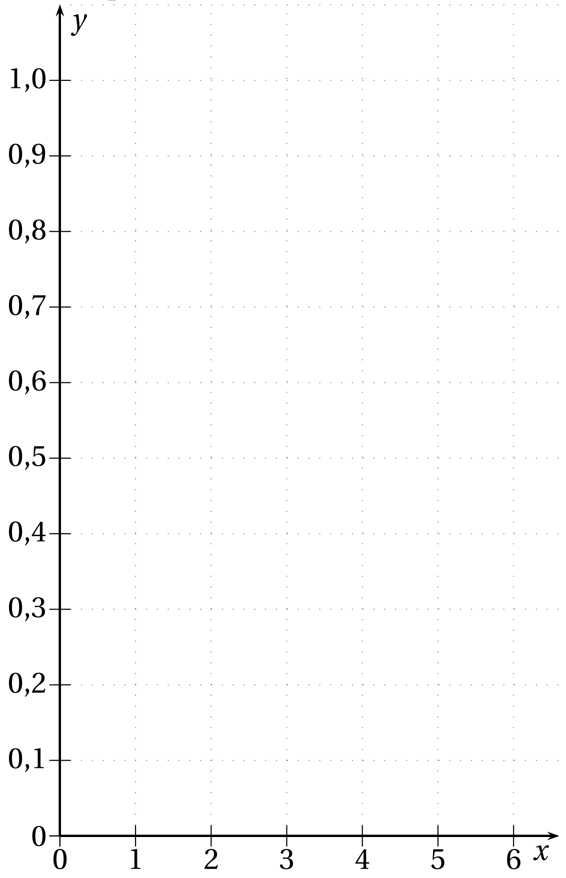
Tableau de fréquence de mon groupe

Face	Effectif	Fréquence
1		
2		
3		
4		
5		
6		
Total		

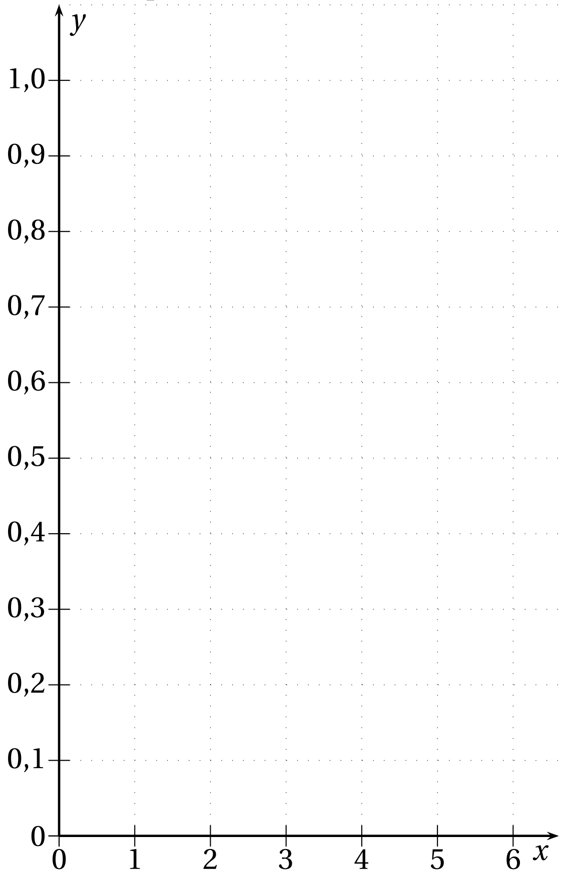
Tableau de fréquence de la classe

Face	Effectif	Fréquence
1		
2		
3		
4		
5		
6		
Total		

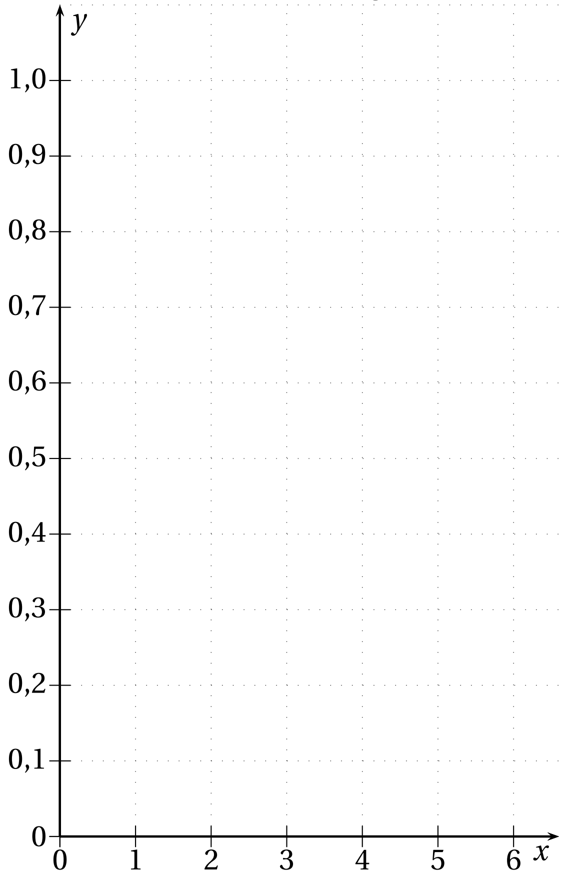
Mes fréquences et celles de mon voisin



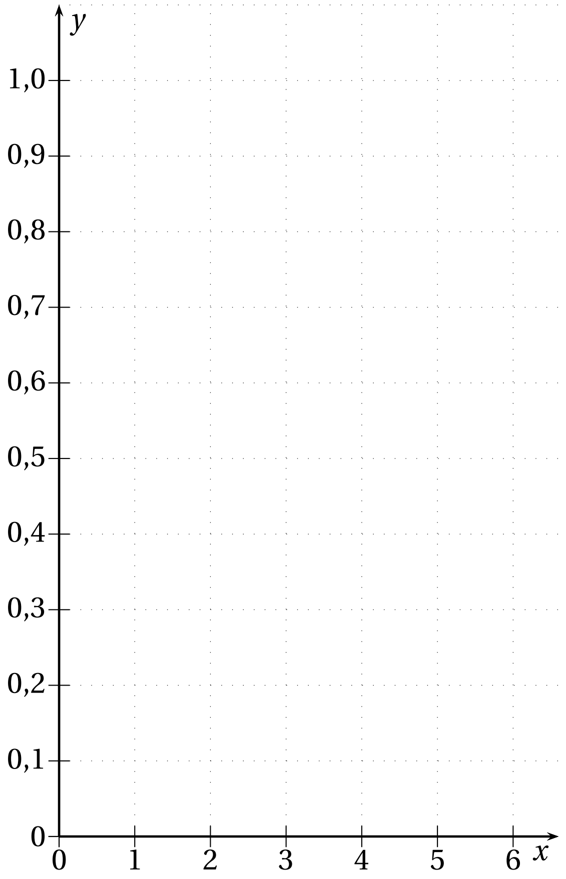
Les fréquences de mon binôme



Les fréquences de mon groupe

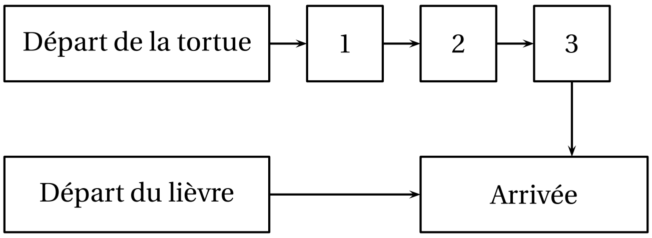


Les fréquences de la classe



ACTIVITÉ 16.2.
Le lièvre et la tortue font la course.
Le lièvre se divertit longuement mais quand il part, il file à l'arrivée. La tortue, quant à elle, avance inexorablement mais lentement vers l'arrivée.
On considère qu'on peut assimiler cette course au lancement d'un dé :

- si le 6 sort, le lièvre avance ;
- sinon la tortue avance d'une case et au bout de 4 cases la tortue a gagné.



1. Indiquer comment simuler plusieurs courses avec la liste de nombres aléatoires suivante et indiquer quel est le protagoniste qui gagne le plus souvent dans cette simulation :

   3 2 4 4 6 4 6 3 4 4 6 6 6 3 4 1 6 2 2 2 3 3 3 1 1 4 3 5 1 4

   Faire vérifier votre simulation par le professeur.

2. (a) Par groupe de 4, en utilisant votre tableau de 50 lancers de l'activité 16.1, simuler le maximum de courses possibles et indiquer la fréquence à laquelle chacun des protagonistes gagne.
   Désigner un représentant pour présenter vos résultats au tableau.
   (b) Comparer vos résultats à ceux des autres groupes de 4.

3. Regrouper toutes les courses de la classe pour déterminer à quelle fréquence chacun des protagonistes gagne.
   Ce résultat est-il loin du vôtre ?

16.2 Loi des grands nombres

Nous avons vu dans l'activité 16.1 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 16.1 (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.

16.3 Intervalle de fluctuation

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 :

Définition 16.1. L'intervalle de fluctuation au seuil de 95 %, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n.

Les propriétés suivantes donnent un intervalle qui contient l'intervalle de fluctuation et qui en sera une bonne approximation dans certaines conditions sur la proportion ou la probabilité et sur n, la taille de l'échantillon.

Propriété 16.2 (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.
Dès lors que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant$, alors, dans 95 % des cas au moins, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$, qui est une bonne approximation de ce qu'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 16.3 (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.*
Dès lors que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, alors, dans 95 % des cas au moins, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}\,;\, p + \frac{1}{\sqrt{n}}\right]$, qui est une bonne approximation de ce qu'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.

16.3.1 Utilisation de l'intervalle de fluctuation

On utilise l'intervalle de fluctuation lorsque la proportion p dans la population est connue ou bien si on fait une hypothèse sur sa valeur :

Représentativité : si la proportion p d'un caractère dans une population est connue, il permet de décider si un échantillon de taille n issu de cette population est représentatif de la population : si la fréquence f du caractère dans l'échantillon appartient à cet intervalle, on considère, au seuil de 95 %, que l'échantillon est représentatif.

Hypothèse sur p : si on émet une hypothèse sur la proportion p d'un caractère dans une population, il permet de savoir si on doit rejetter cette hypothèse : si la fréquence f observée dans un échantillon de taille n appartient cet intervalle, on considère que l'hypothèse selon laquelle la proportion est p dans la population n'est pas remise en question ; sinon on rejette l'hypothèse selon laquelle cette proportion vaut p.

16.4 Intervalle de confiance

On cherche à déterminer la proportion p d'un caractère dans une population, par exemple la proportion d'individus atteints d'une maladie bénigne. Il est souvent difficile pour des raisons à la fois financières et logistiques de pouvoir recueillir des données sur la population toute entière. Le plus souvent on se contente de travailler sur un échantillon de la population, dont on peut parfois vérifier au préalable s'il est représentatif de la population entière (sur d'autres critères, comme la fréquence d'hommes et de femmes par exemple). On sait que d'un échantillon à l'autre la fréquence d'apparition du caractère fluctue autour de la proportion p du caractère dans la population entière. Des simulations permettent d'obtenir qu'environ 95 % des intervalles de la forme $\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]$ contiennent la proportion p. Aussi à partir de la fréquence f d'apparition du caractère dans notre échantillon définit-on l'intervalle suivant :

Définition 16.2. L'intervalle $\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]$ est un *intervalle de confiance* de la proportion inconnue p au niveau de confiance 0,95.

Exemple. On souhaite estimer la proportion de personnes en surpoids, selon les critères de l'OMS, dans une ville quelconque. Pour cela 460 personnes ont été sélectionnées de manière aléatoire et un enquêteur est allé recueillir des informations auprès de ces personnes. La proportion de personnes en surpoids dans cet échantillon étudié est de 29,5 %.
L'intervalle $\left[f-\frac{1}{\sqrt{n}}\,;\,f+\frac{1}{\sqrt{n}}\right]=\left[0,295-\frac{1}{\sqrt{460}}\,;\,0,295+\frac{1}{\sqrt{460}}\right]\approx[0,25\,;\,0,34]$ est l'intervalle de confiance de la proportion de personnes en surpoids dans cette ville au niveau de confiance 0,95.

16.5 Exercices

16.5.1 Fluctuation

EXERCICE 16.1.
On se réfère dans cet exercice aux lancers de dés de l'activité 16.1.

1. Quelle est la probabilité de chacune des faces de ce dé ?
2. Déterminer les intervalles de fluctuations au seuil de 95 % pour des échantillons de taille 50, 100, n (où n est le nombre de lancers dans votre groupe) et p (où p est le nombre de lancers dans la classe pour chacune des faces).

3. Indiquer si les fréquences observées appartiennent à ces intervalles.

EXERCICE 16.2.
Dans la classe de Seconde 14 pour l'année scolaire 2010–2011, il y avait 9 garçons et 28 filles, ce qui paraît disproportionné.
On peut se demander toutefois si, lorsqu'on choisit 37 élèves au hasard dans une population constituée d'une moitié de filles et d'une moitié de garçons, cette distribution est rare.

Partie A : Une simulation.

1. Quelle était la fréquence des filles dans la classe de Seconde 14 ?
2. Expliquer comment simuler le choix de 37 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 filles 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. D'après vos résultats et ceux de la classe, peut-il arriver que le hasard produise une distribution comparable à celle de la Seconde 14 ? Si oui, est-ce fréquent ?

Partie B : Intervalle de fluctuation.

Essayons de répondre à la question suivante :
« Dans le cas de la classe de Seconde 14, 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 expérience.
   (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 Dupuy de Lôme, pour l'année scolaire 2010–2011, il y a en Seconde 524 élèves, dont 329 filles et 195 garçons ».
   (a) Déterminer l'intervalle de fluctuation (toujours pour un échantillon de taille 37).
   (b) La fréquence des filles de la Seconde 14 appartient-elle à cet intervalle ? Qu'en conclure ?

EXERCICE 16.3.
On lance deux dés cubiques et on note **la somme des deux nombres obtenus**.

Partie A : Simulation

1. Quels sont les résultats possibles ?
2. Avec la table de nombres aléatoires entiers de 0 à 9 donnée ci-dessous, simuler 25 lancers en expliquant votre façon de procéder.

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

3. Donner la suite des 25 résultats obtenus.
4. Calculer les fréquences obtenues pour chaque résultat possible.
5. Norbert a procédé lui aussi à une simulation de 25 lancers, avec une autre table de nombres aléatoires, et il a obtenu les résultats suivants :

Face	2	3	4	5	6	7	8	9	10	11	12
Effectif	1	1	2	2	1	4	4	4	3	2	1

Comparer ces résultats à ceux de votre simulation.

6. Une simulation à l'ordinateur a donné les résultats suivants :

Face	2	3	4	5	6	7	8	9	10	11	12
Effectif	24	49	86	103	145	178	139	114	77	55	30

Comparer ces résultats à ceux de votre simulation.

Partie B : Intervalle de fluctuation

1. Quelle est la probabilité de chacune des sommes ?
On pourra s'aider d'un arbre des possibles ou d'un tableau.
2. Déterminer les intervalles de fluctuations au seuil de 95 % pour des échantillons de taille 25 et 1000 pour les sommes dont la probabilité le permet.
3. Indiquer si les fréquences observées appartiennent à ces intervalles.

EXERCICE 16.4.
On lance deux dés cubiques et on note **le plus grand des deux nombres obtenus**.

Partie A : Simulation

1. Quels sont les résultats possibles ?
2. Avec la table de nombres aléatoires entiers de 1 à 6 donnée ci-dessous, simuler 50 lancers en expliquant votre façon de procéder.

2	3	2	5	4	2	1	2	5	3	1	3	5	5	2	6	5	3	2	6
6	2	5	2	6	2	3	5	6	3	4	4	4	1	6	6	5	1	1	5
5	5	6	1	4	2	2	1	2	6	6	3	2	3	4	3	5	1	3	2
6	4	2	1	5	6	2	6	1	6	6	3	1	2	6	6	4	4	1	5
4	1	2	6	3	2	5	2	1	2	1	6	2	3	6	6	1	3	2	2

3. Donner la suite des 50 résultats obtenus.
4. Calculer les fréquences obtenues pour chaque résultat possible.
5. Une simulation à l'ordinateur a donné les résultats suivants :

Face	1	2	3	4	5	6
Effectif	25	79	141	203	234	318

   Comparer ces résultats à ceux de votre simulation.

Partie B : Intervalle de fluctuation

1. Quelle est la probabilité de chacun des résultats ?
   On pourra s'aider d'un arbre des possibles ou d'un tableau.
2. Déterminer les intervalles de fluctuations au seuil de 95 % pour des échantillons de taille 50 et 1000 pour les résultats dont la probabilité le permet.
3. Indiquer si les fréquences observées appartiennent à ces intervalles.

EXERCICE 16.5.
Une urne contient 10 boules : **cinq** rouges, **trois** noires et **deux** blanches. On tire une boule et on regarde sa couleur.

Partie A : Simulation

1. Sur un très grand nombre de tirages, quelle fréquence prévoyez-vous pour le tirage d'une boule rouge ? d'une boule noire ? d'une boule blanche ?
2. Avec la table de nombres aléatoires entiers de 0 à 9 donnée ci-dessous, simuler 25 tirages en expliquant votre méthode.
   Calculer les fréquences obtenues pour chaque couleur.
3. Comparer les résultats obtenus question 2. avec vos prévisions de la question 1.

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

Partie B : Intervalle de fluctuation

1. Quelle est la probabilité de chacune des couleurs ?
2. Déterminer les intervalles de fluctuations au seuil de 95 % pour des échantillons de taille 25 pour les couleurs dont la probabilité le permet.
3. Indiquer si les fréquences observées appartiennent à ces intervalles.

EXERCICE 16.6.
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 16.7.
Les questions 1 et 2 sont indépendantes.

1. À Dupuy de Lôme, pour la session 2009 du baccaulauréat général, il y a eu 290 reçus pour 320 candidats se présentant à l'épreuve. Les fréquences des reçus en Série L, ES 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 16.8.
Au premier tour de l'élection présidentielle française de mai 2007, parmi les suffrages exprimés, les proportions, en pourcentage, pour les candidats ayant obtenu pour 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 16.9.
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 16.10.
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 ?

16.5.2 Confiance

EXERCICE 16.11.
Le responsable de la maintenance des machines à sous d'un casino doit vérifier qu'un certain type de machine est bien réglé sur une fréquence de succès de 0,06.

1. Lors du contrôle d'une machine, le technicien constate qu'elle a fourni 8 succès sur 65 jeux. Doit-il remettre en question le réglage de la machine ?

2. Lors du contrôle d'une autre machine, il constate qu'elle a fourni 12 succès sur 100 jeux. Doit-il remettre en question le réglage de la machine ?

EXERCICE 16.12.
En 2005, il y avait en France, selon l'INSEE, 62 731 milliers d'habitants dont 30 366 milliers d'hommes et 32 365 milliers de femmes. Cette même année en Premières générales à Dupuy de Lôme il y avait 350 élèves dont 218 femmes et 132 hommes. Les élèves de Dupuy étaient-ils représentatifs de la population française ?

EXERCICE 16.13.
On fait l'hypothèse que tous les ans à Dupuy de Lôme il y a, en Seconde, deux élèves sur trois qui sont des femmes soit une proportion $p = \frac{2}{3}$. En Seconde 13, sur 36 élèves il y a 16 femmes.

1. Déterminer si on doit rejetter l'hypothèse de départ, au seuil de 95 %.

2. Après vérification auprès de l'administration, il s'avère que cette hypothèse est juste. Que peut-on dire alors de la Seconde 13 ?

EXERCICE 16.14.

Les enfants sont dits prématurés lorsque la durée gestationnelle est inférieure ou égale à 259 jours. La proportion de ces naissances est de 6 %. Des chercheurs suggèrent que les femmes ayant eu un travail pénible pendant leur grossesse sont plus susceptibles d'avoir un enfant prématuré que les autres. Il est décidé de réaliser une enquête auprès d'un échantillon aléatoire de 400 naissances correspondant à des femmes ayant eu pendant leur grossesse un travail pénible. Les chercheurs décident a priori que si la proportion d'enfants nés prématurés dans cet échantillon est supérieure à la borne supérieure de l'intervalle de fluctuation au seuil de 0,95 alors leur hypothèse sera acceptée. Finalement le nombre d'enfants prématurés est de 50. Quelle est donc la conclusion ?

EXERCICE 16.15.

On admet que dans la population d'enfants de 11 à 14 ans d'un département français le pourcentage d'enfants ayant déjà eu une crise d'asthme dans leur vie est de 13 %.
Un médecin d'une ville de ce département est surpris du nombre important d'enfants le consultant ayant des crises d'asthme et en informe les services sanitaires. Ceux-ci décident d'entreprendre une étude et d'évaluer la proportion d'enfants de 11 à 14 ans ayant déjà eu des crises d'asthme.
Ils sélectionnent de manière aléatoire 100 jeunes de 11 à 14 ans de la ville.
La règle de décision prise est la suivante : si la proportion observée est supérieure à la borne supérieure de l'intervalle de fluctuation au seuil de 95 % alors une investigation plus complète sera mise en place afin de rechercher les facteurs de risque pouvant expliquer cette proportion élevée.

1. Déterminer l'intervalle de fluctuation au seuil de 95 % de la proportion de jeunes de 11 à 14 ans ayant eu une crise d'asthme dans un échantillon de taille 100.

2. L'étude réalisée auprès des 100 personnes a dénombré 19 jeunes ayant déjà eu des crises d'asthme. Que pouvez-vous conclure ?

3. Le médecin n'est pas convaincu par cette conclusion et déclare que le nombre de personnes interrogées était insuffisant pour mettre en évidence qu'il y avait plus de jeunes ayant eu des crises d'asthme que dans le reste du dépertement.
   Combien faudrait-il prendre de sujets pour qu'une proportion observée de 19 % soit en dehors de l'intervalle de fluctuation ?

EXERCICE 16.16.

Le 18 avril 2002, l'institut IPSOS effectue un sondage dans la population en âge de voter. On constitue un échantillon de 1 000 personnes (inscrites sur les listes électorales) que l'on suppose choisies ici de manière aléatoire. Les résultats partiels en sont les suivants : sur les 1 000 personnes

- 135 ont déclaré vouloir voter pour Jean-Marie Le Pen
- 195 ont déclaré vouloir voter pour Jacques Chirac
- 170 ont déclaré vouloir voter pour Lionel Jospin

1. Déterminer les trois intervalles de confiance au niveau de confiance de 95 % correspondant aux proportions d'intention de votes pour chacun des trois candidats.

2. Si on ne donne que le résultat brut du sondage et non l'intervalle de confiance, quel est le degré d'imprécision du résultat ?

3. À l'issue du premier tour Jean-Marie Le Pen, Jacques Chirac et Lionel Jospin ont obtenu, respectivement, 16,9 %, 19,9 % et 16,2 % des suffrages exprimés. Commenter.

4. L'institut CSA donnait en avril 14 % d'intention de votes pour Jean-Marie Le Pen pour un échantillon de taille identique. Déterminer l'intervalle de confiance à 95 % associé à ce nouveau sondage. Commenter.

EXERCICE 16.17.
Les sondages d'intention de vote s'effectuent en général sur des échantillons de taille $n = 1\,000$.

1. Déterminer l'amplitude de l'intervalle de confiance au seuil de 95 % des votants pour l'un des candidats quand le sondage indique des intentions de votes proches de 50 % (cas du second tour de l'élection présidentielle).

2. Déterminer l'amplitude de l'intervalle de confiance au seuil de 95 % des votants pour l'un des candidats quand le sondage indique des intentions de votes proches de 10 % (cas des « petits » candidats du premier tour de l'élection présidentielle).

EXERCICE 16.18.
Dans le but d'évaluer la prise en charge de la bronchiolite dans un hôpital de la région Aquitaine, une étude rétrospective a été mise en place.

1. Il est recommandé de coucher l'enfant de manière très inclinée (couchage en proclive) dans le cadre de la prise en charge de la bronchiolite. On évalue cette pratique à partir d'un échantillon de 134 dossiers. 106 enfants ont été couchés en proclive.
   Déterminer un intervalle de confiance au niveau de confiance de 95 % de la proportion d'enfants dont le couchage respecte la recommandation.

2. Une étude plus fine permet de comparer les pratiques entre les différents services ayant admis des enfants dont les résultats sont dans le tableau 16.3 de la présente page.

TABLE 16.3: Tableau de l'exercice 16.19

Couchage proclive	En service des urgences	En service hospitalier	Total
Oui	45	52	97
Non	29	8	37
Total	74	60	134

   (a) Déterminer un intervalle de confiance au seuil de 95 % de la proportion de couchage en proclive pour chaque type de service.

   (b) Peut-on conclure selon vous au seuil de 95 % que la pratique de couchage n'est pas identique selon le service ?

EXERCICE 16.19.
Un maraîcher achète un lot de semances de tomates pour produire ses plants de tomate. Il lui reste des semences de l'année passée, dont il doit contrôler le taux de germination pour pouvoir les utiliser avec les autres. En effet, des taux de germination trop différents provoquent des trous dans les plates bandes de production, ce qui génère un coût de manutention plus élevé (il faut enlever les pots non germés avant de les conditionner). Il faut donc comparer les taux de germination des semences des deux années.
Une stratégie consiste à calculer et à comparer les intervalles de confiance des taux de germination des plants de l'année et de l'année précédente. Si les deux intervalles ne se recoupent pas, on peut conclure à une différence de taux de germination entre les deux semances d'origines. Il faudra alors les semer séparément.

Pour faire cette comparaison, le maraîcher prélève, aléatoirement dans les semences de l'année, un échantillon de 200 graines qu'il met à germer. Il constate que 185 graines germent.
Il prélève ensuite, aléatoirement dans les semences de l'année précédente, un échantillon de 200 graines qu'il met à germer. Il constate que 150 graines germent.

1. Déterminer un intervalle de confiance, au niveau de confiance de 95 %, du taux de germination p_a du lot de semences de l'année.

2. Déterminer un intervalle de confiance au niveau 95 % du taux de germination p_b du lot de semences de l'année précédente.

3. Conclure.