

A Le programme

Objectifs visés par l'enseignement des statistiques et probabilités à l'occasion de résolutions de problèmes, dans le cadre de l'échantillonnage :

- faire réfléchir les élèves à la conception et la mise en œuvre d'une simulation ;
- sensibiliser les élèves à la fluctuation d'échantillonnage, aux notions d'intervalle de fluctuation et d'intervalle de confiance et à l'utilisation qui peut en être faite.

Contenus	Capacités attendues	Commentaires
<p>Échantillonnage Notion d'échantillon. Intervalle de fluctuation d'une fréquence au seuil de 95 %*. Réalisation d'une simulation</p>	<ul style="list-style-type: none"> - Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. - Exploiter et faire une analyse critique d'un résultat d'échantillonnage. 	<p>Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience. À l'occasion de la mise en place d'une simulation, on peut :</p> <ul style="list-style-type: none"> - utiliser les fonctions logiques d'un tableur ou d'une calculatrice, - mettre en place des instructions conditionnelles dans un algorithme. <p>L'objectif est d'amener les élèves à un questionnement lors des activités suivantes :</p> <ul style="list-style-type: none"> - l'estimation d'une proportion inconnue à partir d'un échantillon ; - la prise de décision à partir d'un échantillon.

* 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 . Cet intervalle peut être obtenu, de façon approchée, par simulation.

Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n \geq 25$ et des proportions p du caractère comprises entre 0,2 et 0,8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ avec une probabilité d'au moins 0,95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais **elle n'est pas exigible**.

B Notre point de vue

Ce chapitre traite la partie « Échantillonnage » du thème « Statistiques et Probabilités » du programme de la classe de Seconde. Les deux objectifs essentiels sont de concevoir et d'exploiter des simulations de situations concrètes, et de savoir exploiter un résultat d'échantillonnage.

Chacune des deux pages de cours est dédiée à un de ces deux objectifs. La première page définit la notion d'échantillon, insiste sur la fluctuation d'échantillonnage et sur la mise en œuvre de simulations : les trois **Savoir – faire** de la page 173 sont consacrés à la mise en œuvre de simulations avec une calculatrice ou un tableur, mais aussi à l'aide d'un algorithme.

La seconde page de cours permet de travailler sur l'exploitation d'un résultat d'échantillonnage : on définit un intervalle de fluctuation à 95 %, conformément au programme, puis on l'utilise dans un problème de prise de décision à partir d'un échantillon. Conformément aux commentaires du programme, on aborde aussi le problème de l'estimation par intervalle d'une proportion.

Une **fiche TICE** permet de faire le point sur toutes les techniques à connaître pour réaliser des simulations sur une calculatrice ou avec un tableur : la connaissance des deux fonctions de la calculatrice utiles pour simuler est indispensable pour que l'élève acquière les capacités du programme sur ce chapitre.

Les exercices utilisent beaucoup la calculatrice, outil pratique dans une classe. Les **deux TP** permettent de réaliser des simulations de plus grande importance avec le tableur. Ces deux TP ont été choisis afin de résoudre expérimentalement deux problèmes bien différents : un problème de santé publique dans une région du Canada et un problème historique posé par le chevalier de Méré au XVII^e siècle.

Les notions abordées dans le chapitre 8

- Échantillons
- Simulation d'expériences aléatoires
- Intervalle de fluctuation d'une proportion à 95 %
- Prise de décision sur un échantillon
- Estimation d'une proportion
- Utilisation de la calculatrice et du tableur pour simuler

C Réactiver les savoirs

Les notions abordées dans ces exercices permettent de réactiver les notions utiles pour ce chapitre : le calcul de fréquences d'une série de valeurs, abordé au collège et revu dans le chapitre 7, la notion d'intervalle, abordée au chapitre 1, et l'utilisation des fonctionnalités de base d'un tableur (formules, fonctions SOMME et SI).

Voir manuel page 332 et le site www.bordas-index.fr pour les corrigés détaillés.

D Activités

Activité 1 Avec une pièce de monnaie

Cette activité a pour but de produire expérimentalement des échantillons en lançant une pièce de monnaie : l'expérience est facile à mener et chaque élève peut la réaliser avec une pièce de monnaie. Il est important que les premiers échantillons produits par les élèves soient indépendants de toute simulation.

- 1. a.** On obtient par exemple : P, P, F, P, F, P, F ; P, F, P.
b. Dans l'exemple ci-dessus, la fréquence de « Pile » est 0,6 et celle de « Face » est 0,4.
- 2. a.** On fait 4 séries de 10 lancers de la pièce, et on note les résultats obtenus.
b. Les distributions de fréquences varient selon les échantillons.
- 3.** On détermine la distribution de fréquences pour l'échantillon de taille 50.

On peut alors regrouper les résultats de groupes d'élèves pour obtenir un échantillon de taille plus importante.

Activité 2 Des simulations

Cette activité permet de découvrir diverses méthodes de simulation à partir d'une expérience simple : le lancer d'un dé. L'outil utilisé ici est la calculatrice, outil le plus facilement disponible pour les élèves.

Fichiers associés sur le site www.bordas-index.fr et le manuel numérique Premium :

08_seconde_activite2.xls (Excel)

08_seconde_activite2.ods (OpenOffice)

08_seconde_activite2.alg (AlgoBox)

1. Pour obtenir 20 lancers successifs du dé, il est inutile de saisir chaque fois la formule donnée dans l'énoncé : une fois le premier résultat obtenu, on obtient d'autres résultats en appuyant simplement sur la touche **EXE** (sur Casio) ou **entrer** (sur Texas).

Cette expérience (comme les suivantes) peut être simulée par le professeur à l'aide du tableur (voir fichiers joints).

2. Avec une calculatrice Casio : **RanInt#(0,1)**.

Avec une calculatrice Texas : **entAléat(0,1)**.

On convient que 0 représente le fait d'obtenir « Pile » et 1 le fait d'obtenir « Face ».

3. Avec une calculatrice Casio : **RanInt#(1,100)**.

Avec une calculatrice Texas : **entAléat(1,100)**.

On convient qu'on obtient une bille rouge si le résultat trouvé est compris entre 1 et 40, et une bille bleue sinon.

4. a. On a simulé l'expérience consistant à lancer une pièce de monnaie équilibrée.

- b. S vaut alors 6 en sortie.
 c. Programmation sur calculatrice.

Casio

```
====ACTI2====
0→S
For 1→I To 100
RanInt#(0,1)→R
S+R→S
Next
S
```

Texas

```
PROGRAM:ACTI2
:0→S
:For(I,1,100)
:entAléat(0,1)→R
:
:S+R→S
:End
:Disp S
```

Activité 3 Un sondage pour une mairie

Cette activité permet de découvrir la notion d'intervalle de confiance.

Fichiers associés sur le site www.bordas-indice.fr et le manuel numérique Premium :

- 08_seconde_activite3.xls (Excel)
- 08_seconde_activite3.ods (OpenOffice)

1. a. La fonction ALEA() renvoie un nombre aléatoire compris entre 0 et 1.

Ainsi, `ALEA()+0,35` renvoie un nombre aléatoire compris entre 0,35 et 1,35. Sa partie entière est ainsi égale à 1 avec une probabilité égale à 0,35.

Cette formule simule donc le tirage au hasard d'un individu dans la population et elle renvoie 1 si cet individu est favorable à la liste Dupont, et 0 s'il est favorable à la liste Duchamp.

b. En recopiant vers la droite cette formule jusqu'en CV3, on obtient les intentions de vote de 100 individus pris au hasard dans la population.

Quand on recopie vers le bas la plage de cellules A3:CV3 jusqu'à la ligne 402, on obtient 400 échantillons d'électeurs pris au hasard de taille 100.

2. La formule `=MOYENNE(A3:CV3)` calcule la fréquence des personnes favorables à la liste Dupont dans le premier échantillon.

3. a. La formule saisie en CV3 est :

`=SI(ET(CW3>=025;CW3<=045);1;0)`. Elle affiche 1 si la fréquence des personnes favorables à la liste Dupont dans le premier échantillon est comprise entre 0,25 et 0,45.

Quand on recopie cette formule vers le bas dans la colonne CY jusqu'en ligne 402, la somme des éléments de la plage CY3:CY402 représente le nombre d'échantillons pour lesquels la fréquence appartient à l'intervalle [0,25 ; 0,45].

b. Pour calculer en CY405 la proportion des fréquences des échantillons appartenant à I, on a saisi en CY405 la formule : `=CY404/400`.

On simule de nouveaux échantillons en appuyant sur la touche F9 avec Excel, et sur la combinaison de touches MAJ Ctrl F9 avec OpenOffice.

On constate que la proportion trouvée est proche de 1, et très souvent supérieure à 0,95.

4. Dans l'onglet « p variable » de la feuille de calcul, on peut choisir la valeur de p dans la cellule B1, puis recommencer le même travail que précédemment.

On fait alors les mêmes constatations qu'à la question 3, même en faisant varier la valeur de la proportion p.

Activité 4 La disparition

Cette activité permet de découvrir la problématique de l'estimation, à partir d'une situation liée à la littérature : quelle est la proportion de « e » dans les textes écrits en langue française ? L'exemple du célèbre roman de Pérec, La disparition, permet de se rendre compte des limites de l'estimation.

1. Il y a 81 lettres dont 14 lettres « e ». La fréquence du « e » dans cette phrase est donc 0,173 (à 0,001 près).

2. Les fréquences de « e » trouvées vont être sensiblement différentes.

3. Chaque élève aura ainsi à sa disposition un intervalle différent.

4. Dans la langue française, la fréquence d'apparition de la lettre « e » est 0,1587 (source : Wikipédia).

On pourra alors examiner combien d'intervalles trouvés dans la classe contiennent effectivement cette valeur de p.

5. Si on utilise l'intervalle donné à la troisième question, on a $n = 1\,225$ et $f = 0$.

Ainsi : $f - \frac{1}{\sqrt{n}} = -\frac{1}{35} \approx -0,028$ et $f + \frac{1}{\sqrt{n}} = \frac{1}{35} \approx 0,029$.

L'intervalle estimant p est [0 ; 0,029] : dans ce cas, p n'appartient pas à l'intervalle de confiance trouvé !

E Exercices

Pour démarrer

- Non, il y a fluctuation des fréquences : elles varient d'un échantillon à l'autre.
- Exemple d'échantillon : Pile – Face – Pile – Pile – Pile.
- Exemple d'échantillon, en notant P le fait qu'il pleuve un jour et NP le fait qu'il ne pleuve pas :

P, P, NP, NP, P, P, NP, P, NP, P, NP, P, NP, NP, NP, P, NP, P, NP, NP, P, NP, P, P, NP, P, NP, P, NP, P, NP.

(Il y a 31 résultats dans cet échantillon).

4 Distribution des fréquences :

Número	1	2	3	4	5	6
Fréquence	0,1	0,14	0,22	0,2	0,16	0,18

- 5** 1. Oui, on peut utiliser une table de nombres au hasard.
2. On utilise la fonction **RanInt#** sur une calculatrice Casio et la fonction **entAléat** sur une calculatrice Texas.
6 Exercice corrigé p. 332 du manuel.
7 Fichiers associés sur le site www.bordas-index.fr

et le manuel numérique Premium :
08_seconde_ex7.xls (Excel)

08_seconde_ex7.ods (OpenOffice)

Exemple : le tirage au hasard d'une boule parmi quatre boules de couleurs différentes dans un sac.

- 8** On saisit **RanInt#(1,32)** sur une calculatrice Casio ou **entAléat(1,32)** sur une calculatrice Texas. Si le résultat trouvé est un entier compris entre 1 et 8, c'est un trèfle, sinon ce n'en est pas un.

9 Fichiers associés sur le site www.bordas-index.fr
 et le manuel numérique Premium :

08_seconde_ex9.xls (Excel)

08_seconde_ex9.ods (OpenOffice)

- 1.** On obtient 0.
2. **ALEA()+0,1** donne un nombre aléatoire compris entre 0,1 et 1,1 : ainsi, dans 10 % des cas, sa partie entière sera égale à 1 (ce qui traduira une partie gagnée).

- 10** 1. $p = 0,4$; $n = 50$.
 2. Non, il y a au moins 95 % des fréquences observées qui sont dans I.

- 11** 1. Non, car $p < 0,2$. 2. Oui.
 3. Non, car $p > 0,8$. 4. Non, car $n < 25$.

12 Exercice corrigé p. 332 du manuel.

13 On doit connaître la taille n de l'échantillon et la fréquence f de ce caractère dans l'échantillon.

- 14** 1. $n = 100$. 2. $f = 0,35$.

3. $f - \frac{1}{\sqrt{n}} = 0,25$ et $f + \frac{1}{\sqrt{n}} = 0,45$.

On estime p par l'intervalle $[0,25 ; 0,45]$.

Pour s'entraîner

- 15** 1. On saisit **Int(Ran# + 0.2)** sur une calculatrice Casio ou **ent(NbrAléat + 0.2)** sur une calculatrice Texas. Si le résultat trouvé est 1, c'est un chocolat noir, sinon c'est un chocolat blanc.

2. En notant N un chocolat noir et B un chocolat blanc, un échantillon de taille 10 peut être : N, N, B, N, B, N, B, N, N, N, soit une fréquence de chocolats blancs de 0,3.

- 16** 1. On peut saisir **RanInt#(1,100)** sur une calculatrice Casio ou **entAléat(1,100)** sur une calculatrice Texas. Si le résultat trouvé est compris entre 1 et 40, alors le fruit est un abricot, sinon c'est une pêche.

2. En notant A un abricot et P une pêche, un échantillon de taille 20 peut être : A, A, P, A, P, P, P, A, A, P, A, P, A, A, P, P, P, A, P, P, soit une fréquence de 0,45 pour les abricots et de 0,55 pour les pêches.

- 17** 1. On saisit **RanInt#(0,37)** sur une calculatrice Casio ou

entAléat(0,37) sur une calculatrice Texas. Le résultat donne le numéro de la case dans laquelle tombe la boule.

- 2.** Ce sont deux listes de taille 30 d'entiers compris entre 0 et 37.
3. Non, il y a fluctuation de la fréquence de sortie d'un numéro pair dans les échantillons de taille 30.

18 1. **Ran#** sur une calculatrice Casio et **NbrAléat** sur une calculatrice Texas.

2. a. **Ran# x 2** avec une calculatrice Casio et **NbrAléat x 2** avec une calculatrice Texas.

b. **Ran# x 4 + 2** avec une calculatrice Casio et

NbrAléat x 4 + 2 avec une calculatrice Texas.

- 19** 1. On peut saisir **RanInt#(1,10)** sur une calculatrice Casio ou **entAléat(1,10)** sur une calculatrice Texas. Si le résultat trouvé est compris entre 1 et 3, alors la bille tirée est rouge, s'il est compris entre 4 et 8, alors la bille tirée est bleue, et sinon elle est jaune.

2. Un échantillon de taille 30 possible est : B, R, R, J, B, B, J, R, B, R, B, R, B, J, R, B, J, B, R, R, B, R, R, J, R, B, B, R, B, B.

- 20** 1. On saisit **RanInt#(1,7)** sur une calculatrice Casio ou **entAléat(1,7)** sur une calculatrice Texas. Le résultat 1 simule le do, le résultat 2 simule le ré, ... etc

2. On donne un échantillon de taille 15. Par exemple : do, mi, fa, si, do, ré, sol, fa, la, la, do, ré, mi, fa, si.

21 Exercice corrigé p. 332 du manuel.

22 Fichiers associés sur le site www.bordas-index.fr
 et le manuel numérique Premium :

08_seconde_ex22.xls (Excel)

08_seconde_ex22.ods (OpenOffice)

1. On saisit dans la cellule A1 la formule : **=ALEA.ENTRE.BORNES(1,6)**, puis on recopie cette formule vers le bas jusqu'à la cellule A200.

2. On calcule la fréquence du 1 à l'aide de la formule :

=NB.SI(A1:A200;1)/200.

23 Fichiers associés sur le site www.bordas-index.fr
 et le manuel numérique Premium :

08_seconde_ex23.xls (Excel)

08_seconde_ex23.ods (OpenOffice)

1. On saisit dans la cellule A1 la formule : **=ENT(ALEA+1/7)**. Si le résultat trouvé est 1, alors la boule tirée est rouge, sinon elle est verte.

2. On recopie cette formule vers le bas jusqu'à la ligne 300.

3. On peut calculer la fréquence de boules rouges dans cet échantillon à l'aide de la formule : **=SOMME(A1:A300)/300**.

On peut aussi utiliser la formule : **=NB.SI(A1:A300;1)/300**.

24 Fichiers associés sur le site www.bordas-index.fr
 et le manuel numérique Premium :

08_seconde_ex24.xls (Excel)

08_seconde_ex24.ods (OpenOffice)

1. On saisit dans la cellule A1 la formule : **=ALEA.ENTRE.BORNES(0,1)+ALEA.ENTRE.BORNES(0,1)**.

2. Pour obtenir un échantillon de taille 1 000, on recopie cette formule jusqu'à la cellule A1000. Pour estimer la probabilité d'obtenir une seule fois PILE quand on lance deux pièces, on

calcule la fréquence du résultat 1 dans la plage **A1:A1000**, par exemple avec la formule : **=NB.SI(A1:A1000;1)/1000**.

25 Exercice résolu p. 179 du manuel.

26 1. La ligne à compléter s'écrit :

« r prend la valeur **Alea(0,1)** », où **Alea(0,1)** renvoie 0 ou 1 de façon aléatoire.

2. Algorithme modifié :

Variables	n, r, P sont des nombres entiers
Entrées	Saisir n
Initialisation	P prend la valeur 0
Traitement	Pour i allant de 1 à n faire r prend la valeur Alea(0,1) S prend la valeur $S + r$
	Fin Pour
Sorties	Afficher S

(La variable S dénombre les « PILE » obtenus).

27 1. Algorithme complété :

Variante	n est un entier naturel
Initialisation	Affecter à n la valeur 1
Traitement	Tant que Alea(1 ; 6) \neq 6 faire Affecter à n la valeur $n + 1$
	Fin Tant que
Sorties	Afficher n

2. Algorithme modifié :

Variables	n et S sont des entiers naturels m est un nombre réel
Initialisation	Affecter à n la valeur 1 Affecter à S la valeur 0
Traitement	Pour i variant de 1 à 100 faire Tant que Alea(1 ; 6) \neq 6 faire Affecter à n la valeur $n + 1$ Fin Tant que Affecter à S la valeur $S + n$
	Fin Pour
Sorties	Affecter à m la valeur $S/100$ Afficher m

28 C'est faux : un échantillon provient d'un tirage de bonbons avec remise ; on peut bien tirer 100 bonbons dans une boîte qui en contient 50 si on remet chaque bonbon dans la boîte après tirage.

29 C'est faux, car **ALEA()+0,9** génère un nombre aléatoire compris entre 0,9 et 1,9, et l'affichage sera 0 avec une probabilité de 0,1.

30 1. Cette proportion est $\frac{20}{64} = 0,3125$.

2. Ici, $n = 400$ et $p = 0,3125$.

Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle : $[0,2625 ; 0,3625]$.

31 On a ici : $n = 400$ et $p = 0,68$. Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle de fluctuation : $[0,63 ; 0,73]$.

32 Ici, $p = 0,5$. Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées.

a. $[0,45 ; 0,55]$. **b.** $[0,475 ; 0,525]$. **c.** $[0,49 ; 0,51]$.

33 1. On a : $n = 100$ et $p = 0,25$. Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle de fluctuation : $[0,15 ; 0,35]$.

2. Dans ce cas, $n = 1\ 600$. On obtient l'intervalle : $[0,225 ; 0,275]$.

34 1. On a : $n = 100$ et $p = 0,3$.

Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle de fluctuation au seuil de 95 % : $[0,2 ; 0,4]$.

2. Règle de décision : « Si la fréquence observée dans un échantillon de taille 100 appartient à l'intervalle $I = [0,2 ; 0,4]$, alors on accepte l'hypothèse $p = 0,3$, sinon on la rejette ».

3. La fréquence observée est : $f = \frac{19}{100} = 0,19$.

On rejette l'hypothèse au seuil de 95 %. L'affirmation donnée ne peut être validée par l'expert.

35 1. On a : $n = 100$ et $p = 0,6$.

Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle de fluctuation au seuil de 95 % : $[0,5 ; 0,7]$.

2. Règle de décision : « Si la fréquence observée dans un échantillon de taille 100 appartient à l'intervalle $I = [0,5 ; 0,7]$, alors on accepte l'hypothèse $p = 0,6$, sinon on la rejette ».

3. La fréquence observée est : $f = \frac{51}{100} = 0,51$.

On accepte l'hypothèse selon laquelle la proportion des personnes dont la couleur des yeux dans cette ville est bleue ou verte est 0,6 au seuil de 95 %.

36 1. On a : $n = 577$ et $p = 0,516$.

Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle de fluctuation au seuil de 95 % : $I = [0,474 ; 0,558]$, en arrondissant les bornes à 0,001 près.

2. La règle de décision est ici la suivante : « si la fréquence observée dans un échantillon de taille 577 appartient à l'intervalle I , alors on accepte l'hypothèse $p = 0,516$ (parité des sexes), sinon on la rejette ».

La fréquence observée à l'Assemblée nationale est $\frac{155}{577} \approx 0,269$: on rejette donc l'hypothèse selon laquelle la parité des sexes est respectée à l'Assemblée nationale au seuil de 95 %.

3. Dans le cas de la Suède, l'intervalle de fluctuation est : $J = [0,451 ; 0,559]$.

La fréquence observée à la Diète Royale est 0,447 : celle-ci n'appartient pas à J , donc la parité des sexes n'est également pas respectée à la Diète suédoise.

37 Exercice corrigé p. 332 du manuel.

38 1. Négation : « Il existe un intervalle de fluctuation de la fréquence de ce caractère au seuil de 95 % dans un échantillon de taille n qui ne contient pas p ».

2. La proposition initiale est fautive.

39 C'est faux, l'intervalle de fluctuation est :

$$\left[0,5 - \frac{1}{10} ; 0,5 + \frac{1}{10}\right] = [0,40 ; 0,60].$$

40 C'est vrai car l'amplitude de l'intervalle de fluctuation est $\frac{2}{\sqrt{n}}$.

41 1. Intervalle de confiance de p au seuil de 95 % : $[0,6 ; 0,8]$.

2. Oui, c'est possible : la proportion p n'appartient pas toujours à l'intervalle de confiance calculé.

- 42** 1. Intervalle de confiance de p au seuil de 95 % : [0,7 ; 0,9].
 2. Oui, c'est possible : la proportion p n'appartient pas toujours à l'intervalle de confiance calculé.

43 Exercice résolu p. 173 du manuel.

44 Exercice corrigé p. 333 du manuel.

- 45** 1. Intervalle de confiance de l'audience du journal au seuil de 95 % : [0,196 ; 0,264], en arrondissant à 0,001 près.

2. $\frac{95}{100} \times 500 = 475$: il y a donc en moyenne 475 de ces intervalles de confiance qui contiennent l'audience réelle du journal.

- 46 Vrai** : p n'appartient pas toujours à l'intervalle de confiance calculé.

- 47 Faux** : p peut aussi bien appartenir à l'intervalle [0,44 ; 0,46] que ne pas y appartenir.

- 48** 1. La fréquence observée dans l'échantillon est :

$$f = \frac{1080}{1600} = 0,675.$$

Un intervalle de confiance au seuil 95 % est : [0,65 ; 0,70].

2. Il est possible que p soit égal à 0,675, mais il peut aussi être égal à beaucoup d'autres valeurs.

- 49** 1. On a : $n = 100$ et $p = 0,24$. Les conditions $n \geq 25$ et $0,2 \leq p \leq 0,8$ sont vérifiées, d'où l'intervalle de fluctuation au seuil de 95 % : [0,14 ; 0,34].

2. Règle de décision : « Si la fréquence observée de fumeurs dans un échantillon de 100 personnes de l'entreprise appartient à l'intervalle $I = [0,14 ; 0,34]$, alors on accepte l'hypothèse $p = 0,24$, sinon on la rejette ».

3. La fréquence observée est : $f = \frac{13}{100} = 0,13$. On rejette cette hypothèse au seuil de 95 %.

Travaux pratiques

TPI Pollution et sex-ratio

L'objectif de ce TP est d'étudier un phénomène remarquable qui s'est produit dans la vallée pétrochimique de Sarnia, au Canada : le déséquilibre dans les naissances de garçons et de filles dans cette vallée est – il lié à la présence d'industries chimiques ? Pour essayer de répondre à cette question, on utilisera deux méthodes : la simulation à l'aide d'un tableur puis l'utilisation des résultats du cours sur les intervalles de fluctuation.

A. Utilisation d'un algorithme

Fichiers associés sur le site www.bordas-indexe.fr et le manuel numérique Premium :

08_seconde_TP1_correction.alg (AlgoBox)

08_seconde_TP1_correction.xls (Excel)

08_seconde_TP1_correction.ods (OpenOffice)

08_seconde_TP1_correction.xws (Xcas).

1. Algorithme complété :

Variation	g est un entier naturel
Initialisation	g prend la valeur 0
Traitement	Pour i allant de 1 à 132 faire g prend la valeur $g + \text{Alea}(0,1)$
	Fin Pour
Sorties	Afficher g

2. • Programme avec AlgoBox :

```

VARIABLES
├── g EST_DU_TYPE NOMBRE
├── i EST_DU_TYPE NOMBRE
└── DEBUT_ALGORITHME
    ├── g PREND_LA_VALEUR 0
    └── POUR i ALLANT DE 1 A 132
        ├── DEBUT_POUR
        │   ├── g PREND_LA_VALEUR g+ALGOBOX_ALEA_ENT(0,1)
        │   └── FIN_POUR
        └── AFFICHER g
FIN_ALGORITHME
  
```

- Programme avec Xcas :

```

g:=0;
pour k de 1 jusque 132 faire
  g:=g+hasard(2);
fpour;
afficher(g)
  
```

3. La plus petite valeur de g obtenue est en général assez éloignée de 46.

B. Utilisation d'un tableur

Fichiers associés sur le site www.bordas-indexe.fr et le manuel numérique Premium :

08_seconde_TP1_correction.xls (Excel)

08_seconde_TP1_correction.ods (OpenOffice)

1. On saisit en **A1** la formule : **=ENT(ALEA()+0,512)**.

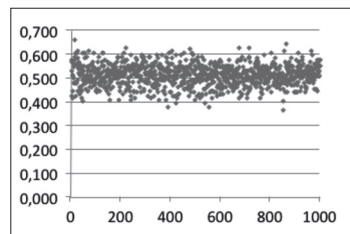
2. Par recopie vers la droite jusqu'à la colonne **EB**, on simule 132 naissances sur la ligne **A**.

3. La formule **=SOMME(A1:EB1)** permet d'afficher dans la cellule **ED1** le nombre de garçons de l'échantillon.

La formule **=ED1/132** permet d'afficher en **ED2** la fréquence de garçons dans cet échantillon.

4. Par recopie vers le bas jusqu'à la ligne **1 000**, on simule 1 000 échantillons de taille 132.

5. a. Représentation graphique des fréquences obtenues :



- b. La fréquence de garçons se trouve le plus souvent dans l'intervalle [0,4 ; 0,6].

- c. Sur ces 1 000 échantillons, on en trouve très peu pour lesquels la fréquence de garçons est inférieure à 0,4 : parfois 2, 3, 4 ou 5 ... On peut générer d'autres échantillons en appuyant sur la touche **F9** avec Excel, ou sur la combinaison de touches **MAJ Ctrl F9** avec OpenOffice.

6. $\frac{46}{32} \approx 0,35$. Si l'on cherche les échantillons tels que la fréquence soit inférieure à 0,35, on en trouve 0, parfois 1. Il semble donc extrêmement rare qu'il y ait 46 garçons sur 132 naissances. On peut donc raisonnablement s'alarmer.

C. Utilisation d'un intervalle de fluctuation

1. Intervalle de fluctuation au seuil de 95 % de la fréquence des garçons dans les échantillons de taille 132 : $I = [0,424 ; 0,600]$, en arrondissant à 0,001 près.
2. Règle de décision : « Si la fréquence de garçons observée dans un échantillon de taille 132 appartient à I , alors on accepte l'hypothèse selon laquelle la proportion de garçons à la naissance est 0,512, sinon on la rejette ».
3. La fréquence observée est 0,34, donc on rejette l'hypothèse faite au seuil de 95 %.

TP2 Problèmes de dés

L'objectif de ce TP est de simuler plusieurs situations de lancers de dés équilibrés afin de résoudre de manière expérimentale le célèbre problème du chevalier de Méré.

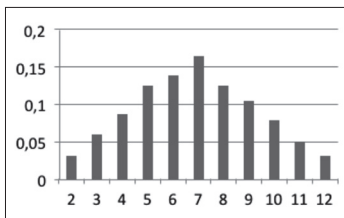
Fichiers associés sur le site www.bordas-indice.fr et le manuel numérique Premium :

08_seconde_TP2_correction.xls (Excel)

08_seconde_TP2_correction.ods (OpenOffice)

A. Lancer de deux dés

1. a. Premier onglet de la feuille de calcul.
- b. On saisit en **A1** et en **B1** la formule : `=ALEA.ENTRE.BORNES(1;6)`.
- c. On saisit en **C1** la formule : `=A1+B1`.
- d. On obtient un échantillon de taille 1 000 en recopiant vers le bas la plage **A1:C1** jusqu'à la ligne 1 000.
2. a. On tape les valeurs dans les cellules **E1** à **O1**.
- b. On saisit en **E2** la formule : `=NB.SI($C1:$C1000;E1)`. On fixe le numéro de colonne afin que, par recopie vers la droite, les valeurs restent inchangées.
- c. On saisit en **E3** la formule : `=E2/1000`.
- d. On recopie ces formules vers la droite.
3. a. Représentation graphique de la distribution des fréquences :



- b. On peut faire d'autres simulations en appuyant sur la touche **F9** avec Excel, ou sur la combinaison de touches **MAJ** **Ctrl** **F9** avec OpenOffice.
- B. Lancer de quatre dés**
1. Deuxième onglet de la feuille de calcul.
2. On saisit en **A1**, en **B1**, en **C1** et en **D1** la formule : `=ALEA.ENTRE.BORNES(1;6)`.
3. Cette formule affiche 1 si le plus grand nombre affiché dans les cellules **A1**, **B1**, **C1** et **D1** est 6, et 0 sinon. En termes de dés, elle affiche 0 si aucun dé ne donne 6, et elle affiche 1 dès qu'un dé affiche 6.
4. On recopie la plage **A1:E1** vers le bas jusqu'à la ligne 1 000.
5. On saisit en **G1** la formule : `=NB.SI(E1:E1000;1)/1000`.

Elle calcule la fréquence de l'événement « obtenir au moins une fois 6 » dans cet échantillon.

C. Le problème du chevalier de Méré

1. a. Troisième onglet de la feuille de calcul.
- b. Quand cette formule donne pour résultat 1, cela signifie que la somme calculée vaut 12 et donc que chaque fonction `ALEA.ENTRE.BORNES(1;6)` a pris la valeur 6 : on obtient 1 quand on obtient 6 sur chaque dé.
- c. On recopie cette formule jusqu'en **X1**.
2. a. On saisit dans la cellule **Z1** la formule : `=SI(SOMME(A1:X1)=0;0;1)`. Si la somme des éléments de la plage **A1:X1** vaut 0, alors c'est que chaque élément de cette plage vaut 0, et donc qu'on n'a jamais obtenu de double-six. Si la somme est différente de 0, c'est qu'on a obtenu au moins un double – six (cette somme peut être 1, 2, ...24).
- b. On recopie la plage **A1:X1** vers le bas jusqu'à la ligne 1 000 pour obtenir un échantillon de taille 1 000.
3. Pour calculer dans la cellule **AB1** la fréquence de l'événement « obtenir au moins un double-six », on saisit la formule `=NB.SI(Z1:Z1000;1)/1000`. En effet, `NB.SI(Z1:Z1000;1)` donne le nombre de fois où on a eu au moins un double – six.
4. Quand on fait plusieurs simulations, on se rend compte que la fréquence trouvée dans la question **B.5** est en général supérieure à celle trouvée dans la question **C.4**. On peut donc conjecturer qu'il est plus avantageux de parier sur l'apparition d'un 6 en lançant 4 dés.

Remarque : un calcul de probabilités du niveau de la classe de Première montre que la première probabilité est $1 - \left(\frac{5}{6}\right)^4 \approx 0,518$, et que la seconde probabilité est $1 - \left(\frac{35}{36}\right)^{24} \approx 0,491$.

Pour approfondir

50 Ici, $p = 0,2$ et $n \geq 25$ pour chaque cas, donc les conditions de validité sont respectées.

– Un intervalle de fluctuation du taux de cancers dans les échantillons de taille 100 est : $[0,1 ; 0,3]$.

– Un intervalle de fluctuation du taux de cancers dans les échantillons de taille 200 est : $[0,129 ; 0,271]$, en arrondissant à 0,001 près.

– Un intervalle de fluctuation du taux de cancers dans les échantillons de taille 10 000 est : $[0,19 ; 0,21]$.

Les fréquences de cancers observés pour chaque traitement sont 0,14.

Cette fréquence appartient aux deux premiers intervalles de fluctuation : on peut donc considérer que les deux premiers traitements sont positifs au seuil de 95 %.

51 $1. f = \frac{78}{160} = 0,4875$.

2. Intervalle de confiance au seuil de 95 % : $[0,408 ; 0,567]$, en arrondissant à 0,001 près.

3. Soit n la taille de l'échantillon.

On doit résoudre : $\frac{2}{\sqrt{n}} < 0,04$, soit $\sqrt{n} > 50$, et $n > 2 500$.

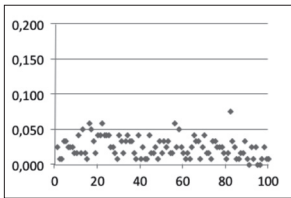
La taille de l'échantillon doit au moins être de 2 501.

52 Fichiers associés sur le site www.bordas-indice.fr et le manuel numérique Premium :

08_seconde_exercice52.xls (Excel)

08_seconde_exercice52.ods (OpenOffice)

- On saisit dans la cellule **A1** la formule : **=ENT(ALEA()+1/40)**
- Pour simuler un échantillon de taille 120 dans la ligne **A**, on recopie la formule précédente jusqu'à la cellule **DP1**.
On calcule la fréquence de décès dans l'échantillon dans la cellule **DR1** par la formule : **=SOMME(A1:DP1)/120**.
- Par recopie vers le bas jusqu'à la ligne 100, on obtient 100 échantillons de taille 120.
- Représentation graphique des fréquences obtenues dans chacun de ces 100 échantillons :



5. On constate que les fréquences sont généralement comprises entre 0 et 0,06.

En 2012, la fréquence observée de décès de bélugas nouveaux – nés est $\frac{16}{120}$, soit environ 0,13 : cette fréquence est nettement supérieure aux fréquences rencontrées dans les échantillons de taille 120 quand on fait l'hypothèse $p = 0,025$. On peut donc penser que ce nombre de décès est anormalement élevé.

53 La proposition exprimant le mieux ce que veut dire le géologue est la proposition C.

54 Fichiers associés sur le site www.bordas-indice.fr et le manuel numérique Premium :

08_seconde_ex54.alg (AlgoBox)

- On complète l'algorithme de la façon suivante :

Si $A = 0$

Alors affecter à X la valeur $X + 1$

Sinon affecter à X la valeur $X - 1$

- Algorithme modifié :

Variables	X et A sont des entiers F est un nombre décimal
Initialisation	Affecter à N la valeur 0 Pour J allant de 1 à 100 faire Affecter à X la valeur 0 Pour I allant de 1 à 20 faire Affecter aléatoirement à A la valeur 0 ou 1 Si $A = 0$ Alors affecter à X la valeur $X + 1$ Sinon affecter à X la valeur $X - 1$ Fin Si Fin Pour Si $X = 0$ Alors affecter à N la valeur $N + 1$ Fin Si Fin Pour Affecter à F la valeur $N/100$
Traitement	
Sorties	Afficher F

55 Soit n la taille de l'échantillon.

On doit avoir : $\frac{2}{\sqrt{n}} < 0,08$ soit $\sqrt{n} > 25$, et $n > 625$. Le nombre minimal de personnes à interroger est 626.

56 Le journal qui fournit le résultat le plus fiable est le journal 3. On peut le justifier :

- l'échantillon considéré a la taille la plus importante (1 000) ;
- les citoyens ont été tirés au hasard ;
- la date du sondage est proche de la date de l'élection.

57 1. a. Un intervalle de fluctuation au seuil de 95 % de la fréquence des femmes est $I = [0,385 ; 0,535]$, en arrondissant à 0,001 près.

b. Un intervalle de fluctuation au seuil de 95 % de la fréquence des compétiteurs est $J = [0,175 ; 0,325]$, en arrondissant à 0,001 près.

c. La fréquence des femmes est $\frac{75}{180} \approx 0,417$.

La fréquence des compétiteurs est $\frac{56}{180} \approx 0,312$.

0,417 appartient à I et 0,312 appartient à J , donc l'échantillon est représentatif de la population étudiée.

2. a. Intervalle de confiance au seuil de 95 % de la proportion des adhérents désirant changer de mur d'escalade : $[0,503 ; 0,653]$.

b. Connaissant cet intervalle, le comité directeur peut prendre la décision de changer de mur d'escalade.

58 Un intervalle de fluctuation de la fréquence des Américains d'origine mexicaine dans les échantillons de taille 870 est :

$$\left[0,79 - \frac{1}{\sqrt{870}} ; 0,79 + \frac{1}{\sqrt{870}} \right], \text{ soit } I = [0,756 ; 0,824], \text{ en}$$

arrondissant à 0,001 près.

La fréquence observée des Américains d'origine mexicaine dans le jury convoqué est $\frac{339}{870} \approx 0,390$.

Cette fréquence n'appartient pas à l'intervalle I , donc au seuil de 95 %, on rejette l'hypothèse selon laquelle la constitution du jury a eu lieu de manière aléatoire.

59 L'amplitude d'un intervalle de fluctuation de taille n est $\frac{2}{\sqrt{n}}$. On cherche la taille N d'un échantillon telle que :

$$\frac{2}{\sqrt{N}} = \frac{1}{10} \times \frac{2}{\sqrt{n}}, \text{ donc } N = 100n.$$

On doit donc multiplier la taille de l'échantillon par 100.

60 On détermine un intervalle de fluctuation au seuil de 95 % de la part de marché du produit dans les échantillons de taille 250 : $I = [0,336 ; 0,464]$, en arrondissant à 0,001 près.

On peut établir la règle de décision : « Si la part de marché constatée dans un échantillon de taille 250 appartient à I , alors on accepte l'hypothèse selon laquelle la part de marché de ce produit est 0,4, sinon on la rejette ».

Ici, 0,46 appartient à l'intervalle I , donc au seuil de 95 % l'hypothèse est acceptée.

Ainsi, cette augmentation peut être due au hasard, et pas particulièrement à la publicité.